Cognitive Psychology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Cognitive Psychology

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

Sense-making under ignorance

Samuel G.B. Johnson *, Greeshma Rajeev-Kumar, Frank C. Keil

Department of Psychology, Yale University, United States

ARTICLE INFO

Article history:
Accepted 20 June 2016
Available online xxxx

Keywords:
Causal reasoning
Categorization
Explanation
Ignorance
Probabilistic reasoning

ABSTRACT

Much of cognition allows us to make sense of things by explaining observable evidence in terms of unobservable explanations, such as category memberships and hidden causes. Yet we must often make such explanatory inferences with incomplete evidence, where we are ignorant about some relevant facts or diagnostic features. In seven experiments, we studied how people make explanatory inferences under these uncertain conditions, testing the possibility that people attempt to *infer* the presence or absence of diagnostic evidence on the basis of other cues such as evidence base rates (even when these cues are normatively irrelevant) and then proceed to make explanatory inferences on the basis of the inferred evidence. Participants followed this strategy in both diagnostic causal reasoning (Experiments 1–4, 7) and in categorization (Experiments 5–6), leading to illusory inferences. Two processing predictions of this account were also confirmed, concerning participants' evidence-seeking behavior (Experiment 4) and their beliefs about the likely presence or absence of the evidence (Experiment 5). These findings reveal deep commonalities between superficially distinct forms of diagnostic reasoning—causal reasoning and classification—and point toward common inferential machinery across explanatory tasks.

1. Introduction

Across perception and cognition, we fill in details missing from our actual experience. In perception, we see illusory contours and infer continuities of forms; indeed, we fill in unattended elements

* Corresponding author at: 2 Hillhouse Ave., New Haven, CT 06520, United States.
E-mail address: samuel.johnson@yale.edu (S.G.B. Johnson).

http://dx.doi.org/10.1016/j.cogpsych.2016.06.004
0010-0285/

of our visual field so successfully that we fail to appreciate the sharp limits of our conscious awareness. Likewise, in cognition, we fill in narratives, scripts, and schemas almost continuously through our daily lives. Although these acts of filling in can create striking illusions and false memories (Loftus & Palmer, 1974; Simons & Levin, 1997), this filling in tendency is an essential tool for cognition: Sound strategies for inferring unknown information allow us to get by with limited information, while still effectively navigating the world.

Here, we argue that this sort of filling in strategy plays a key role in explanatory reasoning, guiding our inferences about causal explanations and likely categorizations of objects, with people reasoning about such explanations based on both the observed and *inferred* evidence. We show at the same time, however, ways in which this strategy can lead to error when people base these inferences on irrelevant information.

1.1. Sense-making under ignorance

We must often make sense of things in the face of incomplete evidence. For example, doctors diagnose diseases when some test results are unavailable or inconclusive, giving the diagnosis they believe most likely or prudent given the evidence at hand. Juries infer the most likely culprit on the basis of often-sketchy evidence, conflicting testimony, and lawyerly doubletalk. People debate about ultimate explanations (e.g., the existence of God or of multiple universes) in the face of these explanations' intrinsically unverifiable predictions (e.g., an afterlife or the splitting of universes). More mundanely but no less remarkably, we all infer other people's mental states on the basis of just a few clues, infer the categories of objects even when some features are indeterminate, and infer causes when some of their potential effects are unknown. Explanation with incomplete evidence is the norm in everyday cognition.

Consider a simple concrete example. Suppose two trial attorneys are presenting two competing theories of a case to the jury (see Fig. 1). If Professor Plum committed the crime (call this hypothesis H_N, because it makes a single, ***n***arrow prediction), then there would be a dent in the candlestick (call this evidence X). Alternatively, if Colonel Mustard committed the crime (hypothesis H_W because it makes two, ***w***ider predictions), then there would be a dent in the candlestick (X), as well as mud on the drawing room carpet (Z). The observations posited by each hypothesis are depicted in Fig. 1.

Clearly, if Plum and Mustard are the only potential culprits, then the key question is whether there was mud in the drawing room (Z), because only this evidence would distinguish between the two hypotheses. That is, learning about the dent in the candlestick (X) is not diagnostic, because this observation would be equally consistent with either hypothesis—learning that this effect was present would tend to confirm both hypotheses (equally) and learning that it was absent would tend to disconfirm both hypotheses (equally). But if we find out that the mud was present, this would be powerful evidence in favor of Mustard, and if we find out that the mud was absent, this would be powerful evidence in favor of Plum. More generally, we rely on diagnostic evidence for telling apart competing explanations, whether the explanations are unobservable mental states, object categories, or causal events.

Sometimes, however, this diagnostic evidence is unavailable. If the jury faces a situation in which the evidence unambiguously indicates a dented candlestick (X), but is silent on the issue of the mud (Z)—say, because the floor had been cleaned before the detectives thought to check for it—then the jury faces incomplete evidence. Here, normative probability theory tells us that we should think the explanations equally likely: If we had no reason to think Plum or Mustard was the more likely culprit before gathering evidence, then we still have no reason after learning about X, but remaining ignorant about Z.

However, human judgments do not always obey probability theory (e.g., Kahneman, Slovic, & Tversky, 1982). Instead, we often use simplifying heuristics that perform reasonably well under ecologically realistic conditions but are prone to error. In cases of incomplete evidence, people tend to choose explanations that do not imply unknown evidence (Khemlani, Sussman, & Oppenheimer, 2011; Sussman, Khemlani, & Oppenheimer, 2014)—that is, people think that Professor Plum is the most likely culprit in the above case, against the dictates of probability theory. This error—known as the *latent scope bias*—is surprising both because it seems to deviate so strikingly from normative

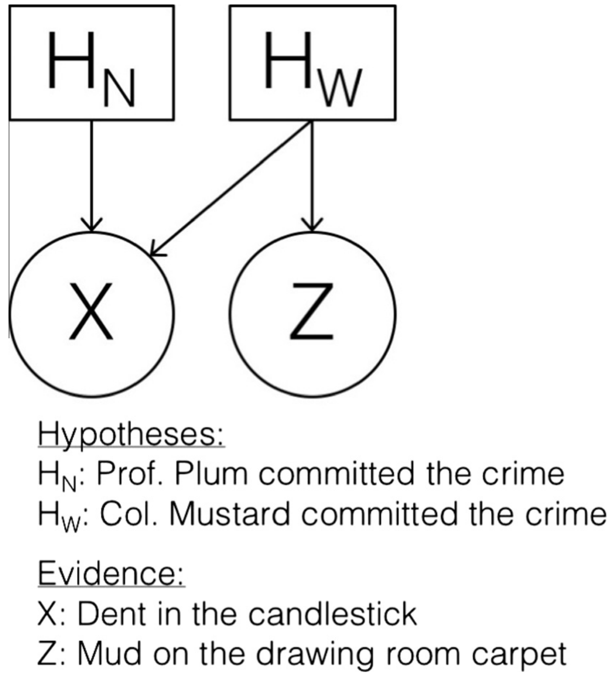

Fig. 1. Example of an explanatory problem where diagnostic evidence might be unavailable (see text for explanation).

judgment and because it is precisely the opposite of the strategy recommended by philosophers of science, to select hypotheses that are subject to falsification (e.g., Popper, 1959/1934).

People have a latent scope bias both when reasoning about causes and about categories, but the psychological mechanisms leading to this bias are unclear. Here, we propose that this bias results, at least in part, because people reason not only using observed evidence, but also *inferred evidence*, which can sometimes be biased in favor of explanations which make fewer predictions. We also contrast this account with several other (not mutually exclusive) mechanisms—*biased priors*, *non-independent evidence*, *representativeness*, and *pragmatic inference*. Before considering these mechanisms, however, some preliminary concepts are needed to place them in a common theoretical framework.

1.2. Explanatory scope

Explanations vary in their *scope*—that is, the range of observations that would be expected if the explanation were true. In our running example, the scope of the Professor Plum theory (H_N) is a dented candlestick (X), whereas the scope of the Colonel Mustard theory (H_W) is a dented candlestick and mud on the floor (X and Z). However, it is not scope alone, but the consistency of an explanation's scope with the available evidence that determines the relative probability of each explanation.

An explanation's scope can be divided into its *positive* scope (confirmed predictions) and *negative* scope (disconfirmed predictions). If we know that the mud was present, then Z is in the positive scope of the Colonel Mustard theory, and provides evidence in favor of that theory because it predicted that effect (whereas its competing theory did not). On the other hand, if we know that the mud was absent, then Z is in the negative scope of the Colonel Mustard theory and provides evidence *against* that theory. Consistent with these intuitions, people favor explanations with relatively *wide* positive scope (making many confirmed predictions) and relatively *narrow* negative scope (making few disconfirmed predictions; Johnson, Johnston, Toig, & Keil, 2014; Johnson, Merchant, & Keil, 2015a; Read & Marcus-Newhall, 1993; Samarapungavan, 1992).

These preferences are broadly consistent with probability theory. Bayes' theorem allows us to compare the relative probabilities of two hypotheses given some evidence, and tells us that our beliefs favoring one hypothesis over the other (our *posterior odds*) should be equal to our previous beliefs about the relative probabilities of the hypotheses (our *prior odds*) times the relative consistency of the evidence with each hypothesis (the *likelihood ratio*), as given by the formula:

$$\frac{P(H_N|Evidence)}{P(H_W|Evidence)} = \frac{P(H_N)}{P(H_W)} \cdot \frac{P(Evidence|H_N)}{P(Evidence|H_W)} \tag{1}$$

On the assumption that we have no reason *a priori* to favor one hypothesis over the other (so that the prior odds equal 1), we are tasked simply with determining which explanation is more consistent with the observed data. Suppose, for example, that the probability of X would be 80% under each hypothesis, and the probability of Z would be 1% under H_N (say, because the family dog spreads mud throughout the manor 1% of the time) but 99% under H_W. If the effects occur independently, conditional on their causes (Pearl, 1988), then the likelihood term can be factorized into a likelihood for X and a likelihood for Z, and the posterior calculated as follows:

$$\frac{P(H_N|X,Z)}{P(H_W|X,Z)} = \frac{P(H_N)}{P(H_W)} \cdot \frac{P(X|H_N)}{P(X|H_W)} \cdot \frac{P(Z|H_N)}{P(Z|H_W)} = \frac{.5}{.5} \cdot \frac{.8}{.8} \cdot \frac{.01}{.99} = \frac{1}{99} \tag{2}$$

Thus, the evidence favors H_W by a ratio of 99 times, and H_W (given our prior beliefs) is 99 times more likely than H_N. This makes intuitive sense, because H_W has wider positive scope and accounts for more of the data. Conversely, suppose that we observed X, and found that Z was absent ($-Z$)—that is, that Z was in the negative scope of H_W. Since $P(-Z|H_i)$ just equals $[1 - P(Z|H_i)]$, we can straightforwardly compute the posterior odds under this new configuration of evidence:

$$\frac{P(H_N|X,-Z)}{P(H_W|X,-Z)} = \frac{.5}{.5} \cdot \frac{.8}{.8} \cdot \frac{.99}{.01} = \frac{99}{1} \tag{3}$$

That is, given $\{X, -Z\}$, hypothesis H_N (with narrower negative scope) is 99 times more likely than hypothesis H_W. Once again, this makes intuitive sense, as explanation H_N accounts for all of the observed evidence, but does not go out on a limb in making disconfirmed predictions.

People's reasoning about positive and negative scope appears to be at least qualitatively consistent with normative Bayesian reasoning (Read & Marcus-Newhall, 1993), suggesting that positive and negative scope preferences may be useful heuristics deployed to realize complex Bayesian computations, much like the simplicity and complexity heuristics (Johnson, Jin, & Keil, 2014; Lombrozo, 2007). One further reason to suppose that people deploy heuristics in scope-based inferences is that people use positive and negative evidence *asymmetrically*. People count negative evidence against an explanation (using a *negative scope* heuristic) far more dramatically than they count positive evidence in favor of an explanation (using a *positive scope* heuristic). Put differently, disconfirmatory evidence is seen as strongly disconfirmatory, whereas confirmatory evidence is seen as only weakly confirmatory, even when there is no clear normative reason for this pattern (Johnson, Kim, & Keil, 2016; Johnson et al., 2015a). The strength of these heuristics thus appears to be independently calibrated—a fact that would be difficult to explain without further assumptions if people are doing straightforward Bayesian inference.

Our main interest in the current article concerns cases where evidence cannot be classified as either belonging to the positive or negative scope of an explanation, but instead is unknown. Observations that would be predicted by an explanation but which are not known to be present or absent fall into that explanation's *latent scope*. If we do not know about the mud one way or the other, then the Colonel Mustard theory has one effect (Z) in its latent scope, whereas the Professor Plum theory has no effects in its latent scope. We would therefore say that the Colonel Mustard theory has a relatively *wide* latent scope, whereas the Professor Plum theory has a relatively *narrow* latent scope. Two explanations can differ in latent scope either in making *some* versus *no* latent predictions (e.g., one versus zero unknown effects, as in our crime example), or by making *more* versus *fewer* latent predictions (e.g., two versus one unknown effects). People appear to reason about these cases similarly (Khemlani et al., 2011), so we focus here on the simpler case of *some* (wide) versus *no* (narrow) latent scope.

