Belief digitization: Do we treat uncertainty as probabilities or as bits?

Samuel G. B. Johnson [1,2]
Thomas Merchant [3,4]
Frank C. Keil [5]

[1] University of Bath, School of Management
[2] University College London, Centre for the Study of Decision-Making Uncertainty
[3] Brown University, Department of Cognitive, Linguistic, and Psychological Sciences
[4] University of Colorado – Boulder, Department of Ecology and Evolutionary Biology
[5] Yale University, Department of Psychology

Corresponding Author: Sam Johnson

Email for Correspondence: sgbjohnson@gmail.com

Word Count: 15,284 (main text, incl. footnotes)

This article is in press at the Journal of Experimental Psychology: General

Acknowledgements: A subset of these results were presented at the Cognitive Science Society (Johnson, Merchant, & Keil, 2015). We thank the reviewers and attendees for their comments, as well as audiences at MIT, Yale, and the Universities of Bath, Chicago, and Warwick.

Abstract

Humans are often characterized as Bayesian reasoners. Here, we question the core Bayesian assumption that probabilities reflect degrees of belief. Across 8 studies, we find that people instead reason in a *digital* manner, assuming that uncertain information is either true or false when using that information to make further inferences. Participants learned about two hypotheses, both consistent with some information but one more plausible than the other. Although people explicitly acknowledged that the less-plausible hypothesis had positive probability, they ignored this hypothesis when using the hypotheses to make predictions. This was true across several ways of manipulating plausibility (simplicity, evidence fit, explicit probabilities) and a diverse array of task variations. Taken together, the evidence suggests that digitization occurs in prediction because it circumvents processing bottlenecks surrounding people's ability to simulate outcomes in hypothetical worlds. These findings have implications for philosophy of science and for the organization of the mind.

Keywords: Reasoning, causal thinking, probability judgment, prediction, categorization

Introduction

Humans are often compared to Bayesian reasoners who use sophisticated probabilistic reasoning to solve all manner of tasks. Here, we pose a fundamental challenge to this proposition: When using probabilities to make predictions, people do not treat them as probabilities at all.

Probabilities quantify uncertainty (Jaynes, 2003) and Bayesian probabilities reflect *degrees* of belief (Jeffreys, 1939). A forest ranger whose lake is infested with snails at a 0.7 probability may use this belief to decide among remedies—a cheap solution to a snail problem may well be worth it since the lake is likelier infested than not, but an expensive solution may be inappropriate given the 30% chance of wasting resources. But probabilities inform other inferences as well as our immediate actions. The ranger may wish to know whether the lake is likely to have a bacteria problem (call this possibility Z), and the odds of a bacteria problem may depend on whether the lake has snails (hypothesis A) or not (hypothesis B). Suppose there is an 80% chance of a bacteria problem if the lake has snails, but only a 20% chance if it doesn't. Probabilities allow us to weight these probabilities appropriately, considering both the possible world in which the lake *is* infested by snails (where a bacteria problem is likely) and the world in which the lake is *not* infested (where a bacteria problem is unlikely). In this example, we set $P(A)=.70$, $P(B)=.30$, $P(Z|A)=.80$, and $P(Z|B)=.20$. (The structure of this problem is summarized in Figure 1.) Then the probability of a bacteria problem can be calculated as:

$$P(Z) = P(Z|A) * P(A) + P(Z|B) * P(B) = 0.80 * 0.70 + 0.20 * 0.30 = 0.62$$

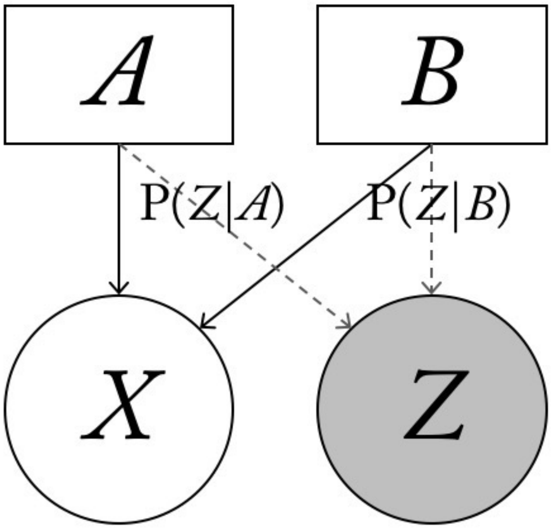

Figure 1. Causal structure where explanatory inference can be used to make predictions.

Note. The X node designates the observed evidence, which can be explained either by explanation A or B. The Z node designates a prediction that differs in probability, depending on whether explanation A or B is correct.

Psychologically, reasoning through a problem like this involves three effortful computations (see Evans, 2007 for a related model). First, the available evidence must be used to evaluate the plausibility of each potential hypothesis through a process of *abduction* or *inference to the best explanation* (Douven, 1999; Lipton, 2004): How likely is it that the lake is versus is not infested with snails? This results in the *hypothesis probabilities* P(A) and P(B). Second, the likely value of Z must be determined for each of these possible worlds through a process of *simulation* or *counterfactual thinking* (Lewis, 1979; Stalnaker, 1976): Given that the lake is (versus is not) infested with snails, what's the chance of a bacteria problem? This results in the *predictive probabilities* P($Z|A$) and P($Z|B$). Finally, these possible values of P(Z) must be weighted by the probabilities of each state through a process of *evidence integration*: All things considered, what is the chance of a bacteria problem? This results in the *integrated probability* P(Z) [= P($Z|A$)*P(A) + P($Z|B$)*P(B)].[1]

There are theoretical reasons to think that people can perform these computations. Categories may be central to cognition precisely because they can facilitate probabilistic predictions (Anderson, 1991). Indeed, the assumption of graded probabilistic reasoning is fundamental to the growing array of computational models that treat humans as rational Bayesian reasoners (e.g., Gershman, Horvitz, & Tenenbaum, 2015; Lake, Ullman, Tenenbaum, & Gershman, 2017; Piantadosi, Tenenbaum, & Goodman, 2012; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). These models capture the insight that navigating the world successfully requires some measure of probabilistic thinking.

There is also empirical research examining how people carry out these component computations, albeit sometimes in a biased or limited way. People readily infer the best explanation for available data through fallible but useful heuristics (e.g., Douven & Schupbach, 2015; Johnson, Valenti, & Keil, 2019; Khemlani, Sussman, & Oppenheimer, 2011; Lombrozo, 2007, 2016). People simulate future outcomes, in part through imagining the workings of physical mechanisms, but do so with small capacity limits (e.g., Escalas, 2004; Forbus, 1984; Hegarty, 2004; Johnson & Ahn, 2015; Kahneman & Tversky, 1981; Markman, Gavanski, Sherman, & McMullen, 1993; Rips, 2010; Sloman, 2005; Taylor & Pham, 1996). And people can combine and weight multiple sources of evidence to generate an overall judgment, although this process often suffers from distortion (e.g., Fisher & Keil, 2018; Griffin & Tverksy, 1992; Kvam & Pleskac, 2016; Manis, Gleason, & Dawes, 1966).

On balance, it is an open question how much people resemble Bayesians in their use of uncertain hypotheses to make predictions. In this article, we question how well people can account for multiple possible hypotheses. For example, our visual systems seem to embody aspects of probabilistic inference (Kersten et al., 2004; Knill & Richards, 1996). Yet, our visual systems do *not* integrate across multiple possibilities: Ambiguous figures such as Necker cubes appear in only a single interpretation

[1] A terminological note. We refer to P(A), P($Z|A$), and P(Z) as the *hypothesis probability*, *predictive probability*, and *integrated probability* throughout this article. We use this terminology instead of more common Bayesian terms such as *prior*, *likelihood*, and *posterior* because these do not map cleanly onto these quantities in this case. In particular, since all of these probabilities are computed in light of the available evidence X, they are all conditioned on X and therefore in a sense "posterior probabilities." For example, the ranger is ultimately interested in P($Z|X$)—the probability of snails conditioned on all the available evidence—and assesses P($A|X$) and P($B|X$) as an intermediate step. We use the simpler notation and terminology to avoid these complications as they do not affect any of the calculations.

at a time (Attneave, 1971). The visual system must adopt one or another belief at a time, rather than delivering both percepts simultaneously.

Higher cognition may work this way too. Experimental studies find that people often have difficulty considering multiple possible categorizations of an object simultaneously (Lagnado & Shanks, 2003; Murphy & Ross, 1994), focus disproportionately on salient causes when making predictions (Fernbach, Darlow, & Sloman, 2010, 2011), and think in terms of particular concrete scenarios (Steiger & Gettys, 1972). More broadly, Evans (2007) proposed a *singularity principle* in hypothetical thinking, such that people can only entertain a single possibility at a time.

Here, we suggest a different computer metaphor: The mind not as a Bayesian machine, but a digital one. We propose that people often represent hypothesis probabilities as 'bits' with values of 0 or 1 when used in downstream inferences. In our earlier example, this would amount to reasoning as though the snail infestation is certain (probability 1) and its absence completely ignorable (probability 0). In that case:

$$P(Z) = P(Z|A) * P(A) + P(Z|B) * P(B) = 0.80 * 1 + 0.20 * 0 = 0.80$$

Everyday experience tells us that people do not *explicitly* ignore uncertainty when reporting hypothesis probabilities. If people were fully incapable of representing hypotheses probabilistically, then commonplace statements like "There is a 70% chance of rain" would be unintelligible because they depend on degrees of belief for their meaning. Clearly, people do make judgments of probability, even if error-prone (Kahneman, Slovic, & Tversky, 1982). The question is how those probabilities are used in other, downstream computations. Even if we can explicitly report a probability of rain, this probability may not appear in a graded manner in later processes such as prediction.

Such a strategy is rather extreme, but consider the opposite extreme of weighting all relevant possibilities according to their probabilities when making predictions. Suppose you come home from work and your door is slightly ajar. Should you go inside? One possibility is that you forgot to fully close the door this morning; a second possibility is that your neighbor, who has a key, borrowed something and forgot to close the door; a third possibility is that a burglar has paid you a visit. If you forgot to close the door in the morning, it's possible that some of the teenagers in your neighborhood went inside and set a trap for you (you forgot to buy Halloween candy this year, and they've been gunning for you ever since); but if your neighbor left the door open, she may have done so recently, in which case the teenagers probably did not have time to get in. If your house was burglarized, the burglar could still be there, in which case it might be dangerous to enter and you should just call the police; but he also could have left, in which case you might as well go in and see what was stolen. Do we assign hypotheses probabilities to each of these possibilities (and more), predictive probabilities conditional on each possibility, and then compute the posterior as a weighted average? At some point, you have to decide whether to go inside or not!

This problem has only three hypotheses and only one layer of possibilities to consider, but if some predictions depend on others in layers of inference (e.g., what is the probability that the burglar has a gun, conditional on having broken in?), then realistic decisions we face are often even more complex. One wonders how plausible—even how adaptive—it would be to follow this procedure. A digitization strategy allows one to compute a quick-and-dirty, if simplistic, answer: What's the likeliest hypothesis? That you left the door open. What's the likeliest prediction to make, conditional on that hypothesis? That the teenagers probably did not visit—perhaps the probability is 5%. You can use that probability to make a decision—perhaps you decide that a 5% chance is low enough not to worry. But since we often overweight small probabilities in risky choice (Tversky & Kahneman, 1992), you may decide a 5% chance is too high and take precautionary measures. This posterior probability is obviously not correct, since it ignores numerous relevant possibilities. But since it is arrived at by taking the most probable path at each point in the decision problem (analogous to Gigerenzer & Goldstein's 1996 "take the best" heuristic), it is a better estimate than any other equivalently simple strategy (e.g., taking the second-most probable branch at every branch point). The computational savings, however, is dramatic: If a prediction depends on *n* mutually dependent layers of possibilities with 2 possibilities at each layer, the Bayesian approach requires adding 2^n terms, each term a multiple of *n* conditional probabilities. The digitization approach requires simply multiplying *n* conditional probabilities once.

Even though a digitization strategy can get us (often reasonable) answers with a huge computational savings, it is nonetheless worrisome in many real-world situations. Suppose you are on a jury and you need to determine how trustworthy a witness is. You might express an 80% chance that the witness is truthful, but then ignore the 20% chance that the witness is lying when assessing the plausibility of her claims—potentially a large enough bias to convert reasonable doubt into a guilty verdict. Suppose you are an investor and need to decide whether a stock will increase or decrease in value. If you think there is a 70% chance of positive movement but ignore the other 30%, you might fail to hedge your bets and risk losing a bundle. Judgments made from digitized beliefs will typically be overconfident, with predicted probabilities too close to 0 or 1.

