Journal of Experimental Psychology: General
2014, Vol. 143, No. 1, 227–246

0096-3445/14/$12.00 DOI: 10.1037/a0032064

Implicit and Explicit Processes in Category-Based Induction: Is Induction Best When We Don't Think?

Stephanie Y. Chen
New York University

Brian H. Ross
University of Illinois at Urbana–Champaign

Gregory L. Murphy
New York University

In category-based induction (CBI), people use category information to predict unknown properties of exemplars. When an item's classification is uncertain, normative principles and Bayesian models suggest that predictions should integrate information across all possible categories. However, researchers previously have found that people often base their predictions on only a single category. In the present studies, we investigated the possible distinction between implicit and explicit processes in CBI. Predictions of an object's motion took the form of either a catching task (implicit) or a verbal answer (explicit). When subjects made predictions implicitly (Experiment 1), they used categories as Bayesian models predict. Explicit predictions (Experiment 2) showed clearly nonnormative use of categories. This distinction between implicit and explicit processes was replicated with a within-subjects design (Experiment 3). When subjects *learned* categories implicitly (categories were never mentioned) in Experiment 4, their explicit predictions did not reflect integration of information across categories but again showed a nonnormative pattern of category use. These results provide support for a distinction between implicit and explicit processes in CBI and furthermore suggest that the same category knowledge may result in normative or nonnormative responding, depending on the response mode.

Keywords: category-based induction, reasoning, implicit processes

Supplemental materials: http://dx.doi.org/10.1037/a0032064.supp

Our ability to infer information about a novel object based on its category is the basis of much intelligent behavior. It aids us in reasoning, social interactions, communication, and predictions. By categorizing an object, we can make predictions about it even though we have never encountered that particular object before. Imagine yourself as a student at a new high school. If categories did not exist, for each chair you came across you would have to learn how to interact with it, what it does, and what you can do with it. Thankfully, we do have categories, and when the teacher tells you to take a seat, since you know that some of the objects in the room belong to the category of chair, you know where to sit. The same goes for making predictions about people. Because you know that the person ordering you to take a seat belongs to the category of teacher, you can predict that he or she has authority over you. Based on this information, you decide to promptly obey. Your behavior would have likely been quite different if you had instead categorized that person as a classmate (or a class clown).

Induction is not always so simple, especially when categorization is uncertain. Now imagine that you are heading to the parking lot of your new school to cut class. You see someone else in the parking lot, but you cannot quite make out who it is. It could be either a teacher or a student. Do you keep going on your mission to ditch class, or do you run back into school before this person sees you? Based on the few characteristics you can observe, you must make a prediction about whether the person will bring you to the principal's office (which, of course, leads to punishment).

To decide whether the unknown person will bring you to the principal's office, normatively you should take into account both the possibility that this person is a teacher and the possibility that he or she is a student. This type of reasoning is consistent with a variety of normative views, including Bayesian approaches to classification and prediction in which people weight different possibilities by their prior likelihoods. Anderson (1991) proposed such a model of category-based induction,[1] in which the probabil-

This article was published Online First March 18, 2013.

Stephanie Y. Chen, Department of Psychology, New York University; Brian H. Ross, Department of Psychology, University of Illinois at Urbana–Champaign; Gregory L. Murphy, Department of Psychology, New York University.

The research was supported in part by National Science Foundation Grants BCS-1128769 and 1128029. We thank Michael Landy for helpful comments.

Correspondence concerning this article should be addressed to Gregory L. Murphy, Department of Psychology, New York University, 6 Washington Place, 8th floor, New York, NY 10003. E-mail: gregory.murphy@nyu.edu

[1] In all our experiments, the categories are novel and equally probable, so we ignore the prior probability component of Bayesian reasoning. We continue to use the term *Bayesian* because of the common feature of Bayesian models of induction that predictions are integrated across multiple categories, weighted by their likelihood.

ity that an object with observed features, *F,* has an unobserved feature, *j,* is the weighted sum of the probabilities across all categories *k*:

$$P(j|F) = \sum_k P(k|F) \times P(j|k). \quad (1)$$

Thus, if you were a Bayesian class cutter, you would take the probability that the unknown person is a teacher and multiply that by the probability that a teacher would take you to the principal. Next, you would take the probability that the person is a student and multiply that by the probability that a student would turn you in. The sum of the two products is the probability that you get sent to the principal. This approach appears normatively correct, since it takes into account your uncertainty and weights the strength of the prediction according to each category's likelihood. If you are very certain that the person is a teacher, you should make a strong prediction about the likelihood of punishment; if uncertain, you should make a weaker prediction.

Past Research on Category-Based Induction Under Uncertainty

Surprisingly, however, previous research on induction when categorization is uncertain has found that people generally do not normatively integrate information across categories. Malt, Ross, and Murphy (1995), Ross and Murphy (1996), and Hayes and Chen (2008) used vignettes about real-life situations to study category-based induction when categorization is uncertain. For example, subjects read a story that described an unknown person who was most likely to be a real estate agent (the *target* category) but who might have been a cable repairman or—in a different story—a burglar (the *secondary* category). Subjects then predicted the likelihood that the unknown person would show a specific behavior—for example, "What is the probability that the man will pay attention to the sturdiness of the doors on the house?" Because this behavior is more consistent with a burglar than a cable repairman, subjects given burglar as the secondary category should give higher probabilities than those given cable repairman as the secondary category. However, in most conditions subjects paid attention only to the target category (the real estate agent) when making predictions and ignored relevant information from the secondary category.

Similar results have been found with artificial categories in which the information known about the categories can be completely controlled (Hayes & Newell, 2009; Murphy & Ross, 1994, 2005; Verde, Murphy, & Ross, 2005). Murphy and Ross (1994) presented subjects with geometric figures said to have been drawn by children. Each figure had two features: shape and shading. Subjects were told one feature of a new item drawn by one of the children (e.g., shading). They then predicted the other feature (e.g., shape) and rated the likelihood that their prediction was correct. These new items could belong to one of two children (categories). Based on the distribution of items with the given feature, the item was more likely to belong to the target category but might belong to the secondary category. Murphy and Ross (1994) compared a neutral and an increasing condition. The target category was identical in the two conditions, but for the increasing condition, the secondary category increased the probability that the most likely feature from the target category was the "correct" answer. In the neutral condition, looking outside the target category did not change the probability of the predicted feature. Thus, if subjects were integrating information across categories, those in the increasing condition should provide higher probabilities than those in the neutral condition. However, in a variety of experiments, Murphy and Ross failed to find any difference in the probability ratings in the neutral and increasing conditions (see also Verde et al., 2005). This suggested that subjects based their inductions on a single category.

This is not to say that people never integrate information across categories when making inductions. Murphy and Ross (2010) and Murphy, Chen, and Ross (2012) characterized individual subjects according to whether they focused on a single category or integrated across categories in their inductions. Their results suggest that a minority of people integrated across categories consistently (about 25%, depending on conditions), with the large majority using a single category most or all of the time. Additionally, certain category structures and question formats promote integrating predictions across categories (Murphy et al., 2012; Murphy & Ross, 2010). Given the variations in normative responding, the more important research question may be understanding when people do and do not integrate information across possibilities. The present research considers the effects of fundamental differences in the task that might lead to more or less Bayesian responding.

We should clarify that general normative principles suggest that people should attend to category certainty when making category-based inferences and that they should combine information from different possible categories in some way. One does not need a Bayesian model in order to derive this prediction, and our own tests of the prediction have been relatively generic—for example, measuring whether there is any effect of a secondary category, rather than the specific effect predicted by Equation 1. Nonetheless, as we explain in the next section, Bayesian models have dominated the analysis of this issue as well as closely related research in perception and motor control. This makes them of particular interest in our investigation.

Bayesian Responding

The finding that many people do not integrate across possible categories during induction is surprising, given that studies of seemingly more complex problems of perception and motor control often find that people do integrate information across possibilities in a Bayesian manner (Kersten, Mamassian, & Yuille, 2004; Tassinari, Hudson, & Landy, 2006; Trommershäuser, Landy, & Maloney, 2006). In studies of perception, Bayesian models are used to explain how the visual system takes ambiguous inputs and returns the most likely percept. People use knowledge about prior probabilities of states of the world and the likelihood of each state, given the visual stimulus, to arrive at the most probable interpretation of the stimulus (Kersten et al., 2004). In motor control, one action may be best suited to achieve a goal, given the state of the world. But because perception is not perfect, the state of the world is uncertain. Models of action propose that people integrate information about the likelihood of the possible states of the world to make near-optimal actions (Haruno, Wolpert, & Kawato, 2001). These actions are sensitive to the payoff structure of a task: Subjects make motor decisions that minimize costs, given the uncertainty of different motor outcomes and the costs

and benefits associated with each outcome (Trommershäuser et al., 2006; Trommershäuser, Maloney, & Landy, 2008). Trommershäuser, Körding, and Landy (2011) summarized many examples of such integration of options in the perceptual domain.

It is surprising that people are able to integrate across possibilities and weigh costs and benefits in complex perception-action tasks but seem unable or unwilling to combine information about two categories in inductive reasoning—a task that seems computationally simpler. To explain this discrepancy, we appeal to the distinction between implicit and explicit reasoning (Sloman, 1996, but see discussion below). Explicit processes tend to be conscious, relatively slow, effortful, and rule-based, whereas implicit processes tend to be unconscious, fast, easy, and associative. Our suggestion is that category-based induction, as studied in the past literature, is an explicit reasoning task in which people overtly consider different options and use heuristics to derive answers. In contrast, perception-action experiments tend to study fast responses that depend on learned associations in which the different options may not be overtly considered. Thus, how people respond may determine the reasoning used for a task (or, perhaps, whether reasoning is used). Our proposal is that the distinction between implicit and explicit induction is tied to the response mode: Verbal, unspeeded responses such as those used in category-based induction tasks tap into the explicit reasoning processes, whereas fast motor responses such as those used in action and perception tasks are made implicitly.

The differences between the explicit induction task and the motor-perception tasks at first glance seem closely related to the implicit/explicit distinction drawn by Sloman (1996) and the System 1 versus System 2 distinction prevalent in the reasoning and decision-making literatures (e.g., Evans, 2007, 2008; Kahneman, 2011; Stanovich, 2010). However, it is not clear that that distinction captures our particular comparison. Sloman (1996, p. 5) appears to have excluded perceptual tasks from his two systems, as not involving reasoning. Examples of System 1 reasoning include processes that are clearly higher level cognition, such as the availability heuristic and anchoring. Perceptual judgments and motor tasks such as pointing seem more automatic and less cognitively penetrable than such heuristics. Therefore, we use the terms *implicit* and *explicit* as descriptive terms rather than implying that these processes fit within the theoretical scheme of dual-system theories of reasoning. Part of our investigation is to find out in more detail what aspects of these different tasks might be responsible for their different results.

Why should explicit processing lead to non-Bayesian responding in induction while seemingly less sophisticated processing leads to more normative use of information in perceptual predictions? Explicit reasoning is subject to a limitation, called the *singularity principle,* that people generally consider only one hypothetical possibility at a time (Evans, 2007; related to Stanovich's, 2009, claim that people are cognitive misers). This does not mean that people cannot consider more than one category but that they are biased not to, which leads to different computations than does the implicit system. The Bayesian solution to category-based induction involves keeping various possibilities in memory and computing probabilities. Given this, focusing on a single category and disregarding alternatives is attractive, especially when induction problems are complex and involve many possible categories. In contrast, implicit processes may be controlled by summation of associations outside working memory and thus are not subject to the singularity principle. Perception and motor control experiments do not ask people to explicitly consider different possibilities; rather, the potential possibilities are only implicit in the situation. A cup might be 10, 11, or 12 inches away, but people do not overtly evaluate all these possibilities when deciding how much force to exert to reach it. Instead, perceptual representations of these distances are activated proportionally to their consistency with the input and transmitted to the motor controllers (Haruno et al., 2001).

