

Water is and is not H₂O

[Kevin Tobia](#)

Yale University

[George Newman](#)

Yale University

[Joshua Knobe](#)

Yale University

The “Twin Earth” philosophical thought experiment has importantly influenced the study of psychological essentialism. The standard philosophical intuition about that thought experiment suggests that it is an entity’s deeper causal properties—and not its superficial features—that are criterial for categorization. Four studies suggest that people do not share this intuition. Instead, people have two distinct criteria for category membership, one based on superficial features and one based on deeper causal properties. Studies 1a and 1b show that people reject the standard Twin Earth intuition, instead endorsing two (opposing) criteria for category membership. Study 2 shows that contextual cues affect categorization of entities in Twin Earth cases. Studies 3a and 3b extend these findings by looking both to a real-world case involving genetically modified organisms and a population of graduate students from elite universities. Together, these studies provide an enriched understanding of essentialized concepts. People do not endorse the standard Twin Earth intuition, categorizing entities solely on the basis of their deep causal properties; instead, people employ two sets of criteria in natural kind categorization.

Keywords: Concepts, essence, essentialism, natural kinds

1. Introduction

Imagine a liquid that is identical to H₂O in terms of its *superficial features*. It has the same color, texture, and taste of H₂O and it quenches thirst in just the same way. However, this liquid is completely different in terms of its *deeper causal properties*. When chemists examine it, they find that it is not composed of H₂O but of some very different chemical compound. Is this liquid water?

This “Twin Earth” thought experiment was first introduced in the philosophical literature. Within that literature, the standard answer is *no*, the entity is not water (e.g. Putnam, 1975; Kripke, 1980). In other words, the standard philosophical view is that when it comes to natural kind concepts like this one, it is not the superficial features but the deeper causal properties that are criterial for category membership.

Cognitive scientists have since drawn on this style of thought experiment to explore certain aspects of *psychological essentialism* (Carey, 1985; Gelman, 2003; Keil, 1989; Wellman & Gelman, 1988). To the extent that theories of psychological essentialism are designed to capture the standard philosophical intuition, these theories should say that when people essentialize a category, they regard the deeper causal properties as criterial for category membership.

In the present studies, we show that when reasoning about such cases, people do not endorse the standard philosophical intuition. Instead, people assent to two distinct claims:

- (1) There is a sense in which the liquid is water.
- (2) Ultimately, if you think about what it really means to be water, you’d have to say there’s a sense in which the liquid is not truly water at all.

In light of this finding, we suggest that theories of psychological essentialism should not be designed to capture the standard philosophical intuition. Instead, the data provide evidence for a more complex account according to which people associate natural kinds with two different sets of criteria. One set of criteria is based on superficial features, the other on deeper causal properties. The complex and ambivalent reaction people have to Twin Earth cases arises from a conflict between these two sets of criteria.

1.1 *Psychological Essentialism*

Psychological essentialism is perhaps best conceived of as a tendency to associate concepts with a particular type of representation. When people essentialize a concept, they associate it with two distinct sets of properties: (1) a variety of observable features and (2) a deeper, unobserved essence.

Consider the concept WATER. This concept is associated with certain observable features (colorless, tasteless, drinkable, etc.), and people could, in principle, represent the concept WATER solely in terms of these features. However, a vast body of evidence suggests that adults and even young children represent concepts like this one not only in terms of superficial features but also in terms of certain deeper causal properties (Gelman, 2003; Keil, 1989; Medin & Ortony, 1989). Thus, adults and even young children seem to represent the concept WATER not only in terms of its superficial features (e.g. color, taste) but also in terms of a deeper property that causes or otherwise explains

these superficial features (e.g., the chemical structure H₂O). This deeper causal property can be understood as the *essence* of the kind.

Research in psychological essentialism has explored people's representations of superficial features, deeper causal properties, and the relationship between two (e.g., Ahn, 2001; Kalish, 1995; Keil, 1989; Murphy & Medin, 1985; Rehder & Hastie, 2001; Rheder & Burnett, 2005; Rips, 1989; Sloman, Love, & Ahn, 1998; but see Strevens, 2000). These representations have been explored in development (Hall et al., 2003; Hirschfeld, 1995; Hirschfeld, 1996; Diesendruck, 2001; Gelman & Markman, 1987; Gelman & Wellman, 1991; Kalish & Gelman, 1992; Newman & Keil, 2008), across a number of different cultures (Inagaki & Hatano, 2002; Waxman, Medin & Ross, 2007), and for a wide-range of concepts including biological and naturally-occurring categories, social categories (Haslam et al., 2000; Haslam, Bastian & Bissett, 2004; Keller, 2005; Bastian & Haslam, 2006; Haslam & Levy, 2006; Rhodes, Leslie & Tworek, 2012), representations of individuals (Schlegel & Hicks, 2011; Schlegel et al., 2013), and even mental entities (Haslam 2000; Haslam & Ernst 2002).

1.2 *The One-Criterion View of Essentialism*

Although essentialism is wide-ranging in its effects, we focus in particular on its impact on categorization judgments. Given that people associate natural kind concepts with both superficial features and deeper causal properties, what representation do they use in determining whether a given entity actually counts as a member of a kind?

One view would be that people's criteria for membership in natural kinds are based *solely* on deeper causal properties. On this view, when people are trying to determine whether a liquid counts as water, their judgments are based entirely on its underlying causal structure (e.g., H₂O). We refer to this hypothesis as the *one-criterion view*.

Existing work consistently finds that deeper causal properties do play a role in categorization judgments (e.g., Keil, 1989; Gelman, 2003). This is a striking phenomenon in itself, but the one-criterion view involves a further claim. It suggests not only that the deeper causal properties play a role in people's criteria for category membership but also that the superficial features (e.g., for water, being colorless, tasteless, potable, etc.) do not play a role.

The standard philosophical intuition in Twin Earth cases seems to provide strong evidence for the one-criterion view. The entity in the thought experiment does not have the deeper causal properties associated with water, but it does have all of the superficial features. Thus, if people do in fact have the intuition that this entity is not water, this intuition would provide strong reason to adopt the one-criterion view.

Strikingly, however, existing experimental work suggests that people do not actually have the standard philosophical intuition about Twin Earth cases. Malt (1994) demonstrates that the presence of H₂O in a liquid is not the only feature that plays a role in categorization as water. For example, compare Malt's results for tea (judged 91% H₂O, but presumably not "water") and salt water (83% H₂O, but presumably "water") (but see Abbott 1997; Ahn et al. 2000). Further, in unpublished data Corcoran (2016) presented participants with a series of Twin Earth cases. In each case, an entity was described as having all of the superficial features associated with a natural kind, but being

fundamentally different in its deeper causal properties. Participants were asked to agree or disagree with a statement about category membership (e.g. “The liquid is water”). Mean responses fell towards the middle of the rating scale, indicating that people did not have a strong opinion either way as to whether the entity was a category member. This result appears to provide further evidence against the one-criterion view.

1.3 The Dual-Character View of Essentialism

An alternative to the one-criterion view is what he will call the *dual character view*. On this view, each natural kind concept is associated with two different sets of criteria. One set of criteria is based on superficial features; the other is based on deeper causal properties. Then people can make categorization judgments using either set of criteria. Thus, if a single entity fulfills one set of criteria but not the other, people should have the intuition that the entity is a category member in one sense but is not a category member in another sense.

This hypothesis may be clarified by analogy to certain sorts of value-based concepts (Knobe, Prasada & Newman 2013). For example, consider 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. When evaluating such a person, participants agree that both:

1. There is a sense in which this person is an artist.
2. Ultimately when you think about what it really means to be an artist, you would have to say that this person is not truly an artist.

This result suggests that the concept ARTIST is actually associated with two different criteria. It is associated with certain relatively superficial criteria (e.g., having a particular sort of job), but it is associated with certain deeper criteria (realizing a specific sort of aesthetic value). When a person fulfills one of these sets of criteria but not the other, participants tend to have a characteristically ambivalent reaction.