Overview of Studies

We test this *digitization* hypothesis across 8 studies. Studies 1 and 2 provide initial tests. Participants choose which of two explanations is likelier, where these inferences are based on either simplicity (Study 1) or perceived fit to the data (Study 2). Across conditions, the implications of these two possible explanations for a prediction are varied, to test whether participants take both possibilities into account when making predictions.

Next, we address possible methodological concerns and study the cognitive processing. Studies 3–7 test whether digitization holds up in within vs. between-subjects designs; with beliefs elicited as a forced-choice, on a probability scale, or not at all; and when moderately high probabilities are explicitly assigned to the less-likely hypothesis. These studies also test people's metacognition about digitization (Study 3), whether digitization depends on explicit abductive reasoning from evidence to hypotheses (Study 5), whether people will digitize a hypothesis that accrues a mere plurality of belief (e.g., three hypotheses, with the leading hypothesis having a 50% probability; Study 6), and whether people can treat a group of possibilities, each with a small probability, as a category and therefore take account of it in predictions (Study 7).

Finally, we probe possible boundary conditions related to how the hypothesis and predictive probabilities are set. Studies 8A–D orthogonally manipulate whether the hypothesis probabilities, P(A) and P(B), versus the predictive probabilities, P($Z|A$) and P($Z|B$) are vaguely or precisely set.

This manipulation helps to identify the cognitive bottleneck leading people to use a digitization strategy, and specifically tests the idea that digitization occurs in part to short-circuit the number of mental simulations that act as input to probabilistic integration across possible worlds.

Studies 1 and 2

The set-up for all studies involves causal systems with the structure depicted in Figure 1. That is, two explanations (*A* or *B*) could account for some data (*X*), and these explanations make different predictions about the probability of a different event *Z* occurring. In all cases, participants were given information that would lead them to believe that *A* is the likelier explanation of *X*. The key question is whether the strength of the *A*→*Z* link [P(*Z*|*A*)] and *B*→*Z* link [P(*Z*|*B*)] influence judgments about *Z*. If people rely on both a higher-probability explanation *A* and a lower-probability explanation *B*, then both links should matter. If they digitize, tacitly placing all weight on *A*, then manipulating the strength of the *B*→*Z* link should make no difference to predictions of *Z*. Studies 1 and 2 provide initial tests of this idea.

Methods

Study 1. Participants read about three causal systems instantiating the relationships in Figure 1. For example, in one item, participants were told that Crescent Lake experienced losses of sculpin fish and crayfish, and that three different kinds of snails can lead to subsets of these effects:

> Freshwater juga snails cause lakes to lose sculpin fish and lose crayfish.
> Freshwater scuta snails cause lakes to lose sculpin fish.
> Freshwater aspera snails cause lakes to lose crayfish.

People usually favor simpler explanations that account equally well for the same data (Johnson, Valenti, & Keil, 2018; Lombrozo, 2007). Thus, for a pond that had loss of both sculpin fish and crayfish, most participants should favor the simpler explanation (juga snails) over the more complex explanation (scuta and aspera snails). Participants also learned about the probability of an additional effect *Z* (e.g., bacteria proliferation) given each of these explanations, which varied across conditions. In the *low*/*low* condition, the probability of this effect was low given either explanation. That is, both the *A*→*Z* and *B*→*Z* links were weak, with P(*Z*|*A*) and P(*Z*|*B*) both low:

> When a lake has juga snails, it occasionally has bacteria proliferation.
> When a lake has both scuta snails and aspera snails, it occasionally has bacteria proliferation.

In the *high*/*low* condition, the *A*→*Z* link was strong (P(*Z*|*A*) was high) but the *B*→*Z* link was weak (P(*Z*|*B*) was low):

> When a lake has juga snails, it usually has bacteria proliferation.
> When a lake has both scuta snails and aspera snails, it occasionally has bacteria proliferation.

Finally, in the *low/high* condition, the *A*→*Z* link was weak (P(*Z*|*A*) was low) while the *B*→*Z* link was strong (P(*Z*|*B*) was high):

> When a lake has juga snails, it occasionally has bacteria proliferation.
> When a lake has both scuta snails and aspera snails, it usually has bacteria proliferation.

Participants completed three such items, with different cover stories, in a random order. Participants always received one item in each of the *low/low*, *high/low*, and *low/high* conditions, counterbalanced across cover stories using a Latin square. All information was on the same screen.

After reading this information, participants first made an explanatory judgment as a forced-choice (e.g., between "Crescent Lake has juga snails" and "Crescent Lake has scuta snails and aspera snails"). On the same screen, they rated P(*Z*) ("What do you think is the probability that Crescent Lake has bacteria proliferation") on a scale from 0 to 100 percent.

Study 2. The method was the same as Study 1, except participants were induced to favor explanation *A* over *B* using a bias generally considered non-normative by Bayesian standards—the tendency to favor explanations with narrow *latent scope*, that is, explanations that do not make unverified predictions (Johnson, Rajeev-Kumar, & Keil, 2016; Khemlani et al., 2011). For example, participants considered the explanations:

> Freshwater juga snails cause lakes to lose sculpin fish.
> Freshwater scuta snails cause lakes to lose sculpin fish and lose crayfish.

Before choosing between these explanations, participants were told that "Crescent Lake has a loss of sculpin fish. We don't know whether or not it has a loss of crayfish." That is, explanation *A* makes one verified prediction (loss of sculpin fish), while explanation *B* makes the same verified prediction in addition to one unverified prediction (loss of crayfish). The latter explanation, therefore, has wider latent scope than the former. The unverified prediction is not diagnostic, so by Bayesian standards it is a mistake that people tend to favor the narrower *A* over the wider *B* (e.g., Khemlani et al., 2011; see Johnson et al., 2016 for the underlying mechanisms and evidence against rational Bayesian accounts of this bias). Nonetheless, we expected that participants who favored explanation *A* would focus on this hypothesis exclusively when predicting *Z*.

Participants. We recruited 120 participants for each of Studies 1 and 2. Participants in all studies were recruited through Amazon Mechanical Turk and were prevented from participating in multiple studies reported here.

The sample size was set *a priori* for all studies. Given that one of the key predictions is a null effect (i.e., $d = 0$), the sample size can usefully be assessed both in terms of power analysis and Bayesian statistics. This sample size can detect an effect of $d = 0.3$ with 90% power. Moreover, a Bayesian *t*-test (Rouder et al., 2009; using JZS priors and $r = 1$) shows that a *t*-score of 0 would imply that the null hypothesis is 13.8 times likelier than the alternative. We also conduct a meta-analysis pooling data across studies to achieve even greater power.

After the main task for all studies in this paper, participants completed a series of 18 check questions and were excluded from analysis if they answered more than one-third incorrectly (for Studies 1 and 2, N = 8 and 11, respectively) to avoid inattentive participants. Additional participants were excluded from the primary analysis reported below, if they did not choose explanation A (i.e., the simple or narrow explanation) for at least one of the items (N = 34 and 59, respectively). These exclusions are important because our predictions are predicated on the assumption that $P(A) > P(B)$, and thus we focus on this more exclusive analysis in the text and in Table 1. However, given the large number of exclusions, the key comparisons were repeated with these participants included, and the significance levels are similar, except as noted below.

Results

All data are available through the Open Science Framework (https://bit.ly/2JRxoc3).

For Study 1, most participants (70%) chose the simple explanation (A) for all three items. The critical question was what prediction participants would make about Z (bacteria proliferation). These judgments across the three conditions are given in Table 1. All theories (i.e., normative probabilistic prediction and belief digitization) predict that participants should attend to the strength of the $A \rightarrow Z$ link in making predictions, because A is the most probable hypothesis. This can be tested by comparing the *high/low* and *low/low* conditions, which differ only in whether $P(Z|A)$ is high or low. Not surprisingly, Z was indeed judged significantly likelier to occur in the *high/low* condition [$t(77) = 7.27$, $p < .001$, 95% CI_d[0.72,1.26]]. (Note that confidence intervals are on the value of Cohen's d, calculated by scaling the CI by the pooled SD.)

Theories differ, however, in their predictions about people's use of the $B \rightarrow Z$ link. By Bayesian norms, people should rely on this link to the extent that some weight is still assigned to B. Thus, manipulating $P(Z|B)$ ought to influence judgments of Z. However, if people digitize and tacitly assign all weight to the likeliest explanation (A), then their predictions about Z would be similar across levels of $P(Z|B)$. This can be tested by comparing the *low/high* and *low/low* conditions, which differ only in whether $P(Z|B)$ is high or low. The digitization hypothesis was supported, as there was no difference at all between these conditions [$t(77) = -0.80$, $p = .43$, 95% CI_d[–0.35,0.15]]. That is, participants behaved as though the simple explanation was not merely likely but *certain* and the complex explanation was not merely improbable but entirely ignorable.

These results suggest that people assume that a single hypothesis is true when making predictions, ignoring the possibility that other explanations could be correct. However, it is unclear whether this finding would be restricted to an explanatory preference as robust as simplicity. Although complex explanations ought to carry some weight, it is true that simpler explanations usually are more probable (Lombrozo, 2007). Would the results extend to predictions made based on a weaker—and, by Bayesian standards, non-normative—explanatory preference?

Participants in Study 2, consistent with previous research, preferred the explanation that did *not* make the unverified prediction (A). Reflecting the lesser robustness of the latent scope bias compared to simplicity, only 46% of participants chose explanation A for all three items (though chance responding would be much lower at 12.5%).

Participants who consistently favored the narrow latent scope explanation digitized this belief (Table 1). Participants did rely on P(*Z*|*A*), leading to a significant difference between the *high*/*low* and *low*/*low* conditions [*t*(49) = 2.25, *p* = .029, 95% CI$_d$[0.04,0.67]]. (This result is not significant when participants are included who did not identify P(*A*) as the likeliest hypothesis for all three problems.) But participants ignored P(*Z*|*B*), failing to differentiate at all between the *low*/*high* and *low*/*low* conditions [*t*(49) = –0.16, *p* = .87, 95% CI$_d$[–0.24,0.21]]. Thus, even though the narrow latent scope bias is relatively weak by comparison with the simplicity bias, participants who had this bias treated the narrow latent scope explanation as though certainly true for the purpose of making predictions.

	Condition				Effect	Effect	Normative
$P(Z\mid A)$ / $P(Z\mid B)$	**Low / Low**	**High / Low**	**Low / High**	**High / High**	**of $P(Z\mid A)$**	**of $P(Z\mid B)$**	**Effect of $P(Z\mid B)$**
Study 1	50.7 (23.6)	71.7 (18.3)	48.5 (21.1)	———	20.9 (25.5)	–2.2 (24.5)	———
Study 2	53.4 (20.8)	60.4 (18.9)	53.0 (19.1)	———	7.0 (22.0)	–0.4 (15.8)	———
Study 3	57.9 (23.5)	83.6 (11.4)	53.7 (20.2)	81.9 (19.5)	27.0 (19.5)	–3.0 (19.5)	———
Study 4	60.2 (22.7)	70.6 (18.6)	59.3 (23.2)	———	10.4 (25.8)	–0.9 (24.1)	5.4
Study 5	54.0 (28.5)	66.6 (16.4)	55.9 (21.0)	———	12.6 (27.8)	1.9 (28.2)	6.8
Study 6	55.7 (28.2)	62.3 (20.0)	55.6 (23.8)	———	6.5 (27.5)	–0.1 (27.5)	3.3
Study 7	40.3 (25.9)	43.0 (24.0)	46.1 (24.6)	———	2.8 (23.0)	5.8 (25.8)	———
Study 8A	66.0 (25.7)	71.5 (19.0)	60.7 (21.0)	———	5.5 (25.6)	–5.3 (28.5)	2.9
Study 8B	32.8 (26.0)	72.1 (19.4)	40.0 (28.0)	———	39.3 (30.2)	7.2 (33.3)	10.9
Study 8C	34.3 (27.9)	64.8 (17.6)	48.0 (23.4)	———	30.4 (21.8)	13.7 (23.7)	16.4
Study 8D	42.2 (27.0)	76.8 (15.7)	40.3 (22.1)	———	34.6 (30.7)	–1.8 (21.8)	11.4

Table 1. Means and SDs across all studies.

