RUNNING HEAD: DUAL CHARACTER CONCEPTS

Dual Character Concepts and the Normative Dimension of Conceptual Representation

[Knobe, J., Prasada, S., & Newman, G. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition*. 127, 242-257.]

Joshua Knobe,[1,2] Sandeep Prasada[3,4] and George E. Newman[5]

[1] *Program in Cognitive Science, Yale University;*
[2] *Department of Philosophy, Yale University;*
[3] *Department of Psychology, Hunter College, City University of New York;*
[4] *Program in Linguistics, Graduate Center, City University of New York.*
[5] *School of Management, Yale University;*

Corresponding Author:
Joshua Knobe
Program in Cognitive Science
P.O. Box 208306
New Haven, CT 06520-8306
joshua.knobe@yale.edu

Abstract

Five experiments provide evidence for a class of 'dual character concepts.' Dual character concepts characterize their members in terms of both (a) a set of concrete features and (b) the abstract values that these features serve to realize. As such, these concepts provide two bases for evaluating category members and two different criteria for category membership. Experiment 1 provides support for the notion that dual character concepts have two bases for evaluation. Experiments 2-4 explore the claim that dual character concepts have two different criteria for category membership. The results show that when an object possesses the appropriate concrete features, but does not fulfill the appropriate abstract value, it is judged to be a category member in one sense but not in another. Finally, Experiment 5 uses the theory developed here to construct artificial dual character concepts and examines whether participants react to these artificial concepts in the same way as naturally occurring dual character concepts. The present studies serve to define the nature of dual character concepts and distinguish them from other types of concepts (e.g., natural kind concepts), which share some, but not all of the properties of dual character concepts. More broadly, these phenomena suggest a normative dimension in everyday conceptual representation..

Keywords: Concepts; natural kinds; teleology.

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). In other words, what we find in this concept is a type of duality: certain concepts seem to involve two ways of characterizing their instances, and thus two ways of determining category membership.

Although these phenomena have been explored with respect to certain specific concepts in philosophy (e.g. Aristotle, 1999/350 BC, on the concept of friendship; Gellner, 1973, on the concept of a gentleman), as far as we know, there has not yet been any systematic work investigating these phenomena empirically. 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 (e.g. Bruner, Goodnow & Austin, 1956), a prototype (e.g. Rosch & Mervis, 1975; Hampton, 1998), salient exemplars (e.g. Medin & Shaffer, 1978; Nosofksy, 1988), or a theory (e.g. Carey, 1985; Gelman & Wellman, 1991; Gopnik & Meltzoff, 1997; Keil, 1989; Murphy & Medin, 1985) (but see Machery & Seppälä, 2010; Smith, Patalano & Jonides, 1998; Weiskopf, 2009). The experiments in this paper provide evidence for a class of cases in which that assumption is violated and a single concept characterizes members of a category using two distinct sets of criteria.

1. Dual character concepts

The experiments seek to demonstrate that there is a class of concepts that are represented via both (a) a set of concrete features and (b) some underlying abstract value. 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. Though the concepts in this class differ from each other in numerous important respects, they share a certain kind of structure that supports dual characterization. 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 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: PHARMACIST, ACQUAINTANCE, RUSTLING NOISE, SECOND COUSIN, and so on. These concepts are not seen as having dual character (at least by most people; Leslie, in press), and we will use them in the experiments below as control concepts.

Of course, it is sometimes possible to use even concepts of this latter type in sentences that in some ways resemble (1) - (3). For example, if a person has been working informally as a pharmacist but is not officially certified to perform that sort of work, one might say: 'There is a sense in which she is a pharmacist, but technically, she is actually not a pharmacist.' The use of sentences like these is well explained by existing theories of hedges (Lakoff 1973; Malt 1990) and task variation (Gelman, 2003), but we will argue that there is something importantly different, and therefore worthy of further examination, at work in people's use of dual character concepts.

2. From concrete features to abstract values

What makes dual character concepts unique? We suggest that it is the fact that each dual character concept contains two different ways of characterizing members of the category to which it applies and that these two ways of characterizing members of the category stand in a particular type of relationship. We now introduce a specific hypothesis about the nature of this relationship.

Consider again the concept SCIENTIST. If you asked someone to explain what it meant to be a scientist, that person might begin by giving you a list of concrete features that scientists typically display:

Conducting experiments
Analyzing data
Developing theories
Writing papers

But when you received this answer, you would immediately notice that you were not simply receiving an arbitrary list of features. On the contrary, it should be clear that all of these features have something important in common. Specifically, they are all ways of realizing the same abstract value: *the pursuit of empirical knowledge*. Hence, you might guess that what the person was trying to communicate to you was not just this list of features but also the abstract value that they all serve to realize.

We propose that dual character concepts have precisely this sort of structure. Like many other concepts, dual character concepts are associated with a list of concrete features (e.g., Murphy, 2002). However, unlike most other concepts, the features associated with dual character concepts can all be seen as ways of realizing the same abstract values. People therefore come to represent the concept not only in terms of the concrete features themselves but also in terms of the abstract values that these features serve to realize.

The structure we are hypothesizing for dual character concepts should be contrasted with the structure to be found in two other classes of concepts. On one hand, it is quite different from the structure found in our control concepts. The concept BUS DRIVER is associated with certain concrete features (*driving*, *transporting passengers*, etc.), but these concrete features would not

normally be seen as ways of realizing any more abstract value. The concept is understood entirely in terms of the concrete features themselves.

On the other hand, dual character concepts should be contrasted with natural kind concepts like TIGER. As a number of researchers have emphasized, these concepts are not merely understood in terms of their superficial features (Gelman, 2003; Keil, 1989; Newman & Keil, 2008). People might associate the concept TIGER with a list of features (*striped*, *ferocious*, etc.), but they see all of these features as the product of an underlying causal factor (the tiger's hidden essence). They then regard this underlying causal factor as the true criterion for category membership (Medin & Ortony, 1989; but see Hampton, Estes & Simmons, 2007).

Though dual character concepts resemble natural kind concepts in being associated with criteria that go beyond superficial features, the structure found in dual character concepts is quite different from the one that has been identified in the existing literature on natural kinds. In natural kind concepts, the observable features are seen as caused by (Ahn, 1998; Keil, 1989) or otherwise dependent on (Sloman, Love & Ahn, 1998) the hidden essence or other 'deeper' features. By contrast, in the case of dual character concepts, the relationship between the concrete features and the abstract values is almost exactly the opposite (see Figure 1). The abstract values are not seen as *causing* the concrete features; rather, the idea is that the concrete features generally *realize* the abstract values. Thus, a given object can fall under a dual character concept even if the relevant abstract values do not in any way serve to explain how it came to have the concrete features it does. For example, even if a song were generated through an entirely random procedure, its guitar chords might serve to realize the abstract values associated with rock music, and we could then say that, in the fullest possible sense, this song counted as rock music (or simply that it 'rocked').

The experiments reported here contrast dual character concepts with both control concepts and natural kind concepts.

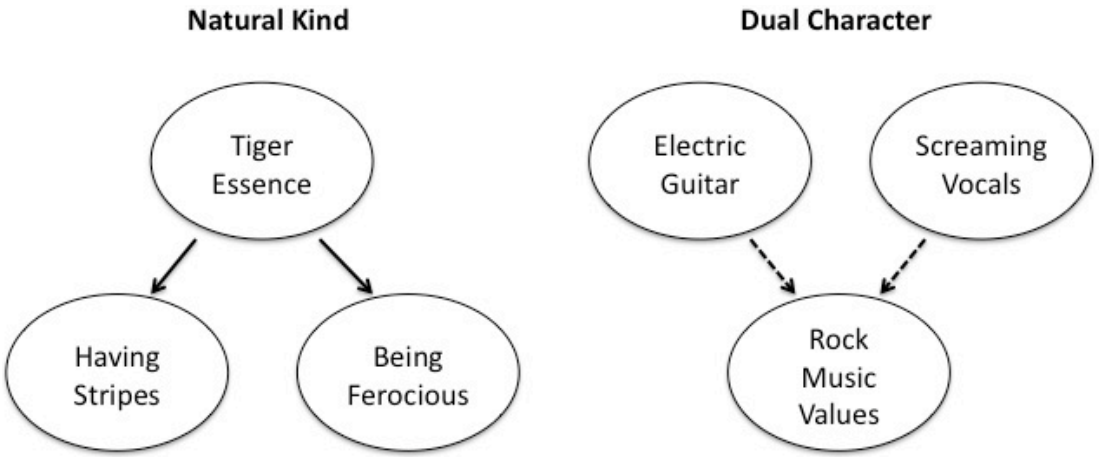

Figure 1. *Natural kind concepts and dual character concepts*. In natural kind concepts, the hidden essence is seen as causing the observable features, whereas in dual character concepts, the concrete features are seen as realizing the abstract values.

3. The role of normative evaluations

A variety of existing studies have examined the ways in which judgments of category membership are affected by normative evaluations. These studies suggest that judgments of typicality can be affected by the degree to which an object is seen as approximating the 'ideal' for goal-derived and role-governed categories and even taxonomic categories in certain circumstances (Barsalou, 1985; Lynch, Coley & Medin, 2000; Bailenson, Shum, Atran, Medin & Coley, 2002; Burnett, Medin, Ross & Blok, 2005; Goldwater, Markman, & Stilwell, 2011; but see Kim & Murphy, in press).