The dual character hypothesis predicts that a similar phenomenon should arise for natural kind concepts. On this hypothesis, each natural kind concept is associated with two different sets of criteria. If a single entity fulfills one set of criteria but not the other, participants should display a characteristically ambivalent reaction. They should say that the entity is a member of that kind in one sense but not in another sense.

Note that dual character is very different from the familiar notion of graded membership. In a typical case of graded membership, people have a single set of criteria that integrates or brings together a wide variety of different features. Then, if an entity has some of the features but not others, people conclude that this entity is a category member to a particular degree. For example, if an entity that has some features associated with the concept CHAIR but lacks other such features, we might conclude ‘This is a borderline case of a chair’ or, more colloquially, ‘This is sort of a chair.’ By contrast, for dual character concepts, there are two distinct criteria of membership. Thus, a single entity can completely satisfy one set of criteria but also completely fail to satisfy the other set of criteria. In such a case, we might conclude ‘There is a sense in which this is clearly a category member,’ but at the same time ‘There is another sense in which this is clearly not a category member.’

This framework makes it possible to formulate a specific hypothesis about natural kind concepts. The hypothesis is not that we have one criterion informed by both deep causal properties and superficial features. Instead, it is that we have two distinct sets of criteria, one based on deep causal properties, the other based on superficial features.

1.4 Dual Character and Situational Context

In most ordinary situations, it is not feasible to express a complex state of ambivalence about whether a given entity falls under a category. People therefore need to have some way of making a single overall determination as to whether the entity falls under the category or not. A question now arises as to how people ordinarily do this.

On the dual character view, people's basic capacity for psychological essentialism does not give them a single criterion for category membership. Rather, psychological essentialism simply makes available two different criteria, without privileging either one over the other. To the extent that people are able to pick out just one of these criteria, they will have to make use of some other psychological process.

We hypothesize that people select among these different criteria by looking to cues from the *situational context*. In some contexts, it seems clear that the most relevant thing to consider is the deeper causal properties (e.g., when having a discussion in a chemistry class). In others, it seems clear that the most relevant thing is the superficial features (e.g., when talking about how to resolve a straightforward practical problem). Perhaps, then, people are able to respond flexibly in a way that takes into account these contextual cues. Since they have two different criteria for category membership, they rely in any given context on the criteria that seem most appropriate to that context.

Numerous different contextual cues could play a role here, but the present studies focus on one cue in particular. A large body of research has shown that formal education in *science* can have a substantial impact on people's judgments (Casler & Keleman 2008; Chi 1981, Shtulman 2006; Shtulman & Valcarcei 2012). A similar phenomenon might be at work here. People might acquire an understanding of a specific type of social context – namely, the context of scientific conversation. They learn that in that specific type of context, one should privilege the deeper causal properties, treating those properties as the sole criteria for category membership.

If this hypothesis is correct, then in the specific case where people see themselves as embedded in a scientific context, they should make category membership judgments that fit the predictions of the one criteria view. However, it should be possible using other methods to see clearly that their concepts show dual character. In particular, (a) when given an opportunity to clarify, people should say that a single entity can be a category member in one sense but not in another and (b) when they are outside of a specifically scientific context, they should be more drawn to rely on other criteria.

2. The Present Studies

In two of the studies reported here, participants received the case of Twin Earth water (Putnam, 1975) and closely parallel cases involving tigers and gold (Kripke, 1972/1980). To eliminate researcher degrees of freedom, we did not write the vignettes describing these cases ourselves but instead used the exact wording of the materials from

Corcoran (2016). These materials were originally designed for a different experimental purpose, without knowledge of our present hypothesis. Thus, the present studies use the exact cases first introduced to support a view that is opposed to ours and also use a way of writing out those cases introduced by another researcher who was not aware of our hypothesis.

Experiments 1a and 1b show that most participants reject the standard philosophical intuition about Twin Earth cases, preferring instead to assent to the claim that there is one sense in which the entity is a member of the natural kind category, but also another sense in which it is not. Experiment 2 shows that categorization is affected by situational context, with participants being more influenced by deeper causal properties when they are in a specifically scientific context. Experiments 3a and 3b extend these findings by replacing philosophical Twin Earth cases with a real-world example (about genetically modified organisms) and also by looking at a more highly educated population (graduate students at elite universities).

2.1 Experiment 1

Participants were presented with different versions of ‘Twin Earth’ scenarios. Specifically, all participants read about entities that had a different underlying essence from the relevant natural kind (e.g., a liquid with a different chemical structure than water). In one condition the entity had all of the same superficial characteristics as the natural kind, while in the other condition the entity had different superficial properties.

If the one-criterion view is correct, there should be no difference in ratings between conditions; in both cases, participants should say that the entity is not a member of the natural kind. By contrast if the dual-character view is correct, participants should show a more complex pattern of judgments. When the entity has different superficial properties, participants should be more inclined to say that the entity is not a member of the natural kind in any sense. When the entity has the same superficial properties, participants should be more inclined to say that the entity is a member of the kind in one sense, but is not a member in another sense.

Experiment 1a tests this prediction by presenting participants with a forced choice between statements. In Experiment 1b participants could endorse two separate statements: one affirmative (is a member of the category) and one negative (is NOT a member of the category).

2.1.1. Experiment 1a

Method

Participants. Six-hundred participants were recruited from Amazon’s Mechanical Turk (56% male, 43% female, 0% non-binary, mean age = 34).

Materials and Procedure. Each participant was presented with one vignette presented in a 3 (Kind: gold, tigers, or water) x 2 (Vignette Structure: same appearance, different appearance) between-subjects design. In the same appearance conditions, participants read vignettes from Corcoran (2016). Those vignettes described an entity

(e.g. a liquid) that had all of the same superficial properties as an entity on Earth (e.g., drinkable, clear), but a different causal property (e.g. not H₂O). In the different appearance condition, participants read vignettes in which the entity did not have any of the superficial properties as an entity on Earth (e.g. is not potable and does not look like water), and also had a different causal property (see Appendix for vignettes).

Then, participants received three statements and were asked to indicate which one they agreed with most. For example, the question for water was: With which of the following do you most agree?

1. The liquid from Twin Earth is water.
2. The liquid from Twin Earth is not water.
3. There's a sense in which the liquid from Twin Earth is water, but ultimately, if you think about what it really means to be water, you'd have to say there's a sense in which the liquid from Twin Earth is not truly water at all.

These three options were presented in a random order. After responding to the forced choice question, participants responded to two comprehension check questions. All vignettes and questions are listed in full in the appendix.

Results

Three hundred and ninety-one participants correctly responded to the two comprehension check questions. Analyses were conducted on these participants. The percentage of participants choosing each option is shown in Figure 1. (Data from all experiments are available on the Open Science Framework: <https://osf.io/f44hq/>.)

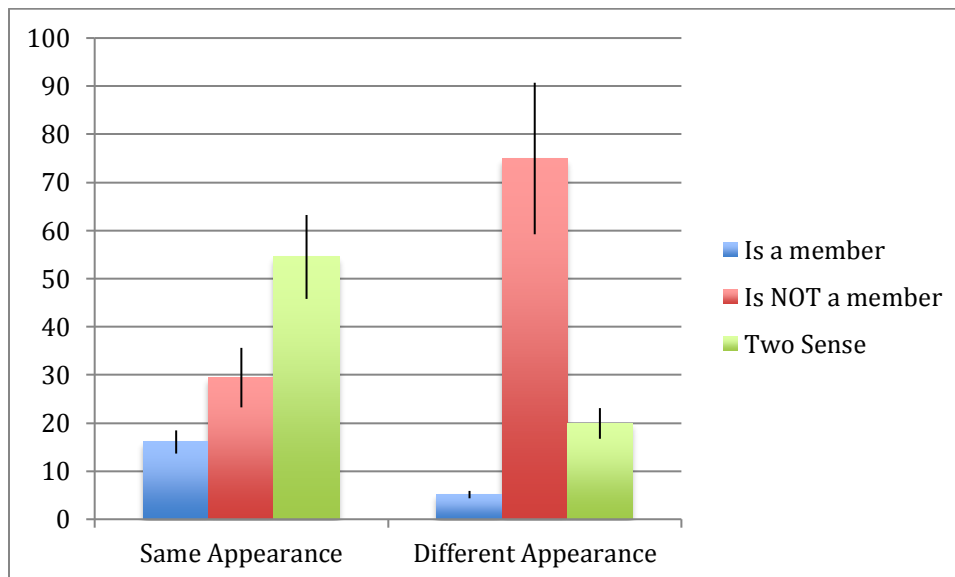


Figure 1. Percentages of participants choosing each statement (Same Appearance vs. Different Appearance), collapsing across Kind (gold, tiger, water). Error bars indicate 95% confidence intervals.

