



Contents lists available at ScienceDirect

Cognition

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

Original Articles

Normality: Part descriptive, part prescriptive

Adam Bear^{a,*}, Joshua Knobe^{b,c}^a Department of Psychology, Yale University, United States^b Program in Cognitive Science, Yale University, United States^c Department of Philosophy, Yale University, United States

ARTICLE INFO

Article history:

Received 15 January 2016

Revised 9 October 2016

Accepted 29 October 2016

Available online xxxx

Keywords:

Normality

Morality

Learning

Concepts

ABSTRACT

People's beliefs about normality play an important role in many aspects of cognition and life (e.g., causal cognition, linguistic semantics, cooperative behavior). But how do people determine what sorts of things are normal in the first place? Past research has studied both people's representations of statistical norms (e.g., the average) and their representations of prescriptive norms (e.g., the ideal). Four studies suggest that people's notion of normality incorporates both of these types of norms. In particular, people's representations of what is normal were found to be influenced both by what they believed to be descriptively average and by what they believed to be prescriptively ideal. This is shown across three domains: people's use of the word "normal" (Study 1), their use of gradable adjectives (Study 2), and their judgments of concept prototypicality (Study 3). A final study investigated the learning of normality for a novel category, showing that people actively combine statistical and prescriptive information they have learned into an undifferentiated notion of what is normal (Study 4). Taken together, these findings may help to explain how moral norms impact the acquisition of normality and, conversely, how normality impacts the acquisition of moral norms.

Published by Elsevier B.V.

1. Introduction

In ordinary life, people often distinguish between the things they regard as normal and those they regard as abnormal. Existing research has explored the downstream consequences of this distinction, and researchers in fields as diverse as linguistics and behavioral economics have examined the ways in which representations of normality play a role in people's cognition (e.g., Cialdini, Reno, & Kallgren, 1990; Dowty, 1979; Peysakhovich & Rand, 2015; Yalcin, 2016).

A further question arises, however, as to how these representations of normality are acquired in the first place. We know that representations of normality have numerous important downstream effects, but how is it that people come to regard certain things as normal and others as abnormal?

According to one obvious hypothesis, representations of normality are acquired through a straightforward process of statistical learning. People have a well-demonstrated capacity to pick up information about the statistical properties of their environments, and they can acquire statistical information about central tendencies both through direct observation and through testimony (see,

e.g., Holland, Holyoak, Nisbett, & Thagard, 1986). It might be thought that people's representations of normality are simply the product of this sort of statistical learning process.

We will be arguing, however, that there is actually something more complex afoot. We suggest that people's normality judgments take into account both descriptive considerations (e.g., the statistical notion of the average) and more prescriptive considerations (e.g., what is morally ideal). Thus, when people are trying to determine whether a given thing is normal or abnormal, they will take into account both information about whether it is statistically average and information about whether it is prescriptively ideal.

This hypothesis opens the door to a whole new topic in research on the way people acquire representations of norms. There has already been a great deal of work on representations of descriptive norms and their acquisition through statistical learning (e.g., Gigerenzer & Todd, 1999; Gweon, Tenenbaum, & Schulz, 2010; Holland et al., 1986; Tenenbaum, Griffiths, & Kemp, 2006; Tversky & Kahneman, 1973), and separately, there has been a great deal of work on representations of moral norms and their acquisition through moral learning (e.g., Bandura, Ross, & Ross, 1963; Blair, 1995; Cushman, 2013; Henrich & Boyd, 1998; Peysakhovich & Rand, 2015). The present claim is that people also have an undifferentiated representation of what is *normal*. This representation is not purely descriptive or purely moral but rather a hybrid of the

* Corresponding author at: Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06511, United States.

E-mail address: adam.bear@yale.edu (A. Bear).

two, and it is therefore acquired through a process that integrates statistical and moral learning.

1.1. A prescriptive theory of normality

In our daily lives, we frequently need to judge what kinds of things are normal or abnormal. A large body of research has invoked normality to explain important aspects of cognition and behavior. Indeed, the notion of normality has played a central role in disciplines as varied as philosophy (e.g., Goldman, 1986), behavioral economics (e.g., Peysakhovich & Rand, 2015), and linguistics (e.g., Dowty, 1979; Yalcin, 2016). Nevertheless, most of these research programs have focused on the downstream consequences of cognition about normality rather than the nature of normality itself. Little work has explored how, exactly, the norms themselves are represented in the mind. (One notable exception comes from the study of causal cognition; e.g., Halpern & Hitchcock, 2015; Hitchcock & Knobe, 2009; see General Discussion for further discussion.)

In this paper, we explore the more basic question of what normality is. At a first pass, this question might seem to have a fairly straightforward answer: what is normal is just what is typical or average. If, for example, you were wondering what a normal height for a man is, you could just seek out statistics about the distribution of male heights in the population and use that value as your judgment about what is normal.

But we suggest that descriptive considerations are not the only important factor influencing how people think about normality. Rather, *prescriptive* considerations also influence what people think is normal. To see this, consider a slightly different case: suppose you are judging what is a normal amount of TV to watch in a day. Would this normal amount just be what is statistically most common? Or might you actually consider this latter amount to be *abnormally* large, such that it would be more normal to watch an amount of TV that is below the statistical average?

Though there are many ways to operationalize descriptive and prescriptive influences, for simplicity we focus on people's judgments about what is descriptively *average* and prescriptively *ideal*. Importantly, these two judgments often come apart. In many aspects of our everyday lives (e.g., how much television we watch or how much we cheat on our taxes), what we do or what happens on average is not what we regard to be ideal. And we propose that normality judgments are influenced by both average and ideal judgments.

We focus on two possible hypotheses about this dual influence of average and ideal. The first hypothesis is that both descriptive beliefs about what is average and prescriptive attitudes about what is ideal influence normality judgments. In other words, if people change their belief about the average, they should show a corresponding change in their belief about the normal; and, likewise, if attitudes about the ideal change, so should beliefs about the normal. For instance, the amount of TV that people watch per day has likely evolved over time, as has perhaps the amount that people regard as ideal. Our theory predicts that both these kinds of variance should affect what people judge to be a normal amount of TV to watch.

A stronger further hypothesis is that people's normality judgments are specifically *intermediate* between what is believed to be average and what is considered ideal. That is, one might predict not only that normality judgments are influenced both by judgments about the average and by judgments about the ideal, but also that people specifically judge that the normal amount lies *between* the average amount and the ideal amount. This further hypothesis would follow fairly straightforwardly if one assumes that normality judgments are influenced by both the average and the ideal and that they are not influenced by any other factors.

1.2. Three case studies of normality

There are a number of ways to explore how both descriptive and prescriptive factors might influence how people represent normality. Here, we focus on three basic measures: (a) explicit use of the word "normal," (b) use of gradable adjectives like "large" and "small," and (c) judgments about the prototypicality of category exemplars.

1.2.1. Use of the word "normal"

The most straightforward way to test people's views about normality is simply to examine their use of the word "normal." For example, consider once again the question of what is a normal amount of TV to watch in a day. At first blush, people's intuitions about this question might just track what they think is the *average* amount that people watch. But recent work suggests there may be more to the story (Wysocki, 2016).

In one study, participants were presented with a vignette about a college student who harbored certain strong political attitudes. Participants judged how normal they thought these opinions were and also subsequently rated (a) how common these attitudes were and (b) how good or bad they were. Both of these judgments predicted participants' assessments of what was normal, with more common and more positively evaluated opinions being rated as more normal (Wysocki, 2016). Thus, people's use of the word "normal" may integrate both statistical and prescriptive considerations.

Along these lines, we suggest that normality judgments are guided both by what people think is average and by what people think is ideal. Thus, our hypothesis opens up the possibility that people might think that the average amount that people watch is actually abnormally large.

Of course, although just asking people what they think is normal is the most obvious first step in testing a theory about how people represent normality, it is also limited in what it can tell us. Even if these judgments are, in fact, influenced by both descriptive and prescriptive information, we could not be sure that such results would not be explained by some idiosyncrasy in how people use this one word, rather than a richer fact about how they reason about what is normal more generally. We must therefore consider other measures of normality, as well.