As mentioned earlier, people generally prefer explanations with narrow latent scope, in both causal reasoning (Khemlani et al., 2011) and categorization (Sussman et al., 2014). That is, most people would

Please cite this article in press as: Johnson, S. G. B., et al. Sense-making under ignorance. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.06.004

think Professor Plum is the more likely culprit. Unlike the positive and negative scope heuristics that people use, however, this latent scope bias is *qualitatively* non-normative from a probabilistic standpoint. Suppose again that the probability of X would be 80% under each hypothesis, and the probability of Z would be 1% under H_N (say, because there is a background cause of Z that is present 1% of the time) but 99% under H_W. Given the evidence $\{X\}$, but without knowledge of Z either way, the posterior odds are equivocal:

$$\frac{P(H_N|X)}{P(H_W|X)} = \frac{.5}{.5} \cdot \frac{.8}{.8} = \frac{1}{1} \tag{4}$$

That is, despite people's preference under these conditions for H_N over H_W, there is no reason to favor one explanation over the other, normatively speaking. Why then do people show these consistent preferences?

2. Making sense of latent scope

2.1. Inferred evidence

Our core proposal is that people perform exploratory reasoning using not only the observed evidence, but also *inferred evidence* (Johnson, Rajeev-Kumar, & Keil, 2014). That is, when some evidence is unavailable but potentially diagnostic, people make a guess as to what that evidence would be, if it were known. This is analogous to filling in strategies used in other areas of cognition, such as filling in gaps in perception (Marr, 1982; Simons & Levin, 1997) and in memory (Bartlett, 1932; Loftus & Palmer, 1974). People might similarly use available information to fill in whether the latent evidence would have been observed, if they were able to look. The latent scope bias occurs, we claim, because people generate this inferred evidence in a biased manner.

At the computational level, this idea can be formalized using an alternative formulation of Bayes' theorem, in which the likelihood term for the unverified prediction Z is broken into likelihood components for when Z is confirmed [$P(Z|H_i)$] and for when Z is disconfirmed [$P(-Z|H_i)$]. If I denotes our state of ignorance about Z and we assume that the evidence is conditionally independent given the causes, the posterior can be written as:

$$\frac{P(H_N|X,I)}{P(H_W|X,I)} = \frac{P(H_N)}{P(H_W)} \cdot \frac{P(X|H_N)}{P(X|H_W)} \cdot \frac{P(Z|H_N) \cdot f^{+Z} + P(-Z|H_N) \cdot f^{-Z}}{P(Z|H_W) \cdot f^{+Z} + P(-Z|H_W) \cdot f^{-Z}} \tag{5}$$

Here, f^{+Z} is a parameter reflecting the degree of bias in estimating the base rate of Z, and f^{-Z} reflects the degree of bias in estimating the base rate of $-Z$. That is, $f^{+Z} = P(Z|I)/P(Z)$ and $f^{-Z} = P(-Z|I)/P(-Z)$, with $P(Z) + P(-Z) = 1$. See Appendix A for a derivation of this result.

Intuitively, one could think of the posterior calculation as proceeding in the following steps (though we do not intend these as processing claims). First, one begins with the prior probabilities (the first term on the right side). These are stipulated to be equal, so there is no bias so far. Second, one updates according to the likelihood of the observed evidence (X), given each hypothesis (the second term on the right side). Since the observed evidence is perfectly consistent with both hypotheses, this ratio still equals one. Because there is no way to know whether Z or $-Z$ is true, except to rely on the base rates of the hypotheses (which are equal), a normative reasoner would stop here and conclude that the hypotheses are equally likely.

However, a reasoner who is motivated to *infer* the state of Z would take two additional steps. Third, she would calculate the likelihoods of Z and $-Z$, given each hypothesis. Given that H_N causes Z and H_W does not, $P(Z|H_N) > P(Z|H_W)$ and $P(-Z|H_W) > P(-Z|H_N)$. Finally, one determines how to weight these likelihoods for Z and $-Z$. If one weights these likelihoods equally (i.e., $f^{+Z} = f^{-Z} = 1$), then the Z likelihood ratio (the term on the far right) collapses to 1 and a normative inference is made. But if one's estimate of the base rate of Z is too low (i.e., $f^{+Z} < 1$ and $f^{-Z} > 1$), then the right-hand term will be less than 1, leading to a posterior biased against latent scope explanations. And if one's estimate of the base rate of Z is too high ($f^{+Z} > 1$), the right-hand term will be greater than 1, leading to a posterior biased *toward* latent scope explanations.

Please cite this article in press as: Johnson, S. G. B., et al. Sense-making under ignorance. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.06.004

Why would people underestimate the base rate of Z? In estimating $P(Z)$, one must evaluate this base rate relative to some reference class. That is, if $P(Z) = 20\%$, this means that Z occurs in 20% of the cases considered in the reference class. The appropriate reference class is the set of worlds where either H_N or H_W is true, because we are interested only in the relative probability of these hypotheses. Further, only cases in which Z is caused by H_N or H_W would be relevant, because Z is not diagnostic when explained by alternative causes. More concretely, imagine 50 worlds in which Plum had committed the crime, and 50 worlds in which Mustard had committed it (i.e., the appropriate reference class where H_N and H_W have equal base rates). In half of these worlds, the carpet is muddy due to Mustard's criminal activities. Thus, the correct base rate to use for $P(Z)$ is 50%.

Normative reasoning that appropriately limits the reference class may be quite difficult in such cases, because it involves three different processes, each of which is known to be effortful and error-prone. First, it requires *extensional reasoning*, to entertain the question of which reference class is relevant. Second, it requires *counterfactual thinking*, to consider only those possible worlds where the relevant hypotheses are true. Third, it requires *disjunctive logic*, because one must consider the union of the set of possible worlds where H_N is true and where H_W is true. Each of these operations is known to be effortful: Extensional reasoning is notoriously error-prone, especially when problems are framed in terms of individual cases rather than a group of cases (Gigerenzer & Hoffrage, 1995; Tversky & Kahneman, 1983), counterfactual thinking is subject to a host of biases (Kahneman & Miller, 1986; Rips, 2010; Rips & Edwards, 2013), and disjunctions are difficult to process (e.g., Bourne, 1970; Shafir, 1994). This strategy is thus likely to be effortful, cognitively unnatural, and highly error-prone.

Hence, people may rely on a simpler strategy—considering *all* possible worlds. Instead of asking themselves to simulate equal numbers of worlds where Plum committed the crime and where Mustard committed the crime, and then counting the number of muddy carpets, people may simply rely on their existing knowledge about carpets—and most carpets are not muddy. That is, if people reason according to Eq. (5) above and need to estimate $P(Z)$, they may not appropriately conclude that $P(Z) = P(-Z) = 50\%$, so that $f^{+Z} = 1$, but instead conclude that $P(Z) < 50\%$ (and $P(-Z) > 50\%$), so that $f^{+Z} < 1$ and $f^{-Z} > 1$. This would lead to a systematic bias against the explanation that predicts Z.

This account makes a clear prediction—that varying the base rate of $P(Z)$ in the world should moderate the size of the narrow latent scope bias, and perhaps even reverse it. On this account, the latent scope bias has been so robust in previous research because previous studies have used effects and features which have low base rates—such as magical transformations (Khemlani et al., 2011), medical abnormalities (Khemlani et al., 2011), and various low-probability category features (e.g., tribe members who carry nets; Sussman et al., 2014). Although such studies are ecologically valid in the sense that most effects and features used for reasoning are likely to have low base rates (at least, less than 50%), cases certainly exist where these base rates are higher. For example, a disease might invariably result in high levels of a protein which are *already* high, by default, in most patients; a form of psychopathology might occur only in individuals with IQs greater than 85. We would predict that relatively high base rates should lead to a weaker latent scope effect, and very high base rates could even reverse it.

Inferring the absence of low base rate evidence changes the nature of the computation to be performed, effectively pushing Z into the negative scope of H_N. Rather than computing the likelihood of each hypothesis given $\{X\}$, these likelihoods must now be computed relative to $\{X, -Z\}$—licensing the inference that H_N is the more likely hypothesis, as computed in Section 1.1. Although this inference is non-normative, the error lies not in the heuristics used to realize the probability computations, but rather in the methods used to arrive at the evidence used in those computations. The latent scope bias may not be an aversion to latent scope at all, but instead a symptom of a broader—and often adaptive—reluctance to accept ignorance about latent evidence, instead filling in details as in perception and memory (Bartlett, 1932; Loftus & Palmer, 1974; Marr, 1982; Simons & Levin, 1997).

2.2. Other potential mechanisms for a latent scope bias

Several other mechanisms, however, could plausibly lead to a latent scope bias. Although we argue that inferred evidence based on base rates contributes over-and-above these other possible mechanisms, it is certainly possible that these mechanisms act in concert. Here, we briefly describe four other potential mechanisms, in terms of the computations postulated in Eq. (5).

Please cite this article in press as: Johnson, S. G. B., et al. Sense-making under ignorance. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.06.004

2.2.1. Biased priors

First, people could believe the priors are not truly equal for wide and narrow latent scope explanations. For instance, if the latent predictions made by the wide scope cause are particularly implausible, this may lead people to assume it has a low base rate. More generally, a wider latent scope cause would lead to more effects than a narrow latent scope cause, and perhaps these more potent causes are thought to be less frequent in the world; alternatively, more potent causes might actually be thought to be *more* frequent in the world (Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008). In terms of Eq. (5), this would lead to a bias in the prior odds, which could lead to either a narrow or wide latent scope bias, depending on what assumptions are made.

2.2.2. Non-independence of evidence

Second, as noted above, Eq. (5) is only valid if the evidence (X and Z) is independent, conditional on its causes. This is what allows the likelihood term to be factorized into a term for each piece of evidence (see Appendix A). Violations of this assumption can lead to either a bias for or against narrow latent scope explanations. Intuitively, if X and Z are positively correlated, the observed evidence (X) is then evidence in favor of the latent evidence (Z), so the wide latent scope explanation (which would explain Z) should be preferred; conversely, if X and Z are negatively correlated, this should lead to a bias favoring the *narrow* latent scope explanation, since X is evidence *against* Z. However, such normative inferences are distinct from the non-normative base rate inferences implicated by our own account.

2.2.3. Pragmatic inference

Third, people could be making pragmatic inferences, assuming that statements such as "We don't know whether or not the carpet is muddy" communicate something more than mere ignorance, by making assumptions about *why* the speaker does not know. For example, people might reason that if the carpet were muddy, the speaker *should* know, hence the carpet probably is not muddy. Although pragmatic inferences are also a form of "inferred evidence," the psychological process is rather different (and potentially normative), relying on reasoners' assumptions about conversational implicature rather than about base rates. Hence, this account makes different predictions from our own. For instance, justifying the speaker's ignorance (so that the reasoner believed that ignorance did not communicate anything about the evidence) should eliminate the latent scope effect if it is caused only by pragmatic effects (e.g., McGarrigle & Donaldson, 1974). If the other accounts contribute to the effect, then an effect should still be observed under these conditions.

In terms of Eq. (5), such pragmatic inferences would occur when reasoners do not assume that Z and I are independent, instead using I to make inferences about Z (e.g., I implies $-Z$). Like our inferred evidence account, this can be modeled by using the parameters $f^{+Z} = P(Z|I)/P(Z)$ and $f^{-Z} = P(-Z|I)/P(-Z)$ to reflect the reasoner's greater belief in Z (or $-Z$) given the speaker's ignorance, relative to the evidence base rates implied by the problem.

2.2.4. Representativeness

Finally, one potential account of the latent scope bias (tentatively offered by Sussman et al., 2014) is *representativeness*. According to this explanation, people simulate what kind of evidence they would expect under each hypothesis, and compare the simulated evidence to the actual evidence. That is, if Professor Plum is culpable, we would expect to observe a dented candlestick, $\{X\}$, and if Colonel Mustard is culpable, we would expect to observe a dented candlestick and mud on the carpet (i.e., $\{X, Z\}$). The actual evidence, $\{X\}$, is more similar to the simulated evidence for the narrow scope explanation,[1] so we conclude that the evidence is more representative (Kahneman & Tversky, 1972) of the narrow

[1] This account would lead to a narrow latent scope bias, so long as the set of features entered into the similarity computations includes only those features that are observed (in the case of the observed feature set) and only those features that are caused by each explanation (in the case of the hypothetical feature sets for each explanation). That is, the observed evidence should be $\{X\}$, the evidence hypothesized by H_N would be $\{X\}$, and the evidence hypothesized by H_W would be $\{X, Z\}$. Then, a similarity model such as Tversky's (1977) contrast model would find the observed evidence to be more similar to H_N's hypothesized evidence than to H_W's.

Please cite this article in press as: Johnson, S. G. B., et al. Sense-making under ignorance. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.06.004

scope explanation and that Professor Plum is thus the likelier culprit. Formally, this would be equivalent to using the similarity ratio to approximate the likelihood ratio in Eq. (1) (see Gigerenzer & Hoffrage, 1995; Tenenbaum & Griffiths, 2001a).