We analyzed the data using two hierarchical binary logistic regression models. In both models, the dependent variable dichotomized participants' responses as either "Is NOT a member" or another response (either "Is a member" or "Two Sense"). In the first model we entered as predictors Vignette Structure (different appearance vs. same appearance), and two dummy codes for the Kind (gold, water). In the second model we also included two interaction terms (gold x vignette structure, water x vignette structure). The comparison of these models indicated that Vignette Structure did not significantly interact with Kind, $X^2(2, N = 391) = 2.63, p = .269$.

The results showed a main effect of Vignette Structure ($B = -1.97, SE = .24, p < .001$, odds ratio (OR) = 7.18, where participants were less likely to choose the "Is NOT a member" statement in the Same Appearance condition than in the Different Appearance condition.¹ Moreover, this pattern significantly replicated for all three Kinds (gold, $X^2(1, N = 113) = 45.04, p < .001$; water, $X^2(1, N = 134) = 19.42, p < .001$; tiger $X^2(1, N = 144) = 4.30, p = .038$). Finally, as seen in Table 1, the Two Sense statement was the most popular response in the Same Appearance condition, while the non-member statement was the most popular in the Different Appearance condition.

Same Appearance – Different causal structure

	Two Sense	Is a member	Is NOT a member
Gold	.68 ***	.06 ***	.26
Tiger	.42 *	.21 *	.37
Water	.54 **	.21 *	.25

Different Appearance – Different causal structure

	Two Sense	Is a member	Is NOT a member
Gold	.19*	.11 ***	.69 ***
Tiger	.25	0 ***	.75 ***
Water	.15 **	.04 ***	.81 ***

Table 1. Percentages of participants (correctly responding to both comprehension questions) choosing each statement (Same Appearance, Different Appearance). Asterisks indicate significance via a binomial comparison to chance (.33), * $p < .05$, ** $p < .01$, *** $p < .001$.

2.1.2 Experiment 1b

Method

Participants. One hundred and eighty-two participants were recruited from Amazon's Mechanical Turk (62% male, 36% female, 2% non-binary, mean age = 34).

¹ We also conducted an analysis on all participants, including those who failed check questions. An inclusive analysis (excluding no participants) also reveals no difference between the two models, $X^2(2, N = 601) = 4.61, p = .100$. There was also a main effect of Vignette Structure ($B = -1.05, SE = .293, \text{odd ratio (OR)} = .35, p < .001$).

Materials and procedure. The design of Experiment 1b was identical to that of Experiment 1a, except that participants responded to two scaled rating questions rather than one forced choice question. For example, for water, participants were asked to rate their level of agreement with two statements:

- i) “There’s a sense in which the liquid from Twin Earth is water.”
- ii) “Ultimately, if you think about what it really means to be water, you’d have to say there’s a sense in which the liquid from Twin Earth is not truly water at all.”

Participants rated both statements on a scale from 1 (disagree) to 7 (agree).

Results

The mean response for each question across vignettes is displayed in Figure 2.

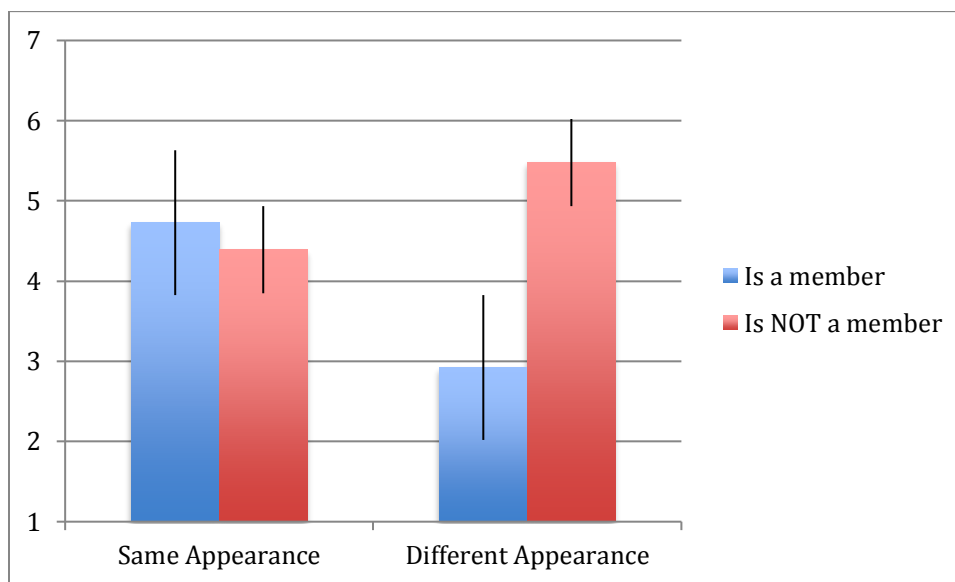


Figure 2. Mean ratings for each statement, collapsing across Kind. Error bars indicate standard error.

Results for the member statement and non-member statement were analyzed separately. For each statement we conducted a 2 (Vignette Structure: different appearance, same appearance) x 3 (Kind: gold, tiger, water) ANOVA.

For the member statement, there was a main effect of vignette structure, such that participants in the same appearance conditions agreed more strongly ($M = 4.73$, $SD = 1.53$) than participants in the different appearance conditions ($M = 2.92$, $SD = 1.82$), $F(1, 180) = 51.90$, $p < .001$, $h_p^2 = .23$. There was no effect of Kind and no interaction.

For the non-member statement, there was a main effect of vignette structure, such that same appearance condition participants agreed less strongly ($M = 4.39$, $SD = 1.72$) than participants in the different appearance condition ($M = 5.48$, $SD = 1.6$), $F(1, 180) = 19.04$, $p < .001$, $h_p^2 = .10$. There was no effect of Kind and no interaction.

To further explore participants’ responses in the same appearance condition, we then conducted one-sample t-tests comparing responses on each of the questions to the

scale midpoint (4). Results indicated that ratings were significantly higher than the midpoint both on the member statement, $t(91) = 4.553, p < .001$, and on the non-member statement, $t(91) = 2.179, p = .032$. In other words, participants agreed with both the statement that there is a sense in which the entity is a member of the kind *and* with the statement that there is a sense in which the entity is not a member of the kind.

Discussion

Two studies showed that participants categorize entities by using two sets of criteria. When an entity lacked both the underlying causal properties and superficial properties, participants were inclined to say it was not a member of the kind in any sense. When an entity lacked the underlying causal properties but shared the superficial properties, participants were inclined to say it was not a member of the kind in one sense, but was a member in one sense.

These results provide evidence against the one-criterion view, which predicts that the entity should not be seen as a member in any sense in either condition. They point instead to a dual-character view, according to which natural kind concepts are seen in terms of two sets of criteria, one involving deep, causal properties and another involving superficial properties.

2.2 Experiment 2

The results thus far suggest that people can use multiple sets of criteria to evaluate different senses of an entity's membership into a natural kind category. We predict that which set of criteria is most relevant varies with context. Thus, when participants are forced to choose whether the entity is or is not a member, they will categorize the entity in line with the set of criteria that the context indicates as most relevant.

Experiment 2 tests whether judgment about the entity's categorization can be directed by presenting participants with contexts that make different sets of criteria more relevant. For instance, consider the Twin Earth liquid. Would participants be less inclined to categorize it as water in the purely scientific context of a chemistry class? What about in a more practical context in which a town has a rule prohibiting residents from creating unapproved pools of water?

Method

Participants. Four hundred and fifty-six participants were recruited from Amazon's Mechanical Turk (62% male, 38% female, 0% non-binary, mean age = 33).