1.2.2. Gradability

The topic of gradability has inspired an enormous literature in linguistic semantics and in cognitive science more generally (e.g., Barner & Snedeker, 2008; Kennedy, 1999; Lassiter & Goodman, 2014), and although controversies remain about certain issues, we now have at least some understanding of how this phenomenon works. The class of gradable predicates includes expressions like 'large,' 'hot,' 'loud,' 'fast,' 'difficult,' and many others. Existing research shows that there is an abstract level at which these different expressions can be understood as being semantically similar.

Roughly speaking, a gradable predicate allows one to characterize an entity in terms of degrees along a scale. For example, entities can be understood in terms of a scale of size, and to the extent that an entity goes beyond a certain threshold on this scale, it can be described using the gradable adjective 'large.' In much the same way, entities can be understood in terms of a scale of temperature, and to the extent that an entity goes beyond a certain threshold on this latter scale, it can be described as 'hot.' The threshold used for the interpretation of gradable predicates like these is usually known as a *standard* (Kennedy, 1999).

One obvious fact about the standard is that it depends in part on the class of entities one is considering. For example, the standard people would use to determine whether an entity is a 'large beetle' is very different from the standard they would use to determine

whether an entity is a ‘large planet.’ A question, then, arises as to how people determine the relevant standard for a given class of entities.

Existing research has shown that people’s intuitions about the relevant standard are determined in part by *descriptive* considerations (Barner & Snedeker, 2008). Thus, when people are trying to determine how large an entity has to be to count as a ‘large beetle’ or a ‘large planet,’ their judgments are influenced in part by beliefs about how large beetles and planets generally tend to be. However, recent studies have also uncovered another effect that is perhaps more surprising. People’s intuitions about the standard can be affected by *prescriptive* considerations (Egré & Cova, 2015). For example, in one study, participants were told that 50% of children died in a fire and 50% survived. Participants were more inclined to agree that “Many children died” than to agree that “Many children survived.” In other words, independent of anything about the actual statistical distribution, people’s intuitions about the standard appear to be influenced by their judgments about whether certain outcomes are good or bad.

We propose that this phenomenon can be explained in terms of normality. That is, when people are trying to determine which point along the scale counts as the standard, they are influenced by intuitions about which point is the *normal* one. Hence, to return to our previous example, suppose that people are trying to determine whether the amount of TV that a given person watches counts as a ‘large amount of TV.’ On the present hypothesis, they do not do so by comparing the amount that he watches to the *average* amount; rather, they compare it to the *normal* amount.

1.2.3. Concept prototypes

Much research over the past several decades, using a variety of methods, suggests that we represent categories in a graded fashion, with some exemplars of a category being judged to be more *prototypical* than others (for reviews, see Murphy, 2002; Smith & Medin, 1981). For example, consider the category *grandmother*. Though a 35-year-old barista who has a daughter with children of her own meets the criterion for being a grandmother, there is a sense in which this woman is a worse example of a grandmother than a much older, retired grandmother who lives in Florida.

What factors influence people’s views about prototypicality? Much work on prototype theory has shown that statistical factors, in one way or another, affect these judgments (Rosch & Mervis, 1975). However, some research suggests that prototypicality judgments can be influenced not only by statistical factors but also by prescriptive considerations (Barsalou, 1985; Lynch, Coley, & Medin, 2000).

In a well-known study of this sort, participants rated different category exemplars on a number of dimensions, such as central tendency (how similar they are to other exemplars within that category), frequency of instantiation, and familiarity. In addition, participants rated how well each exemplar fulfilled some goal that was assigned to them for that category (e.g., for an article of clothing, they were asked to rate how necessary it is to wear it). In conjunction with other factors, participants’ evaluations of how well a given category exemplar fulfilled its goal predicted the extent to which that exemplar was judged to be a “good example” of its category. In other words, not only descriptive, but also prescriptive factors influenced prototypicality judgments (Barsalou, 1985).

We suggest that these results can be subsumed under a more general theory of how people represent normality. Specifically, when people are assessing what is the prototypical grandmother, they are not just thinking about what is an *average* grandmother, but they are thinking about what is a *normal* grandmother. Therefore, for the same reason people’s explicit judgments of normality and use of gradable adjectives may be affected both by what is average and by what is ideal, people’s judgments about prototypi-

cality may be so influenced by both of these factors. As a result, a completely average grandmother may actually be judged to be less prototypical than a slightly less average but more ideal grandmother because the latter is considered more normal.

1.3. The present studies

We explore the above three topics using a similar methodology across all of our studies. For each of these phenomena, we ask participants about what is descriptively average and what is prescriptively ideal and then observe how these variables predict views about normality. In each case, we predict that judgments of normality will be impacted both by descriptive judgments and by prescriptive judgments.

Using nonparametric analyses, we further explore whether there is evidence for the stronger claim that judgments of normality are specifically *intermediate* between average and ideal. In Studies 1, 2, and 4, we analyze the proportion of normality judgments that lie on the ideal (versus non-ideal) side of the average and the proportion of normality judgments that lie on the average (versus non-average) side of the ideal. Judgments that are both on the ideal side of average and average side of ideal are intermediate.

The first three studies explore these questions by examining people’s use of the word “normal” (Study 1), judgments about gradable adjectives (Study 2), and ratings of the prototypicality of various concept exemplars (Study 3). A final study then looks directly at how representations of normality are learned by manipulating statistical and prescriptive information given to participants for a novel category (Study 4).

2. Study 1

In this study, we examine how people’s intuitions about average and ideal amounts relate to what they think are normal amounts. To explore this question, we developed a list of specific domains (behaviors, activities, events, etc.). We hypothesized that people’s normality judgments for these various domains would be influenced not only by their statistical beliefs, but also by their prescriptive beliefs.

2.1. Method

Ninety-two participants from Amazon’s Mechanical Turk (47.8% female, $M = 35.6$ years old) were randomly assigned to judge average, ideal, or normal amounts for a set of 20 domains, which were presented randomly on a single page. (We picked domains that we predicted would have judged averages that were significantly different from their judged ideals.) Thus, for all domains, approximately 30 participants were asked questions like “What would you guess is the average number of hours of TV that a person watches in a day?”; another approximately 30 participants were asked questions like “What do you think is the ideal number of hours of TV for a person to watch in a day?”; and the remaining participants were asked questions like “What is a normal amount of hours of TV for a person to watch in a day?”

2.2. Results

Participants’ responses in each condition were averaged for each of our 20 domains (Table 1). No participants failed the attention check. However, 49 individual responses that were 3 standard deviations away from the mean answer for a given question were excluded.

Table 1
Mean average, ideal and normal judgments (from Study 1) and standard judgments (from Study 2) across domains.

Domain	Average	Ideal	Normal	Standard
Hrs TV watched/day	4.00	2.34	3.03	3.83
Sugary drinks/wk	9.67	3.52	7.30	10.08
Hrs exercising/wk	5.37	7.31	6.77	5.49
Calories consumed/day	2159.26	1757.84	2063.33	2007.57
Servings of vegetables/mnth	34.81	67.67	51.97	35.54
Lies told/wk	24.25	2.75	8.43	23.03
Mins doctor is late/appointment	17.78	3.97	18.47	17.79
Books read/yr	10.07	26.15	9.90	10.00
Romantic partners/lifetime	8.04	4.25	8.47	8.02
International conflicts/decade	19.30	1.59	4.82	18.43
Money cheated on taxes	604.56	136.45	636.60	624.82
Percent students cheat on exam	34.64	3.50	15.97	32.46
Times checking phone/day	45.33	13.12	37.17	40.95
Mins waiting for customer service	15.04	5.78	12.73	14.51
Times calling parents/mnth	6.04	6.00	5.23	6.29
Times cleaning home/mnth	5.57	6.75	4.72	5.45
Computer crashes/mnth	4.78	0.50	1.60	3.82
Percent high school dropouts	12.64	3.82	11.13	12.38
Percent middle school students bullied	27.59	2.31	27.26	23.70
Drinks of frat brother/weekend	16.79	5.91	14.30	13.79

2.2.1. Predicting normality from average and ideal

Since our questions asked about very different kinds of quantities (hours, calories, etc.), assumptions of statistical normality were violated. Specifically, all measures were heavily right-skewed (all $p < 0.001$ from a skewness test). To address this problem, mean responses for each measure were converted to (natural) log scale.

