Cognition 168 (2017) 357–369

Contents lists available at ScienceDirect

Cognition

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

Original Articles

The explanatory effect of a label: Explanations with named categories are more satisfying

Carly Giffin *, Daniel Wilkenfeld, Tania Lombrozo

University of California, Department of Psychology, 3210 Tolman Hall, # 1650, Berkeley, CA 94720, United States

ARTICLE INFO

Article history:
Received 30 March 2017
Revised 27 July 2017
Accepted 29 July 2017
Available online 8 August 2017

Keywords:
Causation
Explanation
Categorization
Category labels

ABSTRACT

Can opium's tendency to induce sleep be explained by appeal to a "dormitive virtue"? If the label merely references the tendency being explained, the explanation seems vacuous. Yet the presence of a label could signal genuinely explanatory content concerning the (causal) basis for the property being explained. In Experiments 1 and 2, we find that explanations for a person's behavior that appeal to a named tendency or condition are indeed judged to be more satisfying than equivalent explanations that differ only in omitting the name. In Experiment 3, we find support for one proposal concerning what it is about a name that drives a boost in explanatory satisfaction: named categories lead people to draw an inference to the existence of a cause underlying the category, a cause that is responsible for the behavior being explained. Our findings have implications for theories of explanation and point to the central role of causation in explaining behavior.

1. Introduction

> First Doctor: "Most learned bachelor whom I esteem and honor, I would like to ask you the cause and reason why Opium makes one sleep?"
>
> Bachelor: "The reason is that in opium resides a dormitive virtue, of which it is the nature to stupefy the senses."
>
> Chorus: "Well, well, well, well has he answered! Worthy, worthy is he to enter into our learned body. Well, well has he answered!"

In this well-known passage from *Le Malade Imaginaire*, Molière invites us to question whether appealing to opium's "dormitive virtue" explains why opium makes a person sleep (Molière, 1673/2012). On the surface, the explanation appears nearly circular: what is it to have a dormitive virtue if not to produce sleepiness when ingested, which is the very property the questioner would like to have explained? A deeper look, however, suggests that the explanation may not be as vacuous as it seems. The medieval scholars whom Molière aimed to mock believed that disposition terms marked particular powers or forces internal to the possessing object (Hutchison, 1991).[1] This example suggests that an explanation that appears to do little more than furnish a label could actually point to a broader network of beliefs that in fact support genuine explanations.

Across three experiments, we investigate whether explanations that invoke a named tendency or condition are considered more explanatory than those that do not, and, if so, why this is the case. For example, is an explanation for someone's abnormal behavior better if it invokes a name (e.g., "she did X because she has *depathapy*, a tendency to X"), than if it appeals to the tendency directly (e.g., "she did X because she has a tendency to X")? And if so, why is this the case? Does a category label support particular inferences (for instance, concerning some stable, causal basis for the behavior being explained?), and do one or more of these inferences offer some reasonable basis for explanation? Below, we review prior work that motivates why a category name could affect the (perceived) quality of an explanation. We then introduce the three experiments we go on to report.

1.1. Psychological background

Several bodies of empirical work shed light on why an explanation that invokes a named category might be judged more explanatory than its unnamed counterpart. First, work with both children and adults suggests that the provision of a category label can have a significant effect on how people conceptualize the category and its relationship to associated properties. Studies find that children prioritize category labels over appearance when making novel

* Corresponding author.
E-mail addresses: carly.giffin@berkeley.edu (C. Giffin), dawilk@gmail.com (D. Wilkenfeld), lombrozo@berkeley.edu (T. Lombrozo).

[1] We thank an anonymous referee for drawing our attention to this historical context.

http://dx.doi.org/10.1016/j.cognition.2017.07.011
0010-0277/

inferences about future behavior (Heyman & Gelman, 2000), and that the use of gender labels for objects increases stereotypically gender-consistent behavior (Zosuls et al., 2009). Gelman and Heyman (1999) found that lexicalization – using a noun label to refer to someone who possesses a certain property – caused children to think of the property as more stable over time and across contexts. For instance, children who were told that a child was a "carrot eater" as opposed to a child who "eats carrots whenever she can" were more likely to believe that the child would eat carrots at a later time, and would do so even if her parents did not encourage her to do so.

Studies with adults reinforce the idea that categorical language can support particularly strong inferences. Yamauchi (2005) found that when a person was described categorically (e.g., "Linda is a feminist") as opposed to descriptively (e.g., "Linda believes in and supports feminism"), participants were more willing to draw inferences concerning other attributes that the person might have. Gelman, Ware, and Kleinberg (2010) found that when category labels were embedded in generic statements (e.g., "Zarpies hate ice cream"), participants represented the category in more "essentialist" terms, as reflected in a battery of subsequent tasks including measures of within-category property generalization and stability. Effects of category labels also extend to classification and memory tasks: adults learn named categories more quickly than unnamed categories in simple category learning experiments (Lupyan, Rakison, & McClelland, 2007), and introducing labels seems to support mental representations that are more categorical (Lupyan, 2012) and prototypical (Lupyan, 2016). Together, these findings suggest that category labels can have a powerful effect on how categories are represented and on the inferences they are taken to support.

One proposal is that the provision of a category label could signal that the category is a *kind*, and that kinds in turn license inferences about underlying causal essences. Ahn, Taylor, Kato, Marsh, and Bloom (2013) put forth this proposal and report evidence supporting its latter component: they found that when a category was introduced as a non-arbitrary classification of people or objects, participants were more likely to agree with the statement that there is something shared by all and only members of the category that *causes* other features of category members. In two of their studies, the provision of a category label was one of the properties used to manipulate whether participants would construe a category as a kind versus an arbitrary category. For example, one of their items was a mental disorder that, in the *kinds* condition, was described with the name "BLV":

> There is a mental disorder called BLV that about 500 people have. The official diagnostic criteria for BLV disorder is to display the following three symptoms: has difficulty remembering new information, requires excessive attention, and always chooses solitary activities.

In the *arbitrary categories* condition, participants instead read:

> There are some people in the world who have difficulty remembering new information. There are others who require excessive attention. And there are others who always choose solitary activities. There are some people who have both the 1st and 2nd symptom, some who have both the 2nd and 3rd symptom, and some who have the 1st and 3rd symptom. And it just so happens that there are about 500 people on Earth who have all three symptoms.

They found that participants in the *kinds* condition reported a significantly greater likelihood that "there is a single cause underlying these three symptoms that all and only [these individuals] have (whether or not we know what that cause is)." While the kind versus arbitrary category manipulation involved several cues beyond the provision of a category label (such as being a known mental disorder with "official diagnostic criteria"), it's plausible that the category label contributed to the belief that the category had some causal basis. If this is correct, then an explanation that appeals to a named category could be judged better because the category is taken to be a causally-essentialized kind that supports causal explanations.

A second body of work sheds light on how adding additional information, even seemingly-vacuous information, could improve the perceived quality of an explanation. Explanations are not only judged better when they are longer (Weisberg, Taylor, & Hopkins, 2015), but also when they contain scientific jargon. In particular, laypeople find circular explanations for psychological behavior significantly better when the explanations additionally contain neuroscience that experts judge to be superfluous (Weisberg, Keil, Goodstein, Rawson, & Gray, 2008; see also Trout, 2008). For example, one group of participants judged an explanation for the curse of knowledge that stated it "happens because subjects make more mistakes when they have to judge the knowledge of others." A second group judged an explanation that additionally stated that it "happens because of the frontal lobe brain circuitry known to be involved in self-knowledge." Adding this additional information led novices, but not experts, to judge the explanation more satisfying.

The effect of adding potentially superfluous claims to a scientific explanation is not restricted to neuroscience; subsequent work has shown that people have a "reductive" bias, generally favoring explanations that contain reductive scientific content over those that do not, with the largest effects found when the augmented explanations are otherwise poor (Hopkins, Weisberg, & Taylor, 2016). Extending these findings to our research questions, it could be that adding a name serves as a cue to implicit reductive content (e.g., signaling that an explanation for behavior is grounded in neuroscience or biology), and/or that there are relevant experts or authorities that underwrite the explanation.

For the domain of mental disorders, additional evidence supports the idea that construing a category in more scientific or reductive terms could affect the way it is conceptualized (Ahn, Proctor, & Flanagan, 2009; Haslam & Ernst, 2002). For instance, Ahn and colleagues found that even trained clinicians have strong beliefs about the biological versus psychological etiology of mental disorders, with more biological disorders more likely to be treated with medication (Ahn et al., 2009). More "biological" disorders are also more strongly essentialized, supporting stronger beliefs in a common cause underlying all cases (Ahn, Flanagan, Marsh, & Sanislow, 2006). Other work finds that phenomena believed to be genetic are thought to be immutable and homogenous (Dar-Nimrod & Heine, 2011). If adding a name in an explanation for abnormal behavior supports a more biological or essentialized construal of the behavior or its basis, it could be that concomitant reductive or essentialist assumptions support an elevated sense of explanatory satisfaction.