In the case of dual character concepts, however, people appear to have two distinct ways of characterizing category members, and thus can associate each dual character concept with two distinct dimensions of normative evaluation. Thus, judgments of category membership for dual

character concepts can potentially be influenced by a distinctive type of normative evaluation that does not play a role in judgments about other concepts.

Consider again the concept SCIENTIST. This concept 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 then play a certain role in intuitions about typicality, as predicted by existing theories (Barsalou, 1985; Lynch et al., 2000; Bailenson et al., 2002; Burnett et al., 2005; Goldwater et al., 2011).

But it seems that there is also another, quite different dimension of normativity to be found here. Specifically, it might be thought that certain people embody, in their whole way of life, the abstract values associated with the scientific enterprise. We could praise a person who embodies these values by saying:

(5) She is a *true* scientist.

The important thing to note here is that these two dimensions of normativity can sometimes come apart. We can imagine a person who has not yet acquired the concrete skills necessary for scientific research but who nonetheless embodies throughout her life the relevant abstract values. Such a person might not be a good scientist, but we could nonetheless praise her by saying 'She is a true scientist.'

More telling perhaps is the fact that 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 as we will see in Experiment 1, the word 'true' is not seen as appropriate in cases like this one. 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, our conceptual systems appear to support at least two types of normative evaluations. One type of evaluation proceeds by looking at certain concrete properties and checking to see whether a given object displays these properties in an ideal form. The second takes the concrete properties as ways of realizing more abstract values and then asks whether a given object embodies those abstract values. This second type of evaluation cannot be applied to all concepts, but it can be applied to concepts that show dual character.

4. Clarifications

At this point, it may be helpful to introduce three quick clarifications. First, we suggested above that the concrete features associated with a given dual character concept should be seen as realizing the relevant abstract value. It should be noted, however, that this relationship only holds in a rough, general way. In other words, it would be wrong to assume that the concrete features *always* realize that abstract value; the point is merely that they *generally* realize it. For example, it is a striking fact that the concrete features associated with science (experiments, statistics, etc.) are ways of realizing a particular abstract value (the pursuit of empirical knowledge), but it is also a striking fact that people can sometimes display all of these concrete features while utterly

failing to realize the corresponding abstract value. Thus, there will be cases in which the two systems of criteria come apart, and these cases will form the basis for the studies we present below.

Second, the claim that the relevant values are 'abstract' raises difficult questions about the very notion of 'abstractness' and the role it plays in theories of concepts. We will not be offering a general answer to those questions here (see Rosen, 2012 for a number of different views). However, we do want to emphasize that our framework does not presuppose any kind of strict dichotomous distinction between the abstract and the concrete. For example, it does not presuppose that the value 'pursuit of empirical knowledge' is completely abstract, while the feature 'conducting experiments' is completely concrete. The only assumption is that the relevant value is *more* abstract than the features it realizes.

Third, it will not always be possible to explicitly describe the abstract value associated with a given concept. Take the concept ROCK MUSIC. It seems that this concept is associated with a list of concrete features (*electric guitars*, *driving beats*, *screaming vocals*) and also with certain more abstract values. Yet it would be extraordinarily difficult to explain in explicit detail what those abstract values are. One might come up with some plausible candidates (*youthful energy*? *cathartic rebellion*?), but no matter what one says, there will always be a sense that one has left out something of vital importance. Perhaps the best way of conveying the abstract value would simply be to talk in detail about the concrete features and then to say something like: 'Listen to those guitars and those vocals. The abstract value I have in mind is the one people can generally realize by making music like that.' (Putting this claim somewhat differently, one might say that dual character concepts can involve 'placeholder values' in much the same way that natural kind concepts have been thought to involve 'placeholder essences'; Medin & Ortony, 1989.)

5. Stimulus construction and overview of experiments

To examine these issues empirically, we need a set of naturally occurring concepts hypothesized to have dual character, and we therefore conducted a brief study designed simply to generate appropriate experimental stimuli. We began by generating a larger list of 55 different concepts, including concepts from a variety of different domains. Twelve participants recruited through Amazon's Mechanical Turk were presented with all 55 of these 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 a scientist [bartender, optician, etc.] at all. In fact, if you think that he is really a scientist [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 or whether it was simply beside the point and didn't make sense. Participants marked their answers on a scale from 1 ('doesn't make sense') to 7 ('makes sense').

We selected the 20 concepts that received the highest scores (e.g., FRIEND, LOVE, POEM) and the 20 that received the lowest scores (e.g., UNCLE, RUSTLING, OBITUARY). The top 20 were hypothesized to be dual character concepts, as references to values were judged to be sensible when determining category membership, and the bottom 20 were used as control concepts. (The two lists are included in full in the Appendix.)

A series of studies then used a variety of measures to provide convergent evidence concerning the nature of dual character concepts. Experiment 1 investigated whether dual character concepts support two types of normative judgments ('good' and 'true') whereas the

control concepts support only one of these types of normative judgment ('good'). Experiments 2-4 explored the idea that dual character concepts support two different criteria for category membership. Finally, in Experiment 5, we used the theory developed here to construct artificial concepts and ask whether participants react to these artificial concepts with the same pattern of responses they show for naturally occurring dual character concepts.

Experiment 1

Experiment 1 tests the hypothesis that dual character concepts provide two bases for evaluation and thus 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

Participants. Twenty-three volunteers participated in the experiment over the Internet. Participants were chosen from Amazon's Mechanical Turk system for human intelligence tasks. All spoke English as their first language.

Stimuli. For each category, we generated one sentence of the form *That is a good x* and one sentence of the form *That is a true x*. Each statement was presented with a 7-point Likert scale with the ends labeled *sounds weird* and *sounds natural*.

Procedure. Each participant received a different random order of the 40 statements. Participants were instructed to rate the sentences as to how natural or weird they sounded. They were asked to make each rating independently and avoid relying on strategies for responding. Two practice trials preceded the experimental trials.

Results

2x2 ANOVAs with concept type (dual character, control) and statement type (good, true) as independent factors and participants' ratings as the dependent measure were performed. We report both participant (F_1) and item analyses (F_2). The mean ratings given in each condition are shown in Figure 2.

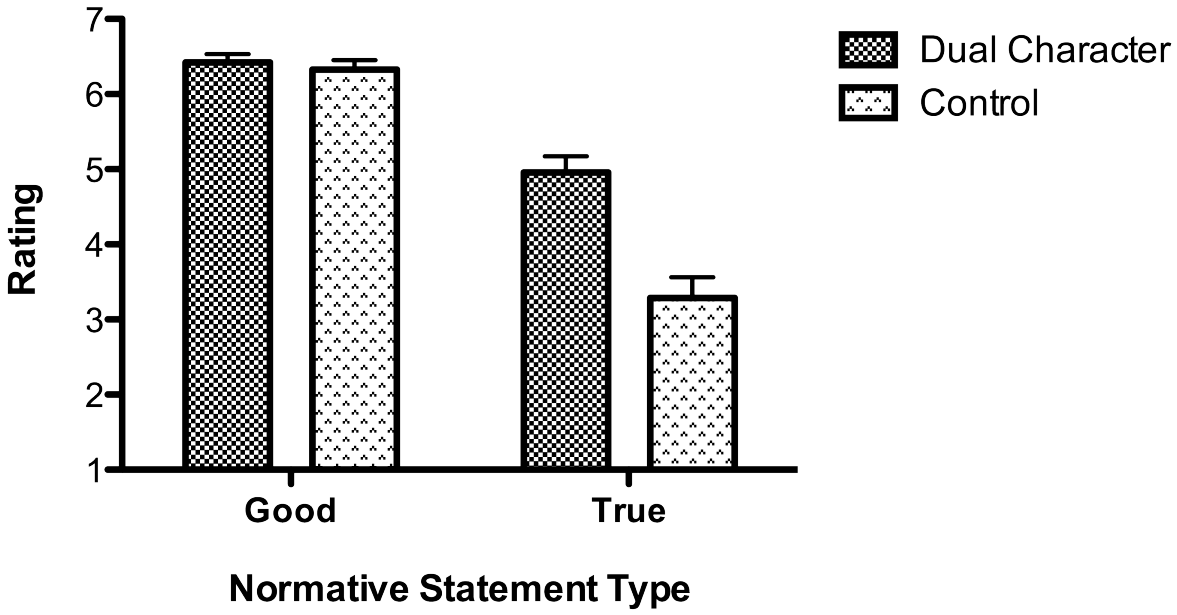

Figure 2. Mean ratings by condition for Experiment 1. (Error bars show SE mean.)

The key prediction was an interaction between concept type and statement type. Specifically, dual character (but not control) concepts were predicted to support normative statements concerning whether a given item may be a true member of the category or not, whereas both dual character and control concepts were predicted to support normative statements concerning how good a member of a category a given item is. As predicted, the interaction was significant (F_1(1,

22)= 74.87, p <.001; F_2 (1. 38) = 14.34, p <.001). Also as predicted, analyses of simple main effects showed that the interaction was due to significantly higher ratings for statements concerning whether an item is a true member of the dual character concepts than for control concepts (t_1 (22) = 9.5, p <.001; t_2 (19) = 5.04, p <.001), but no difference between the two types of concepts for statements concerning how good a member of a category a given item is.