Note. Entries in first four data columns are probabilistic predictions, expressed as percentages (pooled SDs in parentheses). All scores are computed on only participants who did not judge P(*B*) > P(*A*). Effect of P(*Z*|*A*) and P(*Z*|*B*) columns reflect the differences among the prediction columns and SDs of those differences for Studies 1–2 and 4–8; they reflect pooled mean differences and SDs for Study 3. Normative Effect of P(*Z*|*B*) requires estimates of P(*A*) and P(*B*), which were only available for Studies 4–6 and 8A–D.

Discussion

These studies show that when people choose between two hypotheses as explanations for evidence, they not only treat the unchosen hypothesis as unlikely, but ignore it entirely when using those hypotheses to make predictions. We know, based on previous research (Khemlani et al., 2011; Lombrozo, 2007) that, although people dislike explanations that are complex or make unverified predictions, people are nonetheless willing to ascribe some probabilistic weight to such explanations. Despite such ascriptions, participants here ignored these lower-probability possibilities when using the outcome of their hypothetical reasoning to make further predictions.

Study 3

Many previous studies have revealed digitization in category use—one type of hypothetical thinking, in which people hypothesize a category to explain an object's features, and then use that hypothesized category to make predictions about other possible features. Although we suggest that the mind uses digitization as a much broader strategy in high-level cognition, we can use prior studies within categorization to motivate possible boundary conditions (e.g., Murphy, Chen, & Ross, 2012). Studies 3–7 therefore address several potential boundary conditions on digitization, along with clearing up some potential methodological concerns.

Study 3 served two primary purposes. First, we addressed a methodological concern about Studies 1 and 2. We attempted to disguise the manipulation by using multiple items with different cover stories, but it is possible that some participants detected the manipulation. This could induce pressure for consistency—which could tend to shrink the effect of $P(Z|A)$ or $P(Z|B)$—or alternatively pressure to differentiate the three conditions—which could lead participants either to respond more to $P(Z|A)$ or to $P(Z|B)$. Given that we found an effect of $P(Z|A)$ but not $P(Z|B)$, as our digitization theory predicts, it seems unlikely that demand characteristics can explain this pattern. Nonetheless, a purer test would adopt a between-subjects design. We do this in Study 3, also adding a *high/high* condition to complete the 2x2 design.

Second, we addressed a theoretical question. To the extent that digitization can be adaptive, it is because it simplifies complex chains of inference and allows one to arrive at predictions in situations where calculating the normative answer is not feasible, and any equivalently simple strategy will do worse. Thus, it accepts a cognitive bias in exchange for an answer that is likely to be in the right ballpark. But are people *aware* of this bias? That is, do people experience meta-cognitive doubt about their predictions when there is conflicting evidence that they are ignoring? If so, we would expect explicit confidence judgments to be higher when $P(Z|A)$ and $P(Z|B)$ are similar, and lower when they are in conflict. This could permit some hedging at times when the strategy is likeliest to be fallible. Are people cognitively biased but meta-cognitively woke?

Methods

The method was the same as Study 1, with three changes. First, participants were asked to give confidence ratings after estimating the probability of Z ("How confident are you in this probability judgment?") to test the possibility that participants could detect conflicting information (even if they do not use it). Second, a *high/high* condition was included, in which Z was likely given either possibility. Finally, Study 3 used a between-subjects design, with participants completing an item from only one of the four conditions.

We recruited 120 participants. The exclusion criteria were similar to Studies 1 and 2. Participants were excluded from all analyses if they failed the check questions ($N = 1$). Participants were also excluded from the primary analysis if they did not choose explanation A ($N = 21$), but the significance levels are similar with these participants included.

Results and Discussion

Participants in Study 3 digitized, replicating Study 1 in a between-subjects design. The great majority of participants (82%) concluded that the simple explanation *A* was likelier than the complex explanation *B*. Comparing the *high*/*low* and *low*/*low* conditions revealed that these participants robustly relied on the strength of the $A \rightarrow Z$ link [$t(47) = 4.70$, $p < .001$, 95% $\text{CI}_d[0.77,1.93]$]. Since Study 3 also includes a *high*/*high* condition, we can compare this condition to the *low*/*high* condition to get a second estimate of participants' reliance on $P(Z|A)$, which reveals a similarly large effect [$t(47) = 4.96$, $p <$ .001, 95% $\text{CI}_d[0.84,2.00]$].

Analogously, we can test the effect of $P(Z|B)$ in two ways. Comparing the *low*/*high* and *low*/*low* conditions, as in Studies 1 and 2, we once again find no difference [$t(51) = -0.70$, $p = .49$, 95% $\text{CI}_d[-0.74,0.36]$]. Similar conclusions are reached by comparing the *high*/*high* and *high*/*low* conditions [$t(43) = -0.37$, $p = .71$, 95% $\text{CI}_d[-0.71,0.49]$]. As in Studies 1 and 2, participants digitized, considering the likely possibility to the total exclusion of the less-likely possibility. Since the design was between-subjects, experimenter demand cannot explain these effects. In fact, these effects were more consistent with digitization than in any of our other (within-subjects) studies, suggesting that demand effects if anything attenuate the tendency to digitize.

Although participants ignored the less-likely possibility when making predictions, could they nonetheless detect greater conflict when the $A \rightarrow Z$ and $B \rightarrow Z$ links differ in strength? Analyses of the confidence ratings produced inconsistent evidence for this possibility. On the one hand, participants did report lower confidence in the *low*/*high* condition [$M = 51.96$, $SD = 31.75$] than in the *low*/*low* [$M = 69.07$, $SD = 24.41$; $t(51) = 2.20$, $p = .032$, 95% $\text{CI}_d[0.05,1.16]$] or the *high*/*high* condition [$M = 75.83$, $SD = 25.62$; $t(47) = 2.87$, $p = .006$, 95% $\text{CI}_d[0.25,1.40]$]. However, confidence in the *high*/*low* condition [$M = 76.77$, $SD = 21.44$] was *not* lower than in the *low*/*low* or *high*/*high* conditions [ts < 1.2, ps > .25]. Thus, while participants seem to have some glimmer of cognitive conflict in some task conditions in which they are digitizing, this seems to only inconsistently lower confidence.

Study 4

We have shown that the within-subjects nature of our manipulation of conditional probabilities is not producing demand effects that magnify our effects—if anything, it is the opposite. But what about the demand associated with forcing participants to choose one of the two explanations as most probable? Perhaps participants take this as a signal that they are free to ignore the unchosen explanation. Consistent with this, some studies (Hayes & Newell, 2009; Murphy et al., 2012; Murphy & Ross, 2010) have found that participants who explicitly quantify their uncertainty about each of the possible categories prior to making a category-based prediction are more likely to hold multiple categories in working memory and therefore to rely on them for prediction. (We note, however, that it is debatable which design better mirrors real-world prediction.) In Study 4, we address this concern by asking participants to rate the probabilities of each hypothesis instead of making a forced-choice.

This change also helps to clear up a second possible issue. In Studies 1–3, we assumed that participants assign non-trivial probabilities to the unchosen hypotheses, as this would be consistent with prior research on the simplicity and latent scope biases (Khemlani et al., 2011; Lombrozo, 2007), but we did not measure the magnitude of these probabilities. If participants are assigning, say, a 95% probability to the simple hypothesis and a 5% probability to the complex hypothesis, this is almost certainly a non-normatively strong simplicity preference. But it is not where we are locating participants' bias in this task. If participants *explicitly* assign such probabilities to the hypotheses and then rationally use Bayesian methods to make predictions, then people are not digitizing at all—they are simply biased in their assignment of probabilities to explanations. We are claiming, instead, that people assign graded, moderate probabilities to simple and complex explanations, and to narrow and wide latent scope explanations, but that these moderate probabilities get converted, in effect, into 0s and 1s when used to make downstream predictions. Measuring these probabilities explicitly allows us to rule out this alternative locus of error.

Methods

The method was the same as Study 1, except that participants were asked to estimate the probability of A [P(A)] and of B [P(B)] rather than making a forced choice between A and B. This has two advantages. First, this avoids experimenter demand to focus only on one explanation, and, if anything, would seem to encourage participants to weight both explanations. Second, this measurement allows a calculation of how much larger the effect of manipulating the $A \rightarrow Z$ link should be, relative to manipulating the $B \rightarrow Z$ link. This makes it possible to compare performance to this normative benchmark. Ratings of P(A) and P(B) were made on a separate page prior to rating P(Z), and the probability information was repeated at the top of both pages.

We recruited 120 participants. The exclusion criteria were similar to Studies 1–3. Participants were excluded from all analyses either if they failed the check questions or if their summed judgments of P(A) and P(B) were outside the range of 80–120% for at least one item (N = 18). Participants were also excluded from the primary analysis if they did not rate P(A) greater than or equal to P(B) for all three items (N = 30), but the significance levels are similar with these participants included.

Results and Discussion

Participants in Study 4 digitized, despite the task demand to explicitly quantify their uncertainty by estimating P(A) and P(B). Most participants (71%) rated P(A) higher than P(B) for all items. Of greater interest, the mean estimate of P(A) among these participants was 65.9% (SD = 16.3%) and the mean estimate of P(B) was 34.1% (SD = 16.3%). This undermines the deflationary explanation of prior studies that participants assigned such low probabilities to B that P($Z|B$) can rationally be ignored. A one-third probability is an awfully big chance to ignore.

Despite these large probabilities assigned to the complex explanation, participants digitized as they did in Studies 1–3. Participants significantly differentiated between the *high/low* and the *low/low* conditions [t(71) = 3.41, p = .001, 95% CI$_d$[0.21,0.79]], indicating reliance on P($Z|A$). However, participants did *not* differentiate between the *low/high* and *low/low* conditions [t(71) = –0.33, p = .74, 95% CI$_d$[–0.29,0.21]], showing no reliance at all on P($Z|B$). Despite participants' insistence that B had a 1 in 3 chance of being the correct explanation, they completely ignored it in making predictions.

A further critique is that any statistical test will have greater power for detecting an effect of P($Z|A$) than P($Z|B$), since the former effect is normatively larger than the latter. We cannot reject the hypothesis that the effect of P($Z|B$) is 0, but can we reject the hypothesis that the effect of P($Z|B$) is normatively appropriate? Using P(A) and P(B), we can calculate how much larger the effect of

$P(Z|A)$ should be compared to $P(Z|B)$ [i.e., $P(A)/P(B)$], which allows us to establish the normative benchmarks for the effect of $P(Z|B)$ in Table 1. (It is worth noting that this analysis cannot distinguish between *over*weighting A versus *under*weighting B, since the calculation is looking at the ratio of $P(A)$ and $P(B)$.) For Study 4, we calculate that the normative effect of $P(Z|B)$ [i.e., the difference between the *low/high* and *low/low* conditions] should be 52% as large as the effect of empirically observed effect of $P(Z|A)$ (i.e., the difference between the *high/low* and *low/low* conditions). The actual effect of $P(Z|B)$ fell significantly short of this benchmark [$t(71) = 2.22$, $p = .030$, 95% $CI_d[0.03,0.50]$]. Thus, we can statistically reject the hypothesis that people used $P(Z|B)$ to the degree required by Bayesian norms.

The previous analysis is predicated on participants' judgments of $P(A)$ being higher than $P(B)$, using the means of these estimates to reduce noise. However, it is also possible to compare every participants' judgments of $P(Z)$ to Bayesian benchmarks, based on whatever values of $P(A)$ and $P(B)$ they individually reported. This allows us to exclude fewer participants. On this analysis too, the observed effect of manipulating $P(Z|B)$ was less than it ought to have been, relative to $P(Z|A)$ [$t(101) = 2.46$, $p = .016$, 95% $CI_d[0.05,0.48]$]. This analysis of individual participants thus corroborates the overall pattern of means, indicating that, relative to the simple explanation, participants insufficiently weighted (in fact, did not weight at all) the complex explanation in estimating $P(Z)$.

After presenting all studies, we provide an internal meta-analysis (McShane & Böckenholt, 2017) of two key questions: (1) Is there evidence for a non-zero weight on lower-probability hypotheses? and (2) Is this weight (whether zero or non-zero) too low relative to normative benchmarks? The results of Study 4 provide no evidence for a non-zero weight on $P(Z|B)$ and statistically significant evidence for underweighting $P(Z|B)$ relative to $P(Z|A)$. The meta-analysis will allow us to test these questions with a much larger dataset.