2.3. Summary of competing accounts

Table 1 compares the five accounts on offer. The biased priors and non-independence accounts both rely on normative Bayesian reasoning, although the bias manifests in different terms of the Bayesian hypothesis comparison (the priors and likelihood, respectively). These accounts can predict either a narrow or wide latent scope bias, depending on the direction of the biased priors or non-independence (a narrow latent scope bias if the priors favor the narrow latent scope explanation or if the evidence is thought to be negatively correlated; and a wide latent scope bias in the opposite cases). We evaluate these approaches most directly in Experiment 2, where we measure reasoners' assumptions about priors and independence.

Pragmatic accounts make less clear predictions, as their implications for the bias depend on what additional assumptions speakers are thought to be conveying. One way to model these inferences is in terms of f^{+Z} and f^{-Z}, which reflect the assumed probability of the latent evidence relative to its true probability. Our empirical approach to the pragmatic account is to provide plausible reasons for the speaker's ignorance, which should undermine the bias to the extent that pragmatic factors play a role (Experiments 2 and 7). We also measure the effects of pragmatic inference directly in Experiment 3.

Finally, the inferred evidence and representativeness approaches both posit heuristic processes. In the case of representativeness, similarity of the actual to predicted evidence is used to heuristically estimate the likelihoods, whereas in the case of inferred evidence, the base rate of the evidence is used to heuristically estimate what evidence itself will be included in the calculation. Although both approaches are heuristic, they differ in *which* process is said to be heuristic (deciding what evidence to evaluate, or evaluating the evidence), and lead to different predictions. Representativeness always predicts a narrow latent scope bias, because the observed evidence $\{X\}$ will always be more similar to narrow scope prediction$\{X\}$ than to wide scope prediction $\{X, Z\}$. The inferred evidence account, in contrast, predicts strongly narrow latent scope effects when the base rate of Z is low, and a weaker or even reversed effect when the base rate of Z is high. This key prediction is tested in several experiments.

3. Our empirical approach

Here, we report seven experiments designed to test the hypothesis that inferred evidence plays a role in explanatory inference with incomplete evidence, above and beyond the alternative mechanisms described above. Because latent scope effects have been found in both causal reasoning (Khemlani et al., 2011) and in categorization (Sussman et al., 2014), it is possible that these disparate diagnostic reasoning tasks involve similar heuristic mechanisms. To support this claim, we tested for

Table 1
Comparison of five accounts of the latent scope bias.

Account	Term affected	Psychological process	Predicted direction	Tested in experiments
Biased priors	Priors	Normative Bayesian reasoning	Depends on prior odds	1–3, 5–6
Non-independence	Likelihood	Normative Bayesian reasoning	Depends on direction of non-independence	2
Pragmatics	f^{+Z}, f^{-Z}	Conversational implicature	No directional prediction	2–3, 7
Representativeness	Likelihood	Heuristic estimation of likelihood	Always predicts narrow latent scope bias	1–6
Inferred evidence	f^{+Z}, f^{-Z}	Heuristic estimation of evidence	Depends on base rate of Z	1–7

signatures of inferred evidence in both causal reasoning (Experiments 1–4 and 7) and categorization (Experiments 5 and 6).

These experiments tested three main sets of predictions made by the inferred evidence account. First, varying the plausibility that a latent effect Z would be observed should modulate the magnitude of the latent scope bias, and perhaps even its direction—when the latent effect is highly plausible in the token case, one might even expect a *wide* latent scope bias. Several experiments tested this prediction by varying the base rate of Z in the world, using both artificial (Experiments 1–3, and 5) and naturalistic (Experiment 6) stimuli. Based on the idea that more readily imaginable possibilities are assigned higher probabilities (Koehler, 1996), we also manipulated the reason for ignorance about the latent evidence to test for downstream consequences on the size of the bias (Experiment 7).

Second, the inferred evidence account posits the importance of evidence-seeking processes, and predicts in particular that the base rate of the latent effect should be sought after in evaluating explanations under ignorance. This stands in contrast to normative probability theory, according to which the base rates of the causes screen off the relevance of the effect base rates, so that, for example, $P(Z)$ is irrelevant once $P(H_N)$ and $P(H_W)$ are known. Further, because the base rates of the wide latent scope causes are informative about the latent effect base rates (since, for example, H_W causes Z in the example depicted in Fig. 1), the base rates of wide latent scope causes might be seen as more relevant than the base rates of narrow latent scope causes. For example, if $P(H_W)$ is 10%, then $P(Z)$ must be at least 9.9% (since H_W causes Z with 99% probability), but $P(H_N)$ places no constraints on $P(Z)$ (since H_N never causes Z). We therefore predicted that the base rates of wide latent scope explanations would be seen as more relevant than the base rates of narrow latent scope explanations. Both of these predictions were tested in Experiment 4.

Third, the account makes a further processing prediction, that people should infer that the latent effect is relatively unlikely to be observed in circumstances where they have a narrow latent scope bias, compared to circumstances where they show no such bias. Thus, if a situation leads people to infer that a narrow latent scope explanation is more probable than a wide latent scope explanation, we would expect that, if asked, they should report that the latent effect has a less than 50% chance of being observed; in contrast, if they prefer the wide latent scope explanation, they should report that the latent effect has a more than 50% chance of being observed. We tested this prediction in Experiment 5B.

Throughout these experiments, we anticipated that in both causal reasoning and categorization, people would seek, infer, and reason on the basis of additional evidence beyond what was given. In Section 11, we consider the normative status of the inferred evidence strategy and the implications of these results for theories of explanatory reasoning.

4. Experiment 1

As a first test of the inferred evidence account, we varied the base rate of the unknown effect Z, as well as the known effect X. According to probability theory and basic assumptions of graphical causal models (Pearl, 1988, 2000), neither piece of information is relevant if we know the base rates of the causes, $P(H_N)$ and $P(H_W)$. Indeed, for deterministic causal systems such as those described in the current experiments, the posterior odds favoring H_N over H_W are simply equal to the prior odds, $P(H_N)/P(H_W)$ (see Section 1.3).

To see why this is the case, imagine as before that we have observed broken glass, that if Prof. Plum were the culprit there would be broken glass, and that if Col. Mustard were the culprit there would be both broken glass and a muddy carpet (Fig. 1). Given that we do not know whether there is a muddy carpet or not, is it important whether muddy carpets are frequently observed in the manor? First, imagine that only on 1 out of 100 occasions is a muddy carpet observed in the manor *in general*. This does *not* suggest that there is only a 1% chance of observing a muddy carpet if we looked *in this case*, because we are assuming that either Plum or Mustard committed the crime, and they have equal chances of having done so—in *this* case, the probability of observing a muddy carpet is 50% (derived from the prior odds). Second, imagine that on *every* occasion a muddy carpet is observed in the manor. May we then infer that the carpet is surely muddy in the current case, therefore Mustard is the probable culprit? Indeed, if we could look, the carpet *would* be muddy on the current occasion, but due to an alternative cause, such as a dog that tracks mud into the manor every day. In that event, the muddy carpet is simply not diagnostic of who committed the crime, because it is *always* muddy. Once again, the prior odds, not the evidence base rates, determine who is likeliest to have committed the crime.

Even though the use of $P(Z)$ is non-normative, we nonetheless anticipated that people would use this base rate, even when we set $P(H_N) = P(H_W)$ so that the prior odds favor neither hypothesis. This follows from the inferred evidence heuristic, according to which small values of $P(Z)$ would lead people to infer that Z probably did not occur, and therefore that the narrow latent scope explanation H_N was the better explanation. This would be consistent with previous demonstrations of latent scope biases (Khemlani et al., 2011) that used stimuli involving effects with low base rates and few plausible alternative causes (such as magical changes and biochemical abnormalities). However, as $P(Z)$ increases, people would be increasingly likely to infer that Z occurred and therefore to choose the broad latent scope explanation; indeed, when $P(Z) > 50\%$, they might even have a *wide* latent scope preference because they would infer that Z probably did occur. On the other hand, manipulating $P(X)$ would have relatively little effect, because X has already been observed, and therefore the base rate is not needed to infer whether X occurred.

4.1. Method

Participants in all experiments were recruited through Amazon Mechanical Turk and were from the United States. Consistent with studies of the demographics of Mechanical Turk (e.g., Buhrmester, Kwang, & Gosling, 2011), participants tended to be somewhat older ($M = 32.2$ years old in Experiment 1), more female (59% female), and more educated (77% had completed a four-year degree or higher), compared to traditional samples of undergraduate students. Participants were prevented from completing more than one study reported in this paper.

We recruited 100 participants from Amazon Mechanical Turk ($N = 50$ and $N = 50$ for Experiment 1A and 1B, respectively); 32 participants ($N = 17$ and $N = 15$ for Experiment 1A and 1B) were excluded because they had missing data ($N = 1$) or failed more than 30% of a set of check questions (see below; $N = 31$). This threshold was adopted based on past studies, where we found this threshold to adequately protect against inattentive participants without discarding too much data. However, the results are qualitatively the same if all participants are included.

Participants completed four problems in a random order. These included two biological systems (diagnosing a patient's disease and a tree's condition) and two artifact systems (diagnosing a robot's hardware problem and a spaceship's malfunction) for generalizability. For each item, two possible explanations were given: H_N, which always leads to X, and H_W, which always leads to X and Z. For example:

Generator shock always causes <u>reverbitial sonic</u>.
Pulsator damage always causes <u>reverbitial sonic</u> and <u>thermal tear</u>.

The order in which the two causes were listed was randomized for each problem. Participants were told that the base rates of H_N and H_W were equal, but the base rates of Z (in Experiment 1A) and of X (in Experiment 1B) were varied across problems at 5%, 35%, 65%, and 95% using a Latin square. These probabilities were presented in frequency format (e.g., "A study of 200 spaceships found that 70 of them had thermal tear"), and the denominator of the frequency ratio (e.g., 200 in the previous example) was varied across problems in order to make the manipulation less transparent.

Then, participants were told that X was observed but that we did not know whether Z had occurred (e.g., "Spaceship #53 was found to have [X]. We do not know whether or not it has [Z]."). Participants rated how satisfying explanations H_N and H_W would be on a scale from 0 ("Definitely [H_N]") to 10 ("Definitely [H_W]"), with the left/right order of H_N and H_W counterbalanced to match the order in which H_N and H_W were listed in the problem.

At the end of each experiment, several check questions were used to detect any participants who were not attending to the experimental task. These questions were multiple choice or true/false memory questions concerning the experimental stimuli (e.g., checking off from a list those items that were seen on previous screens). In all experiments, participants incorrectly answering more than 30% of the check questions were excluded from analysis.

4.2. Results

In reporting explanation judgments for all experiments, scores were centered so that 0 indicates no preference (the scale midpoint), and oriented so that positive scores (between 0 and 5) indicate a broad latent scope preference and negative scores (between 0 and −5) indicate a narrow latent scope preference.

As shown in Table 2, manipulating the base rate of the unknown evidence, $P(Z)$, in Experiment 1A had a large effect on explanatory preferences, whereas manipulating the base rate of the known evidence, $P(X)$, in Experiment 1B had a much more modest effect.

For statistical tests, we computed a linear contrast for each participant in both experiments. The linear effect of $P(Z)$ was very large [$t(32) = 8.76$, $p < .001$, $d = 1.52$, $BF_{10} > 1000$].[2] Participants had a strong preference for the narrow latent scope explanation (H_N) when $P(Z)$ was 5% ($M = -1.95$, $SD = 2.01$), a weaker preference when $P(Z)$ was 35% ($M = -1.25$, $SD = 1.87$), a weak preference for the *broad* latent scope explanation (H_W) when $P(Z)$ was 65% ($M = 0.86$, $SD = 2.07$), and a strong preference for H_W when the base rate was 95% ($M = 2.80$, $SD = 1.96$). Thus, the base rate of Z not only modulated the size of the latent scope bias, but *reversed* it when $P(Z)$ was very high. This pattern stands in contrast to the normative probability theory (according to which the base rate of Z is irrelevant).

In contrast, manipulating $P(X)$ in Experiment 1B had a far more modest linear effect [$t(33) = -2.46$, $p = .019$, $d = -0.42$, $BF_{10} = 2.02$], with an overall narrow latent scope preference in every condition. Further, the effect of $P(X)$ in Experiment 1B was much smaller than the effect of $P(Z)$ in Experiment 1A [$t(65) = -6.63$, $p < .001$, $d = -1.49$, $BF_{10} > 1000$]. This is consistent with the inferred evidence account, since X was already observed and its base rate is uninformative about Z. Thus, to the extent that the effect in Experiment 1B was due to scaling biases or demand characteristics, it is unlikely that these factors could explain the much larger effect in Experiment 1A.

4.3. Discussion

These results favor the inferred evidence account, from which the non-normative effect of $P(Z)$ was predicted. Might any of the alternative accounts be able to explain these findings? Pragmatic inference triggered by the speaker's supposed ignorance seems an unlikely explanation, as this factor did not vary with $P(Z)$. Likewise, representativeness merely uses the similarity of the observed and predicted evidence to estimate the likelihood term, and the observed evidence did not vary with $P(Z)$. Such accounts would predict only a general bias toward the narrow latent scope explanation, in contrast to the dramatic effect of $P(Z)$, which even led to a *wide* latent scope preference when $P(Z)$ was very high.