Significant main effects were also found for concept type, with higher overall ratings for the dual character concepts than the control concepts (F_1(1, 22)= 83.30, p <.001; F_2 (1. 38) = 27.17, p <.001) and for sentence type, with overall higher ratings for the 'good' statements than the 'true' statements (F_1(1, 22)= 111.80, p <.001; F_2 (1. 18) = 117.24, p <.05).

Discussion

As predicted, participants gave high ratings to 'good' statements for both dual character concepts ('good scientist,') and control concepts ('good cashier'), but when it came to 'true' statements, participants gave high ratings 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 represent two distinct ways of characterizing members of a category, they would allow people to make two independent assessments of category membership. Participants received a series of vignettes in which an object was described as possessing the concrete properties characteristic of a category

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 two statements:

(i) There is a sense in which this person is an artist [pharmacist].

(ii) Ultimately when you think about what it really means to be an artist [pharmacist], you would have to say that this person is not truly an artist [pharmacist].

As discussed above, dual character concepts might be thought to resemble natural kind concepts in certain respects, and we therefore included a series of natural kind concepts in the experiment. For each natural kind concept, participants received a vignette adapted from Keil (1989). These vignettes described an object that displayed the superficial characteristics of a given category, but lacked the crucial underlying causal factors of that category. For example, one vignette described animals that looked and acted exactly like a raccoon but that had skunk insides, skunk parents and skunk children. After receiving these vignettes, participants were asked whether they agreed with the corresponding statements:

(i) There is a sense in which the animals are raccoons.

(ii) Ultimately when you think about what it really means to be a raccoon, you would have to say that these animals are not truly raccoons.

It was predicted that participants would show a complex pattern of judgments across the three types of concepts. 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. The dual character concepts, however, should involve an attention to both types of information, such that if an object has the concrete features but lacks the abstract values, participants will say that it can count as a category member in one sense (leading them to agree with statement (i)) while simultaneously not counting as a category member in another (leading them also to agree with statement (ii)).

Method

Participants. Thirty-one volunteers participated in the experiment over the Internet. Participants were chosen from Amazon's Mechanical Turk system for human intelligence tasks. All spoke English as their first language.

Stimuli. To limit the length of the task, we used the 10 dual character and 10 control concepts that had the highest and lowest ratings on the stimulus selection task. For each of these concepts, we constructed a vignette about an object that was described as possessing the concrete properties characteristic of a category but lacking the relevant normative properties. In addition, we included 10 vignettes of natural kind categories adapted from Keil (1989). These vignettes described things that had the concrete superficial characteristics of a given category, but lacked crucial underlying causal factors of that category. Examples of each type of vignette are given in the Appendix.

Procedure. Each participant received all 30 vignettes in a different random order. After each vignette, participants judged the truth of the following two statements concerning an object's category membership on a 7 point scale. (i) There is a sense in which this person is a scientist [pharmacist/raccoon]; (ii) Ultimately when you think about what it really means to be a

scientist [pharmacist/raccoon], you would have to say that this person is not truly a scientist [pharmacist/raccoon]. We will refer to these statements as the 'member statement' and the 'non-member statement.'

Results

3x2 ANOVAs with concept type (dual character, control, natural kind) and membership statement type (member, non-member) as independent factors and participants' ratings as the dependent measure were performed. The mean ratings given in each condition are shown in Figure 3. A significant main effect of concept type ($F_1(2, 60) = 6.60$, $p < .001$; $F_2(2, 27) = 3.56$, $p < .001$) and a significant interaction were found ($F_1(2, 58) = 83.52$, $p < .001$; $F_2(2, 27) = 42.63$, $p < .001$).

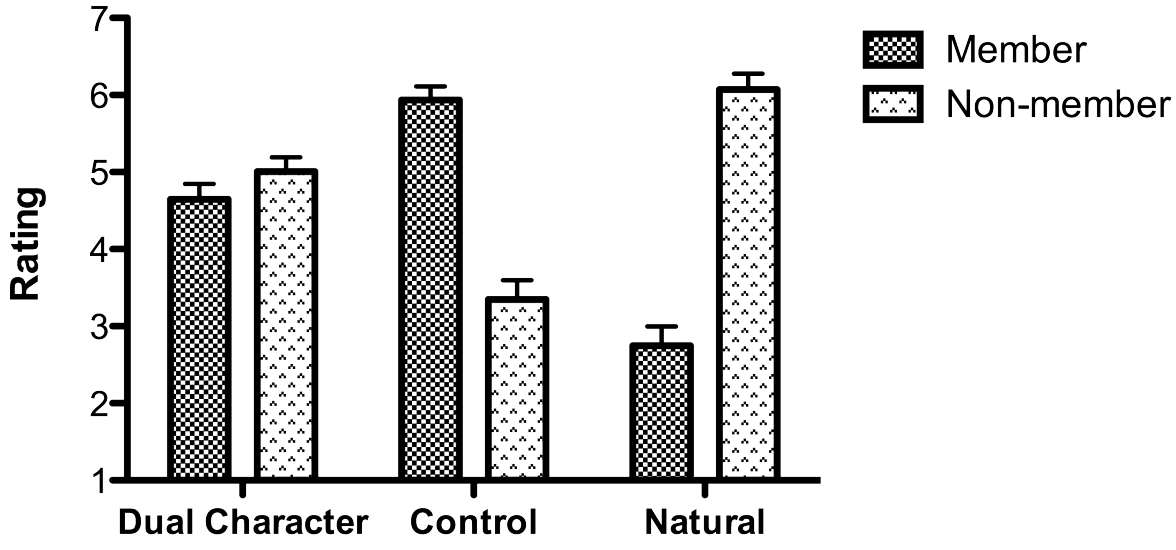

Figure 3. Mean ratings by condition for Experiment 2. (Error bars show SE mean.)

The key predictions were 2x2 interactions between concept type and statement type for dual character and control concepts as well as between dual character concepts and natural kind concepts. Both member and non-member statements were predicted to be judged to be true for dual character concepts, whereas only member statements were predicted to be judged to be true for the control concepts. As predicted, the interaction was significant ($F_1(1, 30) = 88.07$, $p < .001$; $F_2(1, 18) = 16.90$, $p < .001$). Analyses of simple main effects showed that the interaction was due to significantly higher ratings for member statements as compared to non-member statements for the control concepts (t_1 (30) = 6.71. $p < 001$; t_2 (9) = 5.80, $p < .001$), but no difference between the two statement types for the dual character concepts.

In comparing dual character concepts to natural kind concepts, we predicted that both member and non-member statements would be judged to be true for dual character concepts, whereas only non-member statements were predicted to be judged to be true for the natural kind concepts. As predicted, the interaction was significant ($F_1(1, 30) = 56.68$, $p < .001$; $F_2(1, 18) = 21.15$, $p < .001$). Analyses of simple main effects showed that the interaction was due to significantly higher ratings for non-member statements as compared to member statements for the natural kind concepts (t_1 (30) = 8.21, $p < 001$; t_2 (9) = 10.46, $p < .001$), but no difference between the two statement types for the dual character concepts.

Finally, the interaction between concept type and statement type for natural kind and control concepts was also significant ($F_1(1, 30) = 93.44$, $p < .001$; $F_2(1, 18) = 116.40$. $p < .001$). As predicted, analyses of simple main effects showed that the interaction was due to significantly higher ratings for non-member statements as compared to member statements for the natural kind concepts (t_1 (30) = 8.21, $p < 001$; t_2 (9) = 10.46, $p < .001$), but the opposite for the control concepts (t_1 (30) = 6.71, $p < 001$; t_2 (9) = 5.80, $p < .001$).

Discussion

The results of Experiment 2 suggest that people's intuitions regarding dual character concepts show a distinctive pattern that does not arise for concepts of other types. For control concepts and for natural kind concepts, participants appeared to be employing a single, unified set of criteria of category membership. (For control concepts, these criteria involved the concrete features, whereas for natural kind concepts, they involved the hidden causes.) By contrast, for dual character concepts, people appeared to employ two distinct sets of criteria. When a given object met one set of criteria but not the other, participants tended to say that it was a category member in one sense but was not a category member in another sense. As such, the experiment provided evidence that dual character concepts provide two bases for categorization.

One worry that one might have about these results is that the absolute levels of agreement for the statements about dual character concepts were not especially high (4.7 for the member statement, 5.1 for the non-member statement). To determine whether these means were an averaging artifact due to a bimodal distribution, we computed the percentage of responses to the dual character items that fell at each point along the 1-7 scale. As Figure 4 shows, there was no hint of bimodality: on both the member statement and the non-member statement, there were a large percentage of responses indicating agreement, along with a smaller percentage at each other point along the scale. It is possible that the slightly depressed ratings in this study are revealing something important about the structure of dual character concepts. For example, building on work by Kalish (1995, 2002), one might suggest that people's judgments about any given sentence will always be affected at least to some degree by both the concrete features and the abstract values. Alternatively, it might be that these slightly depressed ratings arose for

reasons that are relatively uninteresting at a broader theoretical level. After all, the means did not reach the top of the scale even for the control and natural kind items, and there may well have been a pragmatic effect whereby participants were reluctant to openly endorse two statements that appeared, at least on a superficial level, to be in contradiction with each other.