Study 5

We noted that predictions from uncertain hypotheses depend on three processes: (1) *Abduction* or *inference to the best explanation* (Lipton, 2004) to assign weights to the hypotheses [$P(A)$ and $P(B)$]; (2) *Simulation* or *counterfactual reasoning* (Lewis, 1979) to imagine how those hypotheses impact potential predictions [$P(Z|A)$ and $P(Z|B)$]; and (3) *Evidence integration* to put this information together (Anderson, 1991). In Study 5, we begin exploring the locus of the informational bottleneck that prevents people from integrating across multiple hypotheses (we turn to this task in earnest in Study 8).

In our studies so far, participants have arrived at their beliefs about hypotheses A and B on the basis of inference to the best explanation (Douven, 1999; Lipton, 2004)—that is, participants were presented with some evidence and then asked to account for that evidence. We manipulated the perceived probabilities of these explanations by using simplicity (i.e., A invokes one cause while B invokes two causes) or fit to the evidence (i.e., A does not make unverified predictions while A does). Is there something special about this deliberative process that crystallizes beliefs in our minds? For example, when people seek explanations they often overgeneralize and fail to account for exceptions (Williams, Lombrozo, & Rehder, 2013), suggesting that the process of explanation itself may cause people to over-rely on a dominant explanation. Further, the judged probability that a hypothesis is true depends on its perceived explanatory qualities, over-and-above objective probabilities (Douven & Schupbach, 2015). On the other hand, the processing bottleneck could occur later, in simulating the prediction or integrating across possible worlds, given capacity limits in mental simulation and imperfections in evidence integration (e.g., Griffin & Tverksy, 1992; Hegarty, 2004).

Study 5 begins to test this issue by directly providing the hypothesis probabilities P(*A*) and P(*B*), eliminating the need for participants to infer these probabilities through abduction. If abduction or explanatory inference leads people to digitize, then we should no longer see digitization under these conditions. If the bottleneck lies in later processing, we should still see digitization here.

Methods

The method was the same as Study 4, except that P(*A*) and P(*B*) were set by directly providing the posterior probabilities of each hypothesis given the evidence. For example:

> Freshwater juga snails cause lakes to lose sculpin fish and lose crayfish.
> Freshwater scuta snails cause lakes to lose sculpin fish and lose crayfish.
>
> Of the lakes that have a loss of sculpin fish and a loss of crayfish, 65% of them have juga snails and 35% of them have scuta snails.
>
> When a lake has juga snails, it [*usually* / *occasionally*] has bacteria proliferation.
> When a lake has scuta snails, it [*usually* / *occasionally*] has bacteria proliferation.
>
> Crescent Lake has a loss of sculpin fish and crayfish.

We recruited 120 participants. The exclusion criteria were the same as Study 4. Participants were excluded from all analyses either if they failed the check questions (same criterion as Studies 1–3) or if their summed judgments of P(*A*) and P(*B*) were outside the range of 80–120% for at least one item (*N* = 10). Participants were also excluded from the primary analysis if they did not rate P(*A*) greater than or equal to P(*B*) for all three items (*N* = 37), but the significance levels are similar with these participants included.

Results and Discussion

Participants in Study 5 digitized, despite P(*A*) and P(*B*) being set directly through explicit posterior probabilities. Most participants (66%) correctly identified P(*A*) as higher than P(*B*) for all three items. These participants' judgments of P(*Z*) were again significantly higher in the *high*/*low* compared to the *low*/*low* condition [$t(72) = 3.87$, $p < .001$, 95% CI_d[0.26,0.82]]. Thus, participants relied on P(*Z*|*A*), as they normatively ought to. However, people once again ignored the possibility that *B* was the correct explanation for the observations, despite its one-third chance of being true. Judgments of P(*Z*) were no higher in the *low*/*high* than the *low*/*low* condition [$t(72) = 0.58$, $p = .56$, 95% CI_d[–0.19,0.34]].

As in Study 4, participants' estimates of P(*A*) and P(*B*) allow for a normative calculation of how estimates of P(*Z*) ought to differ across conditions. Once again, participants relied on P(*Z*|*B*), relative to P(*Z*|*A*), significantly less than Bayesian norms [$t(109) = 2.19$, $p = .031$, 95% CI_d[0.02,0.34]]. As in previous studies, there is no evidence that participants place positive weight on $P(Z|B)$ and there is evidence that participants place insufficient weight on $P(Z|B)$.

These results show that digitization is not restricted to cases in which people evaluate hypotheses through explanatory reasoning (e.g., using simplicity as a cue to probability; Lombrozo, 2007), but occurs even for explicitly provided posteriors. As there are theoretical reasons to think that abduction or explanation could lead people to focus disproportionately on particularly satisfying explanations (e.g., Douven, 2015; Lipton, 2004; Williams et al., 2013), it remains possible that abduction is an additional contributing factor to digitization, even if not a necessary one.

Moreover, these results further rule out the possibility that participants digitize because they assign explicitly low probabilities. Here, the explicit posteriors indicated a 35% probability of the less likely hypothesis, which was ignored and treated as a 0% probability. Together with our other studies in which explicit probabilities are provided or measured, it is unlikely that digitization is an artifact of assigning explicitly very low—perhaps rationally ignorable—probabilities.

Study 6

So far we have seen that people act as though they "round up" a probability of 65% to 100%, and "round down" a probability of 35% to 0%. However, these studies looked at situations where only two hypotheses were in competition. Although people have a sharply limited number of hypotheses they can evaluate at a given time and must adopt strategies to limit this number to a manageable size (e.g., Johnson & Keil, 2014; Lagnado, Waldmann, Hagmayer, & Sloman, 2007), many realistic situations would require people to evaluate more than two possibilities. In the example in the introduction, one's open front door could signify one's own mistake, or a neighbor's visit, or a burglary. In such cases, it is possible that *no* hypothesis has a more than 50% credence. For example, one could assign a 50% chance to the personal-error hypothesis, 40% chance to the neighbor hypothesis, and 10% chance to the burglar hypothesis. In this case, would people really be willing to rely solely on a hypothesis with only 50% probability?

If digitization occurs because we have difficulty entertaining and integrating across multiple possibilities at once, then they should be willing to ignore the 40% and 10% chances in this example, and act as though the plurality belief (with 50% probability) is certain. Study 6 tests this possibility.

Methods

The method was the same as Study 5, except that three hypotheses rather than two were enumerated, with the highest probability explanation having a 50% posterior probability and the others having 25% posterior probability. For example:

> There are three types of snails that can each cause a lake to lose sculpin fish and to lose crayfish: juga snails, scuta snails, and aspera snails.
>
> Of the lakes that have a loss of sculpin fish and a loss of crayfish, 50% of them have juga snails; 25% of them have scuta snails; and 25% of them have aspera snails.
>
> When a lake has juga snails, it [*usually* / *occasionally*] has bacteria proliferation.

When a lake has scuta snails, it occasionally has bacteria proliferation.
When a lake has aspera snails, it [*usually* / *occasionally*] has bacteria proliferation.

Crescent Lake has a loss of sculpin fish and crayfish.

We recruited 120 participants. The exclusion criteria were similar to Study 5. Participants were excluded from all analyses either if they failed the check questions or if their summed judgments of P(*A*), P(*B*), and P(*C*) were outside the range of 80–120% for at least one item (*N* = 27). Participants were also excluded from the primary analysis if they did not rate P(*A*) greater than or equal to P(*B*) and P(*C*) for all three items (*N* = 14), but the significance levels are similar with these participants included, with one exception noted below.

Results and Discussion

Participants in Study 6 digitized, despite the likeliest possibility having only a 50% probability. Most participants (85%) correctly identified P(*A*) as higher than P(*B*) or P(*C*) for all three problems. These participants' judgments of P(*Z*) were higher in the *high*/*low* than the *low*/*low* condition [$t(78) = 2.12$, $p = .037$, 95% $\text{CI}_d[0.02,0.52]$], indicating that shifts in P(*Z*|*A*) influenced judgments of P(*Z*). (This result becomes marginally significant when participants are included who did not identify P(*A*) as the likeliest hypothesis for all three problems.) There was no difference between the *low*/*high* and *low*/*low* conditions [$t(78) = –0.04$, $p = .97$, 95% $\text{CI}_d[–0.24,0.23]$], with the means slightly in the opposite direction of normative benchmarks. These results suggest that shifts in P(*Z*|*B*) had no effect.

Unlike Studies 4 and 5, however, the under-reliance on P(*Z*|*B*) relative to P(*Z*|*A*) did not reach significance [$t(92) = 0.87$, $p = .39$, 95% $\text{CI}_d[–0.11,0.28]$]. The evidence for digitization is therefore less clear-cut compared to previous studies, where this under-reliance was statistically significant. That said, this study did not find any evidence suggesting that participants did rely on P(*Z*|*B*), so the results are broadly consistent with Studies 1–5 albeit less robust.

Study 7

Study 6 showed that people can still treat a hypothesis as having all the probability mass even if it has only a plurality of the mass—people relied only on a possibility that had a 50% chance of being true, to the exclusion of two possibilities each with a 25% chance. But in other situations, it seems that we might be willing to acknowledge that *some* low-probability event is likely, even if we do not know which one. People buy insurance because they understand that on any particular day they are unlikely to get into a car accident, but over many years the disjunction of these daily probabilities becomes large. People avoid Russian roulette because a series of 1 in 6 chances adds up to a serious risk sooner or later. In fact, people often *overestimate* the probability of disjunctive events in their explicit judgments (e.g., Rottenstreich & Tversky, 1997; Yousif & Keil, 2019). Could this phenomenon lead to boundary conditions on digitization?

It could, with two caveats. First, we would expect people to spontaneously consider a disjunctive hypothesis when it is possible to consider the component hypotheses as parts of the same category (e.g., Waldmann & Hagmayer, 2006). In Study 7, for instance, participants were provided with a set of several hypotheses belonging to the same category, which collectively have a high probability even though each hypothesis has a low probability. We would expect people to readily lump together the events "has Type G scuta snails" and "has Type S scuta snails" into a broader "has some kind of scuta snails" hypothesis. If people do add up these component hypotheses, the resulting disjunction may actually be assigned a higher-than-normative explicit probability weight (Rottenstreich & Tversky, 1997; Yousif & Keil, 2019).

Second, if people do take account of disjunctive hypotheses, we would expect them to do so digitally—to ignore a non-disjunctive hypothesis (e.g., "30% of ponds have juga snails") in favor of the disjunction, if the disjunction has sufficient probability mass. Therefore, if less than half of the probability mass falls on a single hypothesis and more than half falls on a disjunction of other hypotheses that can be treated as a category, we would expect the disjunction to be used to the exclusion of the single hypothesis in making predictions.

In addition to testing this possibility, Study 7 also addresses a methodological concern. In some of our other experiments, participants were given the posterior probability of each hypothesis, given the evidence. These tended to add up to 100%, which could have led participants to see the hypotheses as mutually exclusive and exhaustive. In Study 7, we provide base rates, which do not sum to 100%. We nonetheless anticipate that participants would digitize the disjunctive probability.

Methods

Similar to the previous studies, participants read three scenarios. For example, the *low/low* version of one item read:

> Crescent Lake has a loss of sculpin fish and crayfish.
>
> There are eight types of snails that can each cause a lake to lose sculpin fish and to lose crayfish:
>
> 30% of lakes have juga snails. When a lake has juga snails, it occasionally has bacteria proliferation.
>
> 5% of lakes have Type C scuta snails. When a lake has Type C scuta snails, it occasionally has bacteria proliferation.
>
> 5% of lakes have Type G scuta snails. When a lake has Type G scuta snails, it occasionally has bacteria proliferation.
>
> 5% of lakes have Type L scuta snails. When a lake has Type L scuta snails, it occasionally has bacteria proliferation.
>
> 5% of lakes have Type M scuta snails. When a lake has Type M scuta snails, it occasionally has bacteria proliferation.
>
> 5% of lakes have Type S scuta snails. When a lake has Type S scuta snails, it occasionally has bacteria proliferation.