Materials and procedure. Participants received one of the Twin Earth vignettes (gold, tiger, water) and then were given information about one of three possible contexts: scientific, legal, or neutral.

Participants in the scientific context conditions were told that a science department in a university had a rule stating that all students must be provided with certain objects (gold, tigers, or water) for their science practical testing. Participants in the legal context conditions were told that a town had a rule stating that certain objects

(gold, tigers, or water) cannot be used for certain purposes without approval (housing additions, pet adoption, home pool creation). Participants in the neutral context were given no information about the context (see Appendix for full materials).

In all conditions, participants were then told there is a controversy about the entity's category members. Participants rated their agreement with a statement about category membership. For example, in the water conditions, the statement was "The liquid from Twin Earth is water" where 1 (disagree) and 7 (agree) Full vignettes and questions are listed in the appendix.

Results

The mean ratings by condition are displayed in Figure 3. The data were analyzed using a 3 (Context: science, neutral, legal) x 3 (Kind: gold, tiger, water) ANOVA. There was a main effect of context, $F(2, 455) = 9.94, p < .001, \eta_p^2 = .043$. There was no effect of kind and no interaction (both $F_s < 1$). Post-hoc Tukey's tests showed that participants were more inclined to rate the object as a member of the category in the legal than in the science context, $p < .001$. Participants were also more inclined to rate the object as a member of the category in the neutral than in the science context, $p = .043$. The neutral and legal context ratings were not significantly different, $p = .112$.

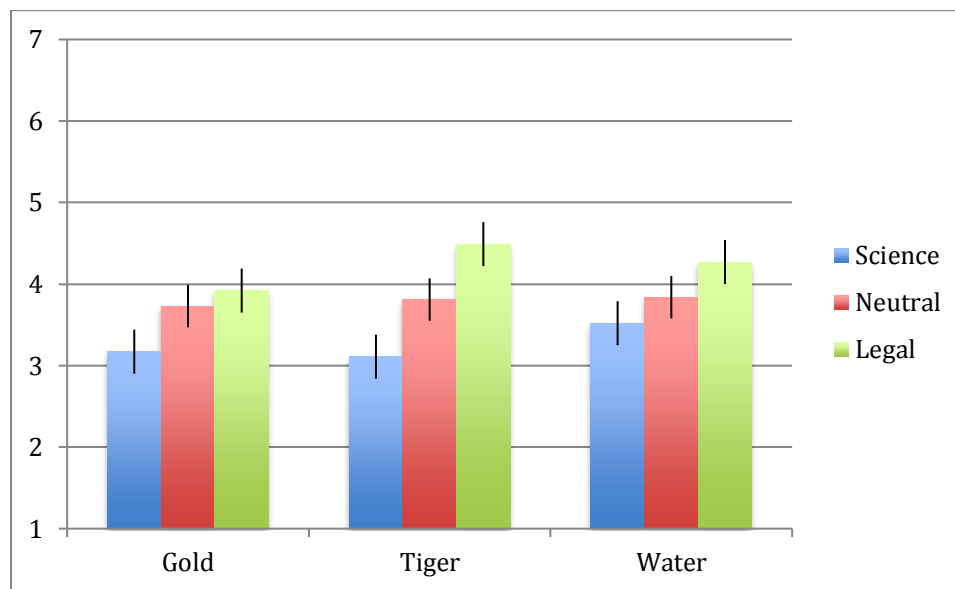


Figure 3. Mean ratings of kind by context. Higher ratings indicate categorization of the particular as the kind. Error bars indicate standard error.

Discussion

Participants' judgments about category membership depended on context. In the science context, participants focus more on the causally central property and in the legal context, participants focus more on the superficial features. The neutral context was intermediate between the two.

This result provides further evidence that people have distinct sets of criteria that determine category membership. Which set of criteria is employed depends on the particular context.

2.3 Experiment 3

Studies 1 and 2 employ Twin-Earth style thought experiments. Since these have been taken as paradigmatic examples in support of a one-criterion view, these studies provide evidence against that view even in the cases introduced to support it.

Although these thought experiments are seminal examples, it might be thought that they are overly philosophical or esoteric. For this reason, Study 3 uses more realistic cases.

Finally, one might worry that the dual-character intuition arises only because participants fail to think clearly and carefully about the questions. For this reason, this final study was conducted on two different populations. Study 3a uses an online sample, while Study 3b turns to a sample of graduate students from elite universities.

2.3.1 Experiment 3a

Methods

Participants. One hundred and fifty participants were recruited from Amazon's Mechanical Turk (57% male, 43% female, 0% non-binary, mean age = 33).

Materials and procedure. All participants read a vignette about genetically modified salmon, fish whose genes have been altered to enable them to grow at faster rates:

The Maxwell Laboratory has made great progress researching fish genetics. They have discovered how to modify the genes of salmon in order to enable the fish to grow year-round instead of only during the summer months. These genes enhance speed of growth but they do not affect any other qualities. The laboratory's fish are identical in all other observable properties to salmon. These properties include appearance, size, taste, and other markers that distinguish salmon from other similar fish.

If one were to perform a genetic analysis of a one of the laboratory's fish, however, one would find the fish does not contain the genes of salmon. Instead, the laboratory fish contains the modified genes.

The modified fish and salmon are completely indistinguishable and interchangeable outside of the laboratory. The laboratory issues a report stating that while the fish do not belong to the same scientific category as familiar salmon, this difference is immaterial for any purpose other than scientific classification.

Participants were randomly assigned to receive information about one of three contexts (as in Experiment 3). In the science context, participants received a story about fish used for testing in a science laboratory. In the legal context, participants received a story about fish sold at a farmer’s market. Participants in the neutral context were given no information about the context (see Appendix).

All participants rated their agreement with a category membership statement, on a scale from 1 (disagree) to 7 (agree): “The fish from the laboratory are salmon.”

Results

The mean ratings by context are displayed in Figure 4. A one-way ANOVA found a significant effect of context, $F(2, 149) = 3.36, p = .028, \eta_p^2 = .047$. Post-hoc Tukey’s tests showed that participants were more inclined to rate the fish as salmon in the legal context than in the science context, $p = .021$. Neutral context ratings did not differ significantly from the legal context ratings, $p = .259$, or the science context ratings $p = .505$.

2.3.2 Experiment 3b

Methods

Participants. One hundred and ninety-three participants were recruited from elite graduate programs (50% male, 47% female, 3% non-binary, mean age = 27). To recruit participants, we emailed department administrators from a diverse selection of graduate programs (Anthropology, Economics, Geology/Geophysics, Neuroscience/Neurobiology, Political Science/Government, Sociology, and Statistics) at elite universities (Harvard, Princeton, Stanford, and Yale University). We planned to continue emailing new departments until any round of emails brought our total participant number past 150. Our first round of emails, to seven departments at four universities, recruited 193 participants. See Table 2 for participants’ graduate degree universities and departments.

	Harvard	Princeton	Stanford	Yale	Other	Total
Anthropology				1		1
Economics		16	18	15		49
Geology/Geophysics				5		5
Neurosci./Neurobio.	22			16	1	39
Political Sci./Gov’t	29	17		1		47
Sociology	21		12			33
Statistics				3		3
Other	3	1	1	3	8	16
Total	75	34	31	44	9	193

Table 2. Number of participants by graduate degree university and department. “Other” includes no response and responses that were ambiguous between categories (e.g. “Public Policy”).

Materials and procedure. The materials and procedure are identical to that described in Experiment 3a.

Results

The mean ratings by context are displayed in Figure 4, above. A one-way ANOVA found a significant effect of context, $F(2, 190) = 6.60$, $p = .002$, $h_p^2 = .065$. Post-hoc Tukey's tests showed that participants were more inclined to rate the fish as salmon in the legal context than in the science context, $p = .001$. Neutral context ratings did not differ significantly from the legal context ratings, $p = .065$, or the science context ratings $p = .308$.