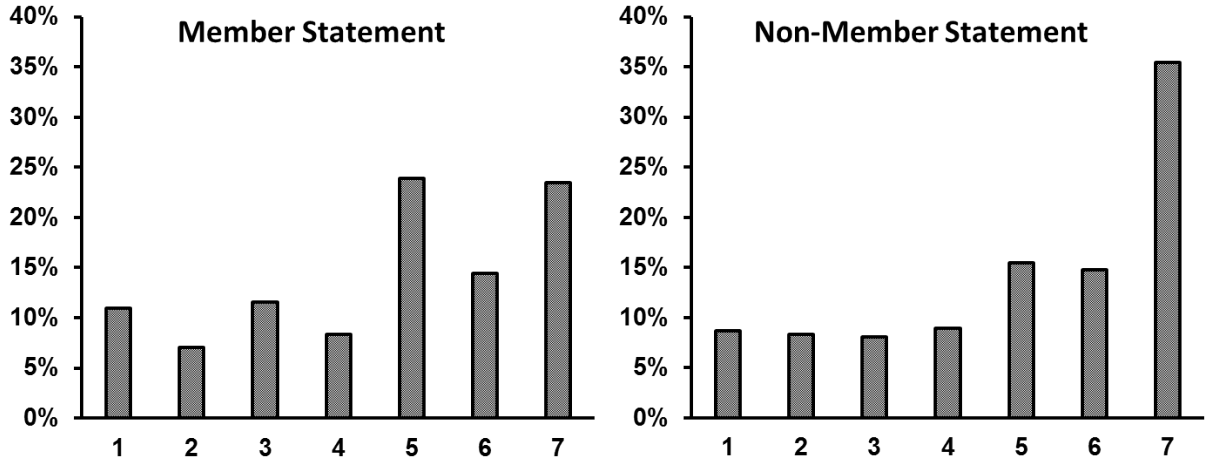

Figure 4. Percentage of responses at each level on a 1-7 scale for dual character statements in Experiment 2.

Another possible concern focuses on the intrinsic limitations of vignette studies (Strickland & Suben, in press). With a vignette study like this one, one might always wonder if the differences found between dual character and control concepts were due simply to an artifact of the vignettes themselves. That is, one might worry that there is actually no difference between dual character and control concepts per se and that the differences we found were merely due to differences in the vignettes that we constructed. (For example, perhaps the vignettes for the control concepts lacked robust enough descriptions of the relevant normative properties.) We sought to address this methodological concern by using a different methodology in Experiment 3.

Experiment 3

In Experiment 3, 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:

> 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.

This statement asserts that the object is ultimately not a category member, and we will therefore refer to it as the *ultimately non-member* statement. We predicted that such statements would sound fine for dual character concepts, but not for the control concepts.

The new methodology also afforded us the opportunity to test another prediction concerning dual character concepts. In Experiment 2, participants revealed that category membership could be granted on the basis of concrete characteristics and denied on the basis of the items lacking the relevant abstract normative values. 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 also examined participants' judgments about statements of the form:

> 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.

This latter statement asserts that the object ultimately is a category member, and we will refer to it as the *ultimately member* statement. We predicted that participants would judge statements of this sort to sound fine for dual character concepts, but not the control concepts.

Method

Participants. Thirty volunteers participated in the experiment over the Internet. Participants were chosen from Amazon's Mechanical Turk system for human intelligence tasks. All spoke English as their first language

Stimuli. We used the 10 dual character concepts and the 10 control concepts used in Experiments 1 and 2 to construct 'ultimately non-member' and 'ultimately member' statements for each concept.

Procedure. Participants were instructed to rate the extent to which the sentences sounded bad/sounded ok on a 1-7 scale. Each participant received all 40 items in a different random order.

Results

2x2 ANOVAs with concept type (dual character, control) and statement type (ultimately non-member, ultimately member) as independent factors and participants' ratings as the dependent measure were performed. The mean ratings given in each condition are shown in Figure 5.

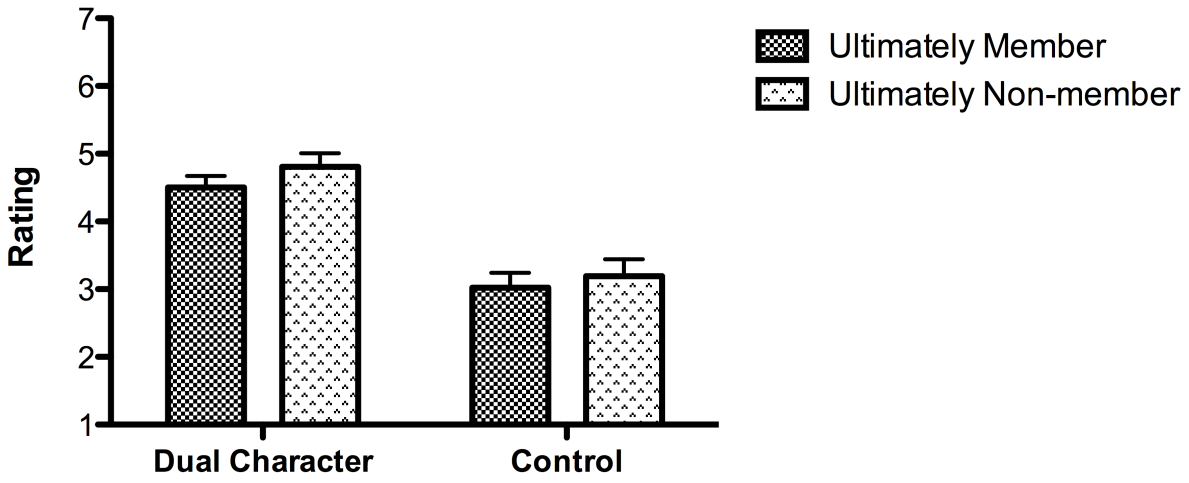

Figure 5. Mean ratings by condition for Experiment 3. (Error bars show SE mean.)

There was a main effect of concept type such that participants were more inclined to accept both ultimately member and ultimately non-member statements for dual character concepts than for the control concepts ($F_1(1, 23) = 45.55$, $p < .001$; $F_2(1, 38) = 71.22$, $p < .001$). There was also a small but significant main effect of statement type, with participants giving slightly higher ratings to the ultimately non-member statements than to the ultimately member statements ($F_1(1, 23) = 6.32$, $p < .02$; $F_2(1, 38) = 21.43$, $p < .001$).

Discussion

The results of this experiment suggest that the effect observed in Experiment 2 was not in any way an artifact of the specific vignettes used there. On the contrary, even when one omits the vignettes, participants are still more inclined to think that certain sentences sound right with dual character concepts than with control concepts. For example, participants leaned toward the view that it sounded right to say: 'There's a sense in which she is clearly an artist, but ultimately, if you

think about what it really means to be an artist, you'd have to say that there is a sense in which she is not an artist at all.' However, they did not think it sounded right to say: 'There's a sense in which she is clearly a bartender, but ultimately, if you think about what it really means to be a bartender, you'd have to say that there is a sense in which she is not a bartender at all.'

Moreover, participants were happier to accept the converse sort of sentence for dual character concepts. Thus, participants were inclined to say that it sounded right to say: 'There's a sense in which she is clearly not an artist, but ultimately, if you think about what it really means to be an artist, you'd have to say that there is a sense in which she is a true artist after all.' Here again, the sentence was only considered acceptable for dual character concepts but not for control concepts.

This last result suggests 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, so too it is possible to fulfill the abstract criteria without fulfilling the concrete ones.

Experiment 4

Because we hypothesized that dual character concepts allow people to simultaneously view a single object both as a category member as a non-member, our dependent measures in Experiments 2 and 3 included various qualifications (‘a sense in which,’ ‘ultimately,’ ‘clearly’). The results of these experiments provide evidence that qualified statements of simultaneous member and non-membership are possible for dual character concepts, but not natural kind or control concepts. However, they do not inform us about the roles of these two ways of

characterizing a category in ordinary unqualified categorization judgments (e.g., for a straightforward sentence like 'Greg is an artist.').

At the broadest level, it seems that there are two basic approaches one might take to answering this issue. One approach would be to suggest that people in some way *integrate* the different ways of characterizing the category, arriving in the end at a system of criteria that involves both concrete features and abstract values. The other would be to say that people *choose between* the different ways of characterizing the category, selecting in any given case either criteria based on concrete features or criteria based on abstract values. These two approaches make distinct predictions about the distribution of responses participants would make for dual character concepts. The feature selection approach predicts a bimodal distribution due to participants choosing to base their categorization judgments on one or another set of features, whereas the feature integration approach predicts no such bimodality.

Future work could examine this issue in more detail, but as an attempt to get some initial evidence concerning these possibilities we ran a modified replication of Experiment 2 in which we eliminated all qualifications ('a sense in which,' 'ultimately,' 'clearly') and simply asked participants to evaluate unqualified statements about whether a given object was a category member or a non-member.

Method

Participants. Forty people filled out a questionnaire through Amazon's Mechanical Turk.

Procedure. Each participant received all 30 vignettes from Experiment 2 in a different random order. After each vignette, participants were asked to evaluate the truth of an unqualified

statement. Participants were assigned either to the member condition or the non-member condition. Participants in the member condition were asked to judge the truth of unqualified statements of category membership (e.g., 'Greg is an artist'), while participants in the non-member condition were asked to judge the truth of unqualified statements of non-membership (e.g., 'Greg is not an artist'). Participants judged each statement on a 1-7 scale.

Results and Discussion.