In sum, while prior work has not investigated the role of category labels in explanations, work on each half of this conjunction (that is, on category labels only or on explanation judgments only) supports several hypotheses. It could be that people find explanations that appeal to named categories more satisfying than those that do not because (a) the name implies greater stability in associated attributes across time or individuals, (b) the name supports stronger inferences about other attributes, (c) the name implies the presence of a (causal) essence, (d) the name supports a representation with more categorical boundaries, (e) the name evokes more prototypical instances of the category, (f) the name implies more reductive content, (g)

the name implies that the category is recognized by some authority, or (for mental disorders in particular) (h) the name supports a more biological construal of the category. The problem with prior work is not that it offers *too few* hints as to why a category label might have explanatory import, it's that it offers *too many*. That's one reason our experiments are of value: they not only investigate whether explanations with named categories are in fact found to be better, but also why this is the case.

1.2. Overview of experiments

Across three experiments, we investigate whether the mere inclusion of a category name in an explanation of some behavior increases people's satisfaction with the given explanation, and if so, why this is the case. We predict that, despite Molière's scorn, people will find an explanation with a named tendency or condition to be more explanatory than an identical explanation that simply lacks the name. We further hypothesize that this explanatory force stems from participants' assumptions about the additional, unstated information potentially implied by a name, including the causal properties of the category and the stability and generalizability of its associated properties.

To test these hypotheses, we employ behavioral categories that are unlikely to be familiar to our participants, but that have the characteristics of "natural" categories created by humans. We do so by basing our stimulus materials on Culture Bound Syndromes (CBSs), which are syndromes recognized in the DSM-V,[2] but that are little-known within the United States, as they are tied to and situated within another culture (American Psychiatric Association, 2013). For example, "latah" is a CBS from Southeast Asia, in which an affected individual will typically engage in behaviors such as screaming, cursing, and mimicry in response to a sudden shock. All of our stimuli describe an individual who engages in a behavior loosely drawn from a real CBS (such as mimicry from latah), and where that behavior is then explained by appeal to a named tendency (the fictional "depathapy"), or with an identical characterization of the tendency that does not include a name.

In Experiment 1, we compare explanations that appeal to a named tendency to those that appeal to the same tendency, but where no name is provided. In Experiment 2, we compare explanations that appeal to a named *condition* to those that appeal to the same condition, but where no name is provided. Finally, in Experiment 3, we compare explanations that appeal to a named condition that is stipulated to *cause* the behavior in question to the same condition (also stipulated to cause the behavior), but without the name. In addition to asking participants to indicate how satisfying they find each explanation, we include a variety of measures designed to capture the inferences that people might draw from the inclusion (versus omission) of a name. These measures are motivated by our review of prior psychological research, and include the stability of the explained behavior over time, its generalizability across individuals with the same tendency or condition, its biological versus psychological basis, whether it involves common causes or symptoms, and whether it reflects special expertise or authority. We also include two more exploratory measures concerning how blameworthy and legally culpable the actor is. With this battery of measures and our experimental manipulations, we can identify whether adding a name improves the perceived quality of an explanation, and if so why this is the case.

[2] While Culture Bound Syndromes remain in the DSM-V, one of our examples was removed from this edition. We did not find this removal troubling as our rationale for the use of CBSs was not their actual inclusion in the DSM but the plausibility inclusion indicated.

2. Experiment 1

The primary aim of Experiment 1 was to determine whether an explanation for some behavior that appeals to a named tendency is judged more explanatory than an explanation that omits the name but is otherwise identical. As described above, we included a variety of additional measures to identify which inferences – if any – are licensed by a name.

2.1. Methods

2.1.1. Participants

One-hundred-and-sixty adults (75 female, 84 male, 1 other/prefer not to specify, mean age = 35, SD = 11) participated in the study through Amazon Mechanical Turk. An additional 8 participants were tested, but were excluded for failing catch questions (7) or to ensure even numbers in all conditions (1). Participation was restricted to workers with IP addresses in the United States and with a prior approval rating of 95% or higher on previous tasks. Participants received monetary compensation for their participation.

2.1.2. Materials & procedure

The experimental stimuli were based on four syndromes, each describing a different CBS: *Latah*, *Ataque de Nervios*, *Gururumba*, or *Pibloktoq*. For each syndrome, two versions were created, one in which the behavior described was given a name, and one in which it was described as a tendency. The name used, Depathapy, was invented by the authors and was the same for all *Named* vignettes, though the symptoms varied to match the corresponding CBS. An example of each version for one syndrome, *Latah*, is excerpted below. (Full stimuli can be found in Appendix A.)

> *Named.* "Randy is a 40-year-old male. Recently, he took a beautiful and expensive painting from his office after one of his co-workers said, 'you should take that painting, you're the only one who ever looks at it.' Randy's co-worker had not been serious.
> It turns out that Randy has Depathapy, a tendency to imitate the actions of others and obey commands directed at them, leading him to take the painting."
> *Tendency.* "Randy is a 40-year-old male. Recently, he took a beautiful and expensive painting from his office after one of his co-workers said, 'you should take that painting, you're the only one who ever looks at it.' Randy's co-worker had not been serious.
> It turns out that Randy has a tendency to imitate the actions of others and obey commands directed at them, leading him to take the painting."

Each participant was randomly assigned to one of the eight vignettes, resulting from a cross of syndrome (4: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) with label condition (4: *Named*, *Tendency*). After reading the assigned vignette, participants answered eleven evaluative questions, one catch question relating to the vignette, one general catch question, and two demographic questions. These questions are reviewed in turn.

The following three evaluative questions were presented on a single screen, in random order. Here we present the text corresponding to the *Latah* syndrome, with the disease name (in brackets) only presented for participants in the *Named* condition.

> *Explanation Satisfaction.* "Suppose someone asks why Randy took the painting. How satisfying do you find the following answer? 'Randy acted this way because he has [Depathapy,] a tendency to imitate the actions of others and obey commands directed at them.'" Rated on a scale of 1 (not at all satisfying) to 7 (very satisfying).

Blame. "How strongly would you agree or disagree that Randy deserves blame for taking the painting?" Rated on a scale of 1 (strongly disagree) to 7 (strongly agree).

Legal Culpability. "Suppose you are a juror in a court case trying Randy for his actions. The judge informs you that you should find Randy not guilty by reason of insanity if you believe that because of a mental disease or defect, he did not know or understand the nature and quality of his act or did not know or understand that his act was morally or legally wrong. How likely would you be to find Randy guilty?" Rated on a scale of 1 (not at all likely) to 7 (very likely).

Participants were then presented with a new screen, with the following questions presented in a randomized order. These questions were asked to test whether a name carried implications about the stability, generalizability, etiology, or treatment of a disorder. The stability questions were modeled after previous research on naming (Gelman & Heyman, 1999). All were rated on a scale from 1 (not at all likely) to 7 (very likely).

Stability-past. "Given Randy's [Depathapy/tendency], how likely do you think it is that he would have obeyed commands directed at him five years ago?"

Stability-future. "Given Randy's [Depathapy/tendency], how likely do you think it is that he might obey commands directed at him five years from now?"

Generalize-others. "How likely is another person with [Depathapy/this tendency] to exhibit behavior resulting from a tendency to imitate the actions of others and obey commands directed at them, similar to that exhibited by Randy (when in a similar position)?"

Generalize-self. "How likely would you be, in Randy's position, to exhibit behavior resulting from a tendency to imitate the actions of others and obey commands directed at you, similar to that exhibited by Randy?"

The next two questions were presented on the same screen, but prefaced by the following paragraph, modeled after previous research that has found that people distinguish between biologically and psychologically caused disorders (Ahn et al., 2009):

"Randy's [Depathapy/tendency] could be caused by biological or psychological factors. Biological factors include any genetic or physiological factors that contribute to or cause the condition. Psychological factors include any behaviors, thoughts, emotions, or identity-related factors that contribute to or cause the condition."

The following were then rated on a scale from 1 (not at all) to 7 (completely/entirely):

Biological. "To what extent is Randy's [Depathapy/tendency] BIOLOGICAL in nature?"

Psychological. "To what extent is Randy's [Depathapy/tendency] PSYCHOLOGICAL in nature?"

The next two questions were still on the same screen, but also prefaced by the following explanatory paragraph (also adapted from Ahn et al., 2009):

"Randy's [Depathapy/tendency] could be treated by either medication or psychotherapy. Medication refers to any psychiatric, psychoactive, or psychotropic drugs. Psychotherapy refers to treatment by psychological means, involving repeated verbal interactions between a clinician and a client."

The following were then rated on a scale from 1 (not at all) to 7 (very effectively):

Medication. "To what extent could Randy's [Depathapy/tendency] be improved, controlled, or managed by medication?"

Therapy. "To what extent could Randy's [Depathapy/tendency] be improved, controlled, or managed by psychotherapy?"

On a new screen, participants were presented with the following questions, also in randomized order. These questions were designed to assess whether any differences in ratings between the *Named* and *Tendency* conditions were due to participants' belief that a name carries an implication about the causal etiology of a disposition or that a medical authority has recognized or diagnosed the disposition. These questions were prefaced by the statement: "Please answer the following questions about Randy's [Depathapy/tendency]."