To examine how judgments of averages and ideals affect normality judgments, we compared a regression model in which only average judgments predict normal judgments, $F(1, 18) = 228.12$, $r^2 = 0.93$, $p < 0.001$, to a model in which both average and ideal judgments predict these judgments, $F(2, 17) = 225.33$, $r^2 = 0.96$, $p < 0.001$. The latter model revealed that both judged averages, $\beta = 0.70$, $SE = 0.09$, $p < 0.001$, and judged ideals, $\beta = 0.33$, $SE = 0.07$, $p = 0.001$, significantly predicted normality judgments.

Moreover, in addition to explaining more variance, the Akaike Information Criterion with finite-sample correction (AIC_c) for this model (17.99) was lower than that for a model in which only judged averages predict normality judgments (29.18), suggesting that it is a more appropriate model of the observed data. We quantified the strength of the evidence in favor of the more complex model by calculating an evidence ratio based on Akaike weights for the two models, as detailed in Wagenmakers and Farrell (2004). This evidence ratio was 269—a decisive result.

2.2.2. Intermediacy of normality

We next examined the extent to which people's normality judgments were intermediate between average and ideal. For a given judgment to be intermediate, it must be both on the ideal side of the average and the average side of the ideal.

We begin by calculating each of these components separately. First, 75% of items had normality judgments that were on the ideal side of the average, diverging from what would be expected by chance (binomial $p = 0.041$). Second, 95% of items had normality judgments that were on the average side of the ideal (binomial $p < 0.001$). Finally, 70% of normal judgments were on both the ideal side of the average and the average side of the ideal and were therefore intermediate. Given that items could be non-intermediate by being either on the non-ideal side of the average or on the non-average side of the ideal, we compared this observed proportion to the null hypothesis that 1/3 of items would be intermediate by chance. The proportion observed significantly differed from this probability (binomial $p < 0.001$).

2.3. Discussion

In this study, people's use of the word "normal" was best explained by considering both descriptive reasoning (what is considered average) and prescriptive judgments (what is considered ideal). For example, they thought the average amount of TV watched per day was four hours, but nevertheless thought the normal amount was around three hours. This normal amount is intermediate between what people thought is average and what people thought is ideal (a little more than two hours).

Of course, this result may reflect something idiosyncratic about people's use of the word "normal," rather than a deeper truth about people's representations of normality. Thus, we turn to other measures of normality in the studies that follow.

3. Study 2

In this study, we examine whether participants' judged averages and ideals from Study 1 predicted their use of gradable adjectives. We assessed this by asking people the degree to which they thought various quantities relating to the domains of Study 1 were large or small amounts. Based on these ratings, we could estimate the amounts at which participants first switch over to begin regarding a quantity as 'large' in each domain (the standards).

As with the use of the word "normal," we hypothesized that these standard amounts would be predicted not just by participants' estimates of averages, but also by what they judged to be ideal.

3.1. Method

One hundred and one new participants (35.6% female, $M = 33.7$ years old) were presented with a single question about each of our 20 domains from Study 1, presented in random order on a single page. The questions had the following format (again taking the TV domain as our example): "Imagine that a person watches y hours of TV in a day. Please rate the extent to which you think this is a large or small number of hours of TV for a person to watch in a day." The number y was a randomly selected integer between 50% of the average and 150% of the average. Participants responded on a 7-point scale, ranging from "very small" to "very large."

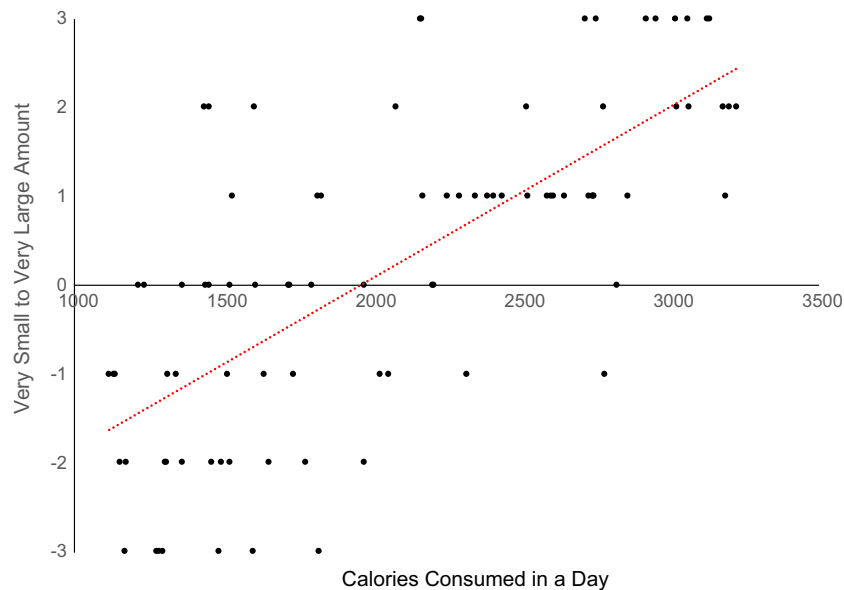


Fig. 1. Participants' ratings of the degree to which various daily calorie amounts are small or large. The standard is estimated to be the point at which the regression line (dotted red line) crosses the x-axis (2007.57 calories). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Results

To calculate standards for each domain, participants' 7-point ratings were mapped into a range from -3 to 3 , such that "very small" corresponded to -3 and "very large" corresponded to 3 . Consequently, the zero point on this scale corresponded to the standard point S , at which some value is judged to be a neither small nor large amount. We estimated this standard point using linear regression according to the following equation:

$$y = b(x - S)$$

where y corresponded to participants' ratings on the -3 to 3 scale, and x corresponded to the randomly queried values that participants were asked about (see Fig. 1).

As before, 15 participants who failed our attention check were eliminated from this analysis. The estimated values for the standard in each domain are shown in Table 1.

3.2.1. Predicting standards from average and ideal

Because the values of these standards were heavily right skewed ($p < 0.001$), we used log-scaled standards. A regression examining the influence of log-converted judged averages and ideals on log-converted standard amounts, $F(2, 17) = 4915.60$, $r^2 = 1.00$, $p < 0.001$, once again found that both averages, $\beta = 0.97$, $SE = 0.02$, $p < 0.001$, and ideals, $\beta = 0.04$, $SE = 0.02$, $p = 0.029$, predicted these standard values.¹ Moreover, in addition to explaining more of the overall variance, the AIC_c for this model (-46.04) was lower than that for a simpler model in which only averages predict

¹ To ensure that the statistically significant influence of ideal on standard values reported here was not simply an artifact of the particular values that subjects were asked about, we conducted a permutation test—shuffling subjects' individual responses and creating new "standards" on the basis of these shuffled values (using the formula described in the Method section). We then reran the regression reported above using 100 variants of these new (log-scaled) "standards" that were generated from subjects' permuted responses. We found that none of these 100 regressions resulted in a coefficient for (the log of) ideal that was equal to or greater than the coefficient we report above for the true standards, and none resulted in a p value for this ideal coefficient that was less than 0.05.

standard amounts (-43.10), $F(1, 18) = 7809.15$, $r^2 = 1.00$, $p < 0.001$ (evidence ratio = 4.36).

3.2.2. Intermediacy of standards

As with our DV measuring people's explicit use of the word "normal," we examined the extent to which standards were intermediate between average and ideal. For a given judgment to be intermediate, it must be both on the ideal side of the average and the average side of the ideal.

Seventy percent of the standards were on the ideal side of the average, which was not statistically different from 50% chance (binomial $p = 0.115$). In contrast, 100% of standards were on the average side of the ideal, which was significantly different from chance (binomial $p < 0.001$). Therefore, 70% of items had standards that were on both the ideal side of the average and the average side of the ideal (and so were intermediate). This proportion significantly differed from the 1/3 probability expected by chance (binomial $p = 0.001$).