Mean ratings for each condition are shown in Figure 6. To test for differences between these means, we conducted a 3 x 2 mixed-model ANOVA, with concept type (natural kind vs. control vs. dual character) as a within-subject factor and statement type (member vs. non-member) as a between-subject factor. There was no main effect of either concept type or statement type, but there was a significant interaction, $F(2, 38) = 97.9$, $p < .001$. Participants gave higher ratings to member than nonmember statements for natural kind concepts $t(38) = 16.9$, $p < .001$ and the opposite for control concepts $t(38) = 8.7$, $p < .001$. For dual character concepts, rating of the two statements did not differ, $t(38) = .1$, $p = .90$, and both were close to the midpoint of the scale.

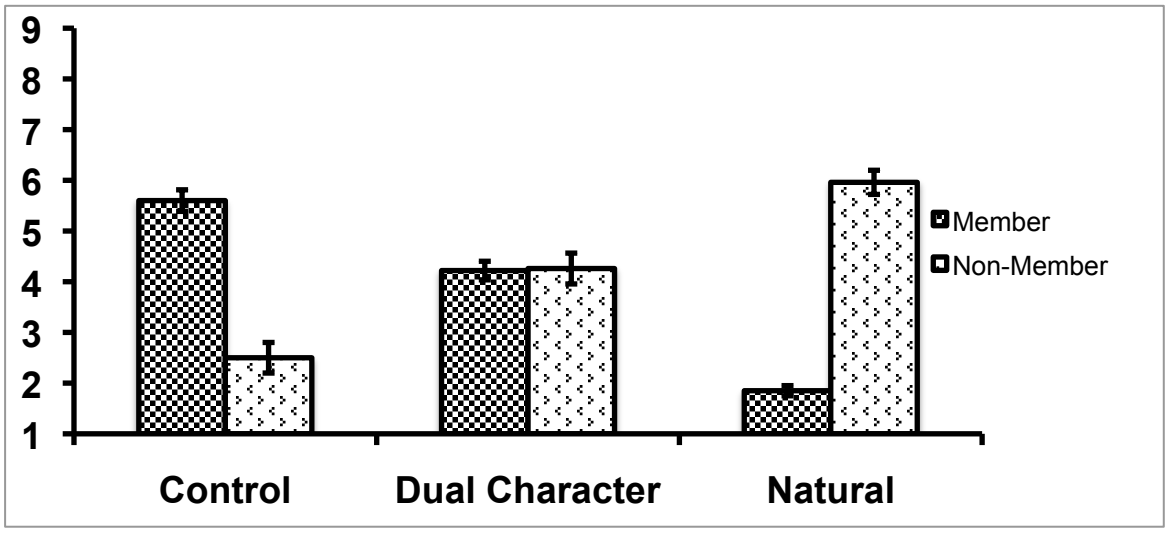

Figure 6. Mean ratings by condition for Experiment 4. (Error bars show SE mean.)

Our primary interest, however, was not in the differences between the means but in the distribution of people's responses for the dual character items. Accordingly, we computed the frequencies with which participants gave each of the possible responses on the 1-7 scale in their judgments on the individual dual character items. As Figure 7 shows, the distributions were bimodal: there were many judgments with high ratings, many with low ratings, but relatively few with intermediate ratings.

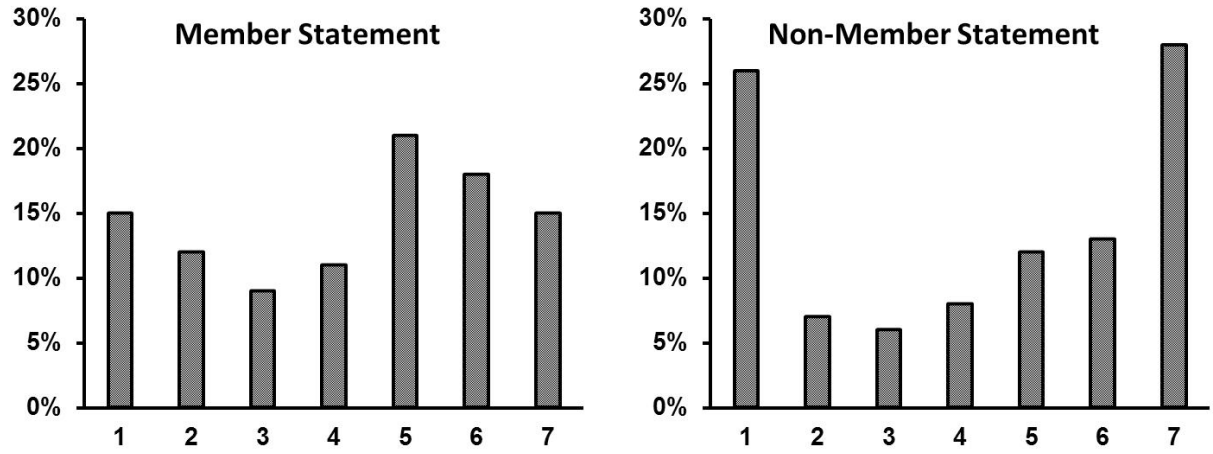

Figure 7. Percentage of responses at each level on a 1-7 scale for dual character statements in Experiment 4.

Note that the distributions found here were quite different from the ones found in Experiment 2. In Experiment 2, we included various qualifiers ('ultimately,' 'clearly,' etc.) that gave participants an indication of whether they were supposed to be using concrete features or abstract values, and the resulting distribution of responses was unimodal. By contrast, in the

present experiment, we used unqualified statements, giving participants no explicit indication of which criteria they should be using, and the distributions were bimodal.

This result provides some initial support for the view that people are making judgments about unqualified statements by *choosing between* the different possible criteria. It seems that people are not converging on a unified and integrated system of criteria that involves both concrete features and abstract values. Rather, on any given occasion, a judgment seems to be based primarily either on one system of features or on the other.

This finding then raises a host of further questions including: How do we choose which set of criteria to use in any given case? Does one set of criteria function as a default? If so, do all dual character concepts have the same default features for categorization? Can contextual information modulate the use of one or another set of criteria? These questions await future research.

Experiment 5

In this final experiment, we generated a series of artificial concepts. Some of the concepts were predicted to have dual character, while others were used as control concepts. The hypothesis was that participants would show the same pattern of judgments for these artificial concepts that they showed in earlier experiments for the naturally occurring concepts.

We introduced each artificial concept simply by providing a list of concrete features. However, some of these lists of features were constructed in such a way that all of the features on a given list would be seen as ways of realizing the same abstract value. We hypothesized that participants would spontaneously ascribe dual character to these concepts, understanding them both in terms of the list of concrete features (described explicitly) and in terms of the abstract

value (inferred from these features).

Method

Participants. One hundred undergraduate students at Yale University filled out a questionnaire packet in exchange for $5. This study was presented along with a series of unrelated surveys and always appeared first in the sequence.

Stimulus construction. We generated four artificial concepts. To ensure generality, concepts were drawn from two different domains (social categories and activities). Within each domain, one concept was hypothesized to show dual character, while the other was used as a control concept.

Each concept was described entirely in terms of a list of concrete features. For the dual character concepts, we generated lists of features that were designed to be seen as ways of realizing an abstract value. For example, the dual character concept in the social category domain was described with the following features:

> Swearing a solemn vow to never to retreat in battle, being the first to volunteer to retrieve necessary food and water, and building homes for people to live in.

For the control concepts, we generated lists of features that did not realize an abstract value. For example, the control concept in the social category domain was described with the features:

> Coming in early every morning to bake the muffins, being in charge of brewing the coffee, and refilling the items in the fruit and salad bar.

The two concepts in the activity domain used a similar structure, but they described activities instead of social categories. (The dual character concept involved a ritual of atonement; the control concept involved a sporting activity).

To verify that participants did indeed see the features of the dual character concepts more as displaying an abstract value, we conducted a stimulus verification study. Twenty-five

participants filled out a questionnaire on Amazon's Mechanical Turk. Each participant received all four lists of features in counterbalanced order. Following each feature list, participants were asked to rate their agreement with the statement: 'These characteristics display a more abstract value.' Statements were rated on a scale from 1 ('disagree') to 7 ('agree').

Results were analyzed using a 2 (concept type: dual-character vs. control) x 2 (domain: social category vs activity) repeated measures ANOVA. As expected, there was a main effect of concept type, F (1, 24) = 20.7, $p < .001$, such that participants showed higher agreement with the statement about abstract values for the dual character concepts (M = 4.6, SD = 1.4) than for the control concepts (M = 3.1, SD = 1.3). There was no main effect of domain and no interaction.

Procedure. Each participant received two concepts – one dual character concept and one control concept. The matching of concept types to domains was then varied, such that some participants received a dual character concept in the social category domain and a control concept in the activity domain, while others received a dual character concept in the activity domain and a control concept in the social category domain. Concepts were presented in counterbalanced order.

Each concept was introduced entirely in terms of the list of concrete features. Participants were then asked about an object that lacked all of these concrete features but which displayed a more abstract value. Thus, for the dual character concept in the person domain, they were told:

> Imagine there is a society in which certain people are known as dalimers. Dalimers are people who swear a solemn vow to never to retreat in battle, are the first to volunteer to retrieve necessary food and water, and who build homes for people to live in.
>
> Now imagine a person named John. In the society that John lives in, people have never heard of dalimers. Therefore, John has not sworn a vow to never retreat in battle, he does not go searching for food and water, and he does not build homes for people to live in. However, in nearly everything that he does, John does show a great love and concern for his community.