Common Cause. "How strongly do you agree or disagree with the idea that there is a common cause that is shared by all and only people with [Depathapy/this tendency] (whether or not we know what that cause is)?" Rated on a scale of 1 (strongly disagree) to 7 (strongly agree).

Common Symptoms. "How strongly do you agree or disagree with the idea that there are common symptoms shared by all and only people with [Depathapy/this tendency] (whether or not we know what all these symptoms are)?" Rated on a scale of 1 (strongly disagree) to 7 (strongly agree).

Recognized. "Experts agree that [Depathapy/this tendency] is a recognized disorder." Rated on a scale of 1 (strongly disagree) to 7 (strongly agree).

Diagnosed. "How likely do you believe it is that Randy's [Depathapy/tendency] has been diagnosed by a medical professional?" Rated on a scale of 1 (very unlikely) to 7 (very likely).

After answering these questions, each participant was presented with a separate screen that contained one true/false question pertaining to the vignette they had read, and a question that asked them whether they were familiar with the tendency described. On a separate screen, participants then answered one additional catch question, an instructional manipulation check, designed to ensure that they were reading instructions carefully, modeled after Oppenheimer, Meyvis, and Davidenko (2009). Participants who answered the true/false question incorrectly, or who failed the instructional manipulation check, were excluded from further analyses. Finally, participants answered demographic questions about their age and gender.

2.2. Results

2.2.1. Explanation satisfaction

Our central prediction was that explanations for behavior would be found more satisfying when they appealed to a named condition. To test this, we performed a 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Tendency*) between-subjects ANOVA with explanation satisfaction as the dependent variable. This analysis revealed the predicted main effect of label condition. Explanations in the *Named* condition were rated as significantly more satisfying than those in the *Tendency* condition, $F(1,152) = 12.04$, $p < 0.001$, $\eta^2 = 0.073$ (see Fig. 1). No other effects were significant.

2.2.2. Blame and legal culpability

The measures of blame and legal culpability were similarly analyzed as dependent variables in 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Tendency*) between-subjects ANOVAs.

For blame, we found a significant main effect of label condition. Blame was significantly higher in the *Tendency* condition than in the *Named* condition, $F(1,152) = 11.64$, $p < 0.001$, $\eta^2 = 0.071$ (see

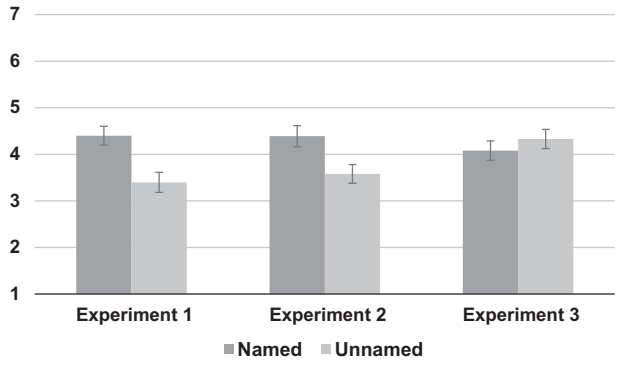

Fig. 1. Explanation satisfaction ratings by label condition across all three experiments. Error bars represent one SEM in each direction.

Table 1). We also found a significant main effect of syndrome, $F(3,152) = 7.49$, $p < 0.001$, $\eta^2 = 0.129$, but no interaction between syndrome and label condition.[3]

For legal culpability, we did not find a significant effect of label. We did find a main effect of syndrome, $F(3,152) = 7.12$, $p < 0.000$, $\eta^2 = 0.123$, but this effect did not interact with label condition.[4] No other significant effects of story nor interactions were found.

Stability and Generalizability. We performed a series of 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Tendency*) ANOVAs on the stability and generalizability variables. We predicted that the *Named* condition would be rated as significantly more stable and generalizable than the *Tendency* condition.

For stability, we averaged the two separate stability scores and treated them as a single variable.[5] Contrary to predictions, we did not find a main effect of label condition on stability. However, we did find a main effect of syndrome, $F(3,152) = 4.00$, $p < 0.009$, $\eta^2 = 0.073$, which did not interact with label condition.[6]

For generalization to others, we found a significant main effect of label condition. Participants in the *Named* condition were significantly more likely than participants in the *Tendency* condition to believe that another person with the disorder/tendency would exhibit the same behaviors, $F(1,152) = 19.00$, $p < 0.001$, $\eta^2 = 0.111$. No other significant effects were found.

For generalization to self, we found the same predicted pattern. Participants in the *Named* condition were significantly more likely than participants in the *Tendency* condition to say that they themselves would exhibit the behavior if they had the disorder/tendency, $F(1,152) = 20.69$, $p < 0.001$, $\eta^2 = 0.120$, with no other significant effects.

[3] Independent *t*-tests revealed that *Latah* ($M = 4.33$, $SD = 1.83$) received significantly less blame than either *Ataque* ($M = 5.30$, $SD = 1.67$), $t(78) = 2.49$, $p < 0.015$, $d = 0.56$, or *Gururumba* ($M = 5.78$, $SD = 1.31$), $t(71) = 4.07$, $p < 0.001$, $d = 0.97$ (corrected for violating Levene's test). *Pibloktoq* ($M = 4.50$, $SD = 1.70$) also received significantly less blame than either *Ataque*, $t(78) = 2.13$, $p < 0.036$, $d = 0.48$, or *Gururumba*, $t(73) = 3.76$, $p < 0.000$, $d = 0.88$ (corrected for violating Levene's test).

[4] Independent *t*-tests revealed that participants gave significantly lower legal ratings for *Latah* ($M = 3.48$, $SD = 1.84$) than *Ataque* ($M = 4.70$, $SD = 1.64$), $t(78) = 3.15$, $p < 0.002$, $d = 0.71$, and *Gururumba* ($M = 4.85$, $SD = 1.64$), $t(78) = 3.53$, $p < 0.001$, $d = 0.80$. Participants also gave significantly lower scores to *Pibloktoq* ($M = 3.55$, $SD = 1.88$) than *Ataque*, $t(78) = 2.92$, $p < 0.005$, $d = 0.66$, or *Gururumba*, $t(78) = 3.29$, $p < 0.001$, $d = 0.75$.

[5] Tests performed on the two stability questions separately revealed the same pattern of results.

[6] *Latah* ($M = 5.18$, $SD = 1.16$) was considered significantly less stable than *Ataque* ($M = 5.76$, $SD = 1.09$), $t(78) = 2.34$, $p < 0.022$, $d = 0.53$, or *Gururumba* ($M = 5.90$, $SD = 0.886$), $t(78) = 3.14$, $p < 0.002$, $d = 0.71$. *Pibloktoq* ($M = 5.40$, $SD = 1.06$) was considered significantly less stable than *Gururumba*, $t(78) = 2.89$, $p < 0.025$, $d = 0.65$.

2.2.3. Biological, psychological, medication, and therapy

We performed a series of 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Unnamed*) ANOVAs on the etiology and treatment questions. We predicted that the *Named* condition would be rated as significantly more likely to be biological and treatable by medication, but that the *Tendency* would be rated significantly more likely to be psychological and treatable with therapy.

Consistent with our predictions, we found a main effect of label condition on biological, with participants in the *Named* condition significantly more likely than those in the *Tendency* condition to indicate that the disorder was biological in nature, $F(1,152) = 5.69$, $p < 0.018$, $\eta^2 = 0.036$. We also found the predicted pattern for psychological, with participants in the *Tendency* condition finding it significantly more likely than participants in the *Named* condition that the disorder was psychological in nature, $F(1,152) = 5.81$, $p < 0.017$, $\eta^2 = 0.037$. No other significant main effects nor interactions were found for these variables, and no significant effects were found for the medication and therapy variables.

2.2.4. Common causes and symptoms

We performed a pair of 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Tendency*) ANOVAs on the common cause and common symptoms questions. We hypothesized that the *Named* condition would be rated as significantly more likely to share a common cause and common symptoms.

As predicted, both a common cause, $F(1,152) = 4.85$, $p < 0.029$, $\eta^2 = 0.031$, and common symptoms, $F(1,152) = 6.30$, $p < 0.013$, $\eta^2 = 0.040$, were rated as significantly more likely for the *Named* than the *Tendency* condition. No other main effects nor interactions were found for these variables.

2.2.5. Authority

We performed a pair of 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Tendency*) ANOVAs on the recognition and diagnosis questions. We hypothesized that participants would be significantly more likely to believe that the *Named* condition constituted a recognized disorder that had been diagnosed by a doctor. Contrary to predictions, no significant differences were found for either variable.

2.2.6. Familiarity

We performed a 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Unnamed*) ANOVA on the familiarity question. We did so to evaluate whether our participants happened to be familiar with some of these syndromes, as this could impact their response to the provided label. We did find a main effect of label condition, $F(1,152) = 8.26$, $p < 0.005$, $\eta^2 = 0.050$, with participants rating the *Named* condition ($M = 1.59$, $SD = 1.78$) less familiar than the *Tendency* condition ($M = 2.19$, $SD = 1.45$), though both were considered quite unfamiliar. We believe this effect was due to the fact that the name "Depathapy" was itself unfamiliar to participants. No significant effect of syndrome was found, and the means for *Latah* ($M = 1.73$, $SD = 1.18$), *Ataque* ($M = 1.68$, $SD = 1.02$), *Gururumba* ($M = 2.25$, $SD = 1.68$), and *Pibloktoq* ($M = 1.90$, $SD = 1.41$) were all low.