3.3. Discussion

This study provides further evidence that people's representations of normality are influenced by prescriptive considerations. As with their use of the word "normal," people's use of gradable adjectives reflected a combination of their judgments of descriptive norms and their judgments of prescriptive norms. For example, the estimated standard amount of television to watch in a day (above which something counted as a large amount to watch) was found to be 3.83 h, which was intermediate between what was judged to be average (4.00 h) and what was judged to be ideal (2.34 h). Thus, people thought that the average amount of television people watch is actually an abnormally large amount.

On the other hand, the effect of ideal on standards in this study was considerably weaker than what was observed with explicit use of the word "normal" in Study 1. This may, in part, be explained by the range of values that participants were asked about (which were always centered on the average). But the weaker effect may also suggest that standards are less influenced by prescriptive norms than explicit judgments of normality. We further explore this question in Study 4, where we measure people's standards after manipulating what is ideal.

4. Study 3

We next examine whether people's prototypicality judgments of a category are influenced by beliefs about both descriptive and prescriptive norms. Specifically, we measured people's judgments about what are average and ideal exemplars of categories using similar methodologies to our previous studies and examined whether both of these judgments influenced judgments about what is a "good," "paradigmatic," or "prototypical" example of that category.

4.1. Method

4.1.1. Participants

Five hundred and forty-two participants from Amazon's Mechanical Turk (45.7% female, $M = 31.3$ years old) completed an online questionnaire.

4.1.2. Materials

Participants were presented with a series of category exemplars and asked to provide ratings for them. In total, there were 8 categories (see Table 2 for list of categories), with 6 exemplars per category. For each category exemplar, participants were provided a short, fictional description. For example, for the category "high-school teacher," one description read, "A 30-year-old woman who basically knows the material she is teaching, but is relatively uninspiring, boring to listen to, and not particularly fond of her job" (see Table A1 for full list). The six different exemplars for a given category were designed to vary independently in how average and ideal they were.

4.1.3. Procedure

Each participant rated one exemplar of each category. Categories were presented in random order. Within each category, each participant was randomly assigned to one of the 8 exemplars. In other words, each participant judged 8 exemplars in total out of a set of 48.

Participants were randomly assigned to judge the exemplars on one of three dimensions: (a) averageness, (b) idealness, or (c) prototypicality. Participants who were assigned to assess prototypicality were, in turn, randomly assigned to judge these exemplars on the degree to which they were a "good example," "paradigmatic example," or "prototypical example" of the category.

The questions from the 5 conditions took the following format and were asked for 8 different categories C (see Table 2):

(Average)	To what extent do you think this is an average C ?
(Ideal)	To what extent do you think this is an ideal C ?
(Prototypicality)	To what extent do you think this is a good example of a(n) C ?
	OR
	To what extent do you think this is a paradigmatic example of a(n) C ?
	OR
	To what extent do you think this is a prototypical example of a(n) C ?

Participants' responded on a 7-point scale ranging from not at all average/not at all ideal/poor example/not at all paradigmatic/not at all prototypical (1) to completely average/perfectly ideal/excellent example/very paradigmatic/very prototypical (7).

Table 2

Effects of average and ideal on normality judgments by category.

Category	Prototypicality composite	
	Average (b)	Ideal (b)
1. High-school teacher	0.64	0.42
2. Dog	0.62	0.23
3. Salad	0.46	0.64
4. Grandmother	0.74	0.29
5. Hospital	0.50	0.29
6. (Set of) stereo speakers	0.48	0.45
7. Vacation	0.51	0.59
8. Car	0.47	0.27

At some random point within the survey, participants were also presented with an attention check that seemed to describe a coffee shop, but then asked them to "please ignore what you just read and answer with the right-most (seventh) response on the question below."

4.2. Results

Participants' responses from each condition were averaged together to generate mean ratings for measures of average, ideal, good example, paradigmatic example, and prototypical example for the 48 category exemplars. Responses from 51 participants, who failed the attention check, were excluded.

The mean ratings on the 3 prototypicality measures ("good," "paradigmatic," and "prototypical" example) were found to have acceptable internal consistency (Cronbach's $\alpha = 0.80$) and were therefore averaged together into a composite prototypicality score.

We examined how ratings about how averageness and idealness of the various category exemplars predicted these prototypicality scores (see Table A2 for mean ratings on each exemplar). As before, we compared regression models in which only averageness influenced these variables to models in which both averageness and idealness played a role.

Judged averages, $b = 0.52$, $SE = 0.02$, $p < 0.001$, and judged ideals, $b = 0.39$, $SE = 0.04$, $p < 0.001$, were both found to significantly predict whether an exemplar was considered prototypical. (All regression analyses used robust standard errors that adjust for eight category clusters, accounting for covariance in within-category ratings. Because of this clustering of errors, standardized regression coefficients were not available, so we report unstandardized coefficients.) As Table 2 shows, this general pattern was observed for all 8 categories.

Moreover, in addition to explaining more of the variance, the AIC_c for this model (63.26), $F(2, 7) = 316.11$, $r^2 = 0.87$, $p < 0.001$, was far lower than that from a model that did not include judgments of the ideal (125.11), $F(1, 7) = 94.67$, $r^2 = 0.50$, $p < 0.001$, suggesting that adding this complexity was justified (evidence ratio $> 10^{13}$).

4.3. Discussion

As with explicit judgments of normality (Study 1) and gradable adjectives (Study 2), here we find evidence that category members are judged to be more prototypical ("good," "paradigmatic," or "prototypical" examples) not only when they are judged to be more average, but also when they are judged to be more ideal. For instance, although the most average grandmother in our sample was "A 70-year-old woman who enjoys baking and reading. Loves her grandchildren, but occasionally gets grumpy and tired and prefers to be by herself," this grandmother was not rated as the most prototypical. A grandmother who "is sweet and pleasant to be around and who enjoys telling stories and knitting in front of

her grandchildren” was rated more prototypical, despite being rated less average.

This finding builds on past work suggesting that category prototypes may have both descriptive and prescriptive components (Barsalou, 1985; Lynch et al., 2000). More importantly, in conjunction with our other studies, it suggests that this past work, which has focused only on concepts, may be explained by a more general theory of how people think about what is normal.

5. Study 4a

Studies 1–3 showed that, across an array of domains, people’s normality judgments are best predicted by a combination of their descriptive and prescriptive judgments. But these studies did not examine how normality judgments are learned and whether this learning is causally influenced by the acquisition of novel descriptive and prescriptive information. Here, we explored this question directly.

We experimentally manipulated the average and ideal sizes of a fictional hunting tool called a “stagnar.” We then examined how the presentation of this information influenced participants’ subsequent judgments about what was a “normal” length for a stagnar. We predicted that participants would acquire a representation of normality that would depend on both the specified statistical and the specified prescriptive information given to them.

5.1. Method

Three hundred and two participants from Amazon’s Mechanical Turk (47.4% female, $M = 42.4$ years old) were presented with the following instructions:

Imagine there is a hunting tool called a “stagnar”, shown in the image below (see Fig. 2).

On the following pages, we are going to show you each of these stagnars one by one. In total, you will see 100 stagnars.

Some stagnars are better than others for hunting. For each stagnar we show you, we will also show you a letter grade on a scale of A–F corresponding to how good that stagnar is for hunting. For example, a stagnar with a letter grade “A” is better for hunting than a stagnar with a letter grade “B.”

Please pay very careful attention, as these images will be up on the screen for only a second each. Afterwards, we will ask you some questions about these stagnars.

Participants were randomly assigned to receive a set of 100 stagnars, whose lengths were sampled from a beta distribution that was either right-skewed ($\alpha = 2$, $\beta = 5$; $M = 0.29$) or left-skewed ($\alpha = 5$, $\beta = 2$; $M = 0.71$). All participants in each of these conditions received the exact same 100 stagnars, and the lengths in the opposite-skew condition were simply reversed (i.e., any given length x was transformed to have length $1 - x$). Lengths were then scaled into a [300,700] pixel range to be presented on the page.

Letter grades were assigned to stagnars in the following way. First, an ideal length was selected independently for each participant by randomly sampling from a set of 101 evenly spaced values between 300 and 700 (i.e., the values were all spaced 4 pixels apart). Stagnars that were within 40 pixels (10% of the total possible range) of the ideal were assigned an “A”; stagnars that were

within 80 pixels were assigned a “B”; and so on. Stagnars that were 160 or more pixels away from the ideal were given an “F”.