> 5% of lakes have Type W scuta snails. When a lake has Type W scuta snails, it occasionally has bacteria proliferation.
>
> 5% of lakes have Type Z scuta snails. When a lake has Type Z scuta snails, it occasionally has bacteria proliferation.

That is, one hypothesis (*A*) had a relatively high base rate (30%), and seven other hypotheses (B_1–B_7) belonging to the same category each had a low base rate (5%). Assuming that these events are independent, there is a probability of 30% that at least one of this disjunction of 7 events occurs ($1 - .95^7 = .30$); thus, the single-possibility and disjunction are equated in probability.

In the *low/low* version, as above, all hypotheses led "occasionally" to the prediction *Z*. In the *high/low* condition, the single-possibility hypothesis *A* led "usually" to *Z*, while the B_i's led "occasionally" to *Z*. In the *low/high* condition, *A* led "occasionally" to *Z*, while the B_i's led "usually" to *Z*. For each item, participants judged the probability of *Z* on the same scale used in previous studies; questions about the probabilities of *A* and the B_i's were omitted to avoid eliciting such a large number of judgments.

We recruited 120 participants. Participants were excluded from analysis if they failed the check questions ($N = 21$). We did not exclude participants based on their judgments of P(*A*) or P(B_i), since we did not collect these data.

Results and Discussion

Participants in Study 7 digitized, but in the opposite manner as previous studies. Participants ignored the high-probability *A* event and instead relied exclusively on the disjunction of low-probability B_i events. That is, participants' judgments of P(*Z*) were similar in the *high/low* and *low/low* conditions [$t(98) = 1.20$, $p = .23$, 95% $\mathrm{CI}_d[-0.07,0.30]$], indicating no significant responsiveness to shifts in P(*Z*|*A*). But participants judged P(*Z*) significantly higher in the *low/high* than in the *low/low* condition [$t(98) = 2.24$, $p = .027$, 95% $\mathrm{CI}_d[0.03,0.43]$], indicating that participants responded to shifts in the P(*Z*|B_i)'s. (As we did not collect judgments of P(*A*), P(B_1), etc., we cannot compare judgments of P(*Z*) to normative benchmarks.)

These results suggest that people can sometimes shift from digitizing a single high-probability hypothesis to digitizing a disjunction of individually low-probability hypotheses. In a sense this is a boundary condition on our effects—people are no longer treating high-probability hypotheses as certain. But in another sense, it is an additional manifestation of digital thinking—since people can sometimes group causes into categories (Waldmann & Hagmayer, 2006) and often overestimate the probability of disjunctive events (Yousif & Keil, 2019), a hypothesis composed of a disjunction of low-probability hypotheses may be used to the exclusion of a single high-probability hypothesis. Studying the processes that lead people to lump together causal events into a single category, facilitating this alternate form of digital thinking, is an important direction for future research.

Interim Discussion

Taken together, Studies 3–7 rule out several deflationary explanations of digitization. This effect occurs in between-subjects (Study 3) and within-subjects designs (Studies 1–2, 4–7). It occurs whether participants commit to one single hypothesis being most likely (Studies 1–3), explicitly quantify their uncertainty (Studies 4–6), or do not report on their beliefs at all (Study 7). It also is not a result of explicitly assigning very low probability to some hypotheses: People report reasonably high values for P(*A*) and P(*B*) (about one-third) when asked explicitly (Study 4), and digitization even persists when explicit—and reasonably high—posterior probabilities (Studies 5 and 6) or base rates (Study 7) are given.

Moreover, these studies provide some suggestions about people's underlying cognition surrounding digitization. Study 3 looked at people's metacognition, finding that people appear to experience some modest cognitive conflict (in the form of lower confidence) when P(*Z*|*A*) and P(*Z*|*B*) differ. This can be a useful check on the overconfident probability assessments systematically produced by a digitization strategy. Study 5 found that digitization does not depend on arriving at beliefs about P(*A*) and P(*B*) through explanatory reasoning processes, but instead persists even when these are set through explicit posterior probabilities. Study 6 suggests that people will even digitize a probability as low as 50% (treating it as certain for making predictions) so long as it represents a plurality of the probabilistic weight (against two hypotheses each assigned 25% weight). And Study 7 found that people will digitize disjunctions of small-probability hypotheses belonging to the same category, ignoring a larger probability from a single hypothesis. In our final set of studies, we further test where the processing bottleneck is, seeking a boundary condition.

Study 8

Study 5 showed that abduction or inference to the best explanation is not required for digitization to occur. That is, even if people are provided P(*A*) and P(*B*) explicitly, rather than generating these values themselves, they are digitized. Studies 8A–D sought to build on this result by more systematically altering the processing demands and observing the effects on digitization.

If the processing bottleneck is not in abduction—estimating P(*A*) and P(*B*)—then it must be either in simulation—estimating P(*Z*|*A*) and P(*Z*|*B*)—or in integration—combining these four quantities to produce an estimate of P(*Z*). Although people are indeed highly imperfect in integrating evidence (Griffin & Tversky, 1992 Kvam & Pleskac, 2016), people are able to account for multiple sources of evidence. Thus, simulation may be the most plausible locus of the effect. If people have sharp limits on their attentional and memory capacity for conducting mental simulations, they may more readily reason *within* one possible world rather than *across* multiple possible worlds. For example, Hegarty (2004) found that people conduct mental simulations of mechanical systems (e.g., a system of gears) by computing piecemeal each individual step in that system (e.g., how one gear's motion impacts an adjoining gear's motion) rather than by simulating the entire system at once. This suggests that people have a very low working memory capacity for simulating possible worlds (see Cavanagh & Alvarez, 2006 for related results in object tracking). People can compute a prediction from within a *single* possible world because they can simulate one piece at a time. But they may be unable to simulate multiple possible worlds simultaneously and integrate the results together. Thus, we argue that people digitize to avoid the computational complexity of imaging consequences simultaneously in multiple possible worlds. (See also Murphy et al., 2012, who argue that digitization effects in category-based induction are due to limitations in the number of categories held in working memory.)

Study 8 sought to distinguish among these possible processing bottlenecks. Although people often have difficulty reasoning about probabilities relative to other information formats (Gigerenzer & Hoffrage, 1995), we anticipated that providing probabilities that participants are already seeking to compute would reduce processing demands by offloading or "shutting off" these computations. By explicitly providing the hypothesis probabilities P(A) and P(B) (as in Study 5), we can offload abduction, since the purpose of abduction is determining the relative plausibility of A and B. Analogously, by explicitly providing the predictive probabilities P($Z|A$) and P($Z|B$) we can offload simulation, since the purpose of simulation is determining the plausibility of Z in each possible world (i.e., where A is true or where B is true). Across Studies 8A–D, we vary whether these probabilities were provided explicitly to see how differing processing demands influence the ability to account for multiple hypotheses.

In Study 8A, we conceptually replicate Study 5 by providing P(A) and P(B) explicitly, while leaving P($Z|A$) and P($Z|B$) vague. We assume that this "shuts off" abduction but still requires simulation and integration. If people now account for both possibilities, this would imply that digitization occurs at the abduction stage—that people accept a single explanation as "best" and ignore all others. However, we anticipated that we would continue to see digitization here (replicating Study 5), which would imply that abduction per se is not to blame for digitization, and the error lies in the simulation or integration processes.

In Study 8B, we provide P($Z|A$) and P($Z|B$) explicitly, while leaving P(A) and P(B) vague. We assume that this "shuts off" simulation, but still requires abduction and integration. If people now account for both possibilities, this would imply that digitization occurs at the simulation stage—that people can only keep one simulation world in mind at a time. If people continued to digitize here, this would imply that digitization occurs either at the abduction or integration stage. Thus, seeing digitization in Study 8A but not in Study 8B would localize digitization to the simulation stage, while seeing digitization in both studies would localize digitization to the final integration stage.

In Study 8C, we provide P(A), P(B), P($Z|A$), and P($Z|B$) explicitly. We assume that this "shuts off" abduction *and* simulation, testing only the integration process. If people still digitize, this would imply that digitization occurs at the integration stage—that people may be able to store the relevant information in their working memory, but cannot put them together. However, if either Study 8A or 8B found that people accounted for multiple hypotheses (localizing digitization to the abduction or simulation stage), we would expect consideration of multiple hypotheses in Study 8C too.

Finally, in Study 8D, we do not specify any of the four probabilities, but instead measure individual participants' judgments of them. This requires all three processes, and we therefore assumed we would see digitization here. But participants' explicit judgments allow us to ensure that Studies 8A–C are roughly equated in terms of explicit probability judgments.

Methods

Study 8A. The method was similar to Study 5, except P(*A*) and P(*B*) were provided directly rather than as base rates, and the procedure was simplified to facilitate comparison with Studies 8B–D. Thus, P(*Z*|*A*) and P(*Z*|*B*) were vague while P(*A*) and P(*B*) were precise. For example:

> When a lake has Juga snails, it [*usually* / *occasionally*] has bacteria proliferation.
> When a lake has Scuta snails, it [*usually* / *occasionally*] has bacteria proliferation.
>
> Crescent Lake has a 65% chance of having Juga snails and a 35% chance of having Scuta snails.

As in previous studies, participants then estimated P(*A*) and P(*B*), followed by P(*Z*) on the subsequent page. The above information was provided on both pages.

Study 8B. The method was the same as Study 8A, except P(*Z*|*A*) and P(*Z*|*B*) were precise while P(*A*) and P(*B*) were vague. Specifically, "about 20% of the time" was included parenthetically after each appearance of "occasionally," while "about 80% of the time" was included parenthetically after each appearance of "usually." Meanwhile, the information about P(*A*) and P(*B*) was replaced with vague probabilities (e.g., "Crescent Lake has a fairly high chance of having Juga snails and a fairly low chance of having Scuta snails").

Study 8C. The method was the same as Study 8A, except P(*Z*|*A*), P(*Z*|*B*), P(*A*), and P(*B*) were all specified precisely, using the phrasing of Study 8B for P(*Z*|*A*) and P(*Z*|*B*) and the phrasing of Study 8A for P(*A*) and P(*B*).

Study 8D. The method was the same as Study 8A, with two changes. First, P(*Z*|*A*), P(*Z*|*B*), P(*A*), and P(*B*) were all specified vaguely, using the phrasing of Study 8B for P(*Z*|*A*) and P(*Z*|*B*) and the phrasing of Study 8A for P(*A*) and P(*B*). Second, on the same page as this information, participants estimated P(*Z*|*A*), P(*Z*|*B*), P(*A*), and P(*B*). The estimates of P(*Z*|*A*) and P(*Z*|*B*) were made first ("When a lake has [*Juga* / *Scuta*] snails, what do you think is the probability that it has bacteria proliferation?") on a 0% to 100% scale, followed by P(*A*) and P(*B*) as in previous studies. Estimates of P(*Z*) were made on a separate page.

Participants. We recruited 120 participants for Study 8A, 8B, and 8D, and 119 participants for Study 8C. The exclusion criteria were the same as Studies 4 and 5. Participants were excluded from all analyses if they failed the check questions or if their summed judgments of P(*A*) and P(*B*) deviated too much from 100% (using the same criteria as previous studies; *N* = 8, 21, 6, and 18 for Studies 8A–D, respectively). As in Studies 4–6, participants were excluded from the primary analysis if they did not rate P(*A*) greater than or equal to P(*B*) for all three items (*N* = 21, 27, 41, and 39 for Studies 8A–D, respectively), but the significance levels are similar with these participants included, with one exception noted below.

Results

Overall, participants digitized when P(*Z*|*A*) and P(*Z*|*B*) were vague (Studies 8A and 8D) but had at least some capacity to integrate across hypothetical worlds when P(Z|A) and P(Z|B) were specified (Studies 8B and 8C). This was especially clear when all four probabilities were specified (Study 8C). Yet, across all of these studies, participants still consistently *underweighted* the lower-probability hypothesis, relative to the higher-probability hypothesis, even when they did not fully *ignore* it. Let's now walk through the results of each study individually.