2.3. Discussion

Experiment 1 confirmed our main prediction: participants considered an explanation for a behavior to be significantly more satisfying if the explanation appealed to a behavioral tendency that was named. We also found that participants ascribed lower blame for the behavior when the behavioral tendency was named. We found these effects even though the name was novel, ensuring that

Table 1
Means and standard deviations for each dependent variable, as a function of label condition and experiment. Pairs of means for a given experiment are in bold when a *t*-test comparing those means is significant. The *p*-values in columns represent the significance of the interaction between label condition and either Experiment 1/2 or 2/3, thus reflecting which manipulations had a significant effect on the influence of labels for the corresponding dependent variable.

Variable	Condition	Experiment 1	Experiment 2	*p*	Experiment 3	*p*
Explanation	Named	**4.40 (1.80)**	**4.39 (2.03)**	0.658	4.08 (1.86)	**0.012**
	Unnamed	**3.40 (1.93)**	**3.58 (1.86)**		3.74 (1.84)	
Blame	Named	**4.55 (1.69)**	4.19 (1.94)	0.131	4.38 (1.81)	0.473
	Unnamed	**5.40 (1.67)**	4.44 (1.78)		4.33 (1.93)	
Legal	Named	4.06 (1.931)	3.51 (2.08)	0.977	3.86 (1.90)	0.353
	Unnamed	4.23 (1.77)	3.69 (1.89)		3.63 (2.05)	
Stability	Named	5.57 (1.12)	5.65 (1.22)	0.144	5.54 (1.17)	0.553
	Unnamed	5.55 (1.05)	5.27 (1.11)		5.13 (1.38)	
Generalize-Others	Named	**5.31 (1.37)**	**5.49 (1.37)**	0.074	5.45 (1.32)	0.585
	Unnamed	**4.29 (1.56)**	**5.04 (1.43)**		5.18 (1.60)	
Generalize-self	Named	**4.28 (1.93)**	3.88 (2.23)	**0.031**	4.04 (2.00)	0.978
	Unnamed	**2.94 (1.81)**	3.50 (1.97)		3.65 (2.05)	
Biological	Named	**4.08 (1.52)**	4.33 (1.38)	0.565	4.00 (1.53)	0.709
	Unnamed	**3.50 (1.53)**	3.94 (1.38)		3.74 (1.67)	
Psychological	Named	**4.78 (1.39)**	4.65 (1.36)	0.616	4.81 (1.42)	0.627
	Unnamed	**5.28 (1.26)**	5.00 (1.33)		5.01 (1.41)	
Medication	Named	4.69 (1.47)	**5.20 (1.29)**	0.086	4.84 (1.50)	0.444
	Unnamed	4.74 (1.33)	**4.74 (1.22)**		4.61 (1.51)	
Therapy	Named	4.88 (1.35)	**5.21 (1.27)**	**0.001**	4.99 (1.36)	**0.040**
	Unnamed	5.14 (1.29)	**4.53 (1.33)**		4.91 (1.33)	
Common Cause	Named	**4.21 (1.38)**	**4.38 (1.44)**	0.968	4.38 (1.46)	0.250
	Unnamed	**3.73 (1.43)**	**3.90 (1.32)**		4.28 (1.58)	
Common Symptoms	Named	**4.71 (1.58)**	**4.75 (1.48)**	0.767	4.65 (1.54)	0.309
	Unnamed	**4.11 (1.46)**	**4.25 (1.51)**		4.50 (1.62)	
Recognized	Named	4.59 (1.70)	4.86 (1.62)	0.887	4.53 (1.66)	0.755
	Unnamed	4.24 (1.48)	4.56 (1.49)		4.34 (1.67)	
Diagnosed	Named	4.36 (1.90)	4.73 (1.74)	0.711	4.88 (1.69)	0.662
	Unnamed	4.01 (1.80)	4.23 (1.77)		4.55 (1.94)	

participants did not have relevant prior experiences or beliefs. Our other variables shed some light on precisely what additional information participants believed the name implied. Participants in the *Named* condition believed the behavior was significantly more likely to generalize to others with the disorder and to generalize to themselves if they had the disorder. They believed the disorder was more likely to be biological, to have a common cause across individuals, and to share common symptoms. Participants in the *Tendency* condition, conversely, found it significantly more likely that the disorder was psychological in nature.

These findings are consistent with the idea that participants find explanations that appeal to named conditions more explanatory *because* the named condition is taken to imply that the disorder supports generalizations across people, is biologically grounded, and shares a common cause. However, the results of Experiment 1 do not support a causal relationship between any of these inferences and judgments of explanatory satisfaction. Moreover, even if there is a causal relationship, they do not reveal which one or more of these inferences is responsible for explanatory satisfaction. To tackle these questions we adopt an experimental approach: in Experiment 2, we introduce an explanation with an unnamed "condition" (as opposed to a mere "tendency") that potentially supports generalizations across people and a supposed biological etiology; in Experiment 3, we introduce causal language. If introducing these modifications to the *Tendency* explanation eliminates the difference between the *Named* and *Unnamed* conditions, that supports the idea that the targeted inference is responsible for the greater explanatory value of explanations that introduce proper names.

3. Experiment 2

In Experiment 1, we confirmed our prediction that participants would find an explanation more satisfying if a behavioral tendency was given a name, even though that name was unfamiliar and the description of the behavior identical across label conditions. We also found that participants judged the named tendency significantly more likely to generalize across individuals, to be biological in nature, and to share a common cause/symptoms. In Experiment 2 we sought to isolate which of these additional inferences, if any, is responsible for the boost in explanatory satisfaction for named tendencies. We did so by replacing the tendency condition with a comparison condition in which the disorder was referred to as a "condition" rather than a "tendency." For instance, some participants evaluated the explanation that Randy acted in a particular way because "he has a condition that is defined by a tendency to…", while others evaluated the explanation that Randy did so because "he has Depathapy, a condition that is defined by a tendency to…"

The change from "tendency" to "condition" was motivated by the idea that labeling something a "condition" signals that it is a well-defined, recognized category. As such, it might support all or a subset of the inferences licensed by a name. If the information conveyed by a named tendency is in fact equivalent to that conveyed by a "condition," we would expect the named and unnamed conditions of Experiment 2 to yield comparable levels of explanatory satisfaction. By contrast, if the named condition continues to support more satisfying explanations than the unnamed condition, that would suggest that the information implied by a name goes beyond the implications of a "condition." Our additional dependent measures could then shed light on what those additional implications might be, helping us winnow down the candidates for what drives the explanatory satisfaction conferred by a name.

3.1. Methods

3.1.1. Participants

One-hundred-and-sixty adults (85 female, 75 male, mean age = 35, SD = 11) participated in the study through Amazon Mechanical Turk. An additional nine participants were tested, but were excluded for failing catch questions (6) or to ensure even numbers in all conditions (3). Participation was restricted to

workers with IP addresses in the United States and with an approval rating of 95% or higher on previous tasks. Participants received monetary compensation for their participation.

3.1.2. Materials and procedure

The experimental stimuli described the same four syndromes from Experiment 1. We created two versions of each, a *Named* version and a *Condition* version. The only adjustment made to the stimuli was that "condition" was added to all the vignettes. The relevant excerpts from the *Latah* story are below (see Appendix B for full stimuli).

> *Named.* "...It turns out that Randy has Depathapy, a condition that is defined by a tendency to imitate the actions of others and obey commands directed at them, leading him to take the painting."
>
> *Condition.* "...It turns out that Randy has a condition that is defined by a tendency to imitate the actions of others and obey commands directed at them, leading him to take the painting."

Each participant was randomly assigned to one of eight conditions, resulting from a cross of syndrome (4: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) with label condition (4: *Named*, *Condition*). All dependent measures in Experiment 2 were identical to those in Experiment 1, except that "tendency" was replaced with "condition" where appropriate.

3.2. Results

3.2.1. Explanation satisfaction

To analyze explanation satisfaction, we performed a 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Unnamed*) between-subjects ANOVA with explanation satisfaction as the dependent variable. This analysis revealed a main effect of label condition, $F(1,152) = 7.54$, $p < 0.007$, $\eta^2 = 0.047$, with participants finding the explanation with a name significantly more satisfying than the explanation that appealed to "a condition" (see Fig. 1). We also found a main effect of syndrome, $F(3,152) = 3.79$, $p < 0.012$, $\eta^2 = 0.070$, but it did not interact with label condition.[7]

To investigate whether the effect of labels on explanation satisfaction in Experiment 2 differed significantly in magnitude from the effect found in Experiment 1, we additionally performed a 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Unnamed*) between-subjects ANOVA with experiment (Experiment 1, Experiment 2) as an additional between-subjects factor. The interaction between label condition and experiment was not significant. Table 1, above, details the results of this comparison for the analyses that follow.