The stagnars (and their associated grades) were presented in random order, one second at a time. After all of the images were presented, participants were randomly assigned to answer one of two possible questions. In the experimental condition, they were asked to adjust a slider to make a stagnar on the page look like a *normal* stagnar. In a control condition, participants adjusted this slider to make a stagnar look like an *average* stagnar. This control served to ensure that any influence of the ideal on normality judgments was due to a genuine difference in perceptions of normality and not just differences in statistical inferences about what was average.

To reduce possible demand characteristics, participants in both conditions were also instructed beforehand that they would be asked “one further question.” In reality, this further question (“Based on what we’ve shown you, what letter grade would you assign to the normal/average stagnar you just created above?”) was unrelated to the study, but was included after the key measures in order to avoid deception.

5.2. Results

5.2.1. Predicting normality from average and ideal

To examine whether both the manipulated average and manipulated ideal affected participants’ “normal” lengths, we regressed these normal lengths on averages and ideals. This analysis revealed that both averages, $\beta = 0.23$, $SE = 0.09$, $p = 0.002$, and ideals, $\beta = 0.32$, $SE = 0.06$, $p < 0.001$, significantly predicted normality judgments.

As in past studies, we compared this model, $F(2, 156) = 13.60$, $r^2 = 0.15$, $p < 0.001$, to a simpler model in which only the average predicts the standard, $F(1, 157) = 8.12$, $r^2 = 0.05$, $p = 0.005$. The AIC_c of the more complex model with the ideal (−14.33) was considerably lower than that for this simpler model (1.12), suggesting it is closer to the true model generating the data (evidence ratio = 2273).

To ensure that this effect was not explained by statistical perceptions about what was average, we ran the same regression on control participants’ judgments about the length of an average stagnar. In this case, although the manipulated average predicted judgments about the average length, $\beta = 0.50$, $SE = 0.07$, $p < 0.001$, the ideal was a significant *negative* predictor of what was judged to be average, $\beta = -0.17$, $SE = 0.05$, $p = 0.021$. Thus, the effect of prescriptive information on normality judgments cannot be explained by its effects on encoding or remembering what the average stagnar length was.

5.2.2. Intermediacy of normality

Seventy-three percent of items had normality judgments that were on the ideal side of the average, diverging from what would be expected by chance (binomial $p < 0.001$). Likewise, 74% of items had normality judgments that were on the average side of the ideal (binomial $p < 0.001$). Forty-seven percent of normal judgments were on both the ideal side of the average and the average side of the ideal and were therefore intermediate. The proportion observed significantly differed from the 1/3 probability expected by chance (binomial $p < 0.001$).



Fig. 2. Example of a “stagnar” presented to participants in Study 4.

6. Study 4b

Study 4b was the same as Study 4a, except we measured normality through gradable adjectives, as done in Study 2.

6.1. Method

An additional 389 participants (45.5% female, $M = 32.7$ years old) performed a study identical to Study 4a, except instead of being asked to produce a “normal” or “average” stagner, they were asked to rate (on separate pages) the degree to which 5 different hypothetical stagnars were small or large. These stagner lengths were sampled randomly without replacement from the set of 101 possible lengths described above. Participants gave ratings on a 7-point scale ranging from “Very small” to “Very large” (as done in Study 2).

6.2. Results

Four participants were excluded from analysis because their judgments of size were negatively correlated with the sizes of stagnars presented, suggesting that they were not answering questions in a coherent fashion, such that larger sizes corresponded to larger judgments. Analyses were conducted on the remaining 385 participants. Using the same method from Study 2, each participant's standard stagner size was calculated based on responses to the five questions described above.

6.2.1. Predicting standards from average and ideal

To examine whether both the manipulated average and manipulated ideal affected participants' standards, we regressed standard stagner sizes on averages and ideals. This analysis revealed that both averages, $\beta = 0.37$, $SE = 0.03$, $p < 0.001$, and ideals, $\beta = 0.21$, $SE = 0.02$, $p < 0.001$, significantly predicted standard amounts.

As in Study 2, we compared this model, $F(2, 386) = 43.08$, $r^2 = 0.18$, $p < 0.001$, to a simpler model in which only the average predicts the standard, $F(1, 387) = 62.98$, $r^2 = 0.14$, $p = 0.004$. The AIC_c of the more complex model with the ideal (-470.10) was much lower than that for this simpler model (-452.41), providing strong evidence that it is the better model (evidence ratio = 6939).

Because we “peeked” at the data after 150 participants and decided to run more participants when the effect of ideal on the standards was not significant in a regression with average ($p > 0.1$), we also calculated $p_{augmented}$, which Sagarin, Ambler, and Lee (2014) recommend for researchers who wish to openly disclose that they collected data from additional participants and thereby avoid engaging in questionable research practices. $p_{augmented}$ is always greater than 0.05, but in this case, the analysis yielded the interval [0.0500007, 0.050001], which easily falls within the range Sagarin and colleagues recommend as providing sufficient evidence for a confident interpretation.

6.2.2. Intermediacy of standards

As before, we examined the extent to which standards were intermediate between average and ideal.

A total of 67% of the standards were on the ideal side of the average, diverging from what would be expected by chance (binomial $p < 0.001$). Likewise, 82% of standards were on the average side of the ideal (binomial $p < 0.001$). Fifty percent of standards were on both the ideal side of the average and the average side of the ideal (intermediate). This proportion observed significantly differed from the 1/3 probability expected by chance (binomial $p < 0.001$).

6.3. Discussion

In this two-part study, participants received varied information about the statistical distribution of sizes of a novel object and the ideal size of this object. Their intuitions about both the word “normal” and gradable adjectives were impacted by both of these types of information, suggesting that the relationship between representations of normality and descriptive and prescriptive norms is not simply correlational. Rather, people actually *learn* what is normal by integrating both acquired descriptive and acquired prescriptive facts.

7. General discussion

Four studies explored how people judge and learn what is normal. The results indicated that a mix of descriptive and prescriptive considerations predicted people's intuitions about the proper use of the word “normal” (Study 1), the standard for gradable adjectives (Study 2) and the prototypicality of concept exemplars (Study 3). Finally, Study 4 examined the learning of normality directly and demonstrated that people's acquisition of both descriptive and prescriptive information causally impacts their representations of normality.

The present studies focused on three specific ways of getting at people's normality judgments, but existing research has invoked notions of normality to explain numerous other phenomena, and it might therefore be fruitful to ask whether the effects observed here could be found for those other phenomena as well. Perhaps most importantly, the notion of normality has played an important role in existing work on moral behavior, with research in a number of domains suggesting that people are more inclined to behave prosocially when they regard such behavior as normal (Bicchieri & Xiao, 2007; Cialdini et al., 1990; Peysakhovich & Rand, 2015). Then, outside the domain of moral psychology, there are also numerous areas in which research has made use of the notion of normality. This notion has appeared in work on everything from linguistic semantics (Yalcin, 2016) to philosophical epistemology (Goldman, 1986) to causal cognition (Halpern & Hitchcock, 2015; Hitchcock & Knobe, 2009). Future research could explore each of these areas, asking whether the effect observed in each is actually best understood in terms of the undifferentiated notion of normality suggested here.

The present results raise many questions not only about the connection that normality has to other phenomena, but also about the processes involved in learning normality. We discuss some of these questions below.

7.1. Average, ideal, normal

The results of the present studies suggest that people's representations of normality are related both to their statistical judgments and to their prescriptive judgments. A key task for future research will be to understand that relation more precisely.

According to one possible view, the normal is simply a weighted combination of the average and the ideal. That is, it might be thought that the normal is always a point that is intermediate between the average and the ideal, perhaps biased toward one or the other. Although this simple view might turn out in the end to be correct, the present results provide at least some evidence against it.

In particular, we find a general tendency for the normal to be intermediate between average and ideal, but we also find certain domains for which this is not the case. For example, in Study 1, the normal was intermediate in 70% of the domains, but it was *not* intermediate in 30%. In subsequent work, we have found

further support for this basic pattern. Specifically, we replicated Study 1 with a larger sample (Bensinger, Bear, & Knobe, 2016). Of the 6 domains that did not show intermediacy in the original study, 3 also did not show intermediacy in the replication. In short, the available data suggest that there truly are domains in which the normal is not intermediate between average and ideal.