In Study 8A, the hypothesis probabilities, P(*A*) and P(*B*), were specified explicitly, but the predictive probabilities, P(*Z*|*A*) and P(*Z*|*B*) were left vague as in previous studies. This set of conditions continued to produce digitization, replicating Study 5. Most participants (81%) correctly identified P(*A*) as higher than P(*B*) for all three items. Participants gave somewhat higher estimates of P(*Z*) in the *high/low* than in the *low/low* condition [$t(90) = 2.03, p = .045$, 95% $\text{CI}_d[0.01,0.48]$], although this effect was smaller than in previous studies. Most importantly, however, estimates of P(*Z*) in the *low/high* condition were no higher than in the *low/low* condition and were, if anything, somewhat lower [$t(90) = –1.78, p = .078$, 95% $\text{CI}_d[–0.48,0.03]$]. That is, once again, people did not take account of the less-likely hypothesis (*B*) when estimating P(*Z*). As in Studies 4 and 5, this led participants to underweight the *B*→*Z* link relative to the *A*→*Z* link, by Bayesian norms [$t(111) = 4.51, p < .001$, 95% $\text{CI}_d[0.22,0.57]$]. Thus, precise hypothesis probabilities are not sufficient to combat digitization. This suggests that digitization occurs somewhere downstream of abduction, since precise hypothesis probabilities forestall the need for abductive inferences.

In Study 8B, the predictive probabilities, P(*Z*|*A*) and P(*Z*|*B*), were specified explicitly, but the hypothesis probabilities, P(*A*) and P(*B*) were left vague. Thus, this should leave abduction intact, but obviates the need for simulating *Z* in each possible world. Unlike previous studies, here we found some evidence for reliance on the *B*→*Z* link. Most participants (73%) correctly identified P(*A*) as higher than P(*B*) for all three items. Among those participants, estimates of P(*Z*) were higher in the *high/low* than in the *low/low* condition [$t(71) = 11.03, p < .001$, 95% $\text{CI}_d[1.40,2.02]$], indicating robust use of P(*Z*|*A*). More importantly, there was also a marginally significant difference between the *low/high* and *low/low* conditions [$t(71) = 1.84, p = .071$, 95% $\text{CI}_d[–0.02,0.56]$]. (This result becomes significant, $p < .01$, when participants are included who did not identify P(*A*) as the likeliest hypothesis for all three problems.) Although this result was not very statistically robust, it seems likely that it reflects genuine use of the *B*→*Z* link rather than statistical noise because the difference between the *low/high* and *low/low* conditions was significantly larger than in Study 8A [$t(161) = 2.59, p = .011$, 95% $\text{CI}_d[0.10,0.72]$]. This suggests that specifying the predictive probabilities explicitly, unlike specifying the hypothesis probabilities, cues attention to uncertainty in making predictions based on those quantities. That said, participants in Study 8B still underweighted the *B*→*Z* link relative to the *A*→*Z* link [$t(98) = 3.28, p < .001$, 95% $\text{CI}_d[0.16,0.65]$]. Thus, precise predictive probabilities may be an important boundary condition on digitization.

If this is the case, then we would expect similar results in Study 8C, where both the hypothesis and predictive probabilities were specified, that is, where P(*A*), P(*B*), P(*Z*|*A*), and P(*Z*|*B*) were all explicit, so that only integration is required. Indeed, participants in Study 8C robustly relied on both P(*Z*|*A*) and P(*Z*|*B*). Most participants (64%) correctly identified P(*A*) as higher than P(*B*) for all three items. Among those participants, estimates of P(*Z*) differed among all three conditions. The *high/low* condition produced higher estimates of P(*Z*) compared to the *low/low* condition [*t*(71) = 11.81, *p* < .001, 95% CI$_d$[1.08,1.52]]. Crucially, however, so did the *low/high* condition [*t*(71) = 4.88, *p* < .001, 95% CI$_d$[0.31,0.75]]. This difference was larger than in Study 8A [*t*(161) = 4.54, *p* < .001, 95% CI$_d$[0.40,1.03]] but similar in magnitude to Study 8B [*t*(142) = 1.34, *p* = .18, 95% CI$_d$[–0.11,0.55]]. Thus, when all four probabilities are presented explicitly, participants robustly account for uncertainty among hypotheses, but this seems to be driven mainly by the specification of the predictive probabilities, P(*Z*|*A*) and P(*Z*|*B*). Together with Studies 8A and 8B, this suggests that abduction is not necessary for digitization, nor is integration sufficient. The simulation process appears to be the critical bottleneck.

Despite some reliance on the *B*→*Z* link, participants in Study 8C under-relied on it relative to the *A*→*Z* link, by Bayesian norms [*t*(112) = 2.72, *p* = .008, 95% CI$_d$[0.07,0.44]]. This indicates that even providing all probabilities explicitly is not a panacea for reasoning errors. It also suggests that higher-probability hypotheses receive a disproportionate share of attention even when people are able to take lower-probability hypotheses into account.

Finally, Study 8D presented participants with all four probabilities vaguely, but asked participants to explicitly quantify them. This problem is formally similar to Study 8C, but imposes more stringent processing demands since the probabilities were self-generated—requiring all three processes of abduction, simulation, and integration—and were not displayed on the same screen at the time P(*Z*) was estimated. Most participants (62%) reported higher P(*A*) than P(*B*) for all three problems. These participants digitized: Their judgments of P(*Z*) were significantly higher in the *high/low* than in the *low/low* condition [*t*(62) = 8.95, *p* < .001, 95% CI$_d$[1.22,1.92]], while judgments of P(*Z*) did not differ between the *low/high* and *low/low* conditions [*t*(62) = –0.66, *p* = .51, 95% CI$_d$[–0.30,0.15]]. This led participants to apply too little weight to the *B*→*Z* relative to the *A*→*Z* link [*t*(101) = 6.22, *p* < .001, 95% CI$_d$[0.45,0.88]], as in the other studies. Overall, participants once again digitized, despite the logical similarity between Studies 8C and 8D.

Indeed, Studies 8C and 8D were not just logically similar, but the mean probability estimates in Study 8D suggest that all of Studies 8A–D were very nearly mathematically equivalent. The mean estimate of P(*A*) was 74.7% (*SD* = 8.8%), compared to the 65% explicitly given in Studies 8A and 8C, while the mean estimate of P(*B*) was 24.6% (*SD* = 9.0%) compared to 35% explicitly given. The mean estimate of P(*Z*|*A*) and P(*Z*|*B*) was 81.6% (*SD* = 10.5%) when the probability was *high* ("usually"), compared to 80% explicitly given in Studies 8B and 8C, while the mean conditional probability estimate was 35.6% (*SD* = 18.6%) compared to 20% explicitly given. Although these participant-generated quantities are not identical to those provided explicitly in Studies 8A–C, they are not so far off that they would produce large divergences in the normative responses across studies. Thus, the differences across Studies 8A–D must be due to differences in cognitive processing rather than normative differences based on different sets of assumptions. Overall, participants were able to resist digitizing only when the predictive probabilities [P(*Z*|*A*) and P(*Z*|*B*)] were provided explicitly, not when they were inferred and reported by the participant. This manifested in similar reliance on the *B*→*Z* link relative to Study 8A [*t*(152) = 0.82, *p* = .41, CI$_d$[–0.19,0.46]] but less reliance relative to Studies 8B [*t*(133) = –1.83, *p* = .069, CI$_d$[–0.66,0.02]] and 8C [*t*(133) = –3.92, *p* < .001, CI$_d$[–1.02,–

0.34]]. Once again, it appears to be the simulation process required to generate the predictive probabilities, $P(Z|A)$ and $P(Z|B)$, that drives digitization.

Discussion

These results point to a boundary condition on digitization—people can take account of lower-probability hypotheses when the predictive probabilities, $P(Z|A)$ and $P(Z|B)$, are given. This is consistent with our theorizing that digitization occurs due to processing constraints associated with simulating multiple hypothetical worlds. Offloading the predictive probabilities through explicit presentations substantially enhances the ability to account for low-probability hypotheses. Nonetheless, participants continued to chronically underweight lower-probability relative to higher-probability hypotheses. This indicates that higher-probability hypotheses receive disproportionate attention even when lower-probability hypotheses are not ignored entirely.

Arguably, it is surprising that participants in Study 8D digitized, considering its very near equivalence to Study 8C, where people accounted for both hypotheses. Both studies included explicit probabilities for $P(A)$, $P(B)$, $P(Z|A)$, and $P(Z|B)$, and these probabilities were similar across studies. One important difference is that the probabilities were experimenter-provided in Study 8C and participant-generated in Study 8D. Perhaps people do not account for uncertainty when it is internally rather than externally estimated—an interesting possibility for future research. However, we think a related procedural difference between studies is the more likely culprit: That Study 8C was designed to offload abduction and simulation, whereas Study 8D was not. Study 8C provided the probabilities on the same screen as $P(Z)$ estimates, meaning that participants were never forced to perform abduction or simulation at all. But Study 8D deliberately did *not* allow these processes to be off-loaded, since participants had to estimate these quantities themselves and these estimates were not available when estimating $P(Z)$ as they occurred on separate screens.

What does Study 8 imply about real-world inference? That is, is it reasonable to say that digitization is the default mode of prediction, with this effect (partially) circumvented in informationally-rich environments providing explicit probabilities (as in Study 8B here)? Or is it more reasonable to say that people are Bayesians whose rational information processing is disrupted in informationally-poor environments? The answer turns on whether one thinks that the real world wears probabilities on its sleeve and we reason explicitly with those probabilities, or whether instead we use heuristic methods to avoid explicit probabilistic inference. We will not adjudicate this debate here, which runs parallel to the century-old debate in the economics profession between modeling *risk* (where probabilities are quantifiable and specifiable) and *Knightian uncertainty* (where they are not; Knight, 1921). Our own position is that most real-world probabilities can be assessed only coarsely and that the world rarely affords the information required to generate precise probabilities (e.g., Budescu & Wallsten, 1995; Yaniv & Foster, 1995). In other work, we have argued that people use a variety of heuristics to circumvent these limits on the specifiability of probabilities (Johnson & Tuckett, 2018; Johnson, Valenti, & Keil, 2019). If we are correct, then digitization is likely to be a pervasive force in real-world decision-making.

Meta-Analysis

A general objection to our methodology is that our studies are by their nature underpowered to detect a difference between the *low/high* and *low/low* conditions, since this difference is necessarily (normatively) smaller than the difference between the *high/low* and *low/low* conditions. One way of addressing this problem was to test, for individual studies, whether participants reported significantly smaller differences between the former two conditions, relative to a Bayesian standard. In all studies where this could be tested, except Study 6, participants did fall short of the normative benchmark.

Here, we use meta-analytic techniques to further buttress this conclusion. We included those studies reported here for which (i) the manipulation was within-subjects (excluding Study 3), (ii) we predicted that people would rely on A rather than B (excluding Study 7), and (iii) the predictive probabilities, $P(Z|A)$ and $P(Z|B)$ were vague (excluding Studies 8B and 8C); overall, this sample consisted of more than 500 participants. Using the *metafor* package in R, the meta-analysis revealed highly significant use of the more-likely hypothesis [$z = 10.84, p < .001$, 95% CI[10.35,14.92]], but not the less-likely hypothesis [$z = -1.21, p = .23$, 95% CI[–3.35,0.79]]. A Bayesian meta-analysis (Rouder et al., 2009) of the same data, using a standard Cauchy prior on the standardized effect size, found overwhelming evidence against the null hypothesis for use of the more-likely hypothesis [$BF_{10} > 1000$] and substantial evidence favoring the null hypothesis for use of the less-likely hypothesis [$BF_{01} = 12.7$].

Moreover, excluding studies for which we did not collect judgments of hypothesis probabilities (Studies 1 and 2), participants in these studies significantly underweighted the less-likely hypothesis relative to the normative standards computed for individual studies [$z = 7.09$, $p < .001$, 95% CI[5.01,8.84]]; this result was also supported by a Bayesian meta-analysis along the lines above [$BF_{10} > 1000$]. This confirms the general findings of previous studies: In the absence of explicit predictive probabilities, participants ignored low-probability hypotheses and significantly underweighted them relative to high-probability hypotheses.