Although various distinctions between explicit and implicit processes have been researched in other forms of reasoning and decision making (Evans & Frankish, 2009; Kahneman, 2011; Kahneman & Frederick, 2002, 2005; Stanovich, 2010) as well as in some key phenomena in category-based induction (Heit & Rotello, 2010; Feeney, 2007; Rotello & Heit, 2009), such distinctions have not been systematically applied to category-based induction when categorization is uncertain. This suggests that deliberate, strategic approaches to induction may lead to fewer "correct" answers than do quick, gut reactions (for similar ideas, see Gigerenzer, 2007; Gigerenzer & Todd, 1999)—at least, when multiple categories are involved. This is an important distinction because, although deliberative predictions are common, inferences sometimes require immediate action, without time to explicitly consider (or exclude) possible categories. Thus, many real-life, category-based inductions are likely made implicitly. In the General Discussion, we consider whether this difference corresponds to two distinct systems of reasoning or is due to only one or two variables (e.g., speed of responding).

There is some evidence that people use categories differently when making implicit and explicit predictions. Shafto, Coley, and Baldwin (2007) taught subjects novel properties about biological categories and found that time pressure weakened inferences to ecologically related categories (e.g., when asking about whether an animal is likely to have the same disease as another animal living in the same environment). Time pressure had no effect when subjects were asked to make the same inferences to a taxonomically related category. As taxonomic categories are usually more readily accessible than ecological ones (Ross & Murphy, 1999), time pressure may restrict analytic processes that select and reason about which categories are appropriate for induction.

There is also evidence from research on induction when categorization is uncertain that time pressure may restrict more strategic category use. Verde et al. (2005) had subjects learn categories of children's drawings like those used in Murphy and Ross (1994). Subjects made predictions in the form of a speeded-verification task or an untimed written response. In the speeded-verification task, subjects were shown one feature (e.g., shape) and they then made a speeded induction as to whether the feature presented after it (e.g., color) was most likely. Verde et al. (2005) found evidence that subjects integrated information across categories when making speeded predictions of this sort—they were faster to respond in the increasing condition than in the baseline condition. Subjects who made their predictions verbally without time pressure, as in Murphy and Ross (1994), did not appear to integrate information (i.e., there was no difference between the likelihood ratings for increasing and baseline conditions). Thus, subjects responding verbally have disregarded information from less likely categories to simplify the induction.

The results of Verde et al. (2005) suggest that quicker responding may lead to different use of categories (though see Newell, Paton, Hayes, & Griffiths, 2010, for a dissenting view). However, the dependent measures of their two tasks were very different and not directly comparable: One was a speeded true-false judgment, and the other was a prediction and probability rating. The present study provides a much stronger test of this account, in which the explicit and implicit tasks are both predictions. We then go further in testing just why these two response modes might lead to different patterns of induction.

These inferences are a key part of our reasoning and often form the basis of decision making. Thus errors in category-based inferences can lead to poor decisions. For example, in the area of medical decision making, if a doctor is unsure of whether a patient has disease A or B, he or she must make a prediction about what treatment will be effective. Focus on a single category could result in poor treatment choices. In fact, this very phenomenon, called *diagnosis momentum* (when an uncertain diagnosis is treated as certain at the exclusion of other possibilities), has been cited as a flaw in medical decision making (Croskerry, 2003).

The distinction between explicit and implicit processing could also help to explain the different conclusions from our past work (summarized above) and that of another line of research (particularly by Tenenbaum and his colleagues), which has made a case for the utility of Bayesian models in categorization and inference (Griffiths & Tenenbaum, 2006; Heit, 1998; Tenenbaum, 1999, 2000). For example, Tenenbaum (1999) presented subjects with a set of values that represented hormone levels of healthy people. They then predicted whether a new value was from the same distribution (i.e., is the new value also healthy?). Tenenbaum's model analyzed this problem as a collection of hypotheses (categories) about the healthy range. They found that people's predictions were fit well by a Bayesian rule that considered each hypothesis, weighted by its likelihood. According to this analysis, subjects are apparently integrating information across multiple categories when making inferences, in contrast to work in our lab with scenarios or visually presented categories.

This normative use of categories may be the result of less explicit thought about categories. The categories in Tenenbaum's (1999) experiment were number ranges that were never presented to the subjects. It seems unlikely that subjects were explicitly considering each possible range—that is, thinking "Is the range 45–55, or is it 47–57, or is it 47–48?" (indeed, there are an infinite number of possible ranges, so they could not all be explicitly considered). In contrast, in experiments in our labs we have always used a small number of categories, which are named or physically presented. It seems likely that subjects explicitly evaluated each of these categories. Thus, the seeming difference in results from the two paradigms could well be due to a difference between explicit and implicit processing, which in turn raises the question of what implicit induction involves. Perhaps, it is not only response mode that leads to implicit processing—the way categories are learned and presented (i.e., whether they are explicitly mentioned) may also lead to implicit processing and normative use of category information. We consider whether implicit induction is determined predominantly by the response mode or also by the way that the categories are learned.

The Present Research

In the present experiments, subjects learned artificial categories of moving geometric figures defined by two features: shape and direction. At test, subjects were presented with a shape and asked to predict its direction either implicitly (Experiment 1) or explicitly (Experiment 2). For the implicit test, we created a novel, game-like motor task that elicited a speeded prediction analogous to those used in perception-action studies (Haruno et al., 2001; Trommershäuser et al., 2006). The explicit test was a formally identical verbal task that tested induction by eliciting a verbal, unspeeded prediction. In Experiment 3, we repeated Experiments 1 and 2 with a within-subjects design to examine whether the same category knowledge can lead to different inductions depending on the mode of prediction. In Experiment 4, we examined the effects of reducing deliberate thought about categories by making learning of categories, rather than the induction itself, implicit.

In sum, these experiments investigate whether the distinction between implicit and explicit processes helps explain when people do and do not use category information in a normative way when making inductions under uncertainty. Our hypothesis is that people will be more likely to use categories normatively when making inductions implicitly, as opposed to explicitly. If this is the case, these results will help explain the apparent discrepancy between perception-action studies, which suggest that people integrate information across various possibilities, and reasoning studies, which suggest that they do not. In these experiments, we also aim to shed light on possible errors in reasoning that affect both everyday and major life decisions and to explain when people are more or less likely to effectively use the information available to them.

Experiment 1

In order to test the hypothesis that people normatively integrate information across categories when making predictions implicitly, we created a novel motor task that elicited a speeded prediction. In Experiment 1, subjects learned artificial categories that consisted of eight moving geometric figures that each had two features, shape and direction of movement. After learning, subjects performed a game-like task in which they caught the moving shapes with their cursor (see online supplemental materials for video of task). They were never explicitly asked to make a prediction, only to catch the objects, and their cursor placement right before the shape moved was used as a proxy for their prediction of direction. Through pretesting, we determined speeds at which it was beneficial for subjects to use category knowledge to catch shapes by placing the cursor along or near the trajectory associated with the categories. If speeds were too slow, subjects could have caught the shapes without having to predict where they would go.

The categories for Experiment 1 each consisted of eight moving geometric figures. There were two critical shapes of interest: squares and hearts. Each of these shapes belonged to two categories, the target category and secondary category. In Condition 1, there was a 66% chance that a square belonged to Category 1 (the target category) and a 33% chance that it belonged to Category 2 (the secondary category). That is, there were eight squares in Category 1 and four in Category 2. In the target category, the critical shape was equally associated with two directions. In the secondary category, the critical shape moved in only one direction,

which was the same as one of the directions from the target category. For example, in Condition 1 half of the squares in Category 1 moved in the 1 o'clock direction, and the other half moved in the 5 o'clock direction. In Category 2, all shapes (including the four squares) moved in the 1 o'clock direction. For hearts, the target category was Category 4. In Category 4, half of the hearts moved in the 11 o'clock direction and half moved in the 7 o'clock direction. The secondary category was Category 3; all its shapes moved in the 7 o'clock direction (see Table 1 for category structure).

Condition 2 served to counterbalance critical directions of the secondary categories. The target categories were exactly the same as in Condition 1, but the secondary categories differed: All the shapes in Category 2 moved in the 5 o'clock direction, and all the shapes in Category 3 moved in the 11 o'clock direction. As a result, if subjects attended to the secondary category, there should be a difference between the predictions made in the two conditions. If they relied only on the target category, there should be no difference between conditions. The question, then, is whether people will shift their predictions depending on what condition they are in (i.e., depending on the less likely category). Such an effect would indicate that people integrate predictions across categories, as in the normative rule of Equation 1.

For the moment, assume that people's predictions (implicit or explicit) tend to be the average observed direction of the moving objects. Because half of Category 1 shapes go to 1 o'clock and half to 5 o'clock, the average prediction for squares based on the target category only would be at 3 o'clock. Similarly, for hearts, the average target prediction would be 9 o'clock. We set these average predictions as 0° and then scored subjects' predictions relative to this zero point. However, people may not have predicted the average of the observed directions but may have instead done something more like probability matching: distributing half their predictions of squares to 1 o'clock and half to 5 o'clock. The mean of such predictions would also be 3 o'clock—our zero point. Therefore, our dependent measure was the predicted direction relative to this zero. (We discuss these two response strategies in the Results.)

Integration of information across categories would be evidenced by a shift away from the direction predicted by attending to only the target category (3 o'clock for squares and 9 o'clock for hearts) toward the direction shared by the secondary category. In Condition 1, attending to the secondary category for squares would mean shifting toward 1 o'clock, and attending to the secondary category for hearts would mean shifting toward 7 o'clock. As Condition 2 serves to counterbalance the directions of the secondary categories, attending to multiple categories in this condition would mean shifting toward the shape's other possible direction (shifting to 5 o'clock for squares and 11 o'clock for hearts).

If implicit prediction does lead to integration of information across categories, we would also expect subjects under time pressure to show greater integration of information, because they would have less time to explicitly consider all the categories and perform calculations. Rather, their response would be pulled toward the secondary category by the association of the shape to the location. To test this hypothesis, we split subjects into two groups, fast and slow. Subjects in the fast group were exposed to the shape for less time than were those in the slow group prior to having to catch it. Thus, subjects in the fast group had less time to employ explicit reasoning strategies and so should show greater integration than the slow group, under the assumption that time pressure encourages implicit processing (Sloman, 1996; Verde et al., 2005).

Method

Subjects. Subjects in all experiments were New York University undergraduates who received course credit for participating. Forty-eight subjects were randomly assigned to conditions in Experiment 1; data were dropped for six subjects who did not follow instructions about cursor placement during the test.

Materials and design. The 2 × 2 design included two between-subject factors: speed of shape movement (fast, slow) and counterbalance of directions (Conditions 1, 2).

Subjects learned four categories of moving shapes. Each category included eight black shapes approximately 1.75 cm to 2.5 cm in length. The category structures (described above) are listed in Table 1. All shapes of the same type were identical (i.e., all squares were identical, all rectangles were identical, etc.). All stimuli were presented on the background of a light gray circle 27 cm in diameter centered on a black computer screen. The stimuli started in the center of the computer screen and then moved off the screen in a straight line, disappearing once they moved beyond the border of the circle. The movement of each shape had a slight random component so that there was variation in each exemplar's direc-

Table 1
Category Structure, Experiment 1

	Category 1		Category 2		Category 3		Category 4	
Exemplar	Shape	Direction	Shape	Direction[a]	Shape	Direction[a]	Shape	Direction
1	Square	1	Square	1/5	Heart	7/11	Heart	7
2	Square	1	Square	1/5	Heart	7/11	Heart	7
3	Square	1	Square	1/5	Heart	7/11	Heart	7
4	Square	1	Square	1/5	Heart	7/11	Heart	7
5	Square	5	Rectangle	1/5	Diamond	7/11	Heart	11
6	Square	5	Rectangle	1/5	Diamond	7/11	Heart	11
7	Square	5	Rectangle	1/5	Diamond	7/11	Heart	11
8	Square	5	Rectangle	1/5	Diamond	7/11	Heart	11

Note. The direction entries are clock directions (1 = 1 o'clock, etc.).
[a] The first number refers to the direction in Condition 1; the second number refers to that in Condition 2.

tion. The precise direction was randomly chosen within a ±2.5° window.

Procedure. The experiment consisted of three phases: observation, learning, and test. A PC presented the instructions and controlled all three phases.