3.2.2. Blame and legal culpability

A 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Unnamed*) between-subjects ANOVA with blame as the dependent variable revealed only a significant main effect of syndrome, $F(3,152) = 10.84$, $p < 0.001$, $\eta^2 = 0.176$, with no significant interaction.[8]

An equivalent ANOVA on legal culpability similarly revealed only a main effect of syndrome, $F(3,152) = 8.64$, $p < 0.001$, $\eta^2 = 0.146$, with no significant interaction.[9]

3.2.3. Stability and generalizability

We performed a series of 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Unnamed*) ANOVAs on the stability and generalizability variables. We predicted the effects of label condition would either disappear or be smaller than those in Experiment 1.

Counter to the findings from Experiment 1, stability revealed a main effect of label condition, $F(1,152) = 4.34$, $p < 0.039$, $\eta^2 = 0.028$, with *Named* being considered significantly more stable than *Condition*. No other main effects nor interactions were found.

Generalization to others also revealed a significant main effect of label condition, $F(1,152) = 4.22$, $p < 0.042$, $\eta^2 = 0.027$, with *Named* being significantly more likely to generalize to others with the disorder. No other significant effects nor interactions were found.

Generalization to self did not show a main effect of label condition, as it had in Experiment 1. In fact, this was one of two instances in which we did find a significant interaction between label condition and experiment, $F(1,304) = 4.79$, $p < 0.029$, $\eta^2 = 0.016$. There was also a main effect of syndrome, $F(3,152) = 3.29$, $p < 0.022$, $\eta^2 = 0.061$.[10] No other significant effects nor interactions were found.

3.2.4. Biological, psychological, medication, and therapy

We performed a series of 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Unnamed*) ANOVAs to test the etiology and treatment variables. As for the previous variables, we predicted that the effects of label condition would either disappear or be reduced from those seen in Experiment 1.

We found no main effect of label condition on biological or psychological. In fact, we found no significant effects for either of these two variables. However, we did find a significant main effect of label for medication, $F(1,152) = 5.66$, $p < 0.019$, $\eta^2 = 0.036$, and therapy, $F(1,152) = 10.89$, $p < 0.001$, $\eta^2 = 0.067$. In both cases, participants in the *Named* condition thought the actor was significantly more likely to respond to the treatment. Therapy was the second of the two cases for which we saw an interaction between label condition and experiment, $F(1,304) = 10.25$, $p < 0.002$, $\eta^2 = 0.033$.

We also found a significant main effect of syndrome for medication, $F(3,152) = 2.80$, $p < 0.042$, $\eta^2 = 0.052$.[11] No other main effects nor interactions were significant.

3.2.5. Common causes and symptoms

We performed a pair of 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Unnamed*) ANOVAs to test the common cause and symptoms

[7] Participants found *Latah* ($M = 4.30$, $SD = 1.80$), $t(78) = 2.41$, $p < 0.018$, $d = 0.55$, and *Pibloktoq* ($M = 4.58$, $SD = 1.89$), $t(78) = 2.99$, $p < 0.004$, $d = 0.68$, significantly more satisfying explanations than *Ataque* ($M = 3.28$, $SD = 2.00$), $t(78) = 2.41$, $p < 0.018$, $d = 0.55$

[8] Independent samples *t*-tests showed that participants gave significantly lower blame for *Latah* ($M = 3.78$, $SD = 1.73$) than *Ataque* ($M = 4.73$, $SD = 1.80$), $t(78) = 2.41$, $p < 0.018$, $d = 0.55$, or *Gururumba* ($M = 5.35$, $SD = 1.49$), $t(78) = 4.36$, $p < 0.000$, $d = 0.99$. Participants also gave significantly lower blame to *Pibloktoq* ($M = 3.40$, $SD = 1.78$) than *Ataque*, $t(78) = 3.31$, $p < 0.001$, $d = 0.75$, or *Gururumba*, $t(78) = 5.31$, $p < 0.000$, $d = 1.20$.

[9] Independent samples *t*-tests showed that participants were significantly less likely to convict *Latah* ($M = 3.08$, $SD = 1.73$) than *Ataque* ($M = 4.08$, $SD = 2.14$), $t(74) = 2.30$, $p < 0.024$, $d = 0.53$ (corrected for violating Levene's test), or *Gururumba* ($M = 4.55$, $SD = 1.75$), $t(78) = 3.79$, $p < 0.000$, $d = 0.86$. Participants were also significantly less likely to convict *Pibloktoq* ($M = 2.70$, $SD = 1.76$) than *Ataque*, $t(73) = 3.14$, $p < 0.002$, $d = 0.74$ (corrected for violating Levene's test), or *Gururumba*, $t(78) = 4.71$, $p < 0.000$, $d = 1.07$.

[10] Independent samples *t*-tests show that participants found it more like that the symptoms of *Pibloktoq* ($M = 4.48$, $SD = 1.87$) would generalize to themselves than those of *Ataque* ($M = 3.10$, $SD = 2.10$), $t(78) = 3.10$, $p < 0.003$, $d = 0.70$, or *Gururumba* ($M = 3.40$, $SD = 2.09$), $t(78) = 2.43$, $p < 0.017$, $d = 0.55$.

[11] Independent samples *t*-tests showed that *Latah* ($M = 5.23$, $SD = 1.12$), $t(78) = 2.75$, $p < 0.007$, $d = 0.62$, and *Pibloktoq* ($M = 5.20$, $SD = 1.29$), $t(78) = 2.47$, $p < 0.016$, $d = 0.56$, were considered significantly more likely to respond than *Gururumba*, ($M = 4.53$, $SD = 1.15$).

variables. Our prediction was that the effects of label condition would either disappear or be decreased from those seen in Experiment 1.

Contrary to our predictions, though consistent with the findings for other variables, we still found a main effect of label condition for common cause, $F(1,152) = 4.82$, $p < 0.030$, $\eta^2 = 0.031$, and for common symptoms, $F(1,152) = 4.99$, $p < 0.027$, $\eta^2 = 0.032$. Participants in the *Named* condition thought it was significantly more likely that those with the disorder shared a common cause and common symptoms. We also found a significant main effect of syndrome on common symptoms, $F(3,152) = 6.74$, $p < 0.000$, $\eta^2 = 0.117$.[12] There were no other significant main effects nor interactions.

3.2.6. Authority

We performed a pair of 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Unnamed*) ANOVAs to test the recognized and diagnosed variables. We found no significant effects nor interactions for either.

3.2.7. Familiarity

We performed a 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Unnamed*) ANOVA on the familiarity question and found no significant effects nor interactions.

3.3. Discussion

In Experiment 2 we compared the evaluations of explanations that contained a named condition to those that contained an unnamed condition. We anticipated that a "condition," even an unnamed one, would support many of the same inferences that Experiment 1 found for a named tendency relative to an unnamed tendency. Consistent with this expectation, we found that the named and unnamed conditions of Experiment 2 were more similar to each other in some respects than the named and unnamed tendencies of Experiment 1. In particular, Experiment 2 eliminated differences in generalization to the self, with generalization to other judgments similarly attenuated, though not significantly. Nonetheless, we still found that the explanations with named conditions were judged significantly more satisfying than those with unnamed conditions. Thus, the belief that a named tendency is treated as "a condition" cannot explain the full pattern of results we obtained in Experiment 1. We therefore sought, in Experiment 3, to target a different inference: that to a common cause.

4. Experiment 3

In Experiment 3 we compared the evaluation of explanations involving a named condition to those involving an unnamed condition that was explicitly stated to *cause* the relevant behavioral tendency. If the named tendencies and conditions from Experiments 1 and 2, respectively, were found to be more explanatory because they licensed particular inferences about the casual basis for the behavior, then we might expect the difference across the named and unnamed conditions to be eliminated in Experiment 3, where even the unnamed condition should provide comparable causal information.

4.1. Methods

4.1.1. Participants

One-hundred-and-sixty adults (70 female, 89 male, 1 other/prefer not specify, mean age = 35, SD = 12) participated in the study through Amazon Mechanical Turk. An additional seven participants were tested, but were excluded for failing catch questions (5) or to ensure even numbers in all conditions (2). Participation was restricted to workers with IP addresses in the United States and with a prior approval rating of 95% or higher on previous tasks. Participants received monetary compensation for their participation.

4.1.2. Materials and procedure

The experimental stimuli consisted of descriptions of the same four syndromes from Experiment 1. We created two versions of each syndrome, a *Named* version and a *Causal Condition* version. The only adjustment made to the stimuli from Experiment 2 was that "is defined by" was replaced with "causes" in all the vignettes. The relevant excerpts from the *Latah* story are below (see Appendix C for full stimuli).

> *Named.* "...It turns out that Randy has Depathapy, a condition that causes a tendency to imitate the actions of others and obey commands directed at them, leading him to take the painting."
>
> *Causal Condition.* "...It turns out that Randy has a condition that causes a tendency to imitate the actions of others and obey commands directed at them, leading him to take the painting."

Each participant was randomly assigned to one of eight conditions, resulting from a cross of syndrome (4: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) with label condition (4: *Named*, *Causal Condition*). The dependent measures in Experiment 3 were identical to those in Experiment 2, except that "defined by" was replaced with "causes" where appropriate.