Finally, we considered the online and graduate student sample together, conducting a 2 (Population: online, graduate) x 3 (Context: science, neutral, legal) ANOVA. There was a main effect of context, $F(2, 337) = 9.86$, $p < .001$, $h_p^2 = .055$. There was no main effect of population, $F < 1$, and no interaction, $F < 1$.

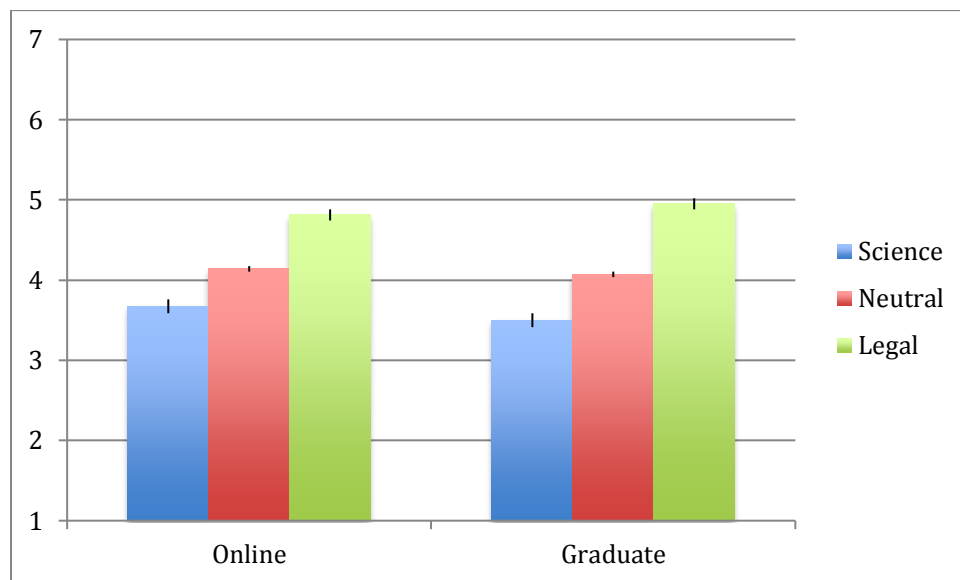


Figure 4. Mean ratings by context. Higher ratings indicate categorization of the entity as a member of the kind. Error bars indicate standard error.

Discussion

Experiment 3 examined participants' intuitions in a more realistic scenario. Once again, there was an effect of context on participants' judgments of category membership. In the scientific context participants were less inclined to categorize the entity as a member of the natural kind. In the legal context they were more inclined to categorize it as a member. The neutral condition was intermediate.

The results also suggest that the dual-character intuition is not simply a result of careless or inattentive thinking. The same effect arose in both the online and graduate student populations. Thus, even participants with extremely high levels of education

categorize these entities in a way that is sensitive to context, in line with the dual-character prediction.

3. General Discussion

Three experiments explored the roles of superficial features and deeper causal properties in people's categorization judgments for natural kinds. Experiments 1a and 1b looked at cases in which an entity does have the superficial features associated with a natural kind but does not have the deeper causal properties. Results indicated that in such cases people think there is one sense in which the entity is a category member and another sense in which it is not. Experiments 2 and 3 examined the impact of context. Categorization judgments for natural kinds were based more on deeper causal properties in some contexts but more on superficial features in others.

Taken together, these results suggest that people's categorization judgments for natural kinds are not simply based on deeper causal properties. Instead, people appear to have two distinct criteria for category membership, one based on deeper causal properties, another based on superficial features. Both of these criteria appear to play important roles in people's categorization judgments.

3.1 The Dual-Character View and Essentialism

Existing research on essentialism holds that essentialized concepts are associated with both (a) superficial features and (b) deeper causal properties. The present findings suggest that both of these elements are very clearly expressed in people's categorization judgments. Thus, the 'dual character' pattern we find in people's judgments seems to capture the distinctive nature of essentialized concepts and distinguish them from concepts of other types.

First, this pattern of judgments distinguishes essentialized concepts from concepts that are understood purely in terms of superficial features. Consider concepts like WALKING, MOTORCYCLE, or ELECTRO SWING. Existing research has emphasized that the pattern of judgments observed for these other concepts is quite different from the one observed for essentialized concepts (Keil, 1989; Gelman, 2003). It might be thought that superficial features matter only for those other concepts and not for essentialized concepts. However, the current findings shed light on how to better understand this difference. For those other concepts, only superficial features matter, but for essentialized concepts *both* the superficial features and the deep causal properties matter.

Second, the findings indicate how essentialized concepts differ from those understood solely in terms of deep causal properties. For a simple example consider the contrast between the concepts H₂O and WATER. All that is relevant to the categorization of H₂O is deeper, causally-central properties. Changes in the appearance of some liquid do not affect whether or not it is H₂O. By contrast, WATER appears to be a more complex concept; a liquid's categorization as WATER depends on both its causal features and its superficial features. Similar remarks apply to the various other concepts people acquire through scientific education (PROTON, RABIES, LIGHT SPEED, etc.). Future research could examine these concepts, but it seems that they might not show the dual character found for people's ordinary essentialized concepts.

Third, the results may have implications for debates about the format of conceptual representation (e.g. prototypes vs. exemplars vs. theories). In particular, some researchers have suggested that the very same category might be associated with representations in more than one of these formats, so that, e.g., the category *water* could be associated both with a prototype and with a theory (Machery, 2009; Weiskopf, 2009). The present results might provide at least some support for a view along these lines (see also Genone & Lombrozo 2012; Nichols, Pinillos & Mallon 2016). That is, if people associate each natural kind both with a set of superficial features and with a deeper causal essence, it might be that these two representations actually have two different formats.

Finally, our results shed light on the relation between natural kind concepts and more value-laden concepts like ARTIST. Existing studies find a pattern for value-laden concepts that is strikingly similar to the one we find here for natural kind concepts. For example, participants think it makes sense 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’ (Knobe, Prasada, and Newman, 2013). This pattern of judgment seems highly analogous to the one obtained here for concepts like WATER and TIGER.

A question now arises as to how to understand this similarity. One possible view would be that the two kinds of concepts are completely different in their structure but simply happen to elicit this same pattern of judgment in certain cases. Another possible view would be that this pattern of judgment is pointing to some deeper similarity in the concepts themselves. For example, it might be thought that the pattern of judgments we find for value-laden concepts is an indication that people are actually essentializing those concepts. Future research could explore this issue more directly.

3.2 The Impact of Situational Context

Studies 2, 3a and 3b suggest that situational context plays a role in determining which criteria people use when applying natural kind concepts. Across all three studies, we observed the same basic pattern. In “scientific” contexts, people were more inclined to use criteria based on deeper, causal properties. In “legal” contexts, they were more inclined to use criteria based on superficial features. Neutral contexts were intermediate between these other two conditions.

Presumably, most of the cognitive processes underlying the context effect observed here will not be specific to this one particular type of case. That is, it seems unlikely that there will be cognitive processes that are devoted solely to the use of situational context in determining the application of natural kind concepts. Rather, there appear to be more general processes that people use when looking to situational context for clues about how to deploy concepts (see, e.g., Sperber & Wilson, 1986; Preyer & Peter, 2005). One strategy for coming to a better understanding of the effect observed for natural kind concepts would therefore be just to focus on the study of these more general processes (see also Nichols, Pinillos & Mallon 2016).

However, there does seem to be at least one factor that is especially relevant to the case of natural kind concepts in particular. The present studies suggest that people’s use of these concepts shifts in certain ways when they are in a *scientific* context. Prior work has shown that scientific education has numerous important effects on people’s judgments

(Casler & Keleman 2008; Chi 1981; Shtulman 2006; Shtulman & Valcarcei 2012). Much of this research has been concerned with the ways in which scientific education can teach people facts about the world, but it seems that such education can also teach people about the norms governing scientific inquiry. Thus, people who have even a passing acquaintance with science may recognize that there is a specific type of context -- the scientific context -- in which distinctive norms apply.