The control concept for the social category domain was then described in a parallel way. After

reading a description of this concept, participants were told about a person who lacked all of the relevant concrete features but who showed a deep interest in preparing and serving food to others.

After reading about each of these objects, participants were asked to rate their agreement with a *non-member statement* ('There is a sense in which John is clearly not a dalimer.') and a *member statement* ('Ultimately, however, there is a deeper sense in which John is a dalimer.'). Agreement with each statement was indicated on a scale from 1 to 9.

Results

Mean responses for each statement type within each concept type are presented in Figure 8. Based on the results of the previous studies, we predicted a significant interaction such that for dual-character concepts (relative to control) participants should be less likely to agree with the non-member statement and more likely to agree with the member statement. This prediction was supported by a 2 (concept type: dual character vs. control) x 2 (statement type: non-member vs. member) x 2 (block-type: person as dual character vs. activity as dual character) mixed-model ANOVA, which indicated a significant two-way interaction between concept type and statement type, $F(1, 98) = 25.9$, $p < .001$. To decompose this interaction, we conducted separate ANOVAs for each concept type. The non-member statements received significantly higher ratings than member statements for control concepts, $F(1, 98) = 64.1$, $p < .001$, but there was no difference in the ratings of the two types of statements for dual character concepts, $F(1, 98) = 2.4$, $p = .13$. Importantly, there was no three-way interaction with block type, indicating that this pattern held for both the social category and activity items.

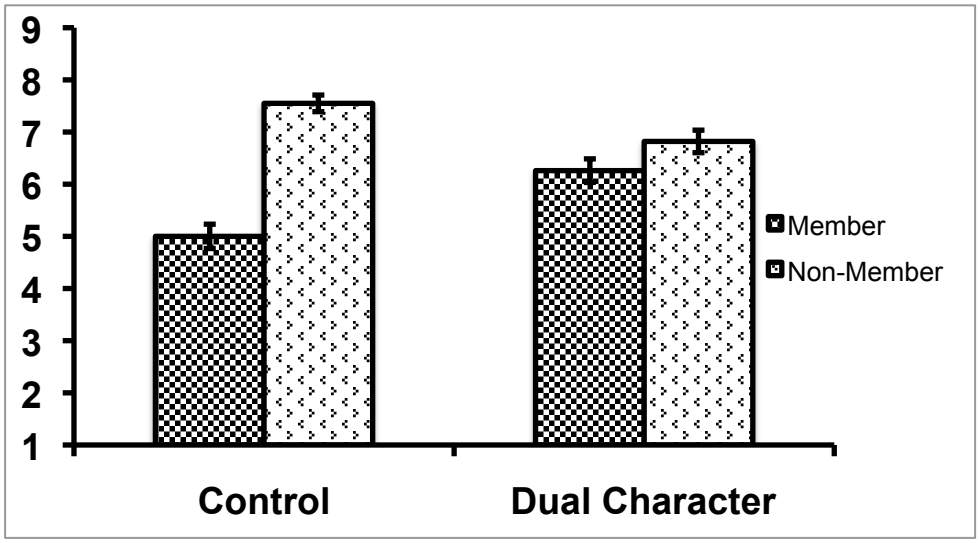

Figure 8. Mean ratings by concept type and statement type for Experiment 5. (Error bars show SE mean.)

We then looked separately at the non-member and member statements using a 2 x 2 mixed-model ANOVA, with concept type (dual character vs. control) as a within-subject factor and block-type (person as dual character vs. activity as dual character) as a between-subject factor.

For the non-member statement, there was a main effect of concept type, $F(1, 98) = 14.1$, $p < .001$, such that participants showed more agreement with the control concepts ($M = 7.6$, $SD = 1.6$) than for the dual character concepts ($M = 6.8$, $SD = 2.2$). There was no main effect of block-type and no significant interaction.

For the member statement, there was also main effect of concept type, $F(1, 98) = 25.8$, $p < .001$, such that participants showed more agreement for the dual character concepts ($M = 6.3$, $SD = 2.3$) than for the control concepts ($M = 5.0$, $SD = 2.3$). There was no main effect of match to domain and no significant interaction.

In sum, in both domains (social categories and activities) we observed that relative to

control concepts, exemplars that lacked all superficial features but retained the abstract value were rated as ‘non-members’ significantly less and ‘members’ significantly more.

Discussion

Participants were presented with a series of artificial concepts. These concepts were defined entirely in terms of lists of concrete features, but some of the lists of features were constructed in such a way that all of the features could be seen as ways of realizing the same abstract value. Participants responded to these artificial concepts in the same way they responded to naturally occurring dual character concepts. Specifically, even when an object lacked all of the concrete features that appeared in the original description of the concept, they were sometimes willing to say that (i) there was a sense in which the object clearly was not a category member but, at the same time, that (ii) ultimately, there was a deeper sense in which this object was a category member.

The fact that participants responded in this way to the artificial concepts provides reason to suspect that the naturally occurring concepts show the same basic structure. Each of these concepts might be represented in terms of a set of concrete features. However, people might note that all of the features associated with a given concept were ways of realizing the same abstract value. They might then conclude that the concept has dual character, being best understood both in terms of the concrete features and in terms of the more abstract value.

General Discussion

The studies reported here were designed to test the hypothesis that certain concepts provide two distinct systems of characterizing their instances: one based on concrete features, the

other based on more abstract values. This hypothesis was tested using both naturally occurring concepts and artificial concepts.

In the studies using naturally occurring concepts, we explored four different tests that distinguished dual character concepts from concepts of other types. Initially, we picked out dual character concepts by selecting concepts that were associated with 'values,' as revealed by judgments about whether it made sense to say:

(1) If you think that he is really an artist [pharmacist], I would have to say that you have some fundamentally wrong values.

Experiment 1 then showed that the concepts picked out by this test were the ones for which people thought it made sense to use the adjective 'true.'

(2) That is a true artist [pharmacist].

Experiments 2 and 3 showed that dual character concepts were also the ones for which people thought it made sense to say that an object could be a category member in a certain sense but not in a more 'ultimate' sense.

(3) There is a sense in which she is clearly an artist [pharmacist], but ultimately, if you think about what it really means to be an artist [pharmacist], you would have to say that she is not an artist [pharmacist] at all.

Experiment 3 showed that dual character concepts also allowed the converse claim, whereby there is a certain sense in which an object is not a category member but then a more 'ultimate' sense in which it actually is.

(4) There is a sense in which she is clearly not an artist [pharmacist], but ultimately, if you think about what it really means to be an artist [pharmacist], you would have to say that she truly is an artist [pharmacist].

Experiment 4 then provided some initial evidence that if participants need to make a simple unqualified categorization judgment, they do so by relying on one or the other set of criteria provided by dual character concepts rather than integrating the two set of criteria.

Finally, Experiment 5 turned to artificial concepts. Participants were introduced to a series of novel concepts, each defined in terms of a list of concrete features that all served to realize the same abstract values. The results showed that participants spontaneously associated these concepts with two distinct criteria for category membership (as revealed by their agreement with sentences like (4) above).

Taken together, these studies provide evidence for a distinctive class of dual character concepts that show three noteworthy properties: (a) each dual character concept provides two distinct ways of characterizing category members, (b) the two ways of characterizing category members provide two distinct bases for evaluation and categorization, and (c) one of these ways of characterizing category members involves abstract values. The remainder of this General Discussion examines these properties in further detail.

1. *Distinct criteria for categorization*

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. This pattern of judgments suggests that dual character concepts provide two different ways of assessing category membership. But how is this duality to be understood? To get a better sense for the answer, it might be helpful to look to comparisons with other phenomena and examine the ways in which the present case might be similar or different.

a. *Degrees of category membership*. In certain cases, people seem not to be using genuinely different criteria but rather to have a single criterion which they can then apply with different levels of stringency. For a simple example, take the concept *hot chocolate*. Faced with a beverage that resembles ordinary hot chocolate in many ways but also has a few unusual features, one might say: 'Loosely speaking, it is hot chocolate, but strictly speaking, it is not.' Yet the fact that people speak like this should not lead us to conclude that they have two entirely different criteria for falling under the concept. Rather, as a number of authors have noted, cases of this type might be well understood using the notion that category membership can come in degrees (Hampton, 1979; Lakoff, 1972; Rosch & Mervis, 1975). Even if we assume that people have only a single basic set of criteria for category membership, we can easily see how there might be cases in which an object does fit the criteria to a certain degree but does not fit the criteria to a higher degree.

Although this notion of degrees of category membership is an interesting and important one, it seems that something more complex is at work in the cases under investigation here. The key thing to notice is the double dissociation. As Experiments 2 and 3 show, there are cases in which participants think that a person is clearly an artist but is not a 'true artist (as in sentence (3)), but there are also cases of the converse type, where participants say that a person is a 'true artist' but that there is clearly a sense in which this person is not an artist at all (as in sentence (4)). The overall pattern of data therefore suggests that a phrase like 'true artist' does not simply pick out people who fulfill the criteria of the concept to an especially high degree. Instead, there appear to be two distinct criteria here, such that a given object can fulfill either one without fulfilling the other.

b. *Heterogeneity.* Turning now to the opposite extreme, defenders of the 'heterogeneity hypothesis' (Machery & Seppälä, 2010; cf. Weiskopf 2009) suggest that categories can actually be associated with two separate conceptual representations. These two representations would be entirely distinct. They would be stored separately, would make use of different representational formats, and people could apply one of them without using the other in any way. (For example, a single category might be associated with a prototype and then, separately, with a set of exemplars.)