Non-independence also seems to be an unlikely explanation. According to this account of latent scope, people tacitly assume that the observed evidence (X) is correlated (positively or negatively) with the unknown evidence (Z), and that X is therefore evidence for Z. Once again, there is no reason to think that the correlation between X and Z would covary with $P(Z)$, so that these variables are negatively related when Z is uncommon but positively related when Z is common. We do acknowledge that participants could have tacit beliefs about the interaction of the effects given the current stimuli

[2] Throughout this article, we supplement all t-tests with Bayes Factors, computed using a default Jeffrey-Zellner-Siow (JZS) prior with a scaling factor of 1, as recommended by Rouder, Speckman, Sun, Morey, and Iverson (2009). Unlike p-values, Bayes Factors quantify evidence either against or in favor of a null hypothesis. When the Bayes Factor favors the null hypothesis, we notate it as BF_{01}, with the value of this factor indicating the probability of the data under the null over its probability under the alternative hypothesis; when the BF favors the alternative, we notate it as BF_{10} and report the reciprocal of BF_{01}, so that higher numbers correspond to greater likelihood of the data under the alternative. For example, $BF_{01} = 3.00$ means that the data is three times likelier under the null than under the alternative hypothesis, whereas $BF_{10} = 6.00$ means that the data is six times likelier under the alternative than under the null hypothesis. For a conceptual comparison of Bayesian versus null hypothesis significance testing, see Dienes (2011), and for computational details, see Rouder et al. (2009).

Please cite this article in press as: Johnson, S. G. B., et al. Sense-making under ignorance. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.06.004

Table 2
Results of Experiment 1.

Experiment 1A		Experiment 1B	
Condition	Explanatory preference	Condition	Explanatory preference
$P(Z) = 5\%$	−1.95 (2.01)	$P(X) = 5\%$	−0.54 (1.70)
$P(Z) = 35\%$	−1.25 (1.87)	$P(X) = 35\%$	−0.50 (1.90)
$P(Z) = 65\%$	0.86 (2.07)	$P(X) = 65\%$	−0.96 (1.45)
$P(Z) = 95\%$	2.80 (1.96)	$P(X) = 95\%$	−1.18 (1.60)

Note. Scores potentially range from −5 to 5 (*SD*s in parentheses), with negative scores indicating a preference for H_N and positive scores indicating a preference for H_W.

(e.g., technological failures, disease symptoms), but these correlations seem more likely to be positive than negative (e.g., one disease symptom making another symptom *more* likely; such positive non-independence was found by Rehder & Burnett, 2005), which would lead to a *wide* latent scope bias. We nonetheless measure these correlations empirically in Experiment 2.

The most plausible alternative explanation is that participants could have assigned higher prior probabilities to causes that generate effects with high base rates, which would indeed lead to the current pattern of results. We took measures to avoid this concern by explicitly stating that the two causes were equally frequent in the problem. However, this statement was rather abstract (phrased in terms of proportions), in contrast to our manipulation of the effect base rates, which used a frequency format. To further rule out concerns about biased priors, Experiment 2 measured participants' estimated base rates of the explanations and Experiment 3 used a frequency format to concretize these base rates.

5. Experiment 2

According to the biased priors account, participants in Experiment 1 assumed that judgments of $P(H_W)$ was higher as $P(Z)$ increased across conditions, and this increase in $P(H_W)$ led to the bias toward the wide latent scope explanation H_W for higher levels of $P(Z)$. According to the non-independence account, a negative correlation between the effects, so that the observed effect would make a latent effect appear less probable, leads to the preference for narrow latent scope explanations in general. To test these accounts, we manipulated $P(Z)$, as in Experiment 1, and measured participants' priors and beliefs about non-independence, as well as their explanatory preferences. We did so using a between-subjects design, as a further way to rule out concerns about demand characteristics in Experiment 1, with each participant assigned to a base rate $P(Z)$ of either 25%, 50%, or 75%. Finally, we also added an explanation for why the latent evidence was unavailable (a blood test had not come back from the lab), to block pragmatic interpretations of the speakers' claim to ignorance.

Including a 50% condition also allowed us to test whether there is still a bias for narrow latent scope even when participants cannot use $P(Z)$ to make any inferences about the latent effects. Since we are controlling statistically for effects of biased priors, non-independence, and inferred evidence—and experimentally for pragmatic inferences—any remaining bias in this condition would potentially be due to representativeness, or the use of similarity to estimate the likelihood term.

5.1. Methods

We recruited 300 participants from Amazon Mechanical Turk for Experiment 2; 7 participants were excluded because they failed more than 30% of the check questions.

Each participant made three judgments about a scenario similar to those used in Experiment 1, pertaining to the *Priors*, the *Independence* of the evidence, and their preferred *Explanation*. The scenario read:

> Imagine that you are a doctor. Below is some information about two diseases.
>
> Vilosa always causes abnormal gludon levels.
> Pylium always causes abnormal gludon and lian levels.
>
> Vilosa and Pylium occur equally often.
>
> A study of 1000 people found that [250/500/750] of them had abnormal lian levels.

The base rate of the unknown effect was varied between-subjects at 25%, 50%, and 75%, as shown in the bracketed text. The participant then completed the *Priors* question:

> Imagine that you took a random sample of people, and you found that a certain number of them had Vilosa. How many would you expect to have Pylium?

Responses were entered on a scale from −5 to 5, anchored at −5 ("Fewer"), 0 ("An Equal Number"), and 5 ("More"). Thus, negative scores indicate prior odds favoring the narrow latent scope explanation, while positive scores indicate prior odds favoring the wide latent scope explanation. Scores of 0 indicate equal priors for each explanation.

On the next page, the scenario was repeated at the top, and below the participant completed the *Independence* question:

> Consider just those people who have neither Vilosa nor Pylium. Some of these people nonetheless have abnormal levels of gludon, of lian, or of both. Now, consider two groups of such people:
>
> Group A: A sample of 100 people who have abnormal levels of gludon, but who have neither Vilosa nor Pylium.
>
> Group B: A sample of 100 people who do not have abnormal levels of gludon, but who have neither Vilosa nor Pylium.
>
> In which group do you think more people would have abnormal levels of lian?

This judgment was made on a scale from −5 to 5, anchored at −5 ("Group A has more"), 0 ("Groups have equal numbers"), and 5 ("Group B has more"). These scores were reverse-coded for analysis so that negative scores indicate non-independence of the evidence because the effects are thought to be *negatively* correlated, and positive scores indicate non-independence because the effects are thought to be *positively* correlated (as found by Rehder & Burnett, 2005). Scores of 0 indicate that the evidence is independent, conditional on the causes.

On the last page, the scenario was repeated once again, and the participant answered *Explanation* question:

> One of your patients, Patient #890, definitely has either Vilosa or Pylium, but you aren't sure which. Therefore, you ordered blood tests for the patient. The tests confirmed that the patient has abnormal levels of gludon. However, the test results for lian levels have not come back from the lab yet, so you don't know whether the patient's lian levels are normal or abnormal.
>
> Which disease do you think Patient #890 is most likely to have?

This judgment was made on a scale from −5 to 5, anchored at −5 ("Definitely Vilosa"), 0 ("Equally Likely"), and 5 ("Definitely Pylium"). Thus, negative scores indicate a preference for the narrow latent scope explanation, and positive scores indicate a preference for the wide latent scope explanation. Scores of 0 indicate equal posteriors for each explanation, which is normative assuming equal prior odds and independence of evidence in response to the previous two questions.

5.2. Results and discussion

As shown in Table 3, explanatory judgments scaled with the base rate of the unknown effect. To test for this effect statistically, while adjusting for the potential confounds of biased priors and independence violations, we used stepwise multiple regression (see Table 4). In Step 1, we found that base rate condition (.25, .50, or .75) significantly affected explanatory judgments [$b = 1.69$, $SE = 0.49$, $p < .001$], as in Experiment 1A. However, judgments of the priors did differ across condition [$b = 2.60$, $SE = 0.56$, $p < .001$], and the independence assumption was violated on average [$M = -0.72$, $SD = 2.58$; $t(292) = 4.74$, $p < .001$]. Thus, Step 2 capitalized on the variance among participants in their priors and independence judgments to test whether these judgments contributed to explanatory preferences. Neither judgment predicted explanatory ratings [for priors, $b = 0.01$, $SE = 0.05$, $p = .87$; for independence, $b = -0.02$, $SE = 0.04$, $p = .62$], while base rate condition continued to predict explanatory judgments just as strongly [$b = 1.68$, $SE = 0.51$, $p = .001$]. Indeed, the Step 2 model was not a significantly better fit than the Step 1 model [$MSE = 2.86$ vs. 2.86; $F(2,289) = 0.13$, $p = .88$] Thus, evidence base rates affect explanatory preferences above and beyond any possible effect on priors or the independence of the evidence.

The regression model can be used to predict explanatory judgments for a hypothetical participant who had precisely equal priors on the wide and narrow latent scope hypotheses and who believed the evidence to be completely independent, by entering '0' for these terms in the regression equation. The predicted response is −0.62 in the 25% condition, favoring the narrow latent scope explanation, and 0.22 in the 75% condition, favoring the wide latent scope explanation, similar to Experiment 1. However, in the 50% condition, a modest preference for the narrow latent scope explanation emerges (−0.20), indicating that factors above and beyond inferred evidence, priors, and independence violations are likely at play in assessing latent scope explanations. Because the cover story makes pragmatic inferences unlikely, the most likely candidate for this additional factor is representativeness. We discuss the relative contribution of all of these possible explanatory factors in Section 11.

It is worth noting that the effect of $P(Z)$ was smaller here than it was in Experiment 2; the Experiment 2 effect size is also more in keeping with subsequent experiments. Several factors likely contributed to the large effect in Experiment 1: That experiment used more extreme base rates (ranging from 5% to 95%); the design was within-subjects rather than between-subjects; it did not include questions probing participants' priors and independence assumptions (which would tend to focus participants on relevant rather than irrelevant cues); and it tested causal reasoning rather than categorization (see Section 8.2). Nonetheless, although the effect size can be modulated by such contextual factors, inferred evidence effects show up across experiments varying along all of these dimensions, testifying to the robustness of these effects.

6. Experiment 3

In addition to establishing that inferred evidence plays a role in the latent scope bias over-and-above the other factors, we also aim to quantify the impact of these other factors. Experiment 2 suggested a modest effect of representativeness (because there is still a bias in the 50% condition of Experiment 2) and little effect of biased priors or non-independence of evidence (because these factors had no effect in the regression model of Experiment 2). However, we experimentally *controlled* for pragmatic inferences, rather than *measuring* their impact. Hence, Experiment 3 directly measured the influence of pragmatic factors on the latent scope bias by varying whether a reason for ignorance was specified (turning off pragmatic inferences) or unspecified (potentially triggering pragmatic inferences).

Table 3
Results of Experiment 2.

Condition	Explanatory preference	Model-adjusted preference
$P(Z) = 25\%$	−0.64 (1.89)	−0.62
$P(Z) = 50\%$	−0.13 (1.48)	−0.20
$P(Z) = 75\%$	0.20 (1.73)	0.22

Note. Scores potentially range from −5 to 5 (*SDs* in parentheses), with negative scores indicating a preference for H_N and positive scores indicating a preference for H_W. The model-adjusted preference column indicates the predicted response for a hypothetical participant with unbiased priors who assumes the evidence to be independent (see main text for model description and Table 4 for coefficients).

Table 4
Multiple regression for Experiment 2, predicting explanatory preferences.

	Step one	Step two
Intercept	−1.03 (0.27)	−1.05 (0.28)
$P(Z)$	1.69 (0.49)***	1.68 (0.51)**
Priors		0.01 (0.05)
Independence		0.02 (0.04)

Note. Entries are unstandardized coefficients (*b*), with standard errors in parentheses, predicting explanatory preferences. For explanatory preferences, higher scores indicate a greater preference for H_W. For priors, higher scores indicate priors biased toward H_W. For independence, higher scores indicate a positive correlation between the observed and inferred evidence (which should lead to a bias toward H_W).
* $p < .05$.
** $p < .01$.
*** $p < .001$.

In addition, Experiment 3 included a condition where the missing evidence is not mentioned at all. In many real-life situations, available evidence will be explicit but missing evidence will simply fail to be observed or mentioned. On the one hand, one might conjecture that the relevance of the missing evidence is not obvious if it is not mentioned, so people may simply ignore it and therefore fail to use an inferred evidence strategy. On the other hand, however, people may automatically see the missing evidence as relevant—it may be the fact that the missing evidence is in fact *missing* that may need to be flagged. In such a case, missing evidence that is not explicitly mentioned may actually trigger inferred evidence even more strongly than missing evidence that is explicitly mentioned.

6.1. Methods

We recruited 299 participants from Amazon Mechanical Turk for Experiment 3; 9 participants were excluded because they failed more than 30% of the check questions.

Each participant read a scenario similar to that used in Experiment 2:

Imagine that you are a doctor. Below is some information about two diseases.