Perhaps one such norm is concerned with the relevance of superficial features vs. deeper causal properties. To the extent that people see themselves as embedded in a specifically scientific context, they may feel that they are supposed to focus more on the deeper causal properties than they would in other, more ordinary contexts. For example, suppose that a person believes that spiders are similar to insects in their superficial features but completely different in their deeper causal properties. Such a person might treat spiders and insects as similar in many contexts, but to the extent that she is engaged in a specifically scientific conversation, she might feel that she should begin treating them as completely different. Future studies could further explore this phenomenon and also the more general question as to how the norms governing scientific contexts might influence people's application of concepts.

Although these experiments focused on “science” and “legal” contexts, future work might further investigate the nature of these contexts. Perhaps there may even be certain “scientific” situations in which the *superficial* features of an entity are most relevant to its categorization; similarly, there might be certain “legal” situations in which an entity’s deep causal property is more relevant. Studies 2, 3a and 3b show more generally that situational context can play a role in determining which sense of natural kind category membership is more salient in categorization. Future work can investigate the way in which specific contexts (e.g. science contexts) might emphasize causal *or* superficial features.

Finally, it should be emphasized that situational context is unlikely to be the only factor that plays such a role here. The core claim is just that people's psychological essentialism gives them two different criteria for natural kind concepts, without privileging either over the other, and some further factor therefore has to determine which set of criteria people use in any given case. One such factor is situational context. Future work could ask whether people's choice of criteria can also be impacted by other, unrelated factors, such as motivational biases or stable individual differences in cognitive style.

3.3. The Twin Earth Thought Experiment

Our primary concern has been with general questions about categorization judgments involving essentialized concepts. However, it is also noteworthy that the present studies explored people’s intuitions regarding the Twin Earth thought experiment. This thought experiment has received a truly enormous amount of attention within existing research: the four-hundred page “The Twin Earth Chronicles” (1996, xi) celebrated twenty years of “Twin Earth and its implications,” and in the twenty years

since “Twin Earth” has received thousands more citations.² The patterns of people’s intuitions concerning Twin Earth are therefore of some interest in themselves.

As we mentioned at the outset, one irony of the present studies is that they provide evidence against the one-criterion view by exploring the very case that was most often used to support that view. In other words, these studies suggest that people’s intuitions about the Twin Earth thought experiment are very different from what they had been assumed to be within prior research. But this finding immediately leads to a new question. Given that these experimental studies suggest that people’s intuitions do not actually conform to the one-criterion view, why did researchers initially assume that they did?

One possible answer is that there was never good reason for this belief in the first place. Perhaps the widespread assumption is simply to be explained in terms of some quirk of academic sociology. Cummins (1998) has argued forcefully for precisely such a claim. As he puts it: “It is a commonplace for researchers in the Theory of Content to proceed as if the relevant intuitions [about Twin Earth] were undisputed... The Putnamian take on these cases is widely enough shared to allow for a range of thriving intramural sports among believers. Those who do not share the intuitions are simply not invited to the games” (Cummins, 1998: 116).

There may be some truth to this suspicion, but we suspect that there is an additional factor at play. Experiments 3 and 4 indicate that people’s intuitions in these cases depend on context. In ‘scientific’ contexts, people tend to focus more on deeper causal properties, whereas in ‘legal’ contexts, they tend to focus more on superficial features. Perhaps the researchers who were investigating questions about the Twin Earth thought experiment were always considering that thought experiment in more scientific contexts. For that reason, it might consistently have appeared to those researchers that the categorization criteria in that thought experiment were simply a matter of deeper causal properties. This would be an understandable conclusion, but as the present studies indicate, people’s ordinary categorization judgments show a more complex pattern.

Finally, it should be noted that the study of intuitions about this thought experiment can potentially have implications that go beyond cognitive science for philosophical questions about the semantics of natural kind terms. Some philosophers have argued that empirical facts about the patterns of people’s intuitions are relevant to these questions (e.g., Corcoran, 2016); others have argued that they are not (e.g., Deutsch, 2015). Future philosophical research could return to these issues and ask whether they can be informed in any way by the present findings.

3.4 Conclusion

The Twin Earth thought experiment has shaped the modern study of essentialism. By distinguishing an entity’s superficial features from its deeper causal properties, it led to the core insight that natural kinds are associated with two different representations: a set of superficial features (e.g., a liquid’s color or smell) and a set of deeper, causal properties (e.g., a liquid’s underlying chemical structure) (Keil 1989; Gelman 2003).

² A Google scholar search reveals 1,130 citations to “Twin Earth” between 1973 and 1996, and 4,520 in the past twenty years (1997-2016)

The present studies suggest that people's ordinary judgments do not conform to the standard philosophical intuition that the deeper causal properties are the *sole* criterion of category membership. Instead, we find that people's actual judgments display a more complex pattern. Entities are categorized into natural kinds according to two different criteria. According to one, the Twin Earth liquid really is water, but according to the other, it is not water at all.

Ultimately, these results suggest that the patterns of people's categorization judgments directly reflect the core insight that motivated essentialism research in the first place. What is most striking about essentialized concepts is that they appear to be associated with two different representations. The present results suggest that both of these representations actually play a role in people's categorization judgments.

References

- Abbott, B. 1997. A Note on the Nature of "Water." *Mind* 106(422): 311-319.
- Ahn, W. 1998. Why are different features central for natural kinds and artifacts? *Cognition* 69: 135-178.
- Ahn, W., Kalish, C., Gelman, A., Medin, D. L., Luhmann, C., Atran, S., Coley, J.D., Shafto, P. (2001). Why essences are essential to the psychology of concepts. *Cognition* 82: 59-69.
- Bastina, B., and Haslam, N. 2006. Psychological Essentialism and Stereotype Endorsement. *Journal of Experimental Social Psychology*, 228-235.
- Carey, S. 1985. *Conceptual Development tin Childhood*. Cambridge: MIT Press.
- Casler, K. and Keleman, D. 2008. Developmental Continuity in Telo-Functional Explanation: Reasoning about Nature Among Romanian Romani Adults. *Journal of Cognition and Development*, 9(3).
- Chi, M.T.H, Feltovich, P.J., and Glaser, R. 1981. Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2): 121-152.
- Cummins, R. (1998). Reflection on Reflective Equilibrium, in DePaul, M. and Ramsey, W. (eds.) *Rethinking Intuition*. Oxford: Rowman & Littlefield Publishers, Inc.
- Deutsch, M. 2015. *The Myth of the Intuitive: Experimental Philosophy and Philosophical Method*.
- Diesendruck, G. 2001. Essentialism in Brazilian Children's Extensions of Animal Names. *Developmental Psychology*, 37(1): 49-60.
- Gelman, S.A. 2003. *The Essential Child: Origins of Essentialism in Everyday Thought*. Oxford University Press.
- Gelman, S.A. and Wellman, H.M. 1991. Insides and essences: Early understandings of the non-obvious. *Cognition*, 38(3): 213-244.
- Gelman, S.A., and Markman, E.M. 1987. Young Children's Inductions from Natural Kinds: The Role of Categories and Appearances. *Child Development*, 58(6): 1532-1541.
- Genone, J. & Lombrozo, T. 2012. Concept possession, experimental semantics, and hybrid theories of reference. *Philosophical Psychology* 25(5): 717-742.
- Hall, D.G., Waxman, S.R., Bredart, S. & Nicolay, A.C. 2003. Preschoolers' use of form class cues to learn descriptive proper names. *Child Development* 74, 1547-1560.
- Haslam, N. and Levy, S. 2006. Essentialist Beliefs about Homosexuality: Structure and Implications for Prejudice. *Personality and Social Psychology Bulletin*, 32(4): 471-485.
- Haslam, N. Psychiatric categories as natural kinds: essentialist thinking about mental disorder. *Social Research*: 1031-1058.
- Haslam, N., and Ernst, D. 2002. Essentialist Beliefs about Mental Disorders. *Journal of Social and Clinical Psychology*, 21(6): 628-644.
- Haslam, N., Bastian, B, and Bisset, M. Essentialist beliefs about personality and their implications. *Personality and Social Psychology Bulletin*, 30(12): 1661-1673.
- Haslam, N., Rothschild, L, and Ernest, D. Essentialist beliefs about social categories. *British Journal of Social Psychology*.
- Hirschfeld, L.A. 1995. Do children have a theory of race? *Cognition*, 54(2): 209-252.
- Hirschfeld, L.A. 1996. *Race in the Making*. MIT Press.