Subjects were told that they would view four categories of moving shapes and were to learn what combination of shapes and directions belonged to each category for a memory test. During the observation phase, all eight shapes from each category were presented singly. Each shape appeared in the center of the screen for 1 s, then moved off the screen in 1.25 s. The name of the shape's category appeared in the center of the screen for the entire time the shape was on the screen. All exemplars from Category 1 were presented, then all exemplars from Category 2, and so on.

Subjects were next told that they would see the same shapes that had been presented in the observation phase. They were to classify each shape into one of the four categories by pressing the number corresponding to the correct category on the keyboard. At the beginning of each trial, a white fixation cross appeared in the center of the screen for 1 s. The shape then appeared in the center of the screen for 1 s and then moved off the screen in 1 s. There was no time limit on responding. After answering, the correct answer appeared for 1.25 s. When a shape was incorrectly classified, subjects viewed a repeat display (without responding) of the moving shape with the correct category displayed. There were four learning blocks in which each of the 32 items was tested in random order. Because of the uncertainty of the critical items' category (e.g., a square could be in Category 1 or 2), subjects could get no more than 75% correct, assuming they chose the most likely category for all presented stimuli. In all experiments, subjects had to reach at least 50% correct during the final block of learning to be included in analysis.

The final phase of the experiment consisted of a 64-trial test in which subjects attempted to catch each shape with their cursor (touch the shape with the cursor before it disappeared from the screen). The cursor appeared as a circle 0.8 cm in diameter (see Figure 1 and online supplemental materials). Subjects were told that they would see the same shapes that they had seen in the previous two phases and that they should use what they had learned about the categories to catch each shape. They were instructed that it was not possible to catch the shape in the center of the screen, so they must place their cursor at the edge of the screen where they thought it would be easiest to catch the shape. Subjects controlled cursor placement and movement with the mouse. At the beginning of each trial, a shape appeared in the center of the screen with the cursor directly on top of it for 0.8 s (fast group) or 2 s (slow group). To prevent subjects from attempting to catch the shape in the center of the screen, effectively not making a prediction about the shape's direction, the shapes momentarily disappeared from the screen after the initial presentation interval and reappeared approximately 5 cm from the center in the direction of movement. For both groups, the shapes moved off the screen in 0.45 s. Subjects were able to move the cursor with their mouse during both the initial presentation of the shape in the center of the screen and while the shape was moving. Subjects saw a 1-s or 2-s feedback message ("Good catch" or "No catch"). Feedback was longer in the fast condition, because the pace of trials was too fast when feedback was only 1 s.

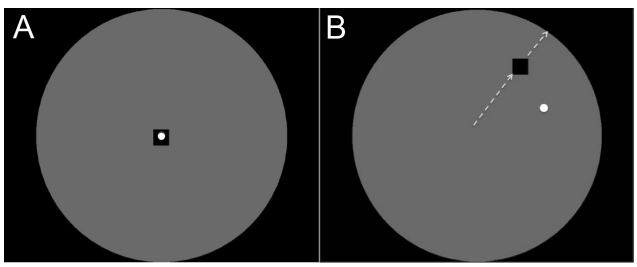

Figure 1. Illustration of the implicit task. The shape appears in the center of the screen with the cursor (the circle) on top of it for either 0.8 s (fast group) or 2 s (slow group), during which the subject cannot catch the shape (A). The shape then momentarily disappears and reappears 5 cm from the center in the direction of its movement and disappears from the screen once it reaches the edge of the gray circle (B). Subjects must touch the cursor to the shape before it disappears. See online supplemental materials for more detail.

Results

Subjects were on average 71.1% correct (chance = 25%) during their last training block, near the 75% maximum, suggesting that they learned the categories quite well. Subjects in the fast group caught 51.4% of trials included in the analysis of cursor position (see below for more explanation of excluded trials), and subjects in the slow group caught 54.6% of these trials. It is difficult to know how to interpret the catching behavior since it involves not just having the cursor in the approximate area (induction) but also some motor variables of no particular interest here. However, it is worth noting that subjects in the fast group did about as well as subjects in the slow group even though they performed a more difficult task (having less time to react to the shape before it disappeared).

Only responses to the two critical shapes (squares and hearts) were included in the main analysis of the test phase. The dependent measure was the placement of the cursor prior to the movement of the shape. As subjects were unable to catch the shape in the center of the screen, any trials where the cursor was within 2.5 cm of the center of the screen were omitted. After these data were omitted, six subjects had responses for fewer than five trials for at least one of the critical shapes. These subjects were dropped from analysis.

Responses were coded such that a position exactly in between the two possible directions of the shape was 0°, and a shift from that point toward the direction reinforced by the secondary category was a positive shift. For example, for the squares in Condition 1 (which might move to 1 o'clock or 5 o'clock), the 3 o'clock position was 0°, a cursor placement at 1 o'clock (which was the direction of the secondary category) was 60°, and a cursor placement at 5 o'clock was −60°. We obtained the mean cursor placement for each subject by averaging the mean cursor placements for squares and hearts. Thus, use of a single category is evidenced by an average prediction of 0°. Normative use of categories is evidenced by a positive average pre-

diction, as this represents a shift from 0° in the direction of the secondary category.[2]

Trials in which the cursor was placed at a position greater than 100° or less than −100° were not included in the analysis, because the cursor was on the opposite side of the screen from where the shape traveled, indicating that the subject either forgot where the shapes went or did not see the shape correctly prior to its movement. (There were 15 subjects with at least one excluded trial. On average, each of these 15 subjects had approximately three excluded trials. Results were similar when such responses were included.)

Cursor placement analysis. As explained above, integration of information across categories is evidenced by a shift from 0° in the direction of the secondary category, which we coded as positive. This is indeed what we found. The average cursor placement ($M = 11.6°$) was significantly greater than 0°, $t(41) = 4.07$, $p < .01$, $d = 0.63$, indicating that people's predictions about the object's motion were integrated across the two categories.

A 2 × 2 (Speed × Condition) analysis of variance (ANOVA) was performed. The main effect of condition and the Speed × Condition interaction were not significant, so Conditions 1 and 2 are collapsed in the following analyses. The main effect of speed was marginally significant, $F(1, 38) = 3.45$, $p = .07$. The fast group's mean cursor placement ($M = 16.8°$, $SD = 18.4$) was significantly different from 0°, $t(20) = 3.72$, $p < .01$, $d = 0.81$. The slow group's mean cursor placement ($M = 6.4°$, $SD = 14.6$) also indicated integration across categories, but the difference from 0° was marginal, $t(20) = 2.00$, $p = .06$, $d = 0.45$. Thus, consistent with our hypothesis, there is a suggestion that the fast group was "more implicit" and showed greater integration of the categories in induction (see Appendix A for details on responses to individual trials).

Performance over time. We asked whether subjects' normative performance might be a result of learning during test by comparing the mean cursor placement from the first block to the second block of test trials. The effect of block was not significant, $F(1, 40) = 0.01$, $p > .05$, suggesting that the amount of shift toward the secondary category did not change over the duration of the test phase. There was no evidence of a difference between the fast and slow groups in this analysis.

Noncritical shape analysis. The design included two shapes, rectangles and diamonds, that were not subject to the experimental manipulations and that did not enter into any of our hypotheses. Nonetheless, we examined their results to ensure that subjects learned about them, showing attention to all the categories. In fact, performance on these items was extremely high—higher than on the test shapes, because they had no category ambiguity. Cursor placement was quite close to the actual location of the shape for both the fast and the slow groups (Ms = 47.5° and 50.4°, SDs = 11.7 and 12.9, respectively; the true direction of motion was 60°). In fact, 88% of all cursor placements for these shapes were within 20° of the correct location. Thus, when subjects were certain of a shape's direction, they moved their cursor close to that location the vast majority of the time.

Individual patterns. Our finding that people shifted toward the secondary category's direction when making implicit predictions might be explained by two different response strategies. First, subjects' predictions could have been like a weighted mean. Cursor placement would have been between the two possible locations toward which the target shape might move but closer to the secondary category's direction. For example, subjects in Condition 1 may have consistently placed their cursor between 1 and 5 o'clock when catching squares but closer to the secondary category's direction of 1 o'clock. Another strategy would be probability matching. If subjects knew that the square could go to either 1 o'clock or 5 o'clock, they could have alternated between placing their cursor at one or the other of these two locations. When the square was more likely to go to 1 o'clock because of the secondary category, they could have placed their cursor at the 1 o'clock location more frequently than the 5 o'clock location. Both strategies would reflect multiple category use.

To investigate which strategy our subjects used, we calculated a center-weighting score: the number of each subject's responses within plus or minus 20° of the subject's mean response (suggesting a weighted mean strategy) divided by the number of total trials. Thus, the center-weighting scores ranged from 0 (all probability matching responses) to 1 (all weighted mean responses). We classified subjects with scores below .5 as probability matchers and subjects with scores above .5 as weighted mean responders. For the slow condition, 17 subjects were probability matchers and four subjects were weighted mean responders. For the fast condition, 12 were probability matchers and nine were weighted mean responders. The mean center-weighting score for the fast group was marginally greater than the mean score for the slow group, $t(40) = 1.88$, $p = .07$ (Ms = .45 and .29). Thus, as the presentation of the shapes (prior to movement) became shorter, subjects tended to go to the same, intermediate spot for a given shape. Perhaps the probability-matching strategy reflects more thought or a more deliberate decision about the direction in which the shape would move. The intermediate position is perhaps the sum of the different associations to the shape, as in motor control models like that in Haruno et al. (2001).

Discussion

The results of Experiment 1 suggest that people integrate information across categories when making predictions in a speeded motor task and that they showed greater integration when under increased time pressure. Taken together, the results provide evidence that people use multiple categories in a normative manner when making predictions implicitly. During the test, categories were not tested or even mentioned—the task was to catch the quickly moving figure. Therefore, this measure of induction is very different from the more usual task in which categories are queried, and subjects have considerable time to choose one to focus on. Across domains, people make category-based inductions under different time pressures, which may lead to differential use of information. As with our class cutter who sees an unknown person in the parking lot, there may not be enough time to consider all the possible categories (e.g., student, teacher, bus driver) and

[2] Note that although the responses for Condition 1 and Condition 2 are coded such that the direction of the secondary category is the same number (60°), these numbers actually represent different responses. For example, for squares in Condition 1, a shift of 30° corresponds to a cursor placement at 2 o'clock. In contrast, for squares in Condition 2, a shift of 30° corresponds to a cursor placement of 4 o'clock. As the only difference between the two conditions is the direction of the secondary categories, this shift suggests that the secondary category influenced predictions.

exclude less likely ones. Because our class cutter must react quickly to avoid punishment, his response may take into account multiple categories. However, if he had more time to think about the unknown person, he might focus on the most likely possibility and continue on his mission to cut class (as there are far more students than teachers).

A key question is what exactly makes an induction task implicit. As previously discussed, experiments on perception and motor control also show that people can integrate information from multiple possibilities. These experiments also tend to use tasks that have an action response, similar to the catching task in Experiment 1. Thus, action selection may be closely associated with implicit processing. We address this question further in Experiment 4.

Experiment 2

The results of Experiment 1 support our hypothesis that people integrate information across categories when making implicit predictions; however, the task we used was different in various respects from the explicit tasks used in our previous work on category-based induction, which consistently found that most people do not use categories normatively. Our purpose in Experiment 2 was to verify that the form of the prediction rather than the category structure, stimuli, or learning procedure accounts for the normative use of categories.

Experiment 2 was identical to Experiment 1 except for the final phase in which subjects made predictions. Instead of catching shapes, subjects viewed static shapes and were asked to verbally state the direction they thought the shape was most likely to travel in. A weakness of Verde et al. (2005) was that the implicit task and the explicit task were quite different. The implicit task involved verification (yes-no judgment) of a presented feature, whereas the explicit task involved producing a feature (the prediction) and a probability rating. In the present experiment, the demands of the tasks corresponded more closely. As we have hypothesized that action may have something to do with what makes a prediction implicit, we wanted to avoid having subjects move a cursor to the direction of their prediction but still allow them to make predictions on the continuous, 360° scale used in Experiment 1. Thus, subjects made verbal predictions using a familiar measure—clock directions. They were asked to imagine the shape's trajectory as the hour hand of a clock and report the time that corresponded to that hour hand (e.g., 5:15 would indicate a position one quarter of the angle between 5 and 6 o'clock).