Vilosa and Pylium are rare diseases. In the United States population, they each occur in 1 in 1000 people.

Vilosa always causes abnormal gludon levels.
Pylium always causes abnormal gludon and lian levels.

A study was conducted of 1000 people randomly selected from the United States population. In that study, 250 of them had abnormal lian levels.

The order of listing the diseases was randomized, and the other information was adjusted to match this order. The base rate of the causes was given in frequency format (unlike Experiments 1 and 2), and the base rate of the unknown effect (lian levels) was always 25%. Based on Experiments 1 and 2, we would expect this base rate to lead to a modest but reliable bias favoring the narrow latent scope explanation (here, Vilosa).

Before answering the explanation question, participants were asked two comprehension questions concerning the base rates of the *causes* and the *effects*, in a random order. For the *cause base rate* question, participants were asked to "Consider a randomly selected American. Is this person more likely to have Vilosa or Pylium?" (options: "More likely to have Vilosa," "More likely to have Pylium," or "Equally likely to have Vilosa or Pylium"). For the *effect base rate* question, participants were asked to "Consider a randomly selected American. What is the probability that this person has abnormal lian levels?" (options: 25%, 50%, or 75%).

For the main task, participants diagnosed three patients (individuated by different patient numbers), one in the *Explanation* condition, one in the *No Explanation* condition, and one in the *No Information* condition. In the *Explanation* condition, the prompt was identical to that of Experiment 2, where the participant was told that "the test results for lian levels have not come back from the lab yet, so you don't know whether the patient's lian levels are normal or abnormal." In the *No Explanation* condition, this information was instead replaced with the sentence "You don't know whether the patient's lian levels are normal or abnormal." In the *No Information* condition, this information was omitted entirely. The three conditions were completed in a random order, and the scale was the same as in Experiment 2.

6.2. Results and discussion

There were no significant differences in explanation preferences between those who answered the cause base rate question correctly or incorrectly [$t(288) = 0.12$, $p = .90$] or between those who answered the effect base rate question correctly or incorrectly [$t(288) = 0.36$, $p = .72$], and there were no significant order effects across conditions [$ts < 1.1$, $ps > .27$], so we collapse across these factors.

To test the effect of pragmatic inference on the latent scope bias, we compared the *Explanation* condition, where a reason was given for the speaker's ignorance, and the *No Explanation* condition, where no reason was given (see Table 5 for means). There was a significant bias toward the narrow latent scope explanation in both conditions [$t(289) = 2.40$, $p = .017$, $d = -0.14$, $BF_{01} = 1.3$ and $t(289) = 2.49$, $p = .013$, $d = -0.15$, $BF_{01} = 1.0$]. Most importantly, these conditions did not differ from each other [$t(289) = 0.37$, $p = .71$, $d = 0.02$], suggesting that pragmatic inferences play a minimal role in producing the latent scope bias, at least for the type of stimuli used in our experiments.

To test whether the latent scope bias would still be found in the absence of information explicitly flagging the unknown effect as unknown, we compared the *No Information* condition to the mean of the other two conditions. Not only did we find a significant bias toward the narrow latent scope explanation in this condition [$t(289) = -7.19$, $p < .001$, $d = -0.42$, $BF_{10} > 1000$], but this bias was larger than in the other conditions [$t(289) = 5.35$, $p < .001$, $d = 0.32$, $BF_{10} > 1000$]. This suggests that even in the absence of explicit flagging, people view the unknown information as relevant. The bias is likely larger at least in part due to pragmatic influences, though more research would be necessary to tease apart potential causal factors.

Altogether, Experiments 1–3 quantify the impact of the factors listed in Table 1 in producing the bias toward narrow latent scope explanations. Biased priors (Experiment 2), non-independence of evidence (Experiment 2), and pragmatic inferences (Experiment 3) seem to have modest influences at most, for the stimuli used in these experiments. In contrast, inferred evidence seems to play the starring role (Experiments 1 and 2), affecting the size of the bias most dramatically (and even reversing the direction). Since there is still a residual bias even when the evidence base rate is 50% (Experiment 2), some additional influences seem to account for some of the variance, which could be representativeness or some as-yet-unidentified factor.

In the remaining experiments, we turn to additional predictions made by the inferred evidence account, including influences on evidence-seeking (Experiment 4), probabilistic inference

Please cite this article in press as: Johnson, S. G. B., et al. Sense-making under ignorance. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.06.004

Table 5
Results of Experiment 3.

Condition	Explanatory preference
Explanation	−0.18 (1.32)
No explanation	−0.20 (1.40)
No information	−0.85 (2.02)

Note. Scores potentially range from −5 to 5 (*SD*s in parentheses), with negative scores indicating a preference for H_N and positive scores indicating a preference for H_W.

(Experiment 5), and categorization (Experiments 5 and 6), as well as the role of the future knowability of the unknown evidence (Experiment 7).

7. Experiment 4

According to the inferred evidence account, when faced with a latent scope explanation, people try to infer whether or not the unknown effect occurred in the case at hand. Because the base rate of the unknown effect, $P(Z)$, can be used in making this inference, people should find $P(Z)$ more relevant than the base rate of the known effect, $P(X)$.

To test this possibility, participants were told about structurally similar situations to Experiments 1–3 (see Fig. 1), where they knew about one effect (X) but not another (Z), and were deciding between a narrow latent scope explanation (H_N, which would only account for the observed X) and a broad latent scope explanation (H_W, which would account for both the observed X and the unknown Z). Participants were asked to rank the base rates of each cause and effect in terms of "how useful" they would be for determining the best explanation—that is, to rank the relevance of $P(X)$, $P(Z)$, $P(H_N)$, and $P(H_W)$.

We anticipated that $P(Z)$ would be seen as more relevant for determining the best explanation compared to $P(X)$, since participants would use $P(Z)$ to assess whether Z occurred in the case at hand. In addition, we anticipated that $P(H_W)$ would be seen as more diagnostic than $P(H_N)$. This is because $P(H_W)$ is informative about $P(X)$ and also $P(Z)$, whereas H_N is informative only about $P(X)$. That is, if $P(H_W)$ is high, then both $P(X)$ and $P(Z)$ must also be high because H_W causes both effects. But if $P(H_N)$ is high, this implies only that $P(X)$ is high, but is not informative about $P(Z)$. Since we expected that $P(Z)$ would be seen as more relevant than $P(X)$, we would expect that likewise $P(H_W)$ would be seen as more relevant than $P(H_N)$. Both of these predictions stand in contrast to normative responding, since it is the prior odds ratio $[P(H_N)/P(H_W)]$ that determines the posterior odds favoring H_N over H_W.

7.1. Method

We recruited 200 participants from Amazon Mechanical Turk; 42 were excluded because they failed more than 30% of the check questions ($N = 18$) or had missing data ($N = 24$). The results are qualitatively the same if all participants are included.

Participants completed four problems in a random order, similar to those used in Experiment 1, but modified to elicit rankings of how useful each base rate would be for deciding between the explanations. For example, for the robot item, participants read the same causal information as in Experiment 1 (with the causes listed in a random order), and were told that "Spaceship #53 was found to have [X]. We do not know whether or not it has [Z]" and that "A study of 200 other spaceships was recently conducted, in which researchers collected measurements of several properties."

They then ranked each base rate in terms of how useful it "would be for determining what malfunction Spaceship #53 has, where '1' is the most useful and '4' is the least useful." The base rates were listed in a random order and worded in the format, "How many out of the 200 spaceships had [Y]," where [Y] was replaced with H_N, H_W, X, or Z.

7.2. Results and discussion

The proportion of times that participants ranked H_N, H_W, X, and Z in each position are shown in Table 6. In absolute terms, the base rate of Z was ranked first more frequently (32%) than any other base rate, and the base rate of X was ranked last more frequently (38%) than any other base rate. Thus, our prediction that $P(Z)$ would be seen as more relevant than $P(X)$ has qualitative support. In addition, $P(H_W)$ was ranked first much more frequently than $P(H_N)$ (29% vs. 15%) and was ranked last less frequently (16% vs. 23%). Again, this is qualitatively consistent with our prediction that $P(H_W)$ would be seen as more relevant than $P(H_N)$.

Statistical analyses confirmed these patterns, both in terms of the overall rank of each base rate, and the frequency of each base rate at the top and bottom ranks. First, we calculated the mean rank of H_N, H_W, X, and Z across all four items for each participant, with '1' representing the first ranked choice and '4' representing the last ranked choice for each item. The mean rank for Z was higher than for X [$M = 2.36$, $SD = 0.91$ vs. $M = 2.69$, $SD = 0.99$; $t(157) = -2.55$, $p = .012$, $d = -0.20$, $BF_{10} = 1.51$], and the mean rank for H_W was higher than for H_N [$M = 2.32$, $SD = 0.81$ vs. $M = 2.63$, $SD = 0.70$; $t(157) = -3.43$, $p < .001$, $d = -0.27$, $BF_{10} = 17.8$]. Thus, Z was seen as more relevant than X and H_W was seen as more relevant than H_N, as predicted.

Second, we performed a series of Chi-squared tests on the frequencies with which each base rate was ranked at each position, to investigate the prevalence of each base rate at the top and bottom rankings. Overall, the frequencies of the first-ranked choices (i.e., the top row of Table 6) differed from chance responding [$\chi^2(3, N = 632) = 39.66$, $p < .001$]. In particular, $P(Z)$ was ranked first more often than $P(X)$ [$\chi^2(1, N = 356) = 5.44$, $p = .020$], and $P(H_W)$ was ranked first more often than $P(H_N)$ [$\chi^2(1, N = 248) = 26.80$, $p < .001$]. The distribution of last-ranked choices (i.e., the bottom row of Table 6) also differed from chance [$\chi^2(3, N = 632) = 59.96$, $p < .001$], and showed precisely the opposite pattern of the first-ranked choices. That is, $P(X)$ was ranked last more often than $P(Z)$ [$\chi^2(1, N = 384) = 21.09$, $p < .001$] and $P(H_N)$ was ranked last more often than $P(H_W)$ [$\chi^2(1, N = 248) = 6.45$, $p = .011$].

Taken together, these results underscore Experiments 1 and 2, where $P(Z)$ was used more strongly than $P(X)$. In Experiment 4, these base rates were also sought out more readily when determining the best explanation. This pattern shows that people actively seek the information they believe to be necessary for inferring unavailable evidence. In addition, Experiment 4 confirmed an additional, novel prediction of the inferred evidence account—that the base rate of the broad latent scope cause (H_W) would be seen as more relevant than the base rate of the narrow latent scope cause (H_N). This overall response pattern—ranking $P(H_N)$ and $P(H_W)$ differentially and $P(Z)$ highest most often—stands in stark contrast to normative responding, since only the ratio of $P(H_N)$ to $P(H_W)$ is relevant to assessing the probability of each explanation.

8. Experiment 5

People are averse to latent scope explanations not only in causal reasoning, but also in categorization (Sussman et al., 2014). When deciding whether an exemplar belongs in one category that predicts an unknown feature (the wide latent scope category H_W) or in another that does not predict that feature (the narrow latent scope category H_N), people prefer to categorize the exemplar in the narrow category. If the inferred evidence strategy is a domain-general aspect of explanatory logic, as we are claiming, then it should explain the latent scope bias not only in causal explanation but also in categorization. Experiments 5 and 6 test this possibility.

As for causal explanation, probability theory tells us that the base rates of unknown features are irrelevant to determining which category it belongs to. When an effect or feature has a higher base rate than its cause, this implies that some alternative causes or categories must exist. For example, suppose that colds cause sneezing and 5% of people have colds at any given time. Then if 10% of people are sneezing, there must be some people who are sneezing even though they do not have colds—that is, there must be alternative causes of sneezing. Similarly, suppose that cheetahs have spots and 8% of African land mammals are cheetahs. Then if 20% of African land mammals have spots, there must be some African land mammals that have spots even though they are not cheetahs—there are alternative categories of African land mammals that have spots. This is why evidence about effect or feature base

Table 6
Results of Experiment 4.

	Base rate (%)			
	$P(H_N)$	$P(H_W)$	$P(X)$	$P(Z)$
First ranked	15	29	25	32
Second ranked	29	27	19	24
Third ranked	33	28	19	21
Fourth ranked	23	16	38	23

Note. Entries indicate the total proportion of times each base rate was ranked in each position across the four problems completed by each participant. Rows may not sum to 100% due to rounding.

rates is irrelevant once the cause or category base rates are known—to the extent that these effect or feature base rates are higher than the cause or category base rate, this simply indicates that a *different* cause or categorization is likely.

In Experiments 1 and 2, people violated this principle in evaluating causal explanations. Even though they knew that two potential causes H_N and H_W had equal base rates, they used the base rate of Z—a latent effect of H_W—in their judgments. When Z had a 5% base rate, participants appear to have reasoned that Z was unlikely to have occurred in the case at hand, so the explanation that did not posit Z was more likely than the explanation that did. Normatively, to the extent that Z is rare, this just means that both H_W and H_N are relatively rare, since they have equal base rates, or that there are preventive causes of Z that mask the relationship between H_W and Z—in neither case does this information help to distinguish the relative probability of H_N and H_W. Similarly, when Z had a 95% base rate, participants appear to have reasoned that Z was very likely to have occurred, so they preferred the explanation that accommodated Z. However, to the extent that Z is very common, this just means that either H_N and H_W are both very common, or that there are other causes of Z.