This result strongly suggests that people's normality judgments should not be understood simply as a function of the average and the ideal. Some further factor must be playing a role. One possibility is that people's normality judgments are being influenced in part by a factor that is not connected to either statistical or prescriptive considerations but truly is just something else entirely. Another is that even if we look just at the roles of statistical and prescriptive considerations, the impact of these factors is not best understood solely in terms of the average and ideal. Instead, it might be that some other aspect of people's statistical or prescriptive representations is playing a role here (e.g., beliefs about the median, mode or some more complex statistical measure). We are currently investigating these questions further in ongoing studies.

7.2. Learning normality

A second key question for future work is how representations of normality are learned. Existing research has explored questions about how people learn descriptive norms (e.g., Gigerenzer & Todd, 1999; Gweon et al., 2010; Holland et al., 1986; Tenenbaum et al., 2006; Tversky & Kahneman, 1973) and, separately, how people learn moral norms (e.g., Bandura et al., 1963; Blair, 1995; Cushman, 2013; Henrich & Boyd, 1998; Peysakhovich & Rand, 2015). The present studies suggest that we also face a further question, namely, how people acquire undifferentiated representations of normality that integrate these two types of considerations.

One extreme view would be that there is a sense in which people never actually do engage in any learning of normality. Perhaps people simply acquire two separate representations (a representation of the average, a representation of the ideal). Then, whenever they need to make a judgment of the normal, they might do so by integrating these two separate representations at that moment. If people recompute normality each time in this way, it might be that they never actually need to learn and store a representation of normality per se.

Admittedly, the results of the present studies do not rule out this extreme view, but there do seem to be some reasons to reject it. First, representations of normality appear to have a pervasive impact on human cognition. The present studies show that these representations impact three quite different kinds of judgments, and further work will presumably identify others along the same lines. Moreover, the notion of normality is quite frequently invoked in ordinary conversations. People describe behaviors as 'normal' or 'weird,' and these expressions are used far more frequently than purely statistical ones ('average,' 'atypical'). It seems unlikely, though still possible, that such a ubiquitous aspect of our cognition would be recomputed again each time.

Second, people do not always acquire information about the normal indirectly by learning about descriptive or prescriptive norms. Although people may sometimes acquire a representation of normality indirectly by integrating information about the average and the ideal (as in Study 4), it seems that there are also times in which they acquire this information directly, through social learning. People can learn about the degree to which a particular behavior is normal by hearing people describe it using terms like 'normal,' 'weird' or 'strange,' as well as through other linguistic cues explored in the recent semantics literature (Yalcin, 2016). Likewise, exposure to certain popular media, like the show "Mod-

ern Family," can influence people's views about whether, say, homosexuality is normal without necessarily changing their views about how common or how desirable it is. Thus, social learning provides a further reason to think that people are actively storing and updating beliefs about what is normal.

If representations of normality are indeed learned, much more work is needed to get a clear sense of how this learning occurs. Such work would of course draw heavily on existing research in statistical learning and in moral learning, but it would also have to address new questions that do not immediately arise in either of those fields considered separately.

On one hand, we would want to know more about how people integrate information about the average and the ideal into an undifferentiated representation of the normal. Study 4 provides evidence that people can integrate information of these two sorts into a unified representation, but future research could further explore the processes whereby this takes place. On the other hand, we would want to know more about how people can acquire representations of normality directly through social learning. Of course, one way in which we might do this is by attending to cases in which other people use linguistic expressions that explicitly mark some behavior as normal or abnormal, but much of this learning presumably takes place through subtler cues. For example, developmental research has explored the ways in which children learn about norms by watching adults model specific behaviors. This research suggests that even when adults do not explicitly label a behavior as normal, children may come to believe that it is normal when they see an adult modeling it without in any way indicating that it is counternormative or out of the ordinary (Schmidt, Rakoczy, & Tomasello, 2011). Future research could explore these processes in more detail.

7.3. Learning morality from normality

The present studies suggest that moral (and other prescriptive) information can influence how people learn what is normal. Specifically, Study 4 demonstrated that learning what is ideal impacts people's normality judgments (as reflected in their use of gradable adjectives). But we did not address whether learning could also unfold in the opposite direction: do people ever learn what is ideal from what they have learned is normal? For example, somebody watching "Modern Family" might come to believe that homosexuality is more normal than previously thought and, in turn, come to view it as more morally acceptable.

Though empirical evidence for this conjecture is limited, some developmental work is suggestive. When young children are told how to use artifacts or how to play a simple game, they spontaneously intervene in protest when a third party acts in a way that violates what they learned (Casler, Terziyan, & Greene, 2009; Rakoczy, Warneken, & Tomasello, 2008), suggesting that these children infer the existence of a prescriptive norm based on the way things are. Of course, in these cases, the pedagogical language used in these experiments (e.g., "This is how *daxing* goes") could have indirectly signaled some sort of prescriptive norm. Interestingly, though, children have been shown to attribute normativity even when they simply *observe* an experimenter confidently performing some action, and they protest actions that go against what they observe (Schmidt et al., 2011). Moreover, studies that have used explicitly non-normative language have come to similar conclusions: children infer what ought to be the case after simply learning what is the case (Roberts, Gelman, & Ho, 2016). Thus, it seems that, at least in some situations, merely learning that an action is conventional or normal is enough to lead children to believe it is wrong to do something else.

Research on children's tendency to "overimitate" the irrelevant and illogical behaviors of others (Horner & Whiten, 2005) may

Table A1
Study 4 list of passages.