4.2. Results

4.2.1. Explanation satisfaction

A 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Causal Condition*) between-subjects ANOVA with explanation satisfaction as a dependent variable failed to find a main effect of label condition (see Fig. 1). Moreover, a 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Causal Condition*) between-subjects ANOVA with experiment (Experiment 2, Experiment 3) as an additional, between-subjects factor revealed a significant interaction between label condition and experiment, $F(1,304) = 6.64$, $p < 0.010$, $\eta^2 = 0.021$, indicating that the effect of label was significantly greater in Experiment 2 than in Experiment 3. Table 1, above, reports the results of additional comparisons across experiments 2 versus 3.

The initial ANOVA additionally revealed a significant main effect of syndrome, $F(3,152) = 3.24$, $p < 0.024$, $\eta^2 = 0.060$, but no interaction between syndrome and label.[13]

4.2.2. Blame and legal culpability

Additional 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Causal Condition*) between-subjects ANOVAs on blame and legal culpability revealed significant main effects of syndrome, $F(3,152) = 7.16$, $p < 0.000$,

[12] Independent samples *t*-tests found that *Latah* ($M = 5.23$, $SD = 1.25$) was considered significantly more likely to have common symptoms than *Gururumba* ($M = 4.35$, $SD = 1.61$), $t(78) = 2.17$, $p < 0.033$, $d = 0.49$.

[13] Independent t-tests on explanation satisfaction found participants rated *Ataque* ($M = 4.53$, $SD = 1.74$), $t(78) = 2.80$, $p < 0.006$, $d = 0.63$, and *Pibloktoq* ($M = 4.68$, $SD = 1.75$), $t(78) = 2.44$, $p < 0.017$, $d = 0.55$, to be significantly more satisfying explanations than *Gururumba*, ($M = 3.53$, $SD = 1.92$).

$\eta^2 = 0.124$, and $F(3,152) = 7.99$, $p < 0.001$, $\eta^2 = 0.136$, but no interactions nor effects of label.[14]

4.2.3. Other variables

We performed a series of 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Causal Condition*) ANOVAs on all remaining variables, excluding familiarity (see below). We found no other significant effects nor interactions.

4.2.4. Familiarity

We performed a series of 4 (syndrome: *Latah*, *Ataque de Nervios*, *Gururumba*, *Pibloktoq*) × 2 (label condition: *Named*, *Causal Condition*) ANOVAs on familiarity ratings. We found a main effect of label condition, $F(1,304) = 6.04$, $p < 0.015$, $\eta^2 = 0.019$, with participants indicating less familiarity with the *Named* condition.

4.3. Discussion

Experiment 3 found that referencing a "condition" that *causes* a given behavioral tendency succeeded in eliminating the effect of a proper name on the perceived quality of an explanation. This suggests that explanations with proper names are found especially satisfying because they license inferences equivalent to those supported by appealing to a "condition that causes" the behavior. The findings from Experiment 2, where the addition of "condition" was insufficient to eliminate the naming advantage, further suggest that the critical feature of the explanations from Experiment 3 was the addition of causal language.

The conclusion that the causal implications of a name are what boost ratings of explanatory satisfaction is supported by additional analyses involving Experiments 1 and 2. First, we found direct experimental evidence for the explanatory role of causal information by comparing explanation ratings for the unnamed *Condition* explanations of Experiment 2 to the unnamed *Causal Condition* explanations of Experiment 3. These explanations differed only in the addition of the causal claim, and yet the *Causal Condition* explanations were judged significantly more satisfying than the *Condition* explanations, $t(158) = 2.60$, $p < 0.010$, $d = 0.41$ (see Fig. 1). Second, we examined whether participants' "common cause" ratings mediated the effect of label condition on explanation satisfaction, combining the data from Experiments 1 and 2. Indeed, we found significant partial mediation, $t = -2.76$, $SE = 0.08$, $p = 0.005$.

While our common cause question almost certainly wasn't a perfect proxy for the suite of causal inferences that our participants may have drawn in response to a named condition, the results of our mediation analysis suggests that causal considerations were a driving factor in boosting named explanation ratings in Experiments 1 and 2. The findings from Experiments 1–3 therefore support the idea that named behavioral tendencies are found explanatory because the name licenses an inference to an underlying cause of the behavior.

5. General discussion

Across three experiments, we find evidence that explanations for behavior that appeal to named tendencies or conditions are found significantly more satisfying than those that omit a name, and that this is because the inclusion of a name is treated as a cue to some underlying cause that is responsible for the behavior being explained. Specifically, Experiment 1 contrasted explanations with named versus unnamed *tendencies*, and found that the former were judged significantly more satisfying. Experiment 2 contrasted explanations with named versus unnamed *conditions defined in terms of a tendency*, and found that the former were still judged significantly more satisfying. Experiment 3 contrasted explanations with named versus unnamed *conditions that caused a tendency*, and succeeded in eliminating the effect of the name. These experimental results suggest that the naming advantage observed in Experiments 1–2 was driven by an inference to the presence of a cause for the specified tendency. This interpretation is bolstered by additional analyses from Experiments 1–2, which revealed that participants' inferences concerning the presence of a common cause underlying instances of the tendency or condition partially mediated the effect of condition on explanatory satisfaction.

Experiments 1–3 also tested, but failed to find support for, a variety of alternative proposals for why explanations with names could be found more satisfying. In particular, they failed to support the ideas that names increase explanatory satisfaction by suggesting that the named tendency or condition is more stable over time, supports stronger generalizations across individuals, is more biologically grounded, or is more scientifically legitimate (in the sense that it is recognized and diagnosed by experts). Of course, it remains a possibility that these inferences play an additional role that was not captured by our measures or analyses. For example, it's plausible that the role of scientific jargon or implied expertise would be greater with different names, different explanations, different measures, or within a different domain.

Our findings have potential implications for both philosophy and psychology. Beginning with philosophy, for over 65 years (since Hempel & Oppenheim, 1948) the question of what explanation amounts to has been one of the most central in the philosophy of science. While traditionally philosophers were only officially interested in a normative theory of how people *should* evaluate explanations, at a certain level even philosophical theories must make contact with data regarding usage (the legitimacy and importance of this strategy is forcefully defended in Waskan et al., 2014).

On many philosophical theories of explanation, the quality of an explanation should not depend on whether some aspect of the explanation involves a name. The Deductive-Nomological account (Hempel & Oppenheim, 1948), for example, tells us that explanations are deductive arguments featuring natural laws; part of what it means to appeal to formal structure is to abstract away from such details as which aspects of the explanation are furnished with a label. As such, it would be difficult for a Deductive-Nomological theorist to make sense of our data. Though the picture is a bit more complicated on unificationist accounts such as Friedman (1974), those accounts generally inherit from Deductive-Nomological accounts the idea that explanations are largely picked out based on their formal structure; this again leaves little room for the explanatory import of names.

Our results are more readily assimilated by accounts of explanation that prioritize the importance of causal relationships (e.g., Woodward, 2003). However, one might worry that the causal structure of the world is — outside of special cases — independent of the way it is described, and so changes in how an explanation is presented should not affect its (perceived) quality. On the other hand, while these accounts of explanation leave little room for a *direct* effect of names on explanations, it could be that names support an inference to some key explanatory content. For instance, names might still be explanatorily relevant to the extent that they

[14] Participants gave significantly higher blame to *Gururumba* ($M = 5.38$, $SD = 1.76$) than *Latah* ($M = 4.00$, $SD = 1.81$), $t(78) = 3.44$, $p < 0.001$, $d = 0.78$, *Ataque* ($M = 4.40$, $SD = 1.61$), $t(78) = 2.58$, $p < 0.012$, $d = 0.58$, or *Pibloktoq* ($M = 3.63$, $SD = 1.86$), $t(78) = 4.31$, $p < 0.000$, $d = 0.98$. Participants' blame ratings for *Ataque* was just significantly higher than *Pibloktoq*, $t(76) = 1.99$, $p < 0.50$, $d = 0.46$ (corrected for Levene's test). participants were significantly more likely to choose guilty for *Gururumba* ($M = 4.88$, $SD = 1.80$) than *Latah* ($M = 3.43$, $SD = 2.04$), $t(78) = 3.37$, $p < 0.001$, $d = 0.76$, *Ataque* ($M = 3.78$, $SD = 1.89$), $t(78) = 2.67$, $p < 0.009$, $d = 0.60$, or *Pibloktoq* ($M = 2.90$, $SD = 1.68$), $t(78) = 5.08$, $p < 0.001$, $d = 1.15$, and *Ataque* than *Pibloktoq*, $t(78) = 2.19$, $p < 0.031$, $d = 0.50$.

offer additional information about the causal structure of the world, and our findings suggest that this is indeed the case.