As in Experiment 1, shifts toward the secondary category were coded as positive values. If subjects were basing their inferences on multiple categories, the mean prediction would be significantly different than 0°. However, we expected to find no such difference, based on previous findings that subjects tend to use categories nonnormatively and focus on a single category when making predictions verbally (Murphy & Ross, 1994; Ross & Murphy, 1996; Verde et al., 2005).

Method

The materials and design were identical to those in Experiment 1. Twenty subjects were randomly assigned to the two counterbalancing conditions. One subject was omitted for failure to categorize any shapes into their target category during the test phase.

The procedures of the observation and learning phases were identical to those in Experiment 1. The test phase of the experiment consisted of a 16-trial test in which subjects were presented with a static shape and asked four questions about it. These are the questions used in most past experiments with this paradigm (Hayes & Newell, 2009; Murphy & Ross, 1994, 2005; Verde et al., 2005), modified to predict direction. Each question was presented on a separate screen. The shape was presented in the middle of the screen directly below the question text:

> Q1: What category do you think the shape most likely belongs to?
>
> Q2: What is the probability that this shape belongs to the category you just identified (0–100)?
>
> Q3: What direction do you think the shape is most likely to travel in? Please input your answer as a time. Please enter the HOUR VALUE (1–12) and PRESS ENTER.
>
> Q4: Now please think about the MINUTE value that corresponds to the direction you expect the shape to travel in. Please enter the MINUTE VALUE (0–59) and PRESS ENTER.

For Q1 and Q2, each screen was identical to that used in the observation and learning phases. Q1 ensures that subjects thought that the target category was most likely. If subjects thought that the secondary category was most likely, a prediction in the direction reinforced by the secondary category could be a result of single-category reasoning. Q2 allows us to be sure that subjects understood that the item's categorization was uncertain. For Q3 and Q4, the numbers from 1 to 12 were presented around the gray circle so that it resembled a clock face. There was no time limit to answer any question. Unlike in the implicit task, then, in the explicit task, there was no moving object, no motor component, and no time pressure. Additionally, predictions were made verbally rather than spatially. There were four blocks in which all four shapes were tested once in random order.

To ensure that people were able to accurately report the shape's trajectory as a time, prior to the test phase subjects completed two practice trials of reporting indicated directions in hour and minute values. All subjects gave an answer within 5 minutes of the correct value on their second practice trial.

Results

Learning performance was again near the maximum: 69% in the last block of learning (chance = 25%). Analysis of responses to the critical shapes is discussed first. Q1 and Q2 were used to verify that subjects obtained similar knowledge about the distribution of the critical shapes in the categories. The likelihood of categorizing the shape into the target category (Q1) was 61.8%. The average probability rating that the shape belonged in the target category (Q2) was 55.7%. Subjects categorized the shape into its secondary category 36.2% of the time, and the average probability that the shape belonged in this category was 45.7%. Subjects rarely categorized the critical shapes into categories to which they did not belong (Categories 3 and 4 for squares and Categories 1 and 2 for hearts). Only one subject made such categorizations. These results show that subjects knew which categories squares and hearts were most

likely to belong to and that their categorization of these shapes into their target category was not certain.

As in the analysis of Experiment 1, the responses from Q3 and Q4 were coded such that the time corresponding to the point exactly in between the two possible directions of the shape was 0° (3 o'clock for squares and 9 o'clock for hearts; a shift toward the direction reinforced by the secondary category was positive). To find the mean prediction (the amount of shift from 0° toward the secondary category) for each subject, we calculated the mean prediction for each shape and took the average of the two. Analyses of explicit prediction data were done twice: once with data including all categorizations (all categorizations analysis) and once with data including only trials in which subjects categorized the critical shapes in their target category (target categorization analysis).

All categorization analysis. The mean prediction ($M = -4.6°$, $SD = 19.1$) was not significantly different than the average observed direction for the shapes in their target category only (0°), $t(18) = 1.0$, $p > .05$, $d = 0.24$. Although the mean prediction is near the center of the two possible directions, this does not mean subjects tended to choose values near 0°. In fact, only one response was within 20° of 0° (see Appendix B).

Target categorization analysis. The above analysis included all categorizations of the critical shapes. This is analogous to the analysis of cursor position in Experiment 1, since in the catching task it was impossible to know what category subjects thought was most likely. Past experiments of this sort, however, have only looked at trials where items are categorized in the most likely category, to ensure that any difference between conditions is a result of integration of information across categories. For example, if subjects in our experiment thought that the object was in the secondary category, a prediction in the direction reinforced by the secondary category could be a result of single-category reasoning. We performed an additional analysis that included only trials in which subjects picked the target category as in all past studies of explicit judgments. The results revealed that the mean prediction ($M = -32.5°$, $SD = 24.6$) was significantly different from 0°, $t(18) = 5.8$, $p < .01$, $d = 1.32$. The negative mean reveals that this effect is in the opposite direction from that of Experiment 1 and that subjects shifted *away from* the direction reinforced by the secondary category rather than integrating the predictions across the two categories as in Equation 1. In other words, subjects seem to be avoiding the features in the secondary category when making their predictions.

Noncritical shape analysis. As in Experiment 1, we examined the results for the shapes that had no category ambiguity (rectangles and diamonds) to ensure that subjects learned and paid attention to the categories. As expected, performance on these items was extremely high. The likelihood of categorizing the shape into the correct category (Q1) was 98.7%. The mean probability rating that the shape belonged in the target category (Q2) was 92.0%. Thus, subjects both knew the category to which these shapes belonged and were confident in their categorizations. The mean prediction for the noncritical shapes was 35.3° ($SD = 27.5$; correct value was 60°). It was clear that subjects learned these shapes' categories and directions.[3]

Discussion

The main finding is that the results from the explicit test in this experiment were very different from the implicit test results in Experiment 1. The implicit test results showed normative integration across categories similar to results from motor prediction tasks, whereas the explicit test results did not. The specifics of the explicit results were somewhat different depending on whether we included all predictions or, as in past research with explicit judgments, analyzed only the trials on which people said that the target category was most likely. The analysis of all trials found no effect of the secondary category, but the analysis of "correct" trials (we use this term without prejudice) found a surprising effect opposite to the expected one.

Bayesian analyses of category-based induction propose that people integrate information about prior likelihoods across categories—that is, the predictions sum across categories. In the past literature, when people showed effects of secondary categories, those effects have always been in the direction predicted by a Bayesian account (e.g., Hayes & Chen, 2008; Hayes & Newell, 2009; Murphy & Ross, 2010; Ross & Murphy, 1996)—shifting the predictions in the direction indicated by the secondary category.

Analysis of trials in which subjects categorized shapes into the target categories, however, showed the opposite pattern: Rather than summing across categories, responses were shifted away from the direction of the secondary category. Features shared between the target and secondary category were predicted less frequently than features only in the target category, rather than more frequently as a Bayesian model would predict. This result does suggest that subjects' predictions are influenced by the secondary categories, because the target categories were identical in Condition 1 and Condition 2. However, it appears subjects were not integrating information across categories but instead avoided predicting features associated with the secondary category. We refer to this as the *avoidance effect*.

This result is a bit confusing, because it suggests that people are influenced by the secondary category when past results with artificial categories have suggested that they focus on a single category. However, a consideration of the entire pattern of results in this condition shows that this difference is not as great as it might appear. It appears that people switched between focusing on the target category and focusing on the secondary category depending on their initial categorization. When subjects chose the secondary category, they chose the only direction associated with it (the direction that all its shapes traveled in); when they chose the target category, they chose the direction that is distinctive to that category—the one not shared with the secondary category, even though only half the shapes go in that direction. When these trials are averaged together, the results are near 0°—the point in between the two possible directions. However, individuals are not actually averaging the two directions; they tend to pick one or the other (see Appendix B for details on responses for individual trials).

[3] In the explicit tests in the rest of the experiments, the classification accuracy of the noncritical shapes was also around 90% or above, and the predictions were also in the correct directions, roughly as above. Because these results do not bear on the hypotheses being tested, we do not provide detailed data on these items for the later experiments.

To explain, imagine that a subject has learned that squares in Category 1 go to 1 o'clock and 5 o'clock and that squares in Category 2 go to 1 o'clock. Apparently, many subjects encode this as "when in Category 1 the shape tends to go to 5 o'clock, and when in Category 2 it tends to go to 1." That is, because 5 o'clock is the distinctive direction of Category 1 o'clock, people tend to choose it over 1 o'clock (which is actually more likely, given the uncertainty of the categorizations) or some kind of average of the two. This results in an "avoidance" of the 1 o'clock direction by our measure. This pattern shows that people are overly influenced by the category they predict that the test object is in (see also Lagnado & Shanks, 2003; Murphy et al., 2012). Further, picking the more distinctive direction associated with Category 1 does not seem very reasonable, given that this direction is not the most likely one in any category, nor overall.

For the present purposes, the most significant point is that the results in the explicit test of Experiment 2 are in contrast to the results from the implicit tests of Experiment 1, where predictions were integrated across categories. In Experiment 2, it seems that subjects did not integrate across the categories but alternated between them, with a suboptimal response in which items in the target category were assumed to go in only one direction. We address this unexpected avoidance effect more fully after we have ensured that it replicates.

One possible explanation for difference in performance between Experiments 1 and 2 unrelated to the response mode is that subjects in Experiment 1 received more exposure to the moving shapes than did those in Experiment 2. Subjects in Experiment 2 never saw the shape move during test. In contrast, during the test phase of Experiment 1, subjects saw the shapes move while performing the catching task and could have learned from this additional information. It seems unlikely that this difference explains the more normative performance of subjects in Experiment 1, as there was no difference in the amount of shift toward the secondary category over the course of the test phase (see Results section of Experiment 1).

It is also worth nothing that subjects in Experiment 2 did not predict directions between the two observed directions. In fact, there was only one response within 20° of 0° (see Appendix B). This was in contrast to the implicit response results of Experiment 1, which found that many subjects did predict such directions. Recall that more than one third of subjects in Experiment 1 used a weighted mean strategy. Subjects who used this strategy placed their cursor between the two possible directions but slightly closer to the one that occurs most frequently. This difference is likely in response to the task goals. Subjects responding explicitly were asked where they think a shape is most likely to travel. Since, for example, they never saw a square move in a 3 o'clock direction, it is not surprising that they did not predict 3 o'clock but rather predicted directions they have seen a square move in (i.e., around 1 or 5 o'clock). However, putting a cursor near 3 o'clock would be useful for subjects responding implicitly, given that their goal is to catch the square, similar to motor control picking a response that is the weighted average of the possibilities (Haruno et al., 2001). Cursor positions near the center ensure that the cursor is somewhat close to both possible directions. Thus, it seems that induction strategies are sensitive to task goals.

Experiment 3

In Experiments 1 and 2, people made very different inductions depending on how they made predictions—via motor response or verbal prediction. Different people served in the two experiments. In real life, the same category knowledge could serve as the basis for both explicit and implicit inductions on different occasions. For example, imagine a hypothetical case in which something is flying toward a person's head at dusk. The object cannot be completely identified but seems to have wings. What should the person do? If the person had enough time to think about it calmly, she might evaluate the possibility that the object is a bird or a bat or perhaps a thrown object and then focus on the most likely possibility (as readers might be doing now, thinking, "it's probably not a bat"). However, in real time, excluding some possible categories might take more time and effort than are immediately available, and her response might take into account multiple such categories (e.g., dropping to the ground rather than swatting away something that she doesn't want to touch). Thus, the same categories could lead to different responses in the same person depending on the task demands on that occasion.

This reasoning assumes that the critical variable determining the different inductive patterns is the mode of response—untimed considered judgment versus immediate reaction. Experiment 3 investigated this question by testing each subject on both implicit and explicit inductions within the same test session, using the same categories. This provides a stringent test of the hypothesis that response mode determines how alternative categories are used, and it also tests the flexibility of people's predictions. It will be interesting to discover whether people will make such "opposite" inductions about the same categories only minutes apart.