Although this hypothesis may provide the correct explanation of certain phenomena, it appears that dual character concepts need not involve completely separate representations in the sense under discussion here. Note, for example, the result obtained in the artificial concepts study. There, participants were given only one of the two criteria (the concrete features) and then spontaneously generated the other one (the abstract values). This result suggests that people do not need to acquire the two criteria independently. Instead, one of the criteria can be derived from the other.

c. *A middle path.* The results obtained here point to a view that steers a middle path between the two extremes. On one hand, it appears that people do associate each dual character concept with two different sets of criteria. But on the other, it seems that these two sets of criteria are not entirely distinct. Instead, the data suggest a view on which people do have two sets of criteria but on which these two sets are nonetheless systematically connected.

Specifically, we propose that the two criteria for each dual character concept can both be derived from the very same set of concrete features. For one of the criteria, people simply check to see whether a given object actually has the concrete features themselves. For the other, they

identify the abstract values that these concrete features serve to realize and then to check to see whether the object displays these abstract values. In short, dual character concepts permit two different criteria for category membership, but these distinct criteria are intimately linked. Typically, these criteria function together and support one another, but if faced with situations in which the criteria come apart, the data from Experiment 4 suggest that we choose one set of criteria to base our categorization decisions rather than integrate the two set of criteria.

2. *Abstract values*

A number of existing studies have demonstrated that normative evaluations can impact intuitions about prototypicality. In particular, it appears that objects are seen as more prototypical to the degree that they approach the ideal for a given category (Barsalou, 1985; Macnamara, 1990; Lynch, Coley & Medin, 2000; Bailenson, Shum, Atran, Medin & Coley, 2002; Burnett, Medin, Ross & Blok, 2005; Goldwater, Markman, & Stilwell, 2011). The present studies do not in any way call into question the conclusions derived from this earlier work, but they do suggest that existing models should be supplemented by the addition of a distinct form of normative evaluation which involves what we have called 'abstract values.'

First, the results of Experiment 1 provide evidence for the existence of two different forms of normative evaluation. Participants seemed willing to apply the word 'good' to a broad array of different concepts ('good soldier,' 'good optician,' 'good table of contents'). When it came to the word 'true,' however, their application was more tightly constrained. Participants did not think this word could be applied to our control concepts ('true optician'? 'true table of contents'?), but they did think it could be applied to dual character concepts ('true soldier'). This result indicates that dual character concepts permit two distinct forms of normative evaluation,

one that can also be applied to a wide variety of other concepts (e.g., 'good soldier') and one that is more specific to dual character concepts in particular (e.g., 'true soldier').

Experiment 5 then shows that this second form of normative evaluation plays a role in judgments of category membership. Participants in that study were introduced to the concept DALIMER. They were told that dalimers engage in three concrete activities: swearing a solemn vow never to retreat in battle, being the first to retrieve necessary food and water, and building homes for people to live in. Now suppose we imagine a person who was highly effective in all of these activities (an excellent soldier, a skilled home-builder, etc.). One might say that such a person was 'good at being a Dalimer,' and existing theories suggest that he or she would be seen as a prototypical category member (e.g., Barsalou, 1985). However, in the actual experiment, participants were instead told about a person who did not engage in any of these activities but who nonetheless embodied a particular abstract value – namely, love and concern for his community. Just as existing theories would predict, participants tended to say that there was clearly a sense in which this person was not a dalimer at all. Yet participants also tended to make a judgment that went in the opposite direction, saying that, ultimately, there was a deeper sense in which the person actually was a dalimer. This result suggests a distinctive role for abstract values in judgments of category membership for dual character concepts.

3. *Comparison with natural kind concepts*

A question now arises about the relationship between dual character concepts and natural kind concepts. On one hand, there are a number of important respects in which the two types of concepts are strikingly similar. In both cases, people show a willingness to go beyond concrete observable features, and in both cases, they seem to be understanding categories in more abstract

theoretical terms (see Keil, 1989; Gelman, 2003, Rips, 1989). Yet, on the other hand, there are also a number of respects in which the two types of concepts are quite different. (Natural kind concepts are characterized in terms of hidden causes, while dual character concepts are characterized in terms of abstract values.) The question now is whether it is possible, despite these obvious differences, to see natural kind concepts and dual character concepts as different forms of some fundamentally unified phenomenon.

On intriguing possibility as to how these two types of concepts may be unified is the following. Consider the natural kind concept SKUNK. People associate this concept with certain superficial features (stripes, smelliness, etc.), but they do not simply treat the category as being defined by this list of features. Instead, a further question arises: 'Why is it that all of these different features have been grouped together and associated with the same category?' The causal essence then provides an answer to this question (Bloom, 2000; Gelman, 2003; Medin & Ortony, 1989). The answer is that these features are united by the fact that they all share the same underlying causes.

Turning now to dual character concepts, we find the same structure at work. People associate the dual character concept ROCK MUSIC with a collection of features, but they then face a further question about why the category is associated with those specific features and not others. Once again, the criteria governing the concept offer people an answer to this question. The difference is just that, this time, the answer is not that all of the features share the same underlying causes but rather that they all embody the same abstract values.

Ultimately, then, it might be possible to understand dual character concepts by generalizing some of the insights that were first introduced in the literature on natural kinds. One of the key insights there was that people's conceptual representations can be shaped by certain

'theories' they hold about the world (Carey, 1985; Diesendruck & Gelman, 1999; Keil, 1989; Gopnik & Melzoff, 1987). We may now need to generalize that claim so that it includes not only scientific theories (about hidden causes) but also normative theories (about abstract values). In the case of natural kind concepts, the features would be unified by a theory about hidden causes, whereas in the case of dual character concepts, the features would be unified by a theory about abstract values. But other than this one difference, it might be that the two kinds of concepts show the same basic structure – though as the experiments in the present paper demonstrate, this one difference can have important functional consequences.

This generalization may at first seem surprising, but it does seem to be in keeping with a more general trend within recent research. It has long been known that causal judgments play an important role in people's use of various concepts (e.g., Ahn, 1998; Sloman, Love, & Ahn, 1998), but a surge of recent research has been pointing to the various ways in which normative judgments also play an important role (e.g., Knobe, 2010; Prasada & Dillingham, 2006, 2009). The present suggestion can be seen as pointing to one further respect in which normative considerations actually figure in judgments that might initially have appeared to be entirely non-normative.

Perhaps it will be possible to take this approach even farther. We have seen that some concepts are unified through hidden causes (natural kind concepts) and others through abstract values (dual character concepts), but perhaps these are just two of the many possibilities, and there are also yet other kinds of concepts that are unified in quite different ways. For example, there might be concepts in which all of the concrete features are unified in that they all tend to make an object suitable for the same basic function (e.g., the concept COMPUTER). People might then associate these concepts with both (a) a list of concrete features and (b) the more abstract

notion of the relevant function. (If so, such concepts would be like the dual character concepts studied here in that they would provide two bases for categorization, but they would be unlike dual character concepts in that they would not provide two bases for normative judgment.)

Since the present experiments were concerned primarily with the role of abstract values, they cannot themselves allow us to evaluate a broader theory along these lines. Evidence for such a theory would require further research.

4. *Concepts or just words?*

One might wonder if the evidence presented in support of dual character concepts is actually evidence that the words that name these concepts are polysemous such that they possess a descriptive sense and a normative sense. Questions concerning whether a word is polysemous, the form of polysemy it displays, and how polysemy is represented are complex questions (e.g., Beretta, Fiorentino & Poeppel, 2005; Murphy, 2007; Rabagliati, Marcus & Pylkkanen, 2011; Srinivasan & Snedeker, 2011). It seems clear, however, that if words like 'scientist' indeed are polysemous, the explanation for *why* they are so, while words like 'bartender' and 'skunk' are not must appeal to the different ways in which we conceive these categories. Furthermore, the form of the putative polysemy as involving a descriptive sense and a normative sense (rather than some other form of polysemy) is readily explained by the fact that dual character concepts characterize category members in terms of both concrete features and abstract values. As such, the data in the present paper clearly reveal differences in the ways in which dual character concepts differ from standard and natural kinds concepts, whether or not they also reveal differences in the ways in which the names for the concepts are represented.

5. *Implications for theories of conceptual representation*

One key task for future work is to integrate this research on dual character concepts within a more general account of conceptual representation. As we noted at the outset, existing research has led to the development of a number of quite general theories of conceptual representation – theories that are concerned not merely with this or that particular type of concept, but rather with the way in which concepts in general are represented (e.g. Bruner et al., 1956; Gelman & Wellman, 1991; Keil, 1989; Medin & Shaffer, 1978; Rosch, 1975). A key task for future work on dual character concepts will be to find a way of integrating them into a broader theory of this sort, understanding the phenomena of dual character concepts within the context of a more general theory of conceptual representation.