Category code	Exemplar code	Passage
1	1	A 30-year-old woman who basically knows the material she is teaching, but is relatively uninspiring, boring to listen to, and not particularly fond of her job
1	2	A 25-year-old woman who captivates her students with exciting in-class demonstrations, grades assignments with remarkable speed, and inspires all of her students to succeed. Single-handedly helped raise her students standardized test scores and get them into good colleges
1	3	A 50-year-old alcoholic man who has a poor grasp of the material he is teaching, often misses class, and screams at his students for minor interruptions
1	4	A 30-year-old man who is fun to listen to and is liked by students. Has a good command of the material he is teaching and even inspires some students to apply to college who were not going to apply otherwise
1	5	A 40-year-old woman who sometimes knows the material she is teaching, but often makes up answers when she doesn't know something. Her students find her boring and don't learn very much from her class
1	6	A 75-year-old man who has a reasonably good grasp of the material he teaches and is generally liked by his students. Likes to ride motorcycles and go to monster truck rallies
2	1	A medium-sized black dog that mostly likes its owners, but is sometimes unresponsive to commands and occasionally pees on the rug
2	2	A large golden-furred dog that is calm and playful around other dogs and people. Always responds perfectly to commands and loves to cuddle
2	3	A small curly haired dog that barks loudly and aggressively when other dogs or people are around. Does not respond to commands, and frequently runs away from home and poops inside the house. Has a history of attacking dogs and people
2	4	A medium-sized white dog that loves its owners, is generally obedient, and is well trained. Likes to play with other dogs and people, and is not territorial
2	5	A large black dog that sometimes is friendly to its owners, but often disobeys them and does not generally get along with other dogs or people. Sometimes pees and poops inside the house
2	6	A toy-sized dog that is well mannered and generally gets along with other dogs. Its fur is purple, and it has gigantic ears. Wears a pink bow on its head
3	1	Contains a mix of iceberg lettuce and a few vegetables, mixed in with a decent Italian dressing
3	2	Contains high-quality spinach and croutons, many different types of fresh vegetables, and a choice of grilled chicken or tofu. Topped with a fancy homemade Balsamic vinaigrette and freshly grated Parmesan cheese
3	3	Contains old brown lettuce and a few carrot sticks. Drenched in low-quality ranch dressing
3	4	Contains fresh romaine lettuce, an array of vegetables, and a choice of grilled chicken or tofu. Dressed with olive oil and red-wine vinegar
3	5	Contains a small amount of iceberg lettuce and croutons, with a few carrot sticks and some Parmesan cheese. Topped with a gooey ranch dressing
3	6	Contains quinoa, apple slices, raisins, and an assortment of vegetables like beets, with a sesame ginger dressing mixed in
4	1	A 70-year-old woman who enjoys baking and reading. Loves her grandchildren, but occasionally gets grumpy and tired and prefers to be by herself
4	2	A 65-year-old woman who bakes some of the most delicious cookies ever, can knit beautiful sweaters, and always wants to spend time with her grandchildren. Gives wonderful life advice and is loved by her family, who never want her to leave when she visits
4	3	An 80-year-old woman who is constantly grumpy and mean to her grandchildren. Detests spending time with other people, but always demands that her children do favors for her. Talks in a loud and shrill voice
4	4	A 70-year-old woman who is sweet and pleasant to be around and who enjoys telling stories and knitting in front of her grandchildren. Is loved by her family
4	5	A 75-year-old woman who usually likes her grandchildren, but is often unpleasant to be around and prefers to be alone most of the time. Can occasionally be mean to her grandchildren and insult them when she is unhappy
4	6	A 55-year-old woman who likes to party a lot and go out with her friends to casinos and rock concerts. Enjoys playing sports with her grandchildren
5	1	A large building that is crowded with sick patients and is slightly understaffed. The nurses keep accurate records and are generally in control of things, but wait times, especially in the emergency room, tend to be long
5	2	A pristine building in a quiet, beautiful area overlooking the mountains. Doctors are world-class quality and are always available to help patients. Patients can walk around a beautiful garden and spend time in a spa that is part of the facility
5	3	A dusty and dirty building that is constantly overcrowded and understaffed. Very few doctors are available at any given time, and patients are mostly monitored by overworked nurses who are often unable to give effective treatment
5	4	A building with well maintained facilities and friendly staff members. Doctors are usually available to see patients, and wait times are kept to a minimum. Patients report receiving good treatment
5	5	An ugly building with old facilities. Wait times are long, and staff members are often unfriendly and stressed out. Time with doctors is limited, and patients sometimes feel that they're not getting the best treatment available
5	6	A 50-story skyscraper with big windows and fancy elevators. Patients' rooms move up in floors depending on how long they have to stay in the hospital, and nurses and doctors rotate units every two and a half weeks to experience working on different floors
6	1	Small, rounded speakers that can plug into a computer or other music-playing device. Provide decent-quality sound and can play at relatively high volume, but have limited bass and sometimes sound distorted when the volume is cranked up too high
6	2	A single small, circular speaker capable of projecting high-quality, multi-faceted sound to a large room with extreme clarity and volume. Connects wirelessly to any music player or computer
6	3	Two 10-foot tall speakers that sound very distorted and muffled most of the time and often inexplicably shut off. Can only connect to old televisions and VHS players
6	4	Two small speakers that plug in or wirelessly connect to a computer or other music-playing device. Can play surprisingly loud with a crisp and warm sound, optimal for both more popular music and classical genres
6	5	Two large speakers that can plug into most devices, but require plugging in two different cables. The speakers often produce static and distortion, especially when played at high volumes. Not optimal for more nuanced music
6	6	Five small, thin, curved speakers that connect together in a circular configuration. Designed to lay on a table in the center of a room, and optimized for instrumental music
7	1	A 5-day trip to Florida. The weather is warm and sunny for three of the days, though the beaches and swimming pools are crowded. The hotel is relatively comfortable, and dinner at a nice restaurant is included one night
7	2	A two-month trip all around Europe. Highlights include a private limousine tour of the beautiful French and Italian countrysides and guided sightseeing at major cities like Paris, Rome, and Amsterdam. Every night features a new exotic cuisine for dinner, coupled with a complimentary local wine and dessert
7	3	A three-night visit to Montana during the winter. The weather is very cold, and the motel room is musty and cramped. The food is

Table A1 (continued)

Category code	Exemplar code	Passage
7	4	mediocre, and movie theaters and bowling alleys provide the only entertainment A two-week trip to Hawaii. Includes tours of the volcanoes and vacationing on the beach. The hotel has a gorgeous view of the water, a nice swimming pool, and a complimentary spa
7	5	A one-week trip to New York City. The weather is mostly cold and rainy, and the hotel is old and smelly. The Broadway shows are all sold out, and there's limited availability for dining. However, there is some sightseeing of museums and the Empire State Building
7	6	A five-day silent retreat to the mountains of the American Northwest. Most of the days are spent hiking and meditating. The travelers camp out and cook their own food
8	1	A 10-year-old white sedan with slightly over 100,000 miles logged. Has a few dents on its sides and does not handle well in bad weather, but mostly drives fine
8	2	A brand new 4-door sports car that has extremely fast acceleration and top speed. Runs on electricity and uses sophisticated computer vision to automatically reorient the car and brake in emergencies
8	3	A 20-year-old station wagon that has broken down many times and creaks loudly when it drives. Sometimes the ignition doesn't work, and the car doesn't start. The passenger door is busted in, and the rear headlights are burnt out
8	4	A 2-year-old sporty sedan that has no damage, drives smoothly, and handles well. Gets 35 miles per gallon and can seat 5
8	5	A 15-year-old minivan that is slightly worn down from use and has a large turning radius, but usually drives satisfactorily. Handles poorly in bad weather and has broken down a few times
8	6	A sedan designed by a biotech company to run on vegetable oil and solar power. The car recycles its own energy to provide heat and air conditioning

provide further basis for thinking that children infer prescriptive content from what they learn is normal. Although some accounts of overimitation have explained this puzzling behavior in terms of brute cognitive limitations (e.g., Lyons, Young, & Keil, 2007), recent work suggests there may be more to the story. Even when it is made salient to them that certain actions are not instrumental to achieving a goal, children still overimitate and protest third parties who fail to overimitate (Kenward, 2012; Keupp, Behne and Rakoczy, 2013). Hence, in these situations, it seems that children may be drawing conclusions about what is prescriptively appropriate on the basis of what they infer is conventional or normal.

How do we explain these surprising developmental patterns? If we started out with the assumption that children simply had two separate kinds of representations – one for descriptive norms, another for prescriptive norms – then we would face a difficult question about how it is that one of these representations could be influencing the other. For example, consider the finding that when children are told that people in a particular social group always eat orange berries, they infer that it would be *wrong* for people in this group to eat a different type of berry (Roberts et al., 2016). If we assume that children just have two completely separate representations, we would have to say that children first conclude that it is frequent for people in this group to eat orange berries (a purely descriptive representation) and that this then leads them to conclude that people in this group ought to eat orange berries (a purely prescriptive representation). Of course, it is possible that the cognitive process at work here proceeds in exactly this way, but at the very least, there is a difficult question about why there should be this sort of link between descriptive representations and prescriptive representations.

By contrast, if we adopt the view that people have an undifferentiated representation of normality, we can begin pursuing a different sort of explanation for results like these. The suggestion would be that children are not simply concluding that it is *frequent* for people in a social group to eat orange berries; they are concluding that it is *normal* for people in this social group to eat orange berries. In other words, the inference we see arising in these developmental findings might be mediated by the undifferentiated notion of normality we have been discussing here. On this account, the children who infer that you should only eat the type of berry that your group eats may be doing so because they have learned that this is what is normal and normality is critically bound up with notions of what is good and bad.

If this broad approach is indeed on the right track, difficult questions arise about how to work it out in detail. One possible view would be that children have three completely separate representations (descriptive norms, prescriptive norms, undifferentiated normality) and that the effects observed in these studies arise from a complex interplay among them. In other words, it could be that children acquire a representation of a descriptive norm, which then leads to a representation of undifferentiated normality, which in turn leads to a representation of a prescriptive norm. An alternative view would be that the ability to keep these three representations distinct is itself something that develops. Thus, it might be that adults find it relatively easy to distinguish purely descriptive norms and purely prescriptive norms but that children find this distinction more difficult and therefore end up relying more heavily on an undifferentiated representation of normality. This latter view would help to explain why children show such a strong tendency to go from the descriptive to the prescriptive while adults show an effect that is greatly attenuated to the extent that it exists at all (Roberts et al., 2016).²

7.4. Conclusion

Existing work, using a variety of methodologies, has explored both how people learn descriptive norms and, separately, how people learn prescriptive norms. The present studies suggest that people may have a representation of normality that takes into account both these kinds of norms. Future research could explore the mechanisms by which people come to acquire this representation and its interconnections with these other types of learning.