Another meta-analytic approach is to add back in the studies using precise probabilities (Studies 8B and 8C) and to test for differences between the *low/high* and *low/low* conditions, using dummy-coded variables as moderators. Studies 5, 6, 8A, and 8C were coded as '1' on an *explicit hypothesis probabilities* variable (all others as '0') because these studies gave explicit probabilities for $P(A)$ and $P(B)$. Studies 8B and 8C were coded as '1' on an *explicit predictive probabilities* variable (all others as '0') because these studies gave explicit probabilities for $P(Z|A)$ and $P(Z|B)$. At the baseline (when both dummy variables are set to 0), there is once again no difference between the *low/high* and *low/low* conditions, even pooling data from 9 studies with approximately 650 participants [$z = -1.28$, $p = .20$, 95% CI[–4.10,0.86]]. There was no moderating effect of explicit hypothesis probabilities [$z = 0.49$, $p = .62$, 95% CI[–2.96,4.96]], but a large and significant moderating effect of explicit predictive probabilities [$z = 4.82$, $p < .001$, 95% CI[7.40,17.55]]. This corroborates the boundary condition seen in individual studies: Participants provided with explicit predictive probabilities, but not explicit hypothesis probabilities, were able to attend to less-likely hypotheses.

General Discussion

Do beliefs come in degrees? Here, we showed that they do not when we use those beliefs to make further predictions—in such cases, probabilities are converted from an 'analog' to a 'digital' format and are treated as either true or false. Compared to Bayesian norms, participants across our studies consistently underweighted low-probability relative to high-probability hypotheses, often ignoring low-probability events completely. This neglect challenges theories of cognition that posit a central role to graded probabilistic reasoning. Here, we discuss where this tendency appears to come from and in what ways it might be limited.

Predictions from Uncertain Beliefs

Many studies have found that when an object's category is uncertain, people rely on the single most-probable category when predicting its other features. Although some studies find individual differences and variability among tasks, single-category use has held up among many different kinds of categorization schemes (e.g., Johnson, Kim, & Keil, 2016; Lagnado & Shanks, 2003; Malt, Murphy, & Ross, 1995; Murphy & Ross, 1994, 1999).

Plausibly, these limitations on probabilistic reasoning are specific to category-based induction tasks. The purpose of categories, after all, is to simplify the world and carve it into discrete chunks. But another possibility is that these previous findings are due to a much broader tendency in our reasoning about uncertain hypotheses and their implications. A categorization of an object is a hypothesis about what kind of object it is, but similarly a causal explanation is a hypothesis about what led something to happen and a mental-state inference is a hypothesis about what someone is thinking. The current studies find that people only think in terms of one hypothesis at a time in a causal reasoning task, suggesting that such *digital thinking* is a broad feature of hypothetical thinking. This is consistent with the *singularity hypothesis* (Evans, 2007), according to which people entertain only a single possibility at a time—an idea with broad explanatory power in higher-level cognition.

Why does digitization occur when making predictions from uncertain beliefs? Such predictions typically require three processes. First, potential hypotheses must be evaluated, given the available evidence, resulting in estimates of the hypothesis probabilities P(A) and P(B) (*abduction*). Second, the prediction needs to be made conditionally on each hypothesis holding, that is, in each relevant possible world, resulting in estimates of the predictive probabilities P($Z|A$) and P($Z|B$) (*simulation*). Finally, these conditional predictions need to be weighted by the plausibility of each hypothesis (*integration*), leading to an estimate of P(Z). Although people are able to perform each of these processes, they each are accompanied by limitations and bias. How do each of these stages contribute to digitization?

Our experiments are most consistent with a model in which abduction leads to more extreme explicit hypothesis probabilities, simulation capacity limits result in digitization, and integration leads people to under-use hypothesis probabilities relative to predictive probabilities. This conclusion is necessarily provisional at this early stage, but here we lay out the best case made by the evidence.

The abduction phase—deciding among potential hypotheses as the best explanation for the data—relies on a variety of heuristics. Although many of these heuristics may adaptively help to circumvent computational limits or even lead to more accurate inferences, these heuristics lead to systematic biases relative to Bayesian norms. Most relevant, people assign a higher probability to a hypothesis that outperforms its competitors, relative to what is implied by objective probabilities (Douven & Schupbach, 2015; see also Lipton, 2004). This sort of process could plausibly give rise to digitization. Moreover, explanation often leads to overgeneralization in the face of exceptions (Williams et al., 2013), consistent with the idea that abduction tends to underweight or ignore lower-probability hypotheses. But in our studies, abduction does not seem to be a necessary ingredient for digitization, since digitization even occurs when hypothesis probabilities P(*A*) and P(*B*) are provided explicitly, avoiding the need for abductive processing (Studies 5 and 8A). The most likely resolution of this puzzle is that abduction leads us to *explicitly* assign more extreme probabilities to hypotheses, relative to Bayesian norms, but not to ignore those less-likely hypotheses altogether.

The simulation phase—imagining the plausibility of the prediction in the possible worlds defined by each hypothesis—is known to have sharp capacity limits (Hegarty, 2004). Indeed, even within a simulation of a single causal system, people imagine each step in that system piecemeal. Thus, it seems unlikely that people can simultaneously simulate multiple possible worlds and store their outputs simultaneously. Consistent with the idea that this is the key processing bottleneck that produces digitization, people do consider multiple possibilities when the predictive probabilities P(*Z*|*A*) and P(*Z*|*B*) are given explicitly, avoiding the need to simulate these outcomes (Studies 8B and 8C).

Yet, this does not seem to be the whole story. The integration phase—putting together multiple pieces of evidence and weighing each by their diagnosticity—is also subject to biases. In particular, people tend to over-rely on information about evidence *strength* (e.g., the proportion of cases consistent with a hypothesis) relative to information about evidence *weight* (e.g., sample size) (Griffin & Tversky, 1992; Kvam & Pleskac, 2016). Although this bias should not be extreme enough to lead people to *ignore* lower-probability hypotheses, it could result in overconfidence—overly extreme probabilities—if people treat predictive probabilities as strength information (how likely the prediction is within each possible world) and hypothesis probabilities as weight information (how much to consider each possible world). This pattern seems to be consistent with the data. Even when both the hypothesis and predictive probabilities are given explicitly, requiring only integration to occur, participants over-rely on the high-probability relative to low-probability hypothesis (Study 8C).

Thus, all three processing steps appear to contribute to overly extreme probability judgments, albeit in different ways. Abduction may result in explicit probabilities that are too extreme, relative to Bayesian norms. Integration seems to result in under-responsiveness to hypothesis probabilities. And simulation seems to lead people to ignore lower-probability hypotheses entirely.

If digitization can lead to systematic errors, relative to Bayesian norms, why might the mind use this principle? Digitization is often necessary to avoid a combinatorial explosion (Bobrow, 2012; Friedman & Lockwood, 2016). Suppose you are unsure whether the Fed will raise interest rates. Depending on this decision, Congress may attempt fiscal stimulus; depending on Congress's decision, the CEO of Citigroup may decrease capital reserves; and depending on the CEO's decision, SEC regulators may tighten enforcement of certain rules. Integrating across such chains of possibilities becomes daunting even for a computer as the number of branches increases. As recently as the 1990s, chess-playing computers used brute force methods to search through trees of possible moves, and even the famous Deep Blue, despite its massive processing power, did not consistently defeat the best human players, such as Garry Kasparov (Deep Blue lost 2.5 out of 6 games in their final match). The computationally efficient way to approach such a problem is precisely the opposite of brute force—to construct plausible scenarios and ignore the rest. Human chess players had, and probably still have, far better heuristics for pruning this huge space of possibilities. Our participants' error was using this strategy even when the normative calculation is straightforward. This strategy may be adaptive in other contexts. Indeed, when the most-likely hypothesis has a probability close to 100%, it may even be a reasonable approximation to the Bayesian solution.

What, then, should we make of probabilistic theories of cognition (Gershman et al., 2015; Tenenbaum et al., 2011)? People clearly can represent analog probabilities at some level ("a 70% chance of rain") but our results show that they cannot use these probabilities to make downstream predictions, instead digitizing them. Because probabilistic models typically characterize the output of reasoning processes rather than the underlying mechanisms, they can be of great value in characterizing the problems that our minds solve. But to the extent that such theories make mechanistic claims involving the processing of analog probabilities within complex computations—even at an implicit level—simpler, heuristic mechanisms may better account for human successes, such as they are, with uncertainty. We look forward to the possibility that computational approaches to the kinds of tasks we model in this paper can help to shed further insight on the underlying cognitive processing.

Limitations and Scope

Digitization is a strong claim, given the wide applications of probabilities in cognitive modeling. But just as there are limits on probabilistic thinking, so are there limits on these limits. Here, we discuss methodological limitations and potential boundary conditions, focusing on how well and when digitization would map onto real-life decision-making, such as medical, legal, or financial decisions.

The chief methodological limitation of this article is that we study a fairly narrow task with a limited number of stimulus items. Yet, we do find that the effects are robust to a variety of methodological alterations, which map onto realistic situations.

First, real-world cases differ in how much the situation cues categorical versus probabilistic thinking. For example, jurors are asked to render a verdict (a categorical judgment) but must consider their verdict relative to a burden of proof (e.g., being confident in their verdict "beyond a reasonable doubt"). As another example, a manager might have a hypothesis about what their firm's competitor is doing, and may have to sell this hypothesis categorically to other stakeholders in order to gain the authority to proceed on this hypothesis. We find digitization across several kinds of task demands, including situations in which prediction is preceded by a categorical decision about the likeliest hypothesis (Studies 1–3), a graded probability judgment about each hypothesis (Studies 4–6, 8A and 8D), or no judgment about the hypotheses at all (Study 7). Previous studies of category-based induction have sometimes found that such manipulations affect the amount of multiple-category use (e.g., Murphy et al., 2012; Murphy & Ross, 2010), but we do not find much evidence for integration across hypotheses in any of these experiments. Although people are more likely to strongly rely on the high-probability hypothesis when they make a forced-choice decision before prediction (Studies 1–3), the other task variations do not produce evidence for use of the low-probability hypothesis.

Second, real-world cases often involve not just two possible hypotheses, but many (e.g., many possible diagnoses of a medical cases, or many explanations for why a stock price changed). These conditions can still produce digitization in two different ways. If one hypothesis is likelier than any other (e.g., 50% probability) whereas its competitors individually have lower probabilities (e.g., 25% probability), people focus on the likely hypothesis to the exclusion of the others (Study 6).

Alternatively, if the less-likely hypotheses can be bundled into one coherent category, then people may collectively focus on that possibility to the exclusion of the hypothesis with individually higher prior probability (Study 7). In neither case, however, did people account for all possibilities as a Bayesian would. Consistent with this, in studies of category-based induction, researchers have even found that participants frequently rely on a single category even when there is no particular reason to favor one category over another (e.g., Murphy et al., 2012).

Third, real-world cases vary in how one arrives at judgments about the hypothesis [P(A) and P(B)] and predictive [P($Z|A$) and P($Z|B$)] probabilities. For example, a medical doctor may rely on parsimony to adjudicate among possible diagnoses, or a financial analyst may assign probabilities to hypotheses based on a quantitative analysis. Likewise, one might rely on intuitive, narrative thought processes to decide what a hypothesis implies about a possible future prediction (e.g., Johnson & Tuckett, 2018), or on a mathematical model that gives precise quantitative forecasts. In our studies, we modeled these types of situations by varying whether the hypothesis and predictive probabilities were given precisely or vaguely (Studies 8A–D). We found that digitization is not dependent on how the hypothesis probabilities were arrived at (Studies 5 and 8A)—for example, digitization was not any stronger when hypotheses were evaluated using simplicity than when posteriors were given directly. But it *was* dependent on how the predictive probabilities were given (Studies 8B–C). Thus, we would expect digitization to be a less important factor in contexts where explicit predictive probabilities are available, rather than decision-makers or reasoners needing to estimate these themselves.

Thus, within the domain of causal reasoning about unfamiliar stimuli, digitization appears to be robust to task alterations. However, we did not vary the stimulus setting itself (e.g., medical vs. legal vs. financial decision-making), instead focusing on unfamiliar, artificial stimuli. We did this to model general as opposed to domain-specific reasoning processes here. But people may well have domain-specific strategies, or even mental modules, that allow them to circumvent these limitations in particular problem domains. Here, we describe three possible ways that domain-specificity could produce boundary conditions, along with what is known so far about each one.