In addition to causal accounts, pragmatic accounts that ground explanation in the sort of understanding they produce (e.g., Achinstein, 1983; Wilkenfeld, 2014) would predict that the way an explanation is phrased — and in particular whether it invokes a name — might affect how it is understood and hence how good an explanation it is. Similarly, some "epistemic" variants of causal-mechanistic accounts (e.g., Bechtel, 2008) argue that the proper criteria for explanation evaluation stem from a combination of facts about the world and how those facts are presented. This model, too, makes room for the explanatory import of names. Moreover, while these accounts do allow for a direct impact of names on explanation quality, the accounts could also accommodate more indirect effects. For example, names could support inferences or representational changes of some kind, where it is these inferences or changes that foster the relevant epistemic consequence. To the extent philosophical accounts of explanation are constrained by intuitive judgments about everyday explanations, our findings therefore lend support to some theories of explanations and present a challenge for others.

For psychology, our results extend previous work on the role of category labels while answering new questions about explanation. Consistent with the work of Gelman and Heyman (1999), Yamauchi (2005), and others, we find that a nominal category label supports a variety of inferences, some of which concern the generalizability and causal basis of attributes associated with the category. Also consistent with the suggestion made in Ahn et al. (2013), we find that a category label supports a causal-essentialist construal of the named category, and we go beyond this suggestion by showing an effect of a name that is not confounded with other cues to kind-hood (such as references to a recognized disorder). Additionally, our work extends this prior research in addressing a new question: our primary aim was not to investigate the inferences supported by a label, but whether an explanation with a named category would be judged more explanatory, and if so, why. Our studies succeed in answering these new questions about the scope and nature of explanation.

Our findings also raise interesting questions for future research. First, why does a name license assumptions about a causal basis? Do people use the presence of a name to signal the presence of an essentialized kind, as suggested by Ahn et al. (2013)? If so, under what conditions and why? Our current categories, based on Culture Bound Syndromes, potentially lend themselves to a causal-essentialist interpretation. Ahn et al. (2006) and Cooper and Marsh (2015) find that laypeople tend to think of mental disorders in more essentialist terms than expert clinicians do, and Wilkenfeld, Gleason, and Lombrozo (in preparation) find that mental disorder categories are considered explanatory (e.g. "he hallucinated because he has schizophrenia"), in large part because laypeople assume that such categories reflect a common cause for the disorder across individuals. It could be that the naming advantage that we observe here is restricted to the kinds of categories that support such causal-essentialist interpretations. If this is the case, then we would not expect to find comparable effects for explanations involving named categories that resist essentialist interpretations (e.g., ad hoc categories) or causal interpretations (e.g., mathematical concepts). More generally, it is worth acknowledging the diversity of categories and dispositional properties. Not all dispositions point to intrinsic properties; when dispositions are relational or highly contingent on the environment, they might be less susceptible to the effects reported here.

Second, might our results offer an alternative explanation for the appeal of reductive scientific content in explanations (Hopkins et al., 2016)? Specifically, it could be that reductive explanations are judged more satisfying in part because laypeople take them to offer (more) underlying causal information. Similarly, scientific jargon (whether it appears as a name or in some other form) could function as a cue to underlying causal structure that supports an explanation, no matter that the structure is opaque from the perspective of a naïve judge.

Although our results offer compelling evidence for the importance of causal assumptions in underwriting explanations, it is worth acknowledging some limitations of our studies. First, it could be that causal information improves the perceived quality of an explanation, and that adding a name improves the perceived quality of an explanation, but that these two effects are independent. We find this possibility unlikely given the mediation results from Experiments 1–2 alongside the results of Experiment 3: a name ceased to have an effect once all explanations were explicitly causal, and the explanations did not yield differences on any of our measures (such as stability or generalizability). If the explanatory effects of names and causal information are distinct, it would be surprising to see them so closely aligned along so many dimensions. That said, it would certainly be valuable to test additional reasons why a name could impact explanation quality. One way in which our named and unnamed conditions could differ is in the mnemonic or processing consequences of adding a name. It's also possible that the presence of a name establishes what Prasada and colleagues call a "principled connection" between the category and the relevant behavior, leading to a boost in explanatory satisfaction that results from the construal of the behavior as an aspect of the category (Prasada & Dillingham, 2006, 2009).

It's also important to note that the scope of our conclusions is constrained by the limited range of experimental materials tested, as well as by the restricted participant population (workers on Amazon Mechanical Turk). Testing a wider variety of explanations in a wider variety of contexts with a diverse sample of participants is an important direction for future research. Discovering whether our findings generalize beyond explanations for human behavior, and identifying real-world implications of the naming advantage – be it in moral or legal judgments, or in assessments of credibility or probability – will be important in establishing whether the effect is of practical importance. Finally, we caution that we do not intend our results to be read with any normative implications: a preference for explanations with named categories may or may not lead to more accurate judgments or decisions. That said, understanding this preference does support a more accurate picture of human judgment and decision-making.

In sum, we find consistent evidence that people find explanations that appeal to named categories more satisfying than matched explanations that differ only in omitting the name, and that this is because named categories are more likely to license inferences to the presence of a cause that is responsible for what is being explained. Under the circumstances, we would be remiss if we did not give our explanation for this phenomenon a satisfying name: we call it the Explanatory Effect of a Label (EEL).

Acknowledgements

This work was partially supported by NSF Career award DRL-1056712, a James S. McDonnell Foundation Scholar Award in Understanding Human Cognition, and the Varieties of Understanding project supported by Templeton Foundation.

Appendix A. Experiment 1 stimuli

A.1. Latah: Named

Randy is a 40-year-old male. Recently, he took a beautiful and expensive painting from his office after one of his co-workers said,

"you should take that painting, you're the only one who ever looks at it." Randy's co-worker had not been serious.

It turns out that Randy has Depathapy, a tendency to imitate the actions of others and obey commands directed at them, leading him to take the painting.

A.2. Latah: Tendency

Randy is a 40-year-old male. Recently, he took a beautiful and expensive painting from his office after one of his co-workers said, "you should take that painting, you're the only one who ever looks at it." Randy's co-worker had not been serious.

It turns out that Randy has a tendency to imitate the actions of others and obey commands directed at them, leading him to take the painting.

A.3. Ataque de Nervios: Named

Laura is a 40-year-old female. Recently, she screamed at and hit her boss when he approached her about a project she was working on. Her boss had to be taken to the hospital. Co-workers reported that after the incident, Laura was trembling.

It turns out that Laura has Depathapy, a tendency to tremble and act verbally and physically aggressive, leading her to hit her boss.

A.4. Ataque de Nervios: Tendency

Laura is a 40-year-old female. Recently, she screamed at and hit her boss when he approached her about a project she was working on. Her boss had to be taken to the hospital. Co-workers reported that after the incident, Laura was trembling.

It turns out that Laura has a tendency to tremble and act verbally and physically aggressive, leading her to hit her boss.

A.5. Gururumba: Named

Mark is a 40-year-old male. Recently, he broke into several of his neighbors' houses, taking various items – from napkin holders to vases. When the police found him, he seemed to believe each object was highly valuable.

It turns out that Mark has Depathapy, a tendency to steal objects believing them to be of high value, even though they seldom are.

A.6. Gururumba: Tendency

Mark is a 40-year-old male. Recently, he broke into several of his neighbors' houses, taking various items – from napkin holders to vases. When the police found him, he seemed to believe each object was highly valuable.

It turns out that Mark has a tendency to steal objects believing them to be of high value, even though they seldom are.

A.7. Pibloktoq: Named

Mary is a 40-year-old female. Recently, she was seen in the street without a shirt or pants on. The police took her into custody. At the police station, she began breaking furniture and objects, and tried to run from the building.

It turns out that Mary has Depathapy, a tendency to remove clothing, break furniture, flee from shelter, and perform other irrational or dangerous acts.

A.8. Pibloktoq: Tendency

Mary is a 40-year-old female. Recently, she was seen in the street without a shirt or pants on. The police took her into custody. At the police station, she began breaking furniture and objects, and tried to run from the building.

It turns out that Mary has a tendency to remove clothing, break furniture, flee from shelter, and perform other irrational or dangerous acts

Appendix B. Experiment 2 stimuli

B.1. Latah: Named

Randy is a 40-year-old male. Recently, he took a beautiful and expensive painting from his office after one of his co-workers said, "you should take that painting, you're the only one who ever looks at it." Randy's co-worker had not been serious.

It turns out that Randy has Depathapy, a condition that is defined by a tendency to imitate the actions of others and obey commands directed at them, leading him to take the painting.

B.2. Latah: Condition

Randy is a 40-year-old male. Recently, he took a beautiful and expensive painting from his office after one of his co-workers said, "you should take that painting, you're the only one who ever looks at it." Randy's co-worker had not been serious.

It turns out that Randy has a condition that is defined by a tendency to imitate the actions of others and obey commands directed at them, leading him to take the painting.

B.3. Ataque de Nervios: Named

Laura is a 40-year-old female. Recently, she screamed at and hit her boss when he approached her about a project she was working on. Her boss had to be taken to the hospital. Co-workers reported that after the incident, Laura was trembling.

It turns out that Laura has Depathapy, a condition that is defined by a tendency to tremble and act verbally and physically aggressive, leading her to hit her boss.

B.4. Ataque de Nervios: Condition

Laura is a 40-year-old female. Recently, she screamed at and hit her boss when he approached her about a project she was working on. Her boss had to be taken to the hospital. Co-workers reported that after the incident, Laura was trembling.