- Ingaki, K., and Hatano, G. 2002. *Young children's naïve thinking about the biological world*. Psychological Press: New York.
- Kalish, C.W. 1995. Essentialism and graded membership in animal and artifact categories. *Memory & Cognition*, 23(3): 335-353.
- Kalish, C.W., and Gelman, S.A. 1992. On Wooden Pillows: Multiple Classifications and Children's Category-based Inductions. *Child Development*, 63(6): 1536-1557.
- Keil, F. 1989. *Concepts, Kinds, and Cognitive Development*. Cambridge: MIT Press.
- Keller, J. 2005. In genes we trust: The biological component of psychological essentialism and its relationship to mechanisms of motivated social cognition. *Journal of Personality and Social Psychology*.
- Knobe, J., Prasada, S., and Newman, G.E. 2013. Dual character concepts and the normative dimension of conceptual representation. *Cognition*, 127(2): 242-257.
- Kripke, S. 1972/1980. *Naming and necessity*. Cambridge: Harvard University Press.
- Machery, E. 2009. *Doing without concepts*. Oxford University Press.
- Malt, B.C. 1994. Water is not H₂O. *Cognitive Psychology*, 27(1): 41-70.
- Medin, D. and Ortony, A. 1989. "Psychological essentialism" in S. Vosniadou and A. Ortony (eds.) *Similarity and Analogical Reasoning*.
- Murphy, G.L. and Medin, D.L. 1985. The role of theories in conceptual coherence. *Psychological Review*, 92(3): 289-316.
- Newman, G.E., and Keil, F.C. 2008. Where is the essence? Developmental shifts in children's beliefs about internal features.
- Nichols, S., Pinillos, N.A. & Mallon, R. 2016. Ambiguous Reference. *Mind*.
- Pessin, A. and Goldberg, S. (eds). 1996. *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's "The Meaning of 'Meaning.'"*
- Preyer, G. and Peter, G. (eds.). 2005. *Contextualism in Philosophy: Knowledge, Meaning and Truth*. Oxford University Press.
- Putnam, H. 1975. The meaning of meaning. *Minnesota Studies in the Philosophy of Science: Language, Mind, and Knowledge*, 7: 131-93.
- Rehder, B., and Hastie, R. 2001. Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General*, 130(3): 323-360.
- Rheder, B, and Burnett, R.C. 2005. Feature inference and the causal structure of categories. *Cognitive Psychology*, 50(3): 264-314.
- Rhodes, M.J., Leslie, S.J. & Tworek, C.M. 2012. Cultural transmission of social essentialism. *Proceedings of the National Academy of the Sciences*, 109(34): 13526-13531.
- Rips, L.J. 1989. "Similarity, typicality, and categorization," in S. Vosniadou and A. Ortony (eds.) *Similarity and Analogical Reasoning*.
- Schlegel, R. J., Hicks, J. A., Davis, W. E., Hirsch, K. A., and Smith, C. M. 2013. The dynamic interplay between perceived true self-knowledge and decision satisfaction. *Journal of Personality and Social Psychology*, 104(3), 542-558.
- Schlegel, R.J. and Hicks, J.A. 2011. The true self and psychological health: Emerging evidence and future directions. *Social and Personality Psychology Compass*, 5(12): 989-1003.
- Schtulman, A. 2006. Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology* 52(2): 170-194.

- Shtulman, A. and Valcarcel, J. 2012. Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition* 124(2): 209-215.
- Sloman, S.A., Love, B.C., and Ahn, W. 1998. Feature centrality and conceptual coherence. *Cognitive Science*, 22(2): 189-228.
- Sperber, D. and Wilson, D. 1986. *Relevance: Communication and Cognition*. Oxford.
- Stevens, M.. 2000. The essentialist aspect of naïve theories. *Cognition*, 74(2): 149-175.
- Waxman, S., Medin, D., and Ross, N. 2007. Folkbiological reasoning from a cross-cultural developmental perspective: Early essentialist notions are shaped by cultural beliefs. *Developmental Psychology*, 43(2): 294-308.
- Weiskopf, D.A. 2009. The plurality of concepts. *Synthese*, 169: 145.
- Wellman, H.M., and Gelman, S.A. 1988. "Children's understanding of the nonobvious," in R.J. Sternberg (ed.) *Advances in the psychology of human intelligence*: 99-135.

Appendix: Experimental Materials

Introduction to all experiments

All participants received this introduction: Different participants in this study will receive different scenarios. For some of these scenarios, answers to the questions are really obvious; for others the questions we ask are more difficult. Just tell us what you think about the scenario you read, whether your answers to the questions seem very obvious or not obvious at all.

Different Appearance Condition Vignettes (Experiment 1)

Gold: The Maxwell Mining Company discovers how to mine for metals on asteroids. It recovers a large amount of metal that tests show is different in all observable properties to the paradigm sample of gold stored as reference M17 in the Paris Department of Precious Metals. These properties include appearance, weight, conductivity, melting point and other markers that distinguish gold from lookalikes. When they perform a chemical analysis of the asteroid's metal, they find out that it does not contain any elemental atoms. The metal from the asteroid is entirely composed of compound molecules. In contrast, scientists long ago discovered that all the samples of gold on Earth are composed of atoms with 79 protons. [Scientists named the element having atomic weight 79 'Au'.] Scientists theorize (correctly) that some compound molecules and atoms will never behave in exactly the same way. This means that they will never be completely indistinguishable and interchangeable outside the lab. Scientists issue a report stating that the pieces of metal that are not at all identical to reference M17 in any observable properties also do not all belong to the same scientific category.

Water: Suppose that in a few years, humans are able to travel to other galaxies. While exploring, they land on a planet that looks nothing like Earth in virtually any respect. It is populated by plants and animals that look totally different from the familiar plants and animals on Earth. Its landscapes and ecosystems look and function totally differently from those on Earth. They dub this planet "Twin Earth". The astronauts remove their helmets and find that they can breathe freely. They drink a liquid not found in any of the

planet's lakes and rivers and find the liquid does not look and taste at all like water. They do not at all quench their thirst on the liquid they collect while they explore the planet. When they perform a chemical analysis of this liquid, they find out that it does not contain any compound molecules. The liquid in Twin Earth's lakes and rivers is entirely composed of elemental atoms. In contrast, scientists long ago discovered that all the samples of water on Earth are composed of a particular compound. [That compound was named 'H₂O'.] Scientists theorize (correctly) that some compound molecules and atoms will never behave in exactly the same way. This means that they will never be completely indistinguishable and interchangeable outside the lab. Scientists issue a report stating that the liquid in Twin Earth's lakes and rivers is not at all identical to the liquid in Earth's lakes and rivers in any observable properties and also does not belong to the same scientific category.

Tiger: Explorers in the mountains of Asia come across a population of animals with no feline characteristics and without any striped orange and black fur. These animals they came across are not 600 pound carnivores and have none of the characteristics of familiar tigers, showing neither the same structural and functional features inside and out. Scientists study the genes of these animals and find that they do not belong to any of the known sub-species of *Panthera tigris*, the species to which all previously recognized tigers belong. In fact, genetic comparisons show that the new population are less closely related to the known members of *Panthera tigris* than are lions (members of *Panthera leo*). Scientists issue a report stating that convergent evolution has led this isolated population to be very different from familiar tigers. The members of the isolated population do not belong to the same scientific category as familiar tigers.

Same Appearance Vignettes (Experiment 1, 2)

Gold: The Maxwell Mining Company discovers how to mine for metals on asteroids. It recovers a large amount of metal that tests show is identical in all observable properties to the paradigm sample of gold stored as reference M17 in the Paris Department of Precious Metals. These properties include appearance, weight, conductivity, melting point and other markers that distinguish gold from lookalikes. When they perform a chemical analysis of the asteroid's metal, they find out that it does not contain any elemental atoms. This is somewhat surprising, because scientists long ago discovered that all the samples of gold on Earth are composed of atoms with 79 protons. [Scientists named the element having atomic weight 79 'Au'.] The metal from the asteroid is entirely composed of compound molecules. Scientists theorize (correctly) that some compound molecules and atoms will behave in exactly the same way. This means that they will be completely indistinguishable and interchangeable outside the lab. Scientists issue a report stating that the pieces of metal identical to reference M17 in all observable properties do not all belong to the same scientific category, but this difference is immaterial for any purpose other than scientific classification.