First, might people integrate across multiple hypotheses in more cognitively encapsulated tasks, such as perception, as opposed to cognitively central tasks such as reasoning? Given that we motivated this paper in part by appealing to the idea of multistable percepts such as Necker cubes that appear in one interpretation at a time (Attneave, 1971), we think it is clear that the visual system often works with a single interpretation at a time. But at least in the domain of category-based predictions, people are often more able to account for multiple categories when the task involves more perceptual processes. For example, people integrate across categories when predicting object trajectories from uncertain categorizations, as indexed by their eye movements (Chen, Ross, & Murphy, 2016) or cursor movements when trying to "catch" virtual objects (Chen, Murphy, & Ross, 2014). Indeed, even in explicit reasoning tasks, people are likelier to use multiple categories when responding in speeded two-alternative forced-choice tasks rather than probability rating tasks (Verde, Murphy, & Ross, 2005), suggesting that encouraging deliberative reasoning may make people *less* Bayesian at this task.

Second, might people integrate across multiple hypotheses in domains with higher stakes? In previous studies of category-based induction, people are likelier to integrate across multiple categories when one of the categories is dangerous (Zhu & Murphy, 2013) or when the cost of ignoring alternatives is highlighted (Hayes & Newell, 2009). Moreover, people sometimes integrate across categories in high-stakes domains, such as legal reasoning (Martin & Hayes, 2012), decision-making (Chen, Ross, & Murphy, 2014), and consumer choice (Gregan-Paxton, Hoeffler, & Zhao, 2005; but see Moreau, Markman, & Lehmann, 2001). Since these studies typically do not compare controlled sets of stimuli from different domains, it is hard to know how much of the multiple-category use in these studies is due to the substantive domain versus methodological variability across studies. Still, this body of evidence from categorization suggests that there may be domain-specific boundary conditions on digitization. Only further research can tell us exactly how probabilistic reasoning differs across domains, in tasks beyond categorization.

Our ongoing research includes two examples of such domain-specific tests of digitization. As a first example, one series of studies tested amateur investors' predictions of the future value of financial securities. Such expectations are critical to the functioning of economic models, which assume that economic agents forecast the future, typically with "rational expectations" (Lucas, 1972). Models in behavioral finance often turn on the assumptions they make about investors' belief updating (e.g., Barberis, Shleifer, & Vishny, 1998), so getting these assumptions right is important. Although it would be plausible if investors integrated across multiple hypotheses given the results described above in other decision-making domains, the evidence so far suggests that investors (at least amateurs in lab experiments) instead digitize (Johnson & Hill, 2017).

As a second example, a different set of studies look at predictions about others' behavior, based on uncertain evidence of moral character. When there are two possible explanations for a behavior, people will take account of a possibility implying bad moral character, even if that explanation is unlikely to be true (Johnson, Murphy, Rodrigues, & Keil, 2019). This use of multiple hypotheses so far does not appear to hold true for inferences based on other, non-moralized traits, such as forgetfulness. This seems consistent with the finding that people consider alternative categories when the cost of error is high (Hayes & Newell, 2009; Zhu & Murphy, 2013), but also could be due to modular cognitive processing in the moral domain (Cosmides, 1989).

Third, might domain expertise promote integration across multiple hypotheses? Although there is much evidence that expertise influences cognitive strategies (e.g., Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981; Medin, Lynch, & Coley, 1997), relatively little is known about expertise effects in prediction tasks such as those studied here, and indeed there is surprisingly little work on expertise effects in causal reasoning more broadly. One exception is a study by Hayes and Chen (2008), finding that mental health clinicians, but not laypeople, were able to integrate across multiple possible diagnostic categories when predicting the behavior of individuals with mental disorders. Moreover, this effect of experience was domain-specific and experts did not consider multiple categories for non-clinical stimuli. Thus, domain experience may well play an important role in leading people to consider multiple hypotheses, but much more research on this topic is needed.

Conclusion

Probabilistic thinking plays an increasingly important role in many areas of decision-making, including managerial strategy, financial decision-making, and public policy. It is critical to understand when and why people are able to think probabilistically, not just about what hypothesis is likeliest to be right, but about what the evidence implies about the future. Given that digitization appears to be a broad strategy in probabilistic thinking, we now need to further understand the cognitive mechanisms and boundary conditions. For example, how does digitization unfold in more complex problems involving longer chains of inference, larger numbers of possibilities at each level, or more complex relationships among those possibilities? Would people continue to focus on a single hypothesis when there is no reason to choose one over another? Would digitization occur in "predictions" about the past as well as the future?

Probability theory was invented to quantify uncertainty about the future—to escape the prison of thinking in terms of one possibility at a time. The current results underscore the historical significance of this development: Intuitive probabilistic thinking is often not probabilistic at all.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.

Attneave, F. (1971). Multistability in perception. *Scientific American*, *225*, 62–71.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, *49*, 307–343.

Bobrow, D. G. (Ed.). (2012). *Qualitative reasoning about physical systems* (Vol. 1). Amsterdam, Netherlands: Elsevier.

Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. In J. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Psychology of learning and motivation, Vol. 32* (pp. 275-318). San Diego, CA: Academic Press.

Cavanagh, P., Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*, 349–354.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81.

Chen, S. Y., Murphy, G. L., & Ross, B. H. (2014). Implicit and explicit processes in category-based induction: Is induction best when we don't think? *Journal of Experimental Psychology: General*, *143*, 227–246.

Chen, S. Y., Ross, B. H., & Murphy, G. L. (2014). Decision making under uncertain categorization. *Frontiers in Psychology*, *5*, 991.

Chen, S. Y., Ross, B. H., & Murphy, G. L. (2016). Eyetracking reveals multiple-category use in induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1050–1067.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187–276.

Douven, I. (1999). Inference to the best explanation made coherent. *Philosophy of Science*, *66*, S424–S435.

Douven, I., & Schupbach, J. N. (2015). The role of explanatory considerations in updating. *Cognition*, *142*, 299–311.

Escalas, J. E. (2004). Imagine yourself in the product: Mental simulation, narrative transportation, and persuasion. *Journal of Advertising*, *33*, 37–48.

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

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2010). Neglect of alternative causes in predictive but not diagnostic reasoning. *Psychological Science*, *21*, 329–336.

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2010). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General*, *140*, 168–185.

Fisher, M., & Keil, F. C. (2018). The binary bias: A systematic distortion in the integration of information. *Psychological Science*, *29*, 1846–1858.

Forbus, K. D. (1984). Qualitative process theory. *Artificial Intelligence*, *24*, 85–168.

Friedman, S., & Lockwood, K. (2016). Qualitative reasoning: everyday, pervasive, and moving forward—a report on QR-15. *AI Magazine*, *37*, 95–97.

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds and machines. *Science*, *349*, 273–278.

Gigerenzer, G., & Goldstein, D. (1996). Reasoning the fast and frugal way: Models of bounded rationalty. *Psychological Review*, *103*, 650–669.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684–704.

Gregan-Paxton, J., Hoeffler, S., & Zhao, M. (2005). When categorization is ambiguous: Factors that facilitate the use of a multiple category inference strategy. *Journal of Consumer Psychology*, *15*, 127–140.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, *24*, 411–435.

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

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, *8*, 280–285.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge, UK: Cambridge University Press.

Jeffreys, H. (1939). *Theory of probability*. Oxford, UK: Clarendon Press.

Johnson, S. G. B., & Ahn, W. (2015). Causal networks or causal islands? The representation of mechanisms and the transitivity of causal judgment. *Cognitive Science*, *39*, 1468–1503.

Johnson, S. G. B., & Hill, F. (2017). Belief digitization in economic prediction. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 2314–2319). Austin, TX: Cognitive Science Society.

Johnson, S. G. B., & Keil, F.C. (2014). Causal inference and the hierarchical structure of experience. *Journal of Experimental Psychology: General*, *143*, 2223–2241.

Johnson, S. G. B., Kim, H. S., & Keil, F. C. (2016). Explanatory biases in social categorization. In A. Papafragou, D. Grodner, D. Mirman, & J.C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 776– 781). Austin, TX: Cognitive Science Society.

Johnson, S. G. B., Merchant, T., & Keil, F.C. (2015b). Predictions from uncertain beliefs. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1003–1008). Austin, TX: Cognitive Science Society.

Johnson, S. G. B., Murphy, G. L., Rodrigues, M., & Keil, F. C. (2019). Predictions from uncertain moral character. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Johnson, S. G. B., Rajeev-Kumar, G., & Keil, F. C. (2016). Sense-making under ignorance. *Cognitive Psychology*, *89*, 39–70.

Johnson, S.G.B., & Tuckett, D. (2018). Narrative decision-making in investment choices: How investors use news about company performance. Available at Social Science Research Network (SSRN): https://ssrn.com/abstract=3037463.

Johnson, S. G. B., Valenti, J. J., & Keil, F.C. (2018). Simplicity and complexity preferences in causal explanation: An opponent heuristic account. *Cognitive Psychology*, *113*.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). Cambridge, UK: Cambridge University Press.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304.

Khemlani, S. S., Sussman, A. B., & Oppenheimer, D. M. (2011). *Harry Potter* and the sorcerer's scope: Latent scope biases in explanatory reasoning. *Memory & Cognition*, *39*, 527–535.

Knight, F. (1921). *Risk, uncertainty, and profit*. New York, NY: Houghton-Mifflin.

Knill, D. C., & Richards, W. (Eds.) (1996). *Perception as Bayesian inference*. Cambridge, UK: Cambridge University Press.

Kvam, P. D., & Pleskac, T. J. (2016). Strength and weight: The determinants of choice and confidence. *Cognition*, *152*, 170–180.

Lagnado, D. A., & Shanks, D. R. (2003). The influence of hierarchy on probability judgment. *Cognition*, *89*, 157–178.

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 154–172). New York, NY: Oxford University Press.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, e253.

Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*, *13*, 455–476.

Lipton, P. (2004). *Inference to the best explanation* (2nd Ed.). New York, NY: Routledge.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology*, *55*, 232–257.

Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences*, *20*, 748–759.

Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, *4*, 103–124.

Manis, M., Gleason, T. C., & Dawes, R. M. (1966). The evaluation of complex social stimuli. *Journal of Personality and Social Psychology*, *3*, 404–419.

Markman, K. D., Gavanski, I., Sherman, S. J., & McMullen, M. N. (1993). The mental simulation of better and worse possible worlds. *Journal of Experimental Social Psychology*, *29*, 87–109.

Martin, A., & Hayes, B. (2012). Inductive reasoning in the courtroom: Judging guilt based on uncertain evidence. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1966–1971). Austin, TX: Cognitive Science Society.

McShane, B. B., & Böckenholt, U. (2017). Single-paper meta-analysis: Benefits for study summary, theory testing, and replicability. *Journal of Consumer Research*, *43*, 1048–1063.

Medin, D. L., Lynch, E. B., & Coley, J. D. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology*, *32*, 49–96.

Moreau, C. P., Markman, A., & Lehmann, D. R. (2001). What is it? Categorization flexibility and consumers' responses to really new products. *Journal of Consumer Research*, *27*, 489–498.

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

Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categorizations. *Cognitive Psychology*, *27*, 148–193.

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

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.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition*, *123*, 199–217.

Rips, L. J. (2010). Two causal theories of counterfactual conditionals. *Cognitive Science.*

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.

Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review*, *104*, 406–415.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t*-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237.

Sloman, S. (2005). *Causal models: How people think about the world and its alternatives.* New York, NY: Oxford University Press.

Stalnaker, R. C. (1976). Possible worlds. *Noûs*, *10*, 65–75.

Steiger, J. H., & Gettys, C. F. (1972). Best-guess errors in multistage inference. *Journal of Experimental Psychology*, *92*, 1–7.

Taylor, S. E., & Pham, L. B. (1996). Mental simulation, motivation, and action. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 219–235). New York, NY: Guilford Press.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–1285.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.

Verde, M. F., Murphy, G. L., & Ross, B. H. (2005). Influence of multiple categories on the prediction of unknown properties. *Memory & Cognition*, *33*, 479-487.

Waldmann, M. R., & Hagmayer, Y. (2006). Categories and causality: The neglected direction. *Cognitive Psychology*, *53*, 27–58.

Williams, J. J., Lombrozo, T., & Rehder, B. (2013). The hazards of explanation: Overgeneralization in the face of exceptions. *Journal of Experimental Psychology: General, 142*, 1006–1014.

Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy–informativeness trade-off. *Journal of Experimental Psychology: General*, *124*, 424–432.

Yousif, S., & Keil, F. C. (2019). *The 'reservoir fallacy': Why we have contradictory intuitions about dichotomous events.* Manuscript in preparation.

Zhu, J., & Murphy, G. L. (2013). The effect of emotionally charged information on category-based induction. *PLoS ONE*, *8*, e54286.