It is unclear whether participants committed these errors because they failed to notice that high base rates of Z imply alternative causes that are uninformative for distinguishing between H_N and H_W, or whether they might instead have noticed this fact but nonetheless used the inferred presence of Z for distinguishing between the hypotheses. In Experiment 5A, we used a categorization task and highlighted this fact, to see whether participants still used the Z base rate for making explanatory inferences. In a categorization task (with a structure analogous to Fig. 1), participants read about situations like the following:

> You come across a deer in a meadow, but you are not sure whether it belongs to species *trocosiens* or species *myronisus*. The meadow contains equal numbers of *trocosiens* and *myronisus* deer, and also contains many other deer. Below is some information you can use to decide which it might belong to:
>
> Deer of the *trocosiens* species have white spots.
>
> Deer of the *myronisus* species have white spots and semi-hollow antlers.
> Most other species of deer also have semi-hollow antlers.
>
> You know that the deer has white spots, but you do not know whether it has semi-hollow antlers.

Some participants read a version of this item, like the above, where feature Z was *common* among other categories of deer ("Most other species of deer also have semi-hollow antlers"), and other participants read a version of this item where feature Z was instead *uncommon* among other categories of deer ("No other species of deer has semi-hollow antlers").

We would expect to find a narrow latent scope preference when the latent feature is rare. This follows from the idea that people use the base rate of the latent feature to infer the probability of that feature in the case at hand, and is consistent with previous findings of narrow latent scope preferences when the features are likely to have low implicit base rates (Sussman et al., 2014). However, we anticipated that participants would be more likely to endorse the wide latent scope category when the feature had a high base rate due to its prevalence among *other* species of deer. This stands in contrast to the dictates of probability theory: Since participants were categorizing this exemplar as belonging to one or the other species (H_N or H_W), facts about the prevalence of semi-hollow antlers among *other* species of deer are irrelevant to interpreting the current evidence. However, it is consistent with the use of inferred evidence: Given an arbitrary deer belonging to any category, it is more likely to have the latent feature if that feature is common among all types of deer.

In addition to testing this prediction of inferred evidence, we tested a further processing prediction: That to the extent that people were more inclined toward the narrow latent scope category, it would be due to their inference that the latent feature was unlikely in the case at hand. In Experiment 5B, we asked participants to rate the probability of observing the latent feature in an exemplar (given that the exemplar belonged to either H_N or H_W) when that feature was either common or uncommon among other categories. We expected that participants would rate the feature more probable when it had a high base rate among other categories, even though they were explicitly told that the exemplar belonged to either H_N or H_W.

8.1. Methods

We recruited 200 participants from Amazon Mechanical Turk ($N = 100$ and $N = 100$ for Experiment 5A and 5B, respectively); 21 participants ($N = 10$ and $N = 11$ for Experiment 5A and 5B) were excluded because they failed more than 30% of the check questions.

In Experiment 5A, participants made categorization judgments based on incomplete information, using items phrased similar to the above deer example. For some items, the latent feature was *common* among other categories (i.e., had a high base rate) and for other items, the latent feature was *uncommon* among other categories (i.e., had a low base rate). Participants made categorization judgments (e.g., "Which species do you think the deer belongs to?") on a scale from 0 ("Definitely *trocosiens*") to 10 ("Definitely *myronisus*").

In Experiment 5B, participants read the same items, but rather than making categorization judgments, they rated the probability of the latent feature being present. Because it was critical that participants know that the exemplar belonged to either H_N or H_W rather than to an alternative category, the first sentence of each item was slightly modified (e.g., "You come across a deer in a meadow, which belongs to either species *trocosiens* or species *myronisus*"). The information was otherwise identical, with the same *common* versus *uncommon* manipulation as in Experiment 5A. Participants were asked to rate the probability that the exemplar had the latent feature on a scale from 0% to 100%.

In both experiments, eight biological categories were used in total, with each participant seeing four items in the *common* version and four in the *uncommon* version (counterbalanced across participants). Items were completed in a random order.

8.2. Results and discussion

Participants in Experiment 5A used the feature base rates in their categorizations, even though these base rates explicitly referred to the prevalence of features in *other* categories (see Table 7). In the *uncommon* condition, participants had a narrow latent scope bias [$M = -0.56$, $SD = 1.06$; $t(89) = 4.97$, $p < .001$, $d = -0.52$, $BF_{10} > 1000$], consistent with Sussman et al.'s (2014) finding of a narrow latent scope bias in categorization, which used low base rate features. However, in the *common* condition, participants had no preference one way or the other [$M = 0.03$, $SD = 1.13$; $t(89) = 0.27$, $p = .78$, $d = 0.03$, $BF_{01} = 11.6$]. This led to a significant difference between conditions [$t(89) = 3.79$, $p < .001$, $d = 0.40$, $BF_{10} = 59.3$], suggesting that participants used the feature base rates in a non-normative way to infer the probability of the feature being present in exemplar being categorized.

Direct evidence for this interpretation came from Experiment 5B, where participants rated the probability of the latent feature being present to be higher when that feature was prevalent in other categories, even though participants were told that the exemplar did not belong to those categories. Participants inferred on average that the exemplar had a 35.8% ($SD = 14.9\%$) chance of having the property in the *uncommon* condition, which is significantly lower than the normative response of 50%

Table 7
Results of Experiment 5.

Condition	Experiment 5A Explanatory preference	Experiment 5B Probability judgment
Uncommon	−0.56 (1.06)	35.8% (14.9%)
Common	0.03 (1.13)	62.7% (17.9%)

Note. For Experiment 5A, scores potentially range from −5 to 5 (*SD*s in parentheses), with negative scores indicating a preference for H_N and positive scores indicating a preference for H_W. For Experiment 5B, probability judgments potentially range from 0% to 100% (*SD*s in parentheses).

[$t(88) = -9.04$, $p < .001$, $d = -0.96$, $BF_{10} > 1000$]. But in the *common* condition, participants inferred that the exemplar had a 62.7% ($SD = 17.9\%$) chance of having the property, which is significantly *higher* than the normative response of 50% [$t(88) = 6.68$, $p < .001$, $d = 0.71$, $BF_{10} > 1000$].

Overall, these results accord with the inferred evidence account. Participants in Experiment 5B used the base rate of the latent feature both to infer the probability of the feature's presence in an exemplar, even though the feature base rate was manipulated by altering its frequency in categories that the exemplar did not belong to. As predicted by the inferred evidence account, this had downstream consequences for participants' explanatory inferences, with a narrow latent scope preference only when the feature was rare among other categories.

One aspect of these results worth noting is the lack of a *wide* latent scope bias in the *common* condition of Experiment 5A. One possible explanation of this result is that the irrelevance of high feature base rates in categorization is more transparent than the irrelevance of high effect base rates in causal reasoning. When a cause does not produce an effect (e.g., H_N not producing Z in Fig. 1), Z can still occur if some alternative background cause is present. For example, suppose a person has one of two equally rare diseases, one of which causes a person's hair to turn brown. Because many people already have brown hair, there is a more than 50% chance that this person will have brown hair, even if she has a 50/50 chance of having each disease—that is, multiple causes can occur simultaneously (a person could have a gene for brown hair *and* one of the diseases). In contrast, when an exemplar's category fails to have a feature (e.g., a category H_N does not have the feature Z), this usually implies that the exemplar *does not have* that feature. For example, suppose that an animal belongs to one of two equally rare subspecies of deer, one of which has brown fur and one of which has white fur. Because most other subspecies of deer have brown fur, it is likely that an arbitrary deer will have brown fur; however, for a deer that definitely belongs to one of the two subspecies with a 50/50 chance, it has precisely a 50% chance of having brown fur. That is, the deer does not belong to multiple subspecies of deer, so the prevalence of brown fur among other subspecies is not relevant. This task difference may make the irrelevance of the latent effect/feature base rate more transparent.

9. Experiment 6

In Experiment 6, we aimed to replicate the effect of latent feature base rates using a more naturalistic task. Here, we asked a group of pretest participants to produce base rates for a range of features for several categories (e.g., "having protruding eyes" was seen as a very prevalent property among frogs, whereas "having a tail" was seen as a less prevalent property). We then used these tacit base rates to test for latent scope biases in a new group of participants, using a task similar to Experiment 5. We anticipated a greater preference for the narrow latent scope category when the latent feature was relatively rare, compared to when the latent feature was relatively common.

9.1. Pretest

We recruited 30 participants from Amazon Mechanical Turk to participate in the norming pretest; no participant incorrectly answered more than 30% of the check questions, so all were included in the data analysis.

Please cite this article in press as: Johnson, S. G. B., et al. Sense-making under ignorance. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.06.004

Participants made judgments about eight categories, covering a variety of natural kinds and artifacts. For each category, participants rated the frequency of features that we expected to have relatively high or relatively low base rates in that category. For example, participants were asked to "think of 100 clocks. Out of those 100 clocks, how many would have the following properties?" and rated properties such as "has a manual setting," "uses roman numerals on the display," "has a pendulum," etc., on separate 0–100 scales. For seven of these items, we were able to select a property with a relatively high base rate (for the clock item, "requires battery") or a relatively low base rate (e.g., "is red in color"). The *high base rate* properties for each category had a mean rating of 80.4 (SD = 10.5) and the *low base rate* properties had a mean rating of 13.5 (SD = 5.3). These items are listed in Table 8.

9.2. Methods

We recruited 100 participants from Amazon Mechanical Turk for the main experiment; 13 participants were excluded because they failed more than 30% of the check questions.

For each of the seven items, participants were randomly assigned to the *high base rate* or the *low base rate* version. The only difference between these versions was whether the latent property possessed by H_W had a high or low implicit base rate in the pretest. For example, the clock item read (differences between conditions in brackets):

You come across a clock in an office, but you are not sure whether it belongs to type *Vermiller* or type *Pomerantz*. The office has equal numbers of clocks of each type. Below is some information you can use to decide which type it might belong to:

Clocks of the *Vermiller* type are <u>rectangular in shape</u>.
Clocks of the *Pomerantz* type are <u>rectangular in shape</u> and [<u>require batteries</u>/<u>red in color</u>].

You know that the clock is <u>rectangular in shape</u>, but you do not know whether it <u>requires batteries</u>.

Which type do you think the clock belongs to?

Participants then made categorization judgments on a scale from 0 ("Definitely Vermiller") to 10 ("Definitely Pomerantz"). The order of listing H_N and H_W was randomized for each item, and the left/right order of the response scale matched this order.

9.3. Results and discussion

Participants used their implicit base rates of the latent effects in making their categorizations. As shown in Table 8, for the *low base rate* versions of each item, participants had a strong preference for the narrow latent scope category [$M = -0.63$, $SD = 0.24$; $t(6) = -12.49$, $p < .001$, $d = -4.74$, $BF_{10} > 1000$]. But for the *high base rate* versions, participants had a comparatively weak preference [$M = -0.25$, $SD = 0.19$; $t(6) = -3.49$, $p = .013$, $d = -1.32$, $BF_{10} = 5.74$], leading to a significant difference between conditions [$t(6) = 4.22$, $p = .006$, $d = 1.60$, $BF_{10} = 11.75$]. Moreover, the pretest ratings of $P(Z)$ were highly correlated with explanatory preferences in the main experiment [$r(12) = .83$, $p < .001$]; this correlation is also of sizable magnitude looking just within the *high base rate* versions [$r(5) = .44$, $p = .33$] and just within the *low base rate* versions [$r(5) = .58$, $p = .17$]. Thus, when evaluating explanations with unknown feature values, participants not only relied on explicit information about evidence base rates, as in our previous experiments, but also on their tacit knowledge about the distribution of features over natural categories. These results suggest that the use of inferred evidence may extend to everyday explanatory reasoning, where explicit base rates are often unavailable.

These effects, though highly consistent (see Table 8), were smaller than those in previous experiments with more explicit manipulations. It is not altogether surprising that our tacit manipulation of base rates here was weaker, since this manipulation requires participants to recruit their prior knowl-

Table 8
Results of Experiment 6.

Category	High base rate version			Low base rate version		
	Feature	Estimated prevalence (%)	Explanatory preference	Feature	Estimated prevalence (%)	Explanatory preference
Fish	Has a jaw	67.4	−0.43	Orange scales	19.7	−0.60
Mushroom	Has a cap	81.3	−0.27	Blue with yellow spots	8.4	−0.52
Frog	Protruding eyes	80.2	−0.38	Has a tail	7.4	−0.84
Bird	Ability to fly	91.7	−0.37	Has teeth	21.2	−0.43
Coat	Full sleeves	90.2	−0.08	Made of silk	10.3	−0.65
Bike	Metal frame	86.6	0.09	Transparent frame	13.3	−0.66
Clock	Requires batteries	65.3	−0.27	Red in color	13.8	−0.71

Note. Prevalence estimates are the mean estimate of category members having each property in the norming pretest. For explanatory preferences from the main experiment, scores potentially range from −5 to 5, with negative scores indicating a narrow latent scope preference and positive scores indicating a wide latent scope preference.

edge and since disagreements among participants' tacit base rates will cause regression to the mean. Further, the same differences between categorization and causal reasoning that we highlighted earlier would also be at work here—multiple causes often occur simultaneously (so they are not mutually exclusive) but exemplars usually do not belong to multiple categories at the same taxonomic level (so they *are* mutually exclusive). As we explained in discussing Experiment 5, this could lead to the irrelevance of feature base rates being more transparent than the irrelevance of effect base rates, resulting in relatively smaller effects of evidence base rates in categorization.