Method

Subjects. Forty-one subjects were randomly assigned to conditions and orders. Data from four subjects who did not follow instructions about cursor placement during the testing phase were dropped. Two subjects were dropped from analysis for not categorizing any of the critical shapes into their target category during the explicit task. One subject was dropped for failing to learn the categories.

Materials and design. The design included two between-subjects factors, condition and order (explicit first vs. implicit first), and one within-subjects factor, test mode (implicit vs. explicit). The category structure and stimulus materials were identical to those in Experiments 1 and 2.

Procedure. The experiment consisted of four phases, (a) observation, (b) learning, (c) implicit test, and (d) explicit test, with the latter two varying in order across subjects. The procedure for the observation and learning phases were identical to those of Experiments 1 and 2. The test phases followed the procedures of Experiment 1 (fast group, implicit) and Experiment 2 (explicit).

Results

Subjects were correct on 67.6% of learning trials in the last block (chance = 25%). For the explicit condition, the likelihood of categorizing the shape into the target category (Q1) was 64.3%. The average estimate of the probability that the shape belonged in

the target category (Q2) was 53.3%. Subjects categorized the shape into its secondary category 34.6% of the time, and the average probability that the shape belonged in the secondary category was 50.2%. Subjects rarely categorized the critical shapes into categories to which they did not belong (only 2 subjects made such categorizations). For the implicit condition, subjects caught 40.2% of trials included in the prediction analysis.

All categorizations analysis. No order effects were found, suggesting that any strategies used during the first test did not change the way subjects used categories in the second test. The main effect of test mode was significant, $F(1, 32) = 13.41$, $p < .01$, revealing that subjects used categories differently in the implicit and explicit conditions. The mean prediction for the implicit condition ($M = 14.5°$, $SD = 17.9$) was significantly different from 0°, $t(33) = 4.7$, $p < .01$, $d = 0.81$, but not for the explicit condition ($M = 0.7$, $SD = 20.7$), $t(33) = 0.21$, $d = 0.04$. These results suggest that subjects were integrating information across categories in the implicit condition but not in the explicit condition. (As in Experiment 1, for the noncritical categories, predictions were within 20° of the correct location 85% of the time, indicating that all categories were learned.)

Target categorization analysis. When only explicit condition trials in which subjects categorized squares and hearts into their target categories were included in the analysis, the mean prediction ($M = -15.8°$, $SD = 27.2$) was significantly different from 0°, $t(33) = 3.4$, $p < .01$, $d = 0.58$. The negative mean reveals that subjects were avoiding picking the direction reinforced by the secondary category when they categorized the critical shapes into their target categories. The main effect of test mode was significant, $F(1, 32) = 32.4$, $p < .01$, revealing that subjects used categories differently in the implicit and explicit conditions. As in Experiments 1 and 2, subjects were (normatively) shifting toward the direction reinforced by the secondary category when making predictions implicitly but were (counternormatively) shifting away from the direction of the secondary category when making predictions explicitly.

This reversal is not an artifact of averaging across subjects, as many individuals showed opposite patterns of induction (integration vs. avoidance) across the two responses. Recall that integration of category information in the implicit task would lead to a positive number, and avoidance in the explicit task would lead to a negative number. Eighteen subjects showed exactly this pattern of response, integrating in the implicit task and avoiding in the explicit task. A weaker form of the task differences would be finding a larger (more positive) prediction in the implicit than in the explicit condition. Another nine subjects showed this weaker form of the task differences. Thus, 27 of the 34 subjects showed evidence of different patterns of induction in the two tasks performed one after the other, suggesting that people's explicit and implicit inductions may be based on different processes even when drawn on the exact same category knowledge.

Discussion

Experiment 3 provided three important results. First, we replicated both the implicit test results from Experiment 1 and the explicit test results from Experiment 2. Subjects normatively integrated information in the implicit task and acted nonnormatively in the explicit task. As in Experiment 2, subjects' explicit predictions of direction were tied to their categorization. When they chose the secondary category they tended to predict the direction associated with the secondary category; when they chose the target category they tended to predict the direction distinct to this category (leading to the avoidance effect). Second, the effects of the implicit and explicit conditions occurred within subjects. The same subject integrated information across categories during the implicit task, even though just moments before (or moments later) he or she segregated information from the two categories. Third, there was no carryover from one task to the other. Performance was about the same on each task, whether it occurred before or after the other task. Taken together, these findings provide strong evidence for the separability of implicit and explicit category-based induction and for the importance of the response mode in determining whether knowledge is integrated across categories when making predictions.

Experiment 4

The results thus far have suggested that implicit predictions integrate information across categories. A critical issue is to understand what aspect of the action task might be leading to these very different results from our usual findings. Perhaps reducing deliberate thought about categories by another means, the implicit learning of categories, would also have this effect. Brooks, Squire-Graydon, and Wood (2007) found that when people acquired categories by learning how different pieces moved in a game on a chessboard, their implicit classifications during the game were more accurate than their verbal descriptions of category properties. Subjects learned two categories of moving pieces by concentrating on how each piece moved (rather than explicitly memorizing exemplars of each category). One category could only move diagonally, and the other could only move in straight lines. The pieces in the two categories had a family resemblance structure, without defining features. When subjects had to implicitly categorize the items to decide how many moves it would take a piece to reach a destination square, they were very accurate. However, the same subjects later often incorrectly claimed that the categories had defining features, suggesting a division between their explicit, verbally stated knowledge and implicit categorization.

As Brooks et al. (2007) combined implicit learning with implicit testing (predictions of moves rather than categorization), these two aspects cannot be separated in their task. Our earlier experiments used explicit learning and compared implicit and explicit testing. Implicit testing led to Bayesian integration across categories. To test whether implicit *learning* of categories also leads to integration of information across categories, we had subjects learn the categories either implicitly or explicitly and then take an explicit test. The explicit learning group learned the categories used in Experiments 1–3 via a classification task, as before. For the implicit learning group, categories were never mentioned, and the object's color replaced category membership—all Category 1 shapes were the same color, all Category 2 shapes were the same color, and so on. Thus, the colors provided the same predictive information as the category labels in the explicit learning condition. Each color was associated with certain shapes and directions. For example, all Category 1 shapes were green, so in Condition 1, a green shape was always a square and moved in a 1 o'clock direction half the time and a 5 o'clock direction half of the time.

All Category 2 shapes were brown, so a brown shape always moved in the 1 o'clock direction and was a rectangle half the time and a square half the time. Subjects in this group learned the categories by attempting to catch the shapes. Even though subjects were never asked about it, paying attention to color was necessary to the task as it provided information about where the shape would move, just as category membership did for the explicit learning group.

Although color provided subjects with useful information for performing the catching task, learning of the (color) categories can be considered implicit, since subjects were never instructed to pay attention to or asked about the colors. In contrast, the explicit learning group was trained to classify the shapes, so the categories were an overt part of the learning task. This is not to say that subjects are not able to explicitly think about color during learning. Since color had predictive information about direction, it is likely that they did. However, we assume that explicit thought about color is significantly less than thought about categories in the explicit group, where category learning was the task.

During the test phase, both groups performed the explicit prediction task from Experiments 2 and 3. Recall that in this task, shapes were presented without the category label, so that there was uncertainty about the category to which squares and hearts belonged. The explicit learning group performed this version of the task. The implicit learning group performed the same task except that, because color played the same role as category for this group, color was removed to create uncertainty. The main question of interest was whether implicit learning leads to a similar effect as does implicit responding at test. If implicit learning also promotes integration of information across categories, predictions of direction should be shifted toward the direction reinforced by the secondary category as the cursor placements in the implicit test were.

Method

Subjects. Twenty-eight subjects were randomly assigned to conditions. Data were not included for four subjects who failed to categorize any of the critical shapes into their target category during the explicit task. One subject was dropped for consistently predicting that hearts would travel toward the right side of the screen (in fact, they went to the left) during the explicit task (this was the same 100° and –100° criterion used in the implicit task in Experiments 1 and 3).

Materials and design. The design included two between-subjects factors, condition and learning mode (explicit vs. implicit). The category structure and stimulus materials were identical to those in Experiments 1 through 3 except for the shapes' color. For the explicit learning group, the shapes were gray with a black outline. For the implicit learning group, during observation and learning there was no mention of categories, and category label was replaced with color (all Category 1 shapes were green, all Category 2 shapes were brown, all Category 3 shapes were blue, and all Category 4 shapes were red). During test and post-test (when subjects were told that the shapes had their color removed), the stimuli were the gray shapes with black outlines used for the explicit learning group. Thus, the two groups were equated in that they saw the identical test items and the color/category information was not present during testing.

Procedure. The experiment consisted of four phases: observation, learning, explicit test, and post-test. During the observation phase, the explicit learning group viewed each exemplar singly on the computer screen with its category label presented in the center of the screen. Each shape appeared in the center of the screen for 1 s, then moved off the screen in 1.25 s. All Category 1 exemplars were presented, followed by all Category 2 exemplars, and so on. The observation phase for the implicit learning group was identical to that of the explicit learning group, except that there were no category labels and the shapes were presented in the colors that corresponded to their category—all green (Category 1) shapes, followed by all brown (Category 2) shapes, and so on.

During the learning phase, the explicit learning group classified each shape into one of the four categories, just as in the earlier experiments. Each trial began with a white fixation cross for 1 s. The shape then appeared in the center of the screen for 1 s and moved off the screen in 1s. There was no time limit for responding. After the subjects answered, feedback was presented on the screen for 2.5 s. There were four learning blocks in which each of the 32 items was tested in random order.

During the learning phase, the implicit learning group performed the catching task used in the implicit test phase from Experiments 1 and 3. Subjects were told that they would see the same shapes that they had seen in the observation phase and that they should use what they had learned about the shapes to catch each one with their cursor. They were instructed that it was not possible to catch the shape in the center of the screen, so they must place their cursor at the edge of the screen where they thought it would be easiest to catch the shape. Subjects controlled cursor placement and movement with the mouse. At the beginning of each trial, a shape appeared in the center of the screen with the cursor directly on top of it for 0.8 s. After the initial presentation interval, the shapes momentarily disappeared and reappeared approximately 5 cm from the center in the direction of movement and moved off the screen in 0.6 s. Subjects saw a 2.5-s feedback message ("Good catch" or "No catch"). The timing of this phase was slower than the implicit test in Experiments 1 and 3, as subjects in this experiment were using this task to learn the categories, whereas subjects in the previous experiments had already learned the categories when they performed the catching task.

The test phase for both groups consisted of a 16-trial test in which subjects were presented with a static shape and asked four questions about it. There were four blocks in which all four shapes were tested once in random order and no shape was asked about in two consecutive trials. The shape was presented in the middle of the screen directly below the question text. Text in brackets was used for the implicit learning group.

> Q1: What category [color] do you think the shape below most likely belongs to [is most likely to be]?
>
> Q2: What is the probability that this shape belongs to the category [is the color] you just identified (0–100)?
>
> Q3: What direction do you think the shape is most likely to travel in? Please input your answer as a time. Please enter the HOUR VALUE (1–12) and PRESS ENTER.

Q4: Now please think about the MINUTE value that corresponds to the direction you expect the shape to travel in. Please enter the MINUTE VALUE (0–59) and PRESS ENTER.

To ensure that they had learned the categories (for the explicit learning group) and the colors (for the implicit learning group) equally well, after the test phase subjects reported what color/category they thought each shape was most likely to be. Subjects in both groups were told that we wanted to get a sense of how much they had learned about the shapes over the course of the experiment. The explicit learning group then performed one block (32 trials) of the classification task that they performed during the learning phase. The implicit learning group performed one block of the same task except that instead of reporting the most likely category, they reported the most likely color.

Results

Subjects in the explicit learning condition were correct on 70.2% of classifications in the last learning block. (There is no measure of accuracy in the implicit task.) For the explicit prediction task, the likelihood of categorizing the shape into the target category/color (Q1) was 62.5% for both the explicit and the implicit learning groups. The average estimate of the probability that the shape belonged in the target category/color (Q2) was 58.7% for the explicit learning group and 59.8% for the implicit learning group. Subjects in both the explicit and implicit learning groups classified the shapes in their secondary/color category 36.4% of the time. The average estimate of the probability that the shape belonged in the secondary category/color (Q2) was 62.2% for the explicit learning group and 58.8% for the implicit learning group. Results from the post-test revealed that subjects in the implicit learning group learned the color as well as subjects in the explicit learning group learned the categories (Ms = 70.3% and 67.0%). Thus, the two groups were very similar in overall learning by all measures.