One important question that arises within a number of existing theoretical frameworks concerns the nature of conceptual coherence (e.g., Murphy & Medin, 1985). A single concept may be associated with a number of different features, but these different features do not appear to be merely an arbitrary list; instead, the different features associated with the same concept appear to be related to each other in some important way. In some cases – as in natural kind concepts – it may be said that the features are united with each other through relationships of causation (Keil, 1989, Gelman & Wellman, 1991; Gopnik & Meltzoff, 1987; Rips, 2001; Sloman, Lombrozo & Malt, 2007; Rehder, 2003). In other cases – as in some of the control concepts used in the present studies – this claim about causation begins to seem less plausible. Still, existing frameworks offer various suggestions about how these concepts might be unified. Some suggest that this unification arises because the different features are statistically correlated with each other (Rogers & McClelland, 2004; Tyler & Moss, 2001; Yoshida & Smith, 2003a,b); others emphasize a formal part-whole relation whereby the different features are represented as

aspects of being a given kind of thing (Prasada & Dillingham, 2009). The present experiments point to a coherence of a somewhat different type. In dual character concepts, the various concrete features cohere because they are all ways of realizing the same abstract values.

At the very least, any correct general theory will have to have some way of accommodating these phenomena. Ideally, however, one would hope for more. One wants a theory that can actually *explain* or *predict* the experimental results reported here. In other words, one wants a theory that helps us to understand the ways in which conceptual features can be connected through causation, correlation or more formal relations but that also directly predicts that under certain circumstances, people will acquire concepts with dual character – concepts that are characterized not only in terms of concrete features but also, at the same time, in terms of more abstract values.

References

Ahn, W. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition*, 69, 135-178.

Aristotle. (1999). *Nicomachean Ethics*. trans. Terence Irwin. Indianapolis: Hackett. First published 350 BC.

Bailenson, J. N., Shum, M. S., Atran, S., Medin, D. L., & Coley, J. D. (2002). A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition, 84,* 1-53.

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.

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: An MEG study. *Cognitive Brain Research*, 24, 57–65.

Bloom, P. (2000). *How children learn the meanings of words.* Cambridge, MA: MIT Press.

Bruner, J., Goodnow, J., & Austin, G. (1956*). A study of thinking*. New Brunswick, NJ: Transaction Publishers.

Burnett, R.C., Medin, D.L., Ross, N.O., Blok, S.V. (2005). Ideal is typical. *Canadian Journal of Experimental Psychology*, *59*, 3-10.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Diesendruck, G., & Gelman, S. A. (1999). Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. *Psychonomic Bulletin and Review*, *6*, 338-346.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: number and the animate-inanimate distinction as examples. *Cognitive Science, 14*, 79-106.

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

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

Goldwater, M. B., Markman, A. M., & Stilwell, C. H. (2011). The empirical case for role-governed categories. *Cognition, 118*, 359-376.

Gopnik, A., & Meltzoff, A.N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18*, 441-461.

Hampton, J.A. (1998). Similarity-based categorization and fuzziness of natural categories, *Cognition*, 65, 137-165.

Hampton, J. A., Estes, Z., & Simmons, S. G. (2007). Metamorphosis: Essence, Appearance and Behavior in the Categorization of Natural Kinds. *Memory & Cognition*, 35, 1785-1800.

Kalish, C. (1995). Essentialism and graded membership in animal and artifact categories, *Memory & Cognition*, 23, 335-353.

Kalish, C. (2002). Essentialist to some degree: Beliefs about the structure of natural kind categories. *Memory & Cognition*, 30, 340-352.

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

Kim, S.W. & Murphy, G. (in press). Ideals and category typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, *33*, 315-329.

Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic, 2*, 458-508.

Leslie, S-J. (in press). 'Hillary Clinton is the only man in the Obama administration': Dual character concepts, generics, and gender. *Analytic Philosophy, Special Issue on Slurs*.

Lombrozo, T. (2009). Explanation and categorization: how "why?" informs "what?". *Cognition, 110,* 248-253.

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.

Machery, E., & Seppälä, S. (2010). Against hybrid theories of concepts. *Anthropology & Philosophy*, *10*, 97-125.

Malt, B.C. (1990). Features and beliefs in the mental representation of categories. *Journal of Memory and Language*, *29*, 289-315.

Manamara, J. (1990). Ideals and psychology. *Canadian Psychology/Psychologie Canadienne*, *31*, 14-25.

Medin, D. L., & Shaffer, M.M.. (1988). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Medin, D., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony, *Similarity and analogical reasoning*. Cambridge, UK: Cambridge University Press.

Murphy, G.L. & Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289-316.

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

Murphy, G. L. (2007). Parsimony and the psychological representation of polysemous words. In M. Rakova, G. Petho, & C. Rákosi (Eds.), *The cognitive basis of polysemy*. Frankfurt am Main, Germany: Peter Lang Verlag.

Newman, G. & Keil, F.C. (2008). Where's the essence? Developmental shifts in children's beliefs about internal features. *Child Development*, 79, 1344-1356.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 700-708.

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

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

Rabagliata, H., Marcus, G. F., & Pylkkanen, L. (2011). Rules, radical pragmatics, and restrictions on regular polysemy. *Journal of Semantics, 28*, 485-512.

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science, 27*, 709-748.

Rips, L. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony, *Similarity and analogical reasoning*. Cambridge, UK: Cambridge University Press.

Rips, L. (2001). Necessity and natural categories. *Psychological Bulletin, 127*, 827-852.

Rogers, T.T. & McClelland, J.L. (2004). *Semantic cognition: A parallel distributed approach*. Cambridge, MA: MIT Press.

Rosch, E. R. & Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.

Rosen, G., (2012). Abstract objects. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/spr2012/entries/abstract-objects/

Sloman, S. A., Lombrozo, T., & Malt, B. C. (2007). Mild ontology and domain-specific categorization. In M. J. Roberts (Ed.). *Integrating the mind.* Hove, UK: Psychology Press.

Sloman, S. A., Love, B. C., & Ahn, W. K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, *22*, 189-228.

Smith, E. E., Patalano, A. L. & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, *65*, 167-196.

Srinivasan, M., & Snedeker, J. (2011). Judging a book by its cover and its contents: The representation of polysemous and homophonous meanings in four-year-old children. *Cognitive Psychology, 62*, 245-272.

Strickland, B. & Suben, A. (in press). Experimenter philosophy: The problem of experimenter bias in experimental philosophy. *Review of Philosophy and Psychology*.

Tyler, L.K., & Moss, H.E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences*, *5*, 244-252.

Uttich, K. & Lombrozo, T. (2010). Norms inform mental state ascriptions: a rational explanation for the side-effect effect. *Cognition, 116,* 87-100

Weiskopf, D. (2009). The plurality of concepts. *Synthese*, *169*, 145–173

Yoshida, H. & Smith, L. B. (2003a). Shifting ontological boundaries: How Japanese- and English-speaking children generalize names for animals and artifacts. *Developmental Science*, *6*, 1-17.

Yoshida, H. & Smith, L. B. (2003b). Correlations, concepts and cross-linguistic differences. *Developmental Science*, *6*, 30-34.

Appendix

Dual character and control concepts

(Within each category, concepts are listed in order by the score they received in the stimulus construction study.)

Dual character concepts. Friend, Criminal, Love, Mentor, Comedian, Minister, Theory, Boyfriend, Artist, Argument, Teacher, Poem, Soldier, Sculpture, Art Museum, Musician, Mother, Rock Music, Scientist, Novel.

Control concepts. Mechanic, Optician, Baker, Blog, Doorman, Mayor, Waitress, Caseworker, Table of Contents, Tailor, Bartender, Rustling, Welder, Catalog, Chair, Firefighter, Uncle, Cashier, Stroller, Obituary, Second Cousin

Sample vignettes for Experiment 2

Scientist. George is employed at Ameritech to run experimental studies and analyze the data. However, he actually has no interest at all in finding the correct answers to the questions he is studying. So although he goes through the motions, he does not actually care in any way about making a contribution to people's understanding of these issues.

Rock Music. The new song 'Born to Rebel' features screaming vocals and electric guitars. However, the song was actually created by a marketing firm that was putting together an advertisement designed for elderly people who are interested in imitating youth culture, and serious music fans always say that it has no real energy or feeling.

Mother. Peggy is a famous celebrity with two young children, whom she is always in the midst of feeding, clothing or otherwise pampering. However, it turns out that Peggy does not have any real feelings for the children and is only taking care of them because she is concerned about publicity and wants the media to portray her as a caring and compassionate person.

Pharmacist. Laura has spent the last 10 years working at the local pharmacy. At work, she wears a white coat and fills medical prescriptions for her customers. She explains to customers how much medicine to take and when to do so. Furthermore, she warns patients about potentially dangerous interactions between drugs. Laura has no interest in medicine or helping people get well, but she likes the pay and benefits of her profession and wants to make sure she doesn't lose her job.

Table of Contents. Laurie has been put in charge of writing out the list of chapters at the beginning of a book. Her lists are usually serviceable, but she has no real passion for the task; it is just something she had to do as a summer job. So the list of chapters ends up looking like a mess, with lots of needless font changes and incorrect line spacing. The people who actually do this for a living all agree that Laurie hasn't gotten the real point of what it is all about.

Second Cousin. Harry and Janet are two American teenagers. It turns out that they are actually related. Harry's grandmother had a brother, who is Janet's grandfather. However, Harry and Janet never spend any time with each other and don't have any warm feelings for each other. In fact, they would have a little bit of difficulty picking each other out in a crowd.

Zebra. Jill loved going to the local zoo. Her favorite part of the zoo was the zebra enclosure. Since it was a small zoo, the enclosure had only one inhabitant. It had beautiful black and white stripes which Jill found mesmerizing. One day when the zebra got sick, the doctor began running tests on it and found that its DNA was unlike that of any previously studied zebras. Instead, the DNA was identical to that of a breed of donkeys.