It is more surprising that a narrow latent scope preference was still found for the *high base rate* versions, however, given a *wide* latent scope preference for the high base rate conditions of Experiment 2. This suggests that some other factors contribute to the latent scope effect, over and above inferred evidence. We parse the relative contribution of the five potential mechanisms—inferred evidence, biased priors, non-independence, pragmatic inference, and representativeness—in Section 11.

10. Experiment 7

In several of the previous experiments, participants were provided with *reasons* that the evidence was unavailable, which would tend to block pragmatic inferences about the speaker's intentions. Experiment 3 specifically measured the effects of such inferences by varying the availability of reasons, and found that pragmatic inference does not play a significant role in the latent scope bias, at least for our experimental materials.

However, the *nature* of the reason for ignorance may have an effect over-and-above pragmatic inferences, if these different reasons lead to inferences about the evidence base rates that differ in strength. In Experiment 7, we contrasted reasons that led to the latent predictions being unknown but *verifiable*, or unknown and *unverifiable*. For example, a *verifiable* reason that test results would be unavailable is that the lab technician's handwriting is illegible. In this case, the lab technician could be contacted or the test could be rerun, so the evidence can be resolved one way or the other in the future. In contrast, an *unverifiable* reason that test results would be unverifiable is that no blood test exists for a particular biochemical. In that case, it is unlikely that the levels of that biochemical could ever be determined, so the predictions of competing diagnoses cannot be verified.

In terms of the inferred evidence account, the verifiability of a prediction may influence inferences about that prediction because people use ease-of-imagining as a heuristic for truth (Koehler, 1991). One way to think about this heuristic formally is in terms of simulation-based mechanisms for estimating probabilities (Bonawitz, Denison, Gopnik, & Griffiths, 2014; Griffiths, Vul, & Sanborn, 2012). According to simulation-based models, hypotheses or evidence are sampled in order to estimate probabilities, and this sampling process can lead to systematic biases (Bonawitz et al., 2014). If ease-of-imagining influences the probability of sampling a particular possibility, then less easily imagined possibilities would be deemed less probable than more easily imagined possibilities, consistent with empirical results (Koehler, 1991). In terms of the inferred evidence model (see Eq. (5)), the weight $f^{+Z} = P(Z|I)/P(Z)$ would be smaller when Z is hard-to-imagine than when Z is easy-to-imagine. This places a larger weight on how well the explanations fare in the event that Z is false, which favors H_N. Since it is easy to imagine finding out that a verifiable prediction is true and difficult to imagine finding out that an unverifiable prediction is true, the bias for H_N should be stronger for unverifiable than for verifiable predictions.

10.1. Method

We recruited 100 participants from Amazon Mechanical Turk; 18 were excluded from data analysis because they failed more than 30% of the check questions.

Participants completed seven items in which they took the role of a doctor diagnosing patients, where the diagnosis options (H_N and H_W) had symptoms corresponding to the causal structure in Fig. 1 (i.e., H_N causes X, and H_W causes X and Z). For each item, a different name was given to the patient, to the symptoms (fictitious names for X and Z), and to the diagnosis options (fictitious names for H_N and H_W). Five of the items consisted of an "excerpt from a medical reference book," stating that one disease (H_N) always caused one biochemical to have abnormal levels (X), while a second disease (H_W) always caused two biochemicals to have abnormal levels (X and Z) but that nothing else was known to lead to those abnormal biochemical levels. Participants then read a "note from the lab," confirming result X but giving various reasons why the value of Z was unknown. Three of these reasons led to Z being unknown but potentially knowable (the *knowable* conditions): (1) the lab technician's handwriting was illegible; (2) the results were misplaced; and (3) the test could not be conducted due to equipment failure. The other two reasons led to Z being unknown and unknowable (the *unknowable* conditions): (4) a blood test for that biochemical has not been developed; and (5) that biochemical is too small to be detected in principle. Two additional problems were used as controls, where Z was known, and was either confirmed or disconfirmed (i.e., was in the positive or negative scope of H_W). Each participant also completed a parallel 'magic diagnosis' scenario. There was no main effect or interaction with scenario, so we collapsed across this variable in our analyses.

For each scenario, a Latin square was used to assign the seven different patients and symptom sets to the seven different problem structures, consisting of the five latent scope problems varying the reason for ignorance, and the two control problems. For each item, participants were asked which explanation they found most satisfying on a scale from 0 ("Definitely [H_N]") to 10 ("Definitely [H_W]"). The order in which participants completed the medical and magic scenarios was counterbalanced, and the order of the seven items was randomized within each scenario.

10.2. Results and discussion

As shown in Table 9, these items led to a robust preference for H_N across every condition. This provides further evidence against pragmatic accounts, since the reasons given for the speaker's ignorance should block participants from making pragmatic inferences.

However, the *magnitude* of the latent scope bias differed across conditions. In cases where the latent effect was potentially *verifiable* (illegible handwriting, misplaced results, equipment failure), the latent scope effect was relatively weak [$M = -0.48$, $SD = 0.85$; $t(81) = -5.04$, $p < .001$, $d = -0.56$, $BF_{10} > 1000$], and did not differ among these reasons [$ts < 1.8$, $ps > .075$, $BF_{01} > 2.4$]. In cases where the latent effect was *unverifiable* (no diagnostic test, unobservable in principle), the latent scope effect was relatively strong [$M = -0.89$, $SD = 1.27$; $t(81) = -6.35$, $p < .001$, $d = -0.70$, $BF_{10} > 1000$], and did not differ between these reasons [$t(81) = 0.38$, $p = .71$, $d = 0.04$, $BF_{01} = 10.7$]. This led to a significant difference between the verifiable and unverifiable reasons [$t(81) = 3.75$, $p < .001$, $d = 0.41$, $BF_{10} = 53.3$].

This sensitivity to ease-of-imagination is consistent with the use of inferred evidence. When Z is unknown and unknowable, it is more difficult to imagine that Z is true (Koehler, 1991), causing an aversion to the broad latent scope explanation (H_W) that predicts Z. In contrast, when Z is unknown

Table 9
Results of Experiment 7.

Reason type	Reason for ignorance	Explanatory preference
Knowable	Illegible handwriting	−0.33 (1.21)
	Misplaced results	−0.53 (1.13)
	Equipment failure	−0.57 (1.11)
Unknowable	No diagnostic test	−0.92 (1.36)
	Unobservable in principle	−0.87 (1.52)

Note. Scores potentially range from −5 to 5 (*SD*s in parentheses), with negative scores indicating a preference for H_N and positive scores indicating a preference for H_W.

but potentially knowable, it is easier to imagine observing *Z* in the future, shifting people relatively more toward the broad latent scope explanation (H_N) that predicts *Z*.

Two alternative possibilities merit consideration. First, participants could be using the reasons for ignorance as a way to estimate their priors on each cause, rather than to evaluate the evidence itself. For example, participants could find diseases with unverifiable symptoms to be implausible. However, this explanation is at best strained for the magic items (diagnosing various 'magical traces' using 'detector spells'). Since the pattern was identical across the medical and magic items, this explanation seems unlikely. Second, participants could think that an effect that is impossible to *detect* also cannot *happen*. However, this interpretation seems unlikely even for the medical items. The *unverifiable* reasons in the medical scenario have plausible and clear physical interpretations (the biomolecule is too small to be detected by any existing test, or that no diagnostic test has been developed for that biomolecule), and we see no reason for participants to think that such molecules could not exist.

11. General discussion

We often must make sense of things with incomplete evidence. Here, we showed that people use *inferred evidence* in both causal reasoning and categorization to try to minimize these unknowns. Although such 'filling in' strategies are broadly adaptive across many areas of cognition (e.g., Bartlett, 1932; Marr, 1982; Simons & Levin, 1997), participants in the current studies used normatively irrelevant cues to make these inferences, such as the base rates of unknown effects or features. Thus, the 'filling in' or inferred evidence strategy can lead to illusory inferences such as the latent scope bias (Khemlani et al., 2011).

We presented two broad kinds of evidence for this thesis. The first kind of evidence concerned the *output* of reasoning processes. Most critically, we expected that people would use the base rate of latent evidence to infer whether the evidence would be present in the case at hand. This would be non-normative, because knowledge of the explanations' base rates screens off information about the base rate of the evidence. Nonetheless, people did use these irrelevant base rates in three experiments, across quite different paradigms and manipulations. In Experiment 1A, participants used the base rates of latent effects in diagnostic causal reasoning, leading to a preference for the wide latent scope cause (i.e., the cause that posits the unknown effect)—a reversal of the many previous findings of narrow latent scope preferences (Khemlani et al., 2011; Sussman et al., 2014). In Experiment 5A, participants preferred a narrower over a wider latent scope categorization when no *other* category posited the latent feature, yet had no preference between the narrow and wide categories when many other categories had that feature. This result is strikingly non-normative given that the problems emphasized the fact that the feature's base rate was driven by categories other than those under consideration as potential categorizations of the exemplar. Finally, Experiment 6 relied on participants' tacit beliefs about the base rates of natural category features, and found a stronger preference for a narrow latent scope categorization when the latent feature had a low tacit base rate (e.g., a clock having the feature "being red in color") rather than a high tacit base rate (e.g., a clock having the feature "requires batteries"). In addition, Experiment 7 found that a completely different manipulation affecting the tendency to infer evidence (verifiable versus unverifiable reasons for ignorance) led to a similar pattern of results.

The second kind of evidence concerned the processing implications of inferred evidence. In Experiment 4, we tested the prediction that people would be especially motivated to seek information about latent effect base rates, and less motivated to seek out information about known effect base rates. This stands in contrast to the laws of probability, according to which neither of these base rates is diagnostic if the base rates of the causes are known. Indeed, we found not only that the latent effect base rate was the most sought-after piece of information, but that the base rate of the wide latent scope cause was also more sought-after than the base rate of the narrow latent scope cause. This latter finding is particularly distinct from normative responding, where it is the *ratio* of the cause base rates that is relevant. However, the wide latent scope cause base rate provides information about the latent effect base rate (since it causes this effect), whereas the narrow latent scope cause base rate provides no such information. Thus, participants' interest in the latent effect base rate appears to trickle up to the wide scope explanation base rate.

Finally, our account predicts that participants should produce inferences about latent observations as they make their explanatory inferences. Experiment 5B found evidence for this prediction, with participants being more likely to infer a feature's presence, given that an exemplar belonged to a wide or narrow latent scope category, when the feature was prevalent in *other* categories, compared to when it was not. This result complements Experiment 5A's finding of a narrow latent scope bias only when the feature was not prevalent among other categories: One would expect a stronger preference for the narrow latent scope explanation when the latent feature was thought unlikely to be present, just as we found.

11.1. Alternative accounts

Taken together, these results support the role of inferred evidence in explanatory reasoning. However, several alternative (in some cases, normative) processes could lead to a bias for narrow latent scope (Table 1). Here, we reconsider these accounts in light of the current findings. Although the current results demonstrate that inferred evidence contributes to the latent scope bias over and above these other accounts, there is reason to think that some of them may play a role.

First, people could have prior probabilities that favor narrow over wide latent scope explanations, and their priors might favor narrow scope explanations more when their predictions have low prior probabilities. In general, adults and even young children are both sensitive to prior probabilities in their explanatory reasoning (Bonawitz & Lombrozo, 2012; Johnston, Johnson, Koven, & Keil, 2015, in preparation; Lombrozo, 2007). It is therefore somewhat surprising that in Experiment 2, participants indicated that their priors *did* favor the narrow latent scope explanation more when the latent prediction had a low base rate, yet these biased priors were not associated with their explanatory judgments. Most critically for the inferred evidence account, the latent scope bias and effect of evidence base rates held up even after adjusting statistically for participants' priors. Although the effect of priors seems to have been swamped by the effect of evidence base rates in this particular experiment, it is certainly possible that biased priors can accentuate or attenuate the latent scope bias, and future work might explore this possibility.

Second, people could believe that the independence assumption is violated—that the observed and latent evidence may be correlated, conditioned on which explanation is true. If the evidence is negatively correlated, then the observed evidence counts as evidence *against* the latent prediction, whereas if it is positively correlated, then the observed evidence counts as evidence *for* the latent prediction. Thus, a negative correlation would lead to a narrow latent scope bias and a positive correlation would lead to a wide latent scope bias. Experiment 2 tested this issue directly, and found *positive* violations of independence, which ought to lead toward a *wide* latent scope bias—against our hypothesis. However, as with the effect of biased priors, this non-independence did not appear to affect judgments in Experiment 2, and the effect of evidence base rates held up after adjusting for violations of independence.