Tiger: Explorers in the mountains of Asia come across a population of animals with feline characteristics and striped orange and black fur. These 600 pound carnivores are indistinguishable from familiar tigers, with exactly the same structural and functional

features inside and out. Scientists study the genes of these animals and find that they do not belong to any of the known sub-species of *Panthera tigris*, the species to which all previously recognized tigers belong. In fact, genetic comparisons show that the new population are less closely related to the known members of *Panthera tigris* than are lions (members of *Panthera leo*). Scientists issue a report stating that convergent evolution has led this isolated population to be indistinguishable from familiar tigers. While the members of the isolated population do not belong to the same scientific category as familiar tigers, this difference is immaterial for any purpose other than scientific classification.

Water: Suppose that in a few years, humans are able to travel to other galaxies. While exploring, they land on a planet that looks exactly like Earth in virtually all respects. It is populated by plants and animals that look exactly like the familiar plants and animals on Earth. Its landscapes and ecosystems look and function exactly like those on Earth. They dub this planet “Twin Earth”. The astronauts remove their helmets and find that they can breathe freely. They drink from the lakes and rivers and find that their contents look and taste just like water. They quench their thirst on what they collect from the lakes and rivers while they explore the planet. When they perform a chemical analysis of this liquid, they find out that it does not contain any compound molecules. This is somewhat surprising, because scientists long ago discovered that all the samples of water on Earth are composed of a particular compound. [That compound was named 'H₂O'.] The liquid in Twin Earth's lakes and rivers is entirely composed of elemental atoms. Scientists theorize (correctly) that some compound molecules and atoms will behave in exactly the same way. This means that they will be completely indistinguishable and interchangeable outside the lab. Scientists issue a report stating that the liquid in Twin Earth's lakes and rivers does not belong to the same scientific category as the liquid in Earth's lakes and rivers, but this difference is immaterial for any purpose other than scientific classification.

Experiment 1 Check Questions

Gold Condition:

Is the metal from the asteroid identical to gold in terms of all its observable properties (e.g. appearance) Yes No

Is the metal from the asteroid identical to gold in terms of its atomic/molecular structure (e.g. number of protons) Yes No

Water Condition:

Is the liquid from Twin Earth identical to water in terms of all its observable properties (e.g. appearance) Yes No

Is the liquid from Twin Earth identical to water in terms of its atomic/molecular structure (e.g. containing H₂O) Yes No

Tiger Condition:

Are the animals the explorers found identical to tigers in terms of all their observable properties (e.g. appearance) Yes No

Are the animals the explorers found identical to tigers in terms of all their genetic structure (e.g. genes) Yes No

[Same Appearance Condition correct response pattern is “Yes, No.”
Different Appearance Condition correct response pattern is “No, No.”]

Experiment 1a Question

With which of the following do you most agree?

The metal from the asteroid is gold.

The metal from the asteroid is not gold.

There’s a sense in which the metal from the asteroid is gold, but ultimately, if you think about what it really means to be gold, you’d have to say there’s a sense in which the metal from the asteroid is not truly gold at all.

With which of the following do you most agree?

The liquid from Twin Earth is water.

The liquid from Twin Earth is not water.

There’s a sense in which the liquid from Twin Earth is water, but ultimately, if you think about what it really means to be water, you’d have to say there’s a sense in which the liquid from Twin Earth is not truly water at all.

With which of the following do you most agree?

The animals the explorers found are tigers.

The animals the explorers found are not tigers.

There’s a sense in which the animals the explorers found are tigers, but ultimately, if you think about what it really means to be a tiger, you’d have to say there’s a sense in which the animals the explorers found are not truly tigers at all.

Experiment 1b Questions

Do you agree or disagree with the following statements?

There's a sense in which the metal from the asteroid is gold. [1 2 3 4 5 6 7]

Ultimately, if you think about what it really means to be gold, you'd have to say there's a sense in which the metal from the asteroid is not truly gold at all. [1 2 3 4 5 6 7]

There's a sense in which the liquid from Twin Earth is water.

Ultimately, if you think about what it really means to be water, you'd have to say there's a sense in which the liquid from Twin Earth is not truly water at all.

There's a sense in which the animals the explorers found are tigers.

Ultimately, if you think about what it really means to be a tiger, you'd have to say there's a sense in which the animals the explorers found are not truly tigers at all.

Experiment 2 Contexts

Neutral context: There is a controversy about whether the [metal from the asteroid is gold, animal from the mountains of Asia is a tiger, liquid from Twin Earth is water].

Legal context: Imagine that the Summerville city council has just convened to establish the new town ordinances. They have thought very hard about these new rules and are now prepared to adopt them.

One of the rules says that no resident can [use gold in any housing or building additions without approval, adopt a pet tiger without approval, create a pool of water in their yard without prior approval].

One of the [Maxwell Miners, mountain explorers, astronauts who went to Twin Earth] is also a resident of Summerville. Upon returning to town, the [asteroid-miner, explorer, astronaut] takes [some of the metal that he found on the asteroid and uses it to build an addition on to his apartment, one of the animals he found in the mountains of Asia and adopts it as his pet, some of the liquid he found on Twin Earth and uses it to create a pool in his yard] without approval. There is a controversy about whether the [metal from the asteroid is gold, animal from the mountains of Asia is a tiger, liquid from Twin Earth is water].

Science context: Imagine that a group of scientists at Summerville University have just convened to establish the new rules for practical science requirements. They have thought very hard about these new rules and are now prepared to adopt them.

One of the rules says that all [engineering laboratory, biochemistry, chemistry laboratory] instructors should provide students with [a sample of gold to use, access to a tiger to study, a sample of water to use] for the students' practical testing requirements.

One of the [Maxwell Miners, mountain explorers, astronauts who went to Twin Earth] is also [an engineering laboratory, biochemistry, chemistry laboratory] instructor at Summerville University. Upon returning to the college, the [asteroid-miner takes some of the metal that he found on the asteroid, explorer takes one of the animals he found in the mountains of Asia, astronaut takes some of the liquid that he found on Twin Earth] and provides it to his students for the students' practical testing requirements. There is a controversy about whether the [metal from the asteroid is gold, animal from the mountains of Asia is a tiger, liquid from Twin Earth is water].

Experiment 2 Question

Do you agree or disagree with the following statement?

[The metal from the asteroid is gold, The animal from the mountains of Asia is a tiger,
The liquid from Twin Earth is water.]

Experiment 3 Contexts

Legal context: Imagine that the Summerville local government has just convened to establish a new tax rate of an additional 2% on certain goods sold in the town. They have thought very hard about these new rates and to what goods they apply, and they are now prepared to adopt these rules.

One of the rules says that all salmon sold are subject to the extra local tax.

One of the Maxwell laboratory workers is also a vendor at a farmer's market in Summerville. Upon returning to town, the laboratory worker takes some of the fish from the laboratory and sells them at the farmer's market, without collecting any amount for the extra local tax. There is a controversy about whether the fish from the laboratory are salmon.

Science context: Imagine that a group of scientists at Summerville University have just convened to establish the new rules for practical science requirements. They have thought very hard about these new rules and are now prepared to adopt them.

One of the rules says that all ichthyology (fish science) laboratory instructors should provide students with a salmon sample to use for the students' practical testing requirements.

One of the Maxwell laboratory workers is also an ichthyology instructor at Summerville University. Upon returning to the college, the laboratory worker takes some of the fish from the laboratory and provides it to his students for the students' practical testing requirements. There is a controversy about whether the fish from the laboratory are salmon.

Neutral context: There is a controversy about whether the fish from the laboratory are salmon.

Experiment 3 Question

Do you agree or disagree with the following statement?

The fish from the laboratory are salmon.