It turns out that Laura has a condition that is defined by a tendency to tremble and act verbally and physically aggressive, leading her to hit her boss.

B.5. Gururumba: Named

Mark is a 40-year-old male. Recently, he broke into several of his neighbors' houses, taking various items – from napkin holders to vases. When the police found him, he seemed to believe each object was highly valuable.

It turns out that Mark has Depathapy, a condition that is defined by a tendency to steal objects because the person believes them to be of high value, even though they seldom are.

B.6. Gururumba: Condition

Mark is a 40-year-old male. Recently, he broke into several of his neighbors' houses, taking various items – from napkin holders to vases. When the police found him, he seemed to believe each object was highly valuable.

It turns out that Mark has a condition that is defined by a tendency to steal objects because the person believes them to be of high value, even though they seldom are.

B.7. Pibloktoq: Named

Mary is a 40-year-old female. Recently, she was seen in the street without a shirt or pants on. The police took her into custody. At the police station, she began breaking furniture and objects, and tried to run from the building.

It turns out that Mary has Depathapy, a condition that is defined by a tendency to remove clothing, break furniture, flee from shelter, and perform other irrational or dangerous acts.

B.8. Pibloktoq: Condition

Mary is a 40-year-old female. Recently, she was seen in the street without a shirt or pants on. The police took her into custody. At the police station, she began breaking furniture and objects, and tried to run from the building.

It turns out that Mary has a condition that is defined by a tendency to remove clothing, break furniture, flee from shelter, and perform other irrational or dangerous acts.

Appendix C. Experiment 3 stimuli

C.1. Latah: Named

Randy is a 40-year-old male. Recently, he took a beautiful and expensive painting from his office after one of his co-workers said, "you should take that painting, you're the only one who ever looks at it." Randy's co-worker had not been serious.

It turns out that Randy has Depathapy, a condition that causes a tendency to imitate the actions of others and obey commands directed at them, leading him to take the painting.

C.2. Latah: Causal condition

Randy is a 40-year-old male. Recently, he took a beautiful and expensive painting from his office after one of his co-workers said, "you should take that painting, you're the only one who ever looks at it." Randy's co-worker had not been serious.

It turns out that Randy has a condition that causes a tendency to imitate the actions of others and obey commands directed at them, leading him to take the painting.

C.3. Ataque de Nervios: Named

Laura is a 40-year-old female. Recently, she screamed at and hit her boss when he approached her about a project she was working on. Her boss had to be taken to the hospital. Co-workers reported that after the incident, Laura was trembling.

It turns out that Laura has Depathapy, a condition that causes a tendency to tremble and act verbally and physically aggressive, leading her to hit her boss.

C.4. Ataque de Nervios: Causal condition

Laura is a 40-year-old female. Recently, she screamed at and hit her boss when he approached her about a project she was working on. Her boss had to be taken to the hospital. Co-workers reported that after the incident, Laura was trembling.

It turns out that Laura has a condition that causes a tendency to tremble and act verbally and physically aggressive, leading her to hit her boss.

C.5. Gururumba: Named

Mark is a 40-year-old male. Recently, he broke into several of his neighbors' houses, taking various items – from napkin holders to vases. When the police found him, he seemed to believe each object was highly valuable.

It turns out that Mark has Depathapy, a condition that causes a tendency to steal objects because the person believes them to be of high value, even though they seldom are.

C.6. Gururumba: Causal condition

Mark is a 40-year-old male. Recently, he broke into several of his neighbors' houses, taking various items – from napkin holders to vases. When the police found him, he seemed to believe each object was highly valuable.

It turns out that Mark has a condition that causes a tendency to steal objects because the person believes them to be of high value, even though they seldom are.

C.7. Pibloktoq: Named

Mary is a 40-year-old female. Recently, she was seen in the street without a shirt or pants on. The police took her into custody. At the police station, she began breaking furniture and objects, and tried to run from the building.

It turns out that Mary has Depathapy, a condition that causes a tendency to remove clothing, break furniture, flee from shelter, and perform other irrational or dangerous acts.

C.8. Pibloktoq: Causal condition

Mary is a 40-year-old female. Recently, she was seen in the street without a shirt or pants on. The police took her into custody. At the police station, she began breaking furniture and objects, and tried to run from the building.

It turns out that Mary has a condition that causes a tendency to remove clothing, break furniture, flee from shelter, and perform other irrational or dangerous acts.

Appendix D. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2017.07.011.

References

Achinstein, P. (1983). *The nature of explanation*. Oxford, U.K.: Oxford University Press.

Ahn, W., Flanagan, E., Marsh, J. K., & Sanislow, C. (2006). Beliefs about essences and the reality of mental disorders. *Psychological Science, 17*, 759–766.

Ahn, W., Proctor, C. C., & Flanagan, E. H. (2009). Mental health clinicians' beliefs about the biological, psychological, and environmental bases of mental disorders. *Cognitive Science, 33*, 147–182.

Ahn, W., Taylor, E., Kato, D., Marsh, J., & Bloom, P. (2013). Causal essentialism in kinds. *Quarterly Journal of Experimental Psychology, 66*(6), 1113–1130.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).
Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Taylor & Francis.
Cooper, J., & Marsh, J. (2015). The influence of expertise on essence beliefs for mental and medical disorder categories. *Cognition, 144*, 67–75. http://dx.doi.org/10.1016/j.cognition.2015.07.016.
Dar-Nimrod, I., & Heine, S. (2011). Genetic essentialism: On the deceptive determinism of DNA. *Psychological Bulletin, 137*(5), 800–818. http://dx.doi.org/10.1037/a0021860.
Friedman, M. (1974). Explanation and scientific understanding. *Journal of Philosophy, 71*(1), 5–19.
Gelman, S., & Heyman, G. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science, 10*(6), 489–493.
Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive Psychology, 61*(3), 273–301.
Haslam, N., & Ernst, D. (2002). Essentialist beliefs about mental disorders. *Journal of Social and Clinical Psychology, 21*(6), 628–644.
Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*(2), 135–175.
Heyman, G., & Gelman, S. (2000). Preschool children's use of trait labels to make inductive inferences. *Journal of Experimental Child Psychology, 77*, 1–19. http://dx.doi.org/10.1006/jecp.1999.2555.
Hopkins, E., Weisberg, D., & Taylor, J. (2016). The seductive allure is a reductive allure: People prefer scientific explanations that contain logically irrelevant reductive information. *Cognition, 155*, 67–76.
Hutchison, K. (1991). Dormitive virtues, scholastic qualities, and the new philosophies. *History of Science, 29*(3), 245–278.
Lupyan, G. (2012). Linguistically modulated perception and cognition: the label feedback hypothesis. *Frontiers in Cognition, 3*(54). http://dx.doi.org/10.3389/fpsyg.2012.00054.
Lupyan, G. (2016). The paradox of the universal triangle: Concepts, language, and prototypes. *Quarterly Journal of Experimental Psychology*. http://dx.doi.org/10.1080/17470218.2015.1130730.
Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Labels facilitate learning of novel categories. *Psychological Science, 18*(12), 1077–1083.
Molière (2012). *Le Malade Imaginere* (H. van Laun, Trans.). Mineola, New York: Dover (Original work published in 1673).
Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*(4), 867–872. http://dx.doi.org/10.1016/j.jesp.2009.03.009.
Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition, 99*(1), 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*(3), 401–448.
Trout, J. D. (2008). Seduction without cause: Uncovering explanatory neurophilia. *Trends in cognitive sciences, 12*(8), 281–282.
Waskan, J., Harmon, I., Horne, Z., Spino, J., & Clevenger, J. (2014). Explanatory anti-psychologism overturned by lay and scientific case classifications. *Synthese, 191*(5), 1013–1035.
Weisberg, D., Keil, F., Goodstein, J., Rawson, E., & Gray, J. (2008). The seductive allure of neuroscience explanations. *J. Cogn. Neurosci., 20*(3), 470–477.
Weisberg, D. S., Taylor, J. C., & Hopkins, E. J. (2015). Deconstructing the seductive allure of neuroscience explanations. *Judgment and Decision Making, 10*(5), 429.
Wilkenfeld, D. A. (2014). Functional explaining: A new approach to the philosophy of explanation. *Synthese, 191*(14), 3367–3391.
Wilkenfeld, D. A., Gleason, J., & Lombrozo, T. (2017). Did he hallucinate because he has schizophrenia? The explanatory power of mental disorder categories (in preparation).
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford, UK: Oxford University Press.
Yamauchi, T. (2005). Labeling bias and categorical induction: Generative aspects of category information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(3), 538.
Zosuls, K., Ruble, D., Tamis-LeMonda, C., Shrout, P., Bornstein, M., & Greulich, F. (2009). The acquisition of gender labels in infancy: Implications for sex-typed play. *Developmental Psychology, 45*(3), 688–701. http://dx.doi.org/10.1037/a0014053.

Further reading

Lupyan, G. (2013). What do words do? Toward a theory of language-augmented thought. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 255–297). Cambridge, MA: Academic Press.