All categorizations analysis. As with the results of Experiments 2 and 3, subjects tended to predict values around 0°. The mean prediction ($M = -5.2°$, $SD = 20.1$) was not significantly different than 0°, $t(22) = 1.2$, $p < .01$, $d = 0.26$. A 2 × 2 ANOVA was performed, with condition and learning mode as between-subjects factors. The main effect of learning mode was not significant, $F(1, 19) = 0.6$, suggesting that subjects did not use categories differently in the two learning conditions.

Target categorization analysis. When only trials in which the target category was selected were analyzed, the mean prediction ($M = -27.1°$, $SD = 25.4$) was significantly less than 0°, $t(22) = 5.1$, $p < .01$, $d = 1.07$. As with the results of the explicit task in Experiments 2 and 3, subjects in both groups avoided the direction reinforced by the secondary category.

A 2 × 2 ANOVA was performed, with condition and learning mode as between-subjects factors. The main effect of learning mode was significant, $F(1, 19) = 5.3$, $p < .05$, revealing that subjects used categories differently in the explicit and implicit learning conditions. The mean prediction for both conditions was negative (Ms = −16.1° and −39.2°, SDs = 24.9 and 20.8, for explicit and implicit). The difference between the mean prediction and 0° was significant for the explicit learning condition, $t(11) = 2.2$, $p < .05$, $d = 0.65$, and was highly significant for the implicit learning condition, $t(9) = 6.2$, $p < .01$, $d = 1.88$. As discussed above, both groups were avoiding predicting the direction reinforced by the secondary category, but the implicit learning group did so more consistently. This is opposite to the prediction that implicit learning would lead to more Bayesian responding and is a replication of the results from the explicit prediction task in Experiments 2 and 3.

Discussion

The results of Experiment 4 suggest that reducing explicit thought about categories during learning does not promote integration of information across categories during induction. Although categories were never mentioned to the implicit learning subjects, they failed to show any evidence of integration but instead showed the same pattern of results found in the explicit learning group (and in Experiments 2 and 3). When they picked the target category, they tended to predict the direction distinct to it, and when they chose the secondary category they tended to choose the direction associated with it. Thus, as suggested by the results of the previous experiment, response mode is critical to how category information is used. Even when subjects have only implicitly learned categories, they use them in the same nonnormative manner seen in the results of the explicit tasks of Experiments 2 and 3—that is, they predict a direction opposite to that of the secondary category.

We speculate that this is because although categories are never mentioned, subjects are still able to explicitly consider them. For category-based induction tasks where Bayesian responding has been found, it seems unlikely that subjects would have been able to consider each category separately. In our implicit prediction task (Experiments 1 and 3), the timing of the task is likely too fast to allow for explicit consideration of the possible categories, and for the tasks used by Tenenbaum and his colleagues, the categories are numerous (or infinite) and somewhat arbitrary, making it unlikely that subjects explicitly consider them (see our discussion of this work in the introduction). In Experiment 4, however, there were only four colors. When asked to predict what direction a square might go in, subjects could easily think about the directions that green shapes go in separately from the directions that brown shapes go in. Thus, during the induction task, when asked what color the shape is most likely to be, subjects may have then considered information about the predicted feature for each category (color) separately or only for the most likely category.

General Discussion

The experiments found that identical category knowledge leads to different predictions in implicit versus explicit induction. See Table 2 for a summary of results. Implicit responses showed integration of information across categories in a Bayesian manner (Experiment 1), whereas explicit responses showed suboptimal use of categories, such that subjects did not integrate information across categories and even avoided features reinforced by the secondary category when categorizing objects into their most likely category (Experiments 2 and 4). The two different patterns were found even within individual subjects during one session (Experiment 3). When under time pressure, people showed greater use of multiple categories, suggesting that the "more implicit" the response, the greater the multiple category use, because there is

Table 2
Mean Predictions for Critical Shapes, in Degrees

	Cursor placement, *M* (*SD*)	Prediction of direction, *M* (*SD*)	
Experiment	Implicit task	Explicit task (all categorizations)	Explicit task (target categorizations)
1, fast group	16.8 (18.4)		
1, slow group	6.4 (14.6)		
2		−4.6 (19.1)	−32.5 (24.6)
3	14.5 (17.9)	0.7 (20.7)	−15.8 (27.2)
4		−5.2 (20.1)	−27.1 (25.4)

Note. Positive cursor placements indicate integration of information across categories. Negative placements indicate the avoidance effect. *SD* = standard deviation.

less chance for interference from explicit strategies (Experiment 1). The results of Experiment 4 suggest that the use of multiple categories found when subjects made implicit predictions was not simply a result of reducing explicit thought about categories but requires implicit responding in particular.

Relation to Previous Results

These findings suggest that the suboptimal use of categories consistently found in previous studies of category-based induction is, at least in part, due to the response mode. In previous studies, subjects categorized an item and then made an explicit verbal prediction. As shown in Experiment 2, this type of responding leads to nonnormative category use (below we discuss the avoidance effect). Implicit inductions used multiple categories in a way that approximates Bayesian models.

The notion that quick, automatic judgments can be more normative than strategic, effortful ones has been echoed in research on decision making under risk, which has found that motor strategies sometimes are superior to decision making in similar situations (Trommershäuser et al., 2008). It seems that more explicit reasoning can interfere with good decision making because of "its sequential and low-capacity nature" (Evans, 2008, p. 267), which, at least in the case of category-based induction, can lead to the disregarding of relevant information. Indeed, our results suggest that implicit processes access and integrate information in a way that explicit processes often do not.

The current findings help explain the discrepancy between studies of induction in reasoning versus perception and action. Previous category-based induction tests have been explicit, whereas perception and action tests are likely implicit. In perception and action studies, subjects are usually not asked to make predictions or give probability ratings but rather to make quick motor responses.

Interestingly, it appears that useful information is also disregarded when perceptual judgments are made in ways similar to our explicit task. Bayesian models have been successful in describing much behavior in perceptual tasks, suggesting that human observers perform them by considering multiple hypotheses about the underlying structure of the sensory evidence. There is evidence, however, that when subjects select a hypothesis prior to making a response (much as our subjects select a category before making their inference), they act as if there is only one hypothesis. For example, in one study (Jazayeri & Movshon, 2007), subjects viewed displays of randomly moving dots with a small proportion moving in the same direction, to be identified by the subject. After the motion stopped, a bar appeared at the edge of the display, and subjects had to decide if the nonrandom motion of the dots was clockwise or counterclockwise relative to the bar (i.e., commit to a hypothesis about the direction). They then marked their estimate of the exact direction of the motion with the mouse. The initial decision task greatly reduced the accuracy of their direction estimates. Stocker and Simoncelli (2008) fit these data to a model in which people ignored hypotheses inconsistent with their initial clockwise–counterclockwise judgment and considered only hypotheses consistent with it. The model fit the data well, suggesting that when an observer chooses one hypothesis as most likely (the dots' direction relative to the reference mark), subsequent judgments assume that this hypothesis is true. Stocker and Simoncelli argued that this disregard for alternatives, which they call *simplification by decision,* is a useful strategy because it frees up resources in complex perceptual tasks.

Like our own task, this one required people to make an initial judgment and then a more detailed one. Although in both tasks subjects were not certain that their initial judgment was correct, they acted as if it was when making subsequent judgments (see below). The conscious, unspeeded nature of these judgments allow the initial decision to warp the evaluation of evidence, even in a purely perceptual task.

Implications for Everyday Inductions

Recall our original example of the class cutter who cannot make out the identity of another person in the parking lot and needs to predict whether that person will send him to the principal's office in order to avoid punishment. As with many real-life examples of category-based induction, our class cutter could have made his prediction either explicitly, by deliberately thinking about how much the person looks like another student or a teacher and then making a strategic decision about his next action, or implicitly, by simply acting instinctively without overtly thinking about the person's identity. Our results show that the same category information may lead to different predictions depending on how they are made. This is important, because many decisions are based on such inferences.

Our paradigm of explicit induction often asks subjects to identify an item's most likely category prior to making inductions. This has most often been the case in experiments with artificial categories that subjects might not spontaneously use (e.g., Murphy et al., 2012). In experiments using real-world categories, such questions can be omitted or placed at the end of the test (e.g., Ross & Murphy, 1996; Zhu & Murphy, 2013). Asking this question in the artificial category case probably increases the rate of single-category use, even though subjects consistently rate that they are uncertain of their answer (Hayes & Newell, 2009; Murphy et al., 2012). In real life, familiar categories are likely to come to mind without any question or instruction. Imagine, for example, that while walking through a park, you observe a sudden rustling in the plants along the walkway, and you wonder whether you should avoid the path near this spot. Two of the authors can confirm that in New York City the question "Is it a squirrel or a rat?" comes to

mind without prompting from any questionnaire. However, it is also possible that such a pedestrian would drift to the far side of the path without necessarily deciding the answer to that question. If you obtain enough evidence to identify the animal as a squirrel, you might then overrule that tendency and walk near the rustling without worry. In short, real-life inductions probably involve both explicit and implicit inductions in a complex interplay.

As with some work in decision making under risk (see Trommershäuser et al., 2008), in our induction task, motor responses seem to outperform explicit, strategic ones. However, whether explicit or implicit processing is more normative clearly depends on task requirements. Associative, intuitive judgments have been cited as leading to errors in prediction that can be corrected by a more reflective reasoning system (Kahneman & Frederick, 2002; Kahneman & Tversky, 1973; Sloman, 1996). In social psychology, a similar distinction has been made between automatic and controlled processes in prejudice. Automatic processes are often associated with stereotype activation (a form of category-based induction), which, in low-prejudice people, conflicts with explicit attitudes and is inhibited in favor of their explicit beliefs (Devine, 1989). Thus, the explicit system's bias to disregard or avoid information from alternative categories (that made it less normative in our task) could, in other cases, lead to more normative responses. Our research shows that this distinction is crucial for understanding when category-based predictions are more likely to be accurate or inaccurate.

In short, we do not conclude that "explicit = inaccurate" and "implicit = accurate" in category-based inductions. Rather, the two kinds of tasks involve different processes that may lead to different outcomes, and whether those outcomes are good or bad will depend on the particular categories and their associated properties.

Explicit Induction and the Avoidance Effect

The main finding of these experiments was that implicit induction led to integration of information across categories, contrary to the predominant use of a single category in previous research using explicit prediction tasks. We found, as expected, no evidence of integration in explicit induction; however, closer examination of these data revealed an interesting and surprising effect. In this section we examine the explicit induction results in more detail.

We provided two analyses of the explicit task results: including all trials or only trials in which the item was categorized as being in the target category. Neither showed integration as the implicit task did, but they showed rather different results that are worth discussion. In one analysis, which was comparable to the analysis of the implicit task, all trials were included. This analysis generally found that predictions on average were near the middle of the two possible directions of the target category and did not differ depending on the direction the shapes moved in the secondary category. Thus, when all the explicit test trials were used, there was no evidence that people used the secondary category in making their judgments.

In previous work, researchers have generally looked only at trials in which subjects selected the target category, because the experiments were designed to contrast single- versus multiple-category use, and the experimental designs allowed a clear test only for such trials. The results in most of this past work showed that people focused on a single category, and their judgments were not influenced by the secondary category. In the present experiments, however, the analysis of only target categorizations showed a very unexpected result, the avoidance effect. This surprising result could suggest that people were influenced by multiple categories during induction but somehow in the wrong direction. Although a definitive analysis of this avoidance effect requires further research, we propose that subjects were actually not using multiple categories during induction per se.