Author note

We are grateful to Paul Egré, Susan Gelman, Julian Jara-Ettinger, Louise McNally, Hannes Rakoczy, David Rand, Nick Stagnaro, Pascale Willemsen, and Seth Yalcin for valuable help and for comments on an earlier draft of the present paper.

² Such an account might also provide the beginnings of an explanation for the tendency, observed in some recent studies, whereby adults show an impact of prescriptive norms in judgments about seemingly non-prescriptive questions when answering under speed (Phillips & Cushman, 2016).

Appendix A

See Tables A1 and A2.

Table A2

Study 4 complete results by passage.

Category code	Exemplar code	Average	Ideal	Good example	Paradigm example	Proto. example	Composite
1	1	5.30	2.41	3.30	4.43	4.78	4.17
1	2	3.00	6.56	6.67	5.00	3.09	4.92
1	3	2.14	1.12	1.47	1.92	1.57	1.65
1	4	3.77	6.56	6.75	3.67	4.67	5.03
1	5	3.48	1.48	2.33	3.75	3.29	3.12
1	6	2.46	5.33	3.00	3.89	3.50	3.46
2	1	4.81	3.18	3.50	4.43	5.60	4.51
2	2	5.06	6.69	6.00	3.83	5.46	5.10
2	3	2.31	1.60	2.13	3.63	2.38	2.71
2	4	5.21	6.25	6.10	5.20	5.82	5.71
2	5	3.82	2.20	2.00	4.09	3.86	3.32
2	6	1.32	3.04	1.70	2.67	2.22	2.20
3	1	5.96	4.21	5.00	5.10	6.00	5.37
3	2	3.54	5.39	5.57	5.86	4.91	5.45
3	3	1.90	1.15	1.36	2.58	1.60	1.85
3	4	5.43	5.16	6.20	6.00	6.33	6.18
3	5	4.79	2.83	4.18	4.43	4.92	4.51
3	6	2.35	4.22	5.20	3.82	2.64	3.88
4	1	6.38	5.21	5.77	4.60	5.45	5.27
4	2	5.03	6.30	6.67	5.13	5.10	5.63
4	3	3.23	1.35	2.67	3.08	2.60	2.78
4	4	5.71	6.30	6.60	5.91	5.50	6.00
4	5	3.55	2.41	2.56	3.00	2.64	2.73
4	6	2.57	5.00	4.00	1.17	2.60	2.59
5	1	5.48	3.17	4.78	4.47	5.90	5.05
5	2	1.95	6.27	5.08	3.63	3.15	3.95
5	3	2.58	1.35	2.13	4.80	3.67	3.53
5	4	4.40	6.38	6.60	5.17	4.71	5.49
5	5	3.74	1.35	2.14	3.00	3.70	2.95
5	6	2.56	4.44	3.14	3.14	4.75	3.68
6	1	5.73	3.35	3.80	5.33	5.33	4.82
6	2	3.64	6.32	6.14	5.17	4.29	5.20
6	3	2.15	1.41	1.73	3.33	2.31	2.46
6	4	4.08	5.71	6.13	4.46	5.33	5.31
6	5	3.50	1.62	2.83	3.27	2.58	2.89
6	6	2.50	5.04	4.86	3.60	3.11	3.86
7	1	5.70	4.21	5.43	5.20	5.67	5.43
7	2	2.30	6.19	6.18	4.10	3.36	4.55
7	3	2.90	1.78	2.10	2.89	3.29	2.76
7	4	4.20	6.52	6.83	6.00	5.44	6.09
7	5	3.56	2.70	2.29	3.56	3.22	3.02
7	6	2.74	4.73	5.83	5.40	2.64	4.63
8	1	4.89	2.65	5.00	3.89	4.33	4.41
8	2	2.08	5.64	6.09	2.33	2.55	3.66
8	3	2.50	1.43	2.50	4.36	2.30	3.05
8	4	4.40	6.07	5.80	4.44	5.30	5.18
8	5	3.56	1.85	3.14	3.50	3.80	3.48
8	6	1.42	5.78	5.11	3.18	3.00	3.76

Note. Composite was calculated by taking the mean of Good Example, Paradigm Example, and Prototypical Example.

References

- Bandura, A., Ross, D., & Ross, S. A. (1963). Imitation of film-mediated aggressive models. *The Journal of Abnormal and Social Psychology*, 66(1), 3.
- Barner, D., & Snedeker, J. (2008). Compositionality and statistics in adjective acquisition: 4-Year-Olds interpret tall and short based on the size distributions of novel noun referents. *Child Development*, 79(3), 594–608.
- Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(4), 629.
- Bensinger, S., Bear, A., & Knobe, J. (2016). *Normality judgments and explicit sampling*. Unpublished raw data.
- Bicchieri, C., & Xiao, E. (2007). Do the right thing: But only if others do so. *Journal of Behavioral Decision Making*, 22, 191–208.
- Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57(1), 1–29.
- Casler, K., Terziyan, T., & Greene, K. (2009). Toddlers view artifact function normatively. *Cognitive Development*, 24(3), 240–247.
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6), 1015.
- Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review*, 17(3), 273–292.
- Dowty, D. (1979). *Word meaning and montague grammar*. Dordrecht: Reidel.
- Egré, P., & Cova, F. (2015). Moral asymmetries and the semantics of many. *Semantics and Pragmatics*, 8, 1–45.
- Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. USA: Oxford University Press.
- Goldman, A. I. (1986). *Epistemology and cognition*. Harvard University Press.
- Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, 107(20), 9066–9071.
- Halpern, J. Y., & Hitchcock, C. (2015). Graded causation and defaults. *The British Journal for the Philosophy of Science*, 66(2), 413–457.
- Henrich, J., & Boyd, R. (1998). The evolution of conformist transmission and the emergence of between-group differences. *Evolution and Human Behavior*, 19(4), 215–241.

- Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, 106(11), 587–612.
- Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). Induction: Processes of inference, learning, and discovery. *Computational Models of Cognition and Perception*.
- Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). *Animal Cognition*, 8(3), 164–181.
- Kennedy, C. (1999). *Projecting the adjective: The syntax and semantics of gradability and comparison*. Routledge.
- Kenward, B. (2012). Over-imitating preschoolers believe unnecessary actions are normative and enforce their performance by a third party. *Journal of Experimental Child Psychology*, 112(2), 195–207.
- Keupp, S., Behne, T., & Rakoczy, H. (2013). Why do children overimitate? Normativity is crucial. *Journal of Experimental Child Psychology*, 116(2), 392–406.
- Lassiter, D., & Goodman, N. D. (2014). Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Semantics and linguistic theory* (pp. 587–610).
- Lynch, E. B., Coley, J. D., & Medin, D. L. (2000). Tall is typical: Central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory & Cognition*, 28(1), 41–50.
- Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences*, 104(50), 19751–19756.
- Murphy, G. L. (2002). *The big book of concepts*. MIT Press.
- Peysakhovich, A., & Rand, D. G. (2015). Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science*, 62(3), 631–647.
- Phillips, J. & Cushman, F. (2016). *Multiple systems for modal cognition*. Harvard University. Unpublished manuscript.
- Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology*, 44(3), 875.
- Roberts, S. O., Gelman, S. A., & Ho, A. K. (2016). *So it is, so it shall be: Group regularities license children's prescriptive judgments*. University of Michigan. Unpublished manuscript.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573–605.
- Sagarin, B. J., Ambler, J. K., & Lee, E. M. (2014). An ethical approach to peeking at data. *Perspectives on Psychological Science*, 9(3), 293–304.
- Schmidt, M. F., Rakoczy, H., & Tomasello, M. (2011). Young children attribute normativity to novel actions without pedagogy or normative language. *Developmental Science*, 14(3), 530–539.
- Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press, p. 89.
- Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, 10(7), 309–318.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207–232.
- Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1), 192–196.
- Wysocki, T. (2016). *Normality: A two-faced concept*. Unpublished manuscript.
- Yalcin, S. (2016). Modalities of normality. In N. Charlow & M. Chrisman (Eds.), *Deontic modals* (pp. 230–255). Oxford University Press.