In Condition 1, Category 1 stimuli went equally often to 1 and 5 o'clock; Category 2 stimuli all went to 1 o'clock. Our proposal is that after learning these categories, subjects came to associate Category 2 stimuli with 1 o'clock and Category 1 with 5 o'clock. During the learning phase, only Category 1 stimuli ever went to 5 o'clock. Furthermore, since all Category 2 stimuli went to 1 o'clock, that direction is the only sensible prediction for items in that category. Thus, even though two directions were equally frequent in Category 1, people came to think of it as the 5 o'clock category, due to that direction's distinctive relationship with the category. The phenomenon seems related to the "base-rate neglect" phenomenon discovered by Gluck and Bower (1988), in which features that in fact occurred equally often in two categories (i.e., the cue validity was equal for both categories) became more associated with the category for which they were a distinctive feature.

As a result of this learning, when people decided that a square was most likely in Category 1, they often chose its distinctive direction (5 o'clock); when they decided that a square was in Category 2, they chose its direction (1 o'clock). On this account, neither induction reflects considering both categories but rather consideration of only the category initially selected and the direction associated with it. This is in line with the singularity principle, which suggests that people are biased to consider only one possibility at a time. Indeed, past work in explicit induction has shown that which category people choose has an enormous effect on their inductions even when they are very uncertain that their categorization is correct (Lagnado & Shanks, 2003; Murphy et al., 2012, Experiment 4).

If this explanation is correct, the reason that the avoidance effect has not been found in previous studies probably relates to the category structure and learning procedures followed here. The secondary category structure in previous research generally had a variety of feature values. In the present studies, the secondary category had a single, universal direction. Our suggestion is that this direction became strongly associated with the secondary category and, by contrast, not the expected direction of the target category (as in other category contrast effects, Krueger & Rothbart, 1990).

Furthermore, the feedback-based category learning of the present experiments may have caused the association of the distinctive feature to the target category, due to processes of error-driven learning (as in Gluck and Bower's learning task). We conducted one test of this idea in a follow-up experiment in which there was no feedback-based category learning. Using the same category structure as the experiments reported here, we presented 21 subjects with visual displays corresponding to the four categories, with shapes and arrows indicating the direction with which each shape was associated. That is, all items were simultaneously present in one display. We asked subjects to predict the direction of

new shapes (e.g., a square, which occurred more often in Category 1 than in Category 2), as in our main experiment. Here we found very few avoidance responses: Subjects picked the direction distinctive to the target category only 12% of the time. Thus, the avoidance effect in the present data probably requires feedback-based learning to manifest itself, as Gluck and Bower (1988) suggest.

Nature of the Implicit/Explicit Distinction

We have referred to our primary manipulation as changing the response mode from explicit to implicit. This is a shorthand to refer to multiple differences between the two response modes (described below). As we commented in the introduction, it is not clear that this distinction corresponds to Systems 1 and 2 referred to by two-system theorists (Evans, 2007; Kahneman, 2011). There is some similarity, such as System 1's automatic operation with no sense of voluntary control (Kahneman, 2011, p. 105), like our implicit inductions. But our explicit subjects also seemed to operate in ways consistent with System 1, by focusing on a single, most likely possibility. Furthermore, it is not clear that our distinction requires one to posit separate systems. What we do claim, however, is that making inductions in these ways leads to different induction processes in which category information is used differently.

In many years of studying explicit inductions in this task, we have found that most people do things such as count up the number of items in each category, overtly choose one or two categories on which to focus, and even do probability calculations to estimate likelihoods. These can be seen in markings and calculations subjects write on response forms. People seem to think things like, "There are seven squares in Category 1 but five in Category 2, so it's most likely in Category 1. The squares in Category 1 are mostly red, so I'm going to predict red." Subjects confirm such strategies when systematically questioned (Murphy et al., 2012). Finally, many of the phenomena of induction under uncertainty suggest that only one category is in working memory, and manipulations that place two categories in working memory then lead to Bayesian reasoning (Murphy et al., 2012; Ross & Murphy, 1996).

Our implicit task discourages a single-category focus and encourages the use of learned associations, through a number of means. First, time pressure prevents the counting and calculation process. Indeed, the older authors found the fast condition of Experiment 1 (used in all the subsequent implicit conditions) to be almost too fast to successfully perform. A square appears, and you must immediately move the cursor to where it might go. Retrieval of past exemplars, calculations of probabilities, and even simple counting are not possible. Second, and related to this, the speed seems to preclude the use of working memory and attention that permit subjects to focus on a single category or select which exemplars to input to the calculations. Third, the implicit condition did not ask about categories, whereas the explicit task did. Fourth, the motor task essentially replicates the past motions of the objects, so that the learned associations between shape and direction can be directly used to make predictions.

We do not now know whether these individual variables would be sufficient to create the differences we observed, nor can we state for certain that the implicit responses come from completely different psychological and neural systems than the explicit responses. One likely important variable is that the categories were well learned, allowing prior associations to direct motor actions. Category-based induction has also been tested in novel categories that are (partially) presented during the test itself (Murphy & Ross, 1994). It is hard to see how the implicit responding seen in our task could take place in such a situation: Because the exemplars are examined at the time of induction, the objects' properties are not yet associated in memory. However, in real-world situations with familiar categories, learned associations could well render a Bayesian response.

In summary, our goal was to show that people may be more normative when they must respond quickly using learned associations than when they carefully consider the categories. This was clearly achieved, as our results show that the inductions are very different in these two conditions—correctly integrating information across categories or showing nonnormative use of category information. Exactly how our tasks might correspond to other proposed distinctions will require further research.

Related to this issue, not everything that might make induction implicit in some respect will necessarily improve induction. The results of Experiment 4 suggest that response mode is critical, as reducing explicit thought about categories during learning did not lead to normative category use when responses were made explicitly. It is clearly a question for future research to investigate other aspects of implicit processing.

As discussed earlier, Tenenbaum and his colleagues have successfully modeled people's category-based inferences with Bayesian models (Griffiths & Tenenbaum, 2006; Tenenbaum, 1999, 2000). These inferences were not reported with an action response, yet people seemed to be integrating information across categories in a normative manner as they do with action tasks. Our explanation of this is that these subjects did not consider each category separately. As discussed in the introduction, it seems highly unlikely that subjects in these tasks are able to explicitly consider the possible categories, which were numerous and somewhat arbitrary (e.g., many ranges of numbers). Thus, although we would not argue that those inductions use the same processes as our own implicit task, they do share the property of not encouraging explicit consideration of potential categories. It might be interesting to ask people to choose from among possible categories in such tasks and see if their responses become less Bayesian, as in Jazayeri and Movshon's (2007) perceptual task.

Future Directions

The inductions made in these experiments were all relatively simple. Exemplars had only two features and belonged to only one of two potential categories. Real-world items, however, have more than two features and could be ambiguous among more than two possible categories. It is unclear precisely how more complex circumstances might affect both implicit and explicit induction. Would implicit induction still be able to come to a Bayesian answer given the greater number of associations now involved? When making explicit inductions, it seems subjects might be even more likely to disregard information to simplify calculations, as having to consider more than two categories or features is more taxing on mental resources. That might have the effect of reducing both the avoidance effect and integration across categories. Further investigation is necessary to examine these issues, as understand-

ing how we use category information and whether the implicit/explicit distinction holds when making predictions about complex items is critical for understanding real-world category-based induction.

Our results likely also have implications for the related situation of cross-classification: When you know an item belongs to more than one category, what category or categories do you base your inductions on? The ice cream that you are thinking about eating belongs to the categories of dairy foods and desserts. The student that you are meeting with belongs to the category of student, but also to female and teenager. Previous research has shown that people often show a single-category focus when making inductions about cross-classified items (Murphy & Ross, 1999), and it has been suggested that the activation of one category inhibits activation of other possible categories (Macrae, Bodenhausen, & Milne, 1995). As when reasoning with uncertain categories, people are able to use category information normatively when making inferences about a cross-classified item (Hayes, Kurniawan, & Newell, 2011), but they often do not. The distinction between implicit and explicit processes may also be useful in characterizing situations when people use category information more normatively.

The implicit–explicit distinction also has implications for our interactions with people. People are the most readily cross-classified objects (Bodenhausen, Todd, & Becker, 2007), so predictions about people likely also do not take into account all the relevant information available. In fact, research in social psychology has suggested that a single category comes to dominate the perception of an individual (Bodenhausen & Macrae, 1998; Bodenhausen & Peery, 2009). However, judgments about people without explicit categorization no doubt occur as well (e.g., deciding how far to sit from someone). Understanding predictions about people is particularly important, because each different identity leads to different expectations about a person, which in turn influence interactions with him or her.

Conclusion

When people make category-based inductions with uncertainty, they often use category information suboptimally and do not integrate information across categories following normative principles. Our results suggest that the suboptimal use of categories is, at least in part, due to the explicit response mode normally used in category-based induction experiments. When subjects in our experiments made inductions implicitly, they were able to appropriately integrate information across categories. These results demonstrate that in some cases our category-based inferences are not necessarily determined only by category knowledge but also by how these inferences are made.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429. doi:10.1037/0033-295X.98.3.409

Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype activation and inhibition. In R. S. Wyer, Jr. (Ed.), *Advances in social cognition* (Vol. 11, pp. 1–52). Mahwah, NJ: Erlbaum.

Bodenhausen, G. V., & Peery, D. (2009). Social categorization and stereotyping *In vivo*: The VUCA challenge. *Social and Personality Psychology Compass, 3,* 133–151. doi:10.1111/j.1751-9004.2009.00167.x

Bodenhausen, G. V., Todd, A. R., & Becker, A. P. (2007). Categorizing the social world: Affect, motivation, and self-regulation. *Psychology of Learning and Motivation, 47,* 123–155. doi:10.1016/S0079-7421(06)47004-3

Brooks, L. R., Squire-Graydon, R., & Wood, T. J. (2007). Diversion of attention in everyday concept learning: Identification in the service of use. *Memory & Cognition, 35,* 1–14. doi:10.3758/BF03195937

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine, 78,* 775–780. doi:10.1097/00001888-200308000-00003

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18. doi:10.1037/0022-3514.56.1.5

Evans, J. S. B. (2007). *Hypothetical thinking: Dual processes in reasoning and judgment*. New York, NY: Psychology Press.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 255–278. doi:10.1146/annurev.psych.59.103006.093629

Evans, J. S. B., & Frankish, K. (2009). *In two minds: Dual processes and beyond*. Oxford, England: Oxford University Press. doi:10.1093/acprof:oso/9780199230167.001.0001

Feeney, A. (2007). How many processes underlie category-based induction? Effects of conclusion specificity and cognitive ability. *Memory & Cognition, 35,* 1830–1839. doi:10.3758/BF03193513

Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. New York, NY: Viking.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford, England: Oxford University Press.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117,* 227–247. doi:10.1037/0096-3445.117.3.227

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science, 17,* 767–773. doi:10.1111/j.1467-9280.2006.01780.x

Haruno, M., Wolpert, D., & Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Computation, 13,* 2201–2220. doi:10.1162/089976601750541778

Hayes, B. K., & Chen, T.-H. J. (2008). Clinical expertise and reasoning with uncertain psychodiagnoses. *Psychonomic Bulletin & Review, 15,* 1002–1007. doi:10.3758/PBR.15.5.1002

Hayes, B. K., Kurniawan, H., & Newell, B. R. (2011). Rich in vitamin C or just a convenient snack? Multiple-category reasoning with cross-classified foods. *Memory & Cognition, 39,* 92–106. doi:10.3758/s13421-010-0022-7

Hayes, B. K., & Newell, B. R. (2009). Induction with uncertain categories: When do people consider the category alternatives? *Memory & Cognition, 37,* 730–743. doi:10.3758/MC.37.6.730

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 248–274). Oxford, England: Oxford University Press.

Heit, E., & Rotello, C. M. (2010). Relations between inductive reasoning and deductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 805–812. doi:10.1037/a0018784

Jazayeri, M., & Movshon, J. A. (2007, April 19). A new perceptual illusion reveals mechanisms of sensory decoding. *Nature, 446,* 912–915. doi:10.1038/nature05739

Kahneman, D. (2011). *Thinking fast and slow*. New York, NY: Farrar, Straus & Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511808098.004

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgement. In

K. G. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). New York, NY: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251. doi:10.1037/h0034747

Kersten, D., Mamassian, P., & Yuille, A. L. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55,* 271–304. doi:10.1146/annurev.psych.55.090902.142005

Krueger, J., & Rothbart, M. (1990). Contrast and accentuation effects in category learning. *Journal of Personality and Social Psychology, 59,* 651–663. doi:10.1037/0022-3514.59.4.651

Lagnado, D. A., & Shanks, D. R. (2003). The influence of hierarchy on probability judgment. *Cognition, 89,* 157–178. doi:10.1016/S0010-0277(03)00099-4

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in social perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology, 69,* 397–407. doi:10.1037/0022-3514.69.3.397

Malt, B. C., Ross, B. H., & Murphy, G. L. (1995). Predicting the features for members of natural categories when categorization is uncertain. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 646–661. doi:10.1037/0278-7393.21.3.646

Murphy, G. L., Chen, S. Y., & Ross, B. H. (2012). Reasoning with uncertain categories. *Thinking & Reasoning, 18,* 81–117. doi:10.1080/13546783.2011.650506

Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categorizations. *Cognitive Psychology, 27,* 148–193. doi:10.1006/cogp.1994.1015

Murphy, G. L., & Ross, B. H. (1999). Induction with cross-classified categories. *Memory & Cognition, 27,* 1024–1041. doi:10.3758/BF03201232

Murphy, G. L., & Ross, B. H. (2005). The two faces of typicality in category-based induction. *Cognition, 95,* 175–200. doi:10.1016/j.cognition.2004.01.009

Murphy, G. L., & Ross, B. H. (2010). Uncertainty in category-based induction: When do people integrate across categories? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 263–276. doi:10.1037/a0018685

Newell, B. R., Paton, H., Hayes, B. K., & Griffiths, O. (2010). Speeded induction under uncertainty: The influence of multiple categories and feature conjunctions. *Psychonomic Bulletin & Review, 17,* 869–874. doi:10.3758/PBR.17.6.869

Ross, B. H., & Murphy, G. L. (1996). Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 736–753. doi:10.1037/0278-7393.22.3.736

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology, 38,* 495–553. doi:10.1006/cogp.1998.0712

Rotello, C. M., & Heit, E. (2009). Modeling the effects of argument length and validity on inductive and deductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1317–1330. doi:10.1037/a0016648

Shafto, P., Coley, J. D., & Baldwin, D. (2007). Effects of time pressure on context-sensitive property induction. *Psychonomic Bulletin & Review, 14,* 890–894. doi:10.3758/BF03194117

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22. doi:10.1037/0033-2909.119.1.3

Stanovich, K. (2009). *What intelligence tests miss: The psychology of rational thought*. New Haven, CT: Yale University Press.

Stanovich, K. (2010). *Rationality and the reflective mind*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195341140. 001.0001

Stocker, A. A., & Simoncelli, E. P. (2008). A Bayesian model of conditioned perception. *Advances in Neural Information Processing Systems, 20,* 1409–1416.

Tassinari, H., Hudson, T. E., & Landy, M. S. (2006). Combining priors and noisy visual cues in a rapid pointing task. *Journal of Neuroscience, 26,* 10154–10163. doi:10.1523/JNEUROSCI.2779-06.2006

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. *Advances in Neural Information Processing Systems, 11,* 59–68.

Tenenbaum, J. B. (2000). Rules and similiarity in concept learning. *Advances in Neural Information Processing Systems, 12,* 59–65.

Trommershäuser, J., Körding, K. P., & Landy, M. S. (Eds.). (2011). *Sensory cue integration*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195387247.001.0001

Trommershäuser, J., Landy, M. S., & Maloney, L. T. (2006). Humans rapidly estimate expected gain in movement planning. *Psychological Science, 17,* 981–988. doi:10.1111/j.1467-9280.2006.01816.x

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning and statistical decision theory. *Trends in Cognitive Sciences, 12,* 291–297. doi:10.1016/j.tics.2008.04.010

Verde, M. F., Murphy, G. L., & Ross, B. H. (2005). Influence of multiple categories on property prediction. *Memory & Cognition, 33,* 479–487. doi:10.3758/BF03193065

Zhu, J., & Murphy, G. L. (2013). Influence of emotionally charged information on category-based induction. *PLoS ONE, 8,* e54286. doi:10.1371/journal.pone.0054286

Appendix A

Experiment 1 Individual Response Histograms

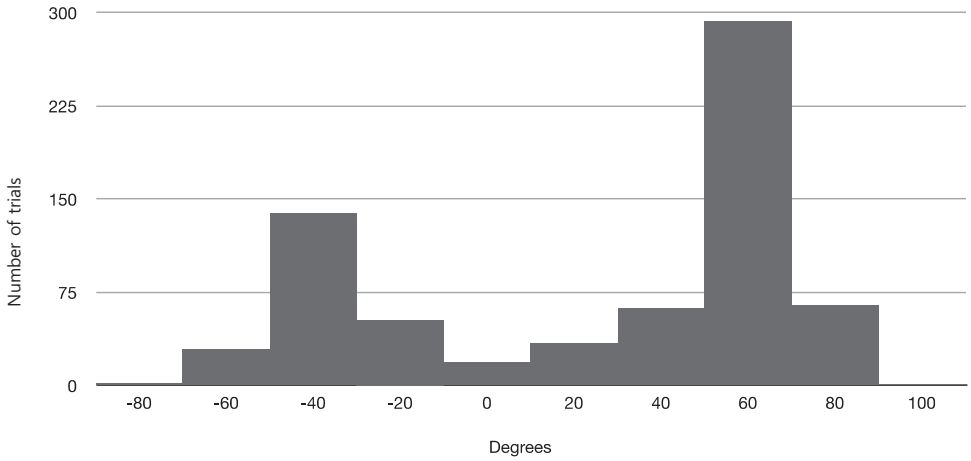

Figure A1. Histogram of cursor placements for the fast group performing the implicit induction task. Normative use of categories is evidenced by positive shifts from 0°. Each bin includes a 20° range and is labeled by the largest degree it includes.

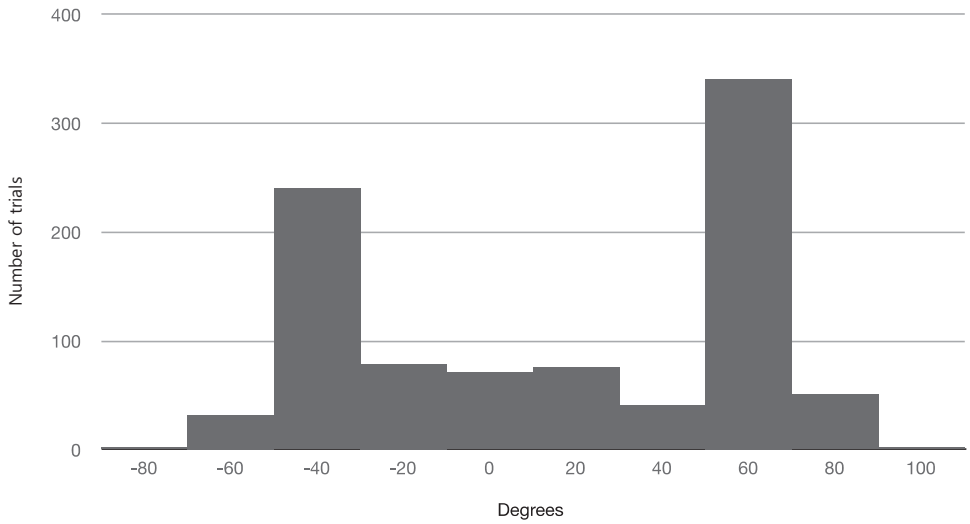

Figure A2. Histogram of cursor placements for the slow group performing the implicit induction task. Normative use of categories is evidenced by positive shifts from 0°. Each bin includes a 20° range and is labeled by the largest degree it includes. Note that the *y*-axis scale is different from that of Figure A1.

(*Appendices continue*)

Appendix B

Experiment 2 Individual Response Histograms

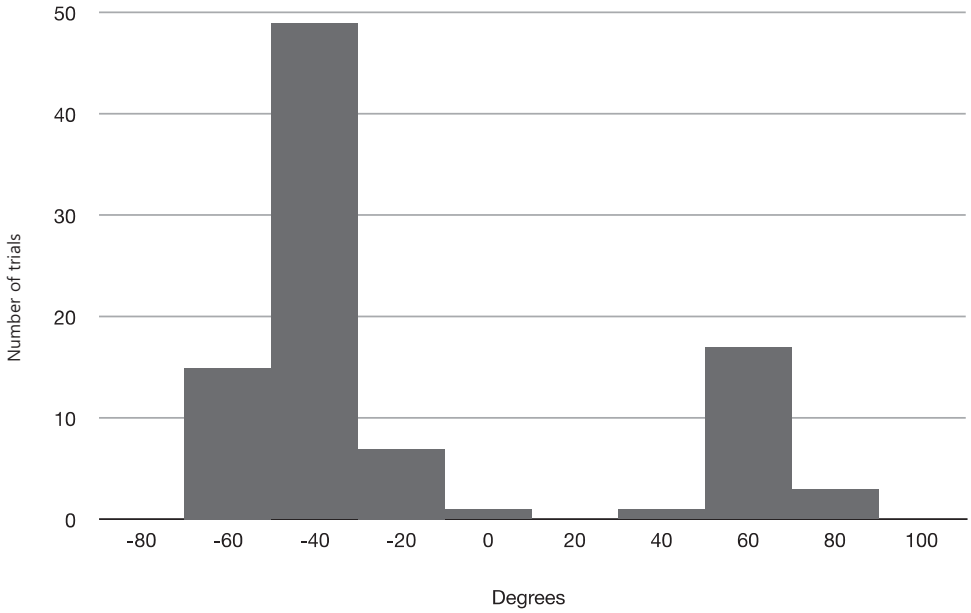

Figure B1. Histogram of predictions when target category was picked in the explicit induction task. Normative use of categories is evidenced by positive shifts from 0°. Each bin includes a 20° range and is labeled by the largest degree it includes.

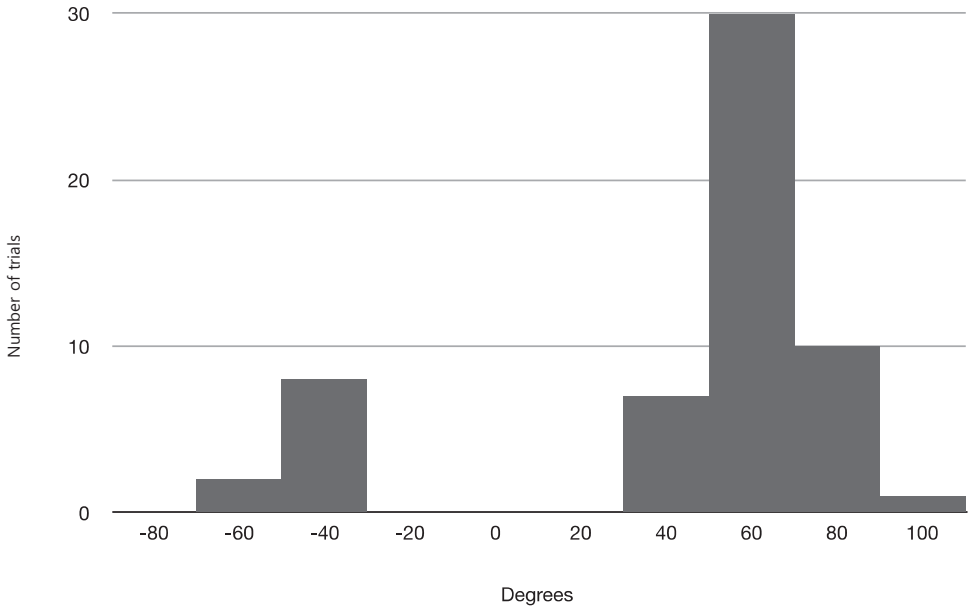

Figure B2. Histogram of predictions when the target category was not picked in the explicit induction task. Normative use of categories is evidenced by positive shifts from 0°. Each bin includes a 20° range and is labeled by the largest degree it includes. Note that the *y*-axis scale is different from that of Figure B1.

Received April 5, 2012
Revision received January 24, 2013
Accepted January 27, 2013 ■