Third, people might be using inferred evidence of a different sort, based not on base rates but on *conversational* implicature. For example, "we don't know about *Z*" could be interpreted to mean "we don't know about *Z*, but we probably would have observed *Z* if it existed, so *Z* is probably false." In that case, pragmatic inferences could lead to a bias favoring narrow latent scope. Alternatively, "we don't know about *Z*" could be interpreted to mean that the speaker is hiding relevant information from the participant. That inference would lead to a bias favoring *wide* latent scope. However, neither of these inferences appears to be a primary factor driving the results. Several experiments included plausible reasons for the speaker's ignorance, and Experiment 3 directly compared cases with and without such reasons. These studies did not support an important role for pragmatic inference in the latent scope bias.

Finally, the observed evidence $\{X\}$ might be seen as more similar to (or representative of) the hypothesized evidence under the narrow scope explanation, $\{X\}$, than to the hypothesized evidence under the wide scope explanation, $\{X, Z\}$. If people use similarity or representativeness to estimate the fit between data and hypothesis (Tenenbaum & Griffiths, 2001a), then such a mechanism could lead to a bias toward the narrow latent scope explanation. However, this cannot be a full explanation because it would not predict any effect of $P(Z)$, since Z is not known to be present in the case at hand regardless of its base rate (cf. Experiments 1–6).

We do not necessarily claim, however, that inferred evidence captures all of the variation in judgments. The regression model in Experiment 2 adjusted for the effects of $P(Z)$, as well as biased priors and non-independence, in an experimental setting that would minimize pragmatic inferences. There was still a slight bias toward the narrow latent scope explanation even when $P(Z) = .5$, suggesting that some factor is at play above and beyond these others. Likewise, Experiment 5 found that there was no bias even when the base rate of Z was high, and Experiment 6 even found a slight bias toward the narrow latent scope explanation even when participants had high tacit base rates of Z. One possibility is that representativeness plays a key role in these biases, since it was the only factor not controlled in Experiment 2. A second, not mutually exclusive, possibility is that even when the base rate of Z is 50%, and participants think that they would be equally likely to observe positive and negative evidence, that they overweight the *importance* of the negative evidence. As noted in the introduction, an explanation's negative scope (or disconfirmed predictions) counts against an explanation more than its positive scope (or confirmed predictions) counts in its favor (Johnson et al., 2015a, 2016). It could be that when scope is ambiguous, the possibility of disconfirmation looms larger than the possibility of confirmation, leading to a narrow latent scope bias that can be reversed only with very strong inferred positive evidence (as in Experiment 2). We regard this as an interesting direction for future work.

11.2. The adaptive value of inferred evidence

Our participants' judgments violated the laws of probability in striking and consistent ways. Yet, explanation with incomplete evidence is ubiquitous in everyday cognition: Are our inferences really so maladaptive as the violations suggest?

We are often confronted both with too little and too much information—too little in the sense that useful information is often unavailable, and too much in the sense that much of the available information is irrelevant or beyond our computational capacity to analyze. To the extent that we can selectively infer diagnostic evidence, such strategies can assist with both horns of this informational dilemma: We can single out those pieces of evidence for inference that are unavailable from the environment but that are especially diagnostic.

Inferred evidence is commonly used in adaptive ways in perception, such as when people infer contours (Kanizsa, 1976) and continuities of objects (Michotte, Thinès, & Crabbé, 1964), and more generally when we infer the three-dimensional world from a two-dimensional retinal array (Marr, 1982). But such strategies are just as ubiquitous—and usually, just as adaptive—in higher-level cognitive tasks, even though we have focused here on non-normative strategies that people use. For example, if Detective Colombo is trying to distinguish between Professor Plum (who just came from his ivory tower office) and Colonel Mustard (who just came from a muddy battlefield) as culprits, then it is perfectly rational to reason from the observed evidence (e.g., chemical signatures of dirt on the carpet) to inferences rendered likely by that evidence (e.g., the carpet was muddy at the time of the crime), and to use those inferences for distinguishing among perpetrators. Colombo might reason, "I know that there is a positive chemical test for mud, so there was likely to have been mud on the floor at the time of the crime. Since Colonel Mustard had muddy shoes, he is the more likely culprit." This reasoning is perfectly valid—that is, people can safely make inferences from observed evidence, to make educated guesses about other diagnostic evidence. Indeed, this reasoning is more than valid: Such inferences are needed to solve the mind's informational dilemma. Were it not for such reasoning, we would be hopelessly bound to the observed.

Our participants, however, appear to have overgeneralized this ordinarily useful heuristic. Instead of making inferences from one piece of evidence to another, they made inferences from the evidence *base rates* to the evidence itself. They behaved more like Inspector Clouseau, who does not know about the chemical signatures of mud, but does know that the family dog often spreads mud throughout the house. He might reason, "I know that the dog often has muddy paws, so there was likely to have been mud on the floor at the time of the crime. Since Colonel Mustard had muddy shoes, he is the more likely culprit." The error here is subtle, because Clouseau's first inference *is* valid—the carpet probably *was* muddy. The argument goes wrong, however, in failing to recognize this fact as irrelevant to determining the culprit.

In both of these cases, both Colombo and Clouseau correctly inferred an unobservable fact from the information available—the fact that the carpet was probably muddy at the time of the crime. If it came from evidence base rates, it will not be diagnostic after all—and it is participants' failure to recognize this fact that makes their inferences non-normative. It is not the inferred evidence strategy itself, then, but its indiscriminate application that is at fault.

11.3. Implications for theories of explanation

The need to make sense of the world drives much of cognition. Categories allow us to bundle features together coherently to support inference, pragmatic inference allows us to interpret others' utterances, theory of mind allows us to infer others' mental states, and causal reasoning allows us to understand present events in terms of the past. These various sense-making capacities can be referred to, collectively, as *abductive cognition* (Peirce, 1997/1903; see Lombrozo, 2012). To what extent do these abductive faculties rely on distinct psychological mechanisms, and to what extent do they share a common logic? Our own view, consistent with the current results, can be contrasted with two other possible positions.

First, one could take a more fine-grained view of abductive cognition—a view that seems to be implicit in the division of labor of cognitive science (see Danks, 2014 for related discussion). For example, categorization has long been an object of intense scrutiny by the cognitive science community (Murphy, 2002; Smith & Medin, 1981). After waves of research ruled by various theoretical traditions (e.g., the classical view of concepts, prototype theories, exemplar theories), many researchers came to adopt the view that concepts are linked to reasoners' tacit theories (Murphy & Medin, 1985)—that our categorizations of objects are intimately linked to our explanatory models. Similar conclusions have been reached independently in many other abductive domains—in theory of mind (Gopnik & Wellman, 1992), pragmatics (Grice, 1989), causal reasoning (Kelley, 1973), perception (Von Helmholtz, 2005/1867), memory (Bartlett, 1932), and even emotion (Schachter & Singer, 1962). Despite these acknowledged links between sense-making and these various domains, their study has proceeded in relative isolation, signaling little confidence that they share an underlying logic. If these diverse faculties make sense of experience in diverse ways, then abductive cognition is highly fine-grained, justifying the intellectual isolationism of their study.

More recently, a much more general, Bayesian view has emerged. This view captures the key insight that these abductive processes have a common informational structure—inferring hypotheses from observations. Many inferential tasks can be understood as modifying beliefs based on new information according to the normative principles of Bayes' theorem. Rational probabilistic models have been applied to such diverse phenomena as causal reasoning (Griffiths & Tenenbaum, 2005), categorization (Tenenbaum & Griffiths, 2001b), language acquisition (Xu & Tenenbaum, 2007), visual perception (Kersten, Mamassian, & Yuille, 2004), and even motor control (Körding & Wolpert, 2004), speaking to the broad applicability of this framework. Although much of the work in these models comes from the specification of the prior probabilities and the likelihood functions, the inference mechanism always relies on the same Bayesian updating principles—not just a single set of principles across abductive tasks, but a single *principle* across these tasks.

Here, we advocate a third approach that falls between these extremes. Whereas we argue, alongside the Bayesians, that abductive cognition is likely to share a set of common mechanisms, we suspect that they rely more on heuristic machinery rather than normative probabilistic inference as such. For example, people use a simplicity heuristic (Lombrozo, 2007) and a complexity heuristic (Johnson et al., 2014) to approximate normative Bayesian inference in evaluating explanations. Specifically, simpler explanations are assigned higher prior probabilities, whereas more complex explanations are assigned higher likelihoods. On the one hand, these heuristics appear to be used quite generally (in causal explanation and categorization, as well as some visual tasks; Johnson et al., 2014, 2016; Lombrozo, 2007). Yet both principles can lead to illusory inferences, suggesting that they are heuristics rather than emergent principles from normative Bayesian calculations.

The current results contribute to this larger project of understanding the inferential machinery of explanation, and underscore in particular the overlaps between the inferential processes involved in categorization and in causal reasoning. For example, people seem to adopt beliefs in an all-or-none manner in both categorization (Malt, Ross, & Murphy, 1995) and causal diagnosis (Johnson, Merchant, & Keil, 2015b). Teleological or function-based reasoning is widespread in both causal explanation (Lombrozo & Carey, 2006) and classification (German & Johnson, 2002; Lombrozo & Rehder, 2012). And people evaluate both putative categorizations and causes using common principles such as simplicity (Johnson, Kim, & Keil, 2016; Lombrozo, 2007; Pothos & Chater, 2002), diversity (Kim & Keil, 2003; Osherson, Smith, Wilkie, López, & Shafir, 1990), and belief utility (Johnson, Rajeev-Kumar, & Keil, 2015). The current findings further make the case for common inferential processes in categorization and causal reasoning, documenting a non-normative use of inferred evidence consistent across these superficially distinct cognitive processes. Although the differing task demands of causal reasoning and categorization led to different magnitudes of the effect (e.g., wide scope preferences were found in causal reasoning but not in categorization), the underlying mechanism was strikingly similar across these processes. This suggests that, given the abstract similarities in data-to-hypothesis reasoning between categorization and causal reasoning, other sorts of data-to-hypothesis inferences may likewise rely on analogous computations.

Although our approach differs from the Bayesian approach, these two frameworks are not inherently in tension. Bayesian theories are generally posed at the computational level, aiming to characterize the problem that people are solving on the assumption that people solve it in an optimal manner given the laws of probability. Although our view—and the current empirical findings—speak against any theory on which people behave in a fully optimal way in local contexts, heuristic strategies such as inferred evidence can be broadly adaptive, and thus potentially rational from a wider point of view. In fact, Bayesian models have had great success in explaining apparently non-normative behavior, given that participants understand their task differently from the experimenters or are adopting strategies that work at a more global level (e.g., Griffiths & Tenenbaum, 2005; Oaksford & Chater, 2007). We look forward to the possibility that such models might help to clarify the rational basis of the inferred evidence strategy, perhaps building on our own formalization of the reasoning processes involved (see Section 2.1 and Appendix A).

12. Conclusion

Both in science and in everyday life, we must weigh explanations consistent with untested predictions, and we often cannot verify more than a small subset of these predictions. In this sense, *most* explanations are latent scope explanations. Here, we showed that rather than accepting ignorance about diagnostic evidence, people attempt to infer what they would observe if they were able to look. Although it may often be possible to make educated guesses from background knowledge, the present results show that people will also use irrelevant information in the service of inferring evidence: We do not settle for ignorance when apparent truth is within reach.

Acknowledgments

Experiments 1, 3, and 7 were presented at the 36th Annual Meeting of the Cognitive Science Society, and we thank the conference attendees and reviewers for their suggestions. We thank Sunny Khemlani, Greg Murphy, and the members of the Cognition and Development Lab for helpful discussion. This research was funded by grant R37-HD23922 from the National Institute of Health, awarded to F.C. Keil.

Appendix A. Derivation of Eq. (5)

First, we apply Bayes' theorem to calculate the posterior odds of H_L (a cause which leads to X) over H_W (a cause which leads to both X and Z), given that X is observed and Z is not observed (call this state of ignorance I):

$$\frac{P(H_N|X,I)}{P(H_W|X,I)} = \frac{P(H_N)}{P(H_W)} \cdot \frac{P(X,I|H_N)}{P(X,I|H_W)} \tag{A.1}$$

By the causal Markov assumption—a standard assumption in graphical causal models (Pearl, 1988, 2000; Spirtes et al., 1993)—we assume that X and I are conditionally independent, given H_i. This means that knowing about X does not tell us anything further about I (and vice versa), assuming that we know whether H_L or H_W is true. Therefore, the likelihood term can be factorized into:

$$\frac{P(H_N|X,I)}{P(H_W|X,I)} = \frac{P(H_N)}{P(H_W)} \cdot \frac{P(X|H_N)}{P(X|H_W)} \cdot \frac{P(I|H_N)}{P(I|H_W)} \tag{A.2}$$

If we assume that ignorance (I) is equally likely given either hypothesis, then the rightmost term should collapse to 1, and the reasoner would have no bias. (Assuming that this ratio is different from 1 is one way to model pragmatic inferences.)

However, to explore the possibility of inferred evidence, we break down this rightmost likelihood term:

$$\frac{P(I|H_N)}{P(I|H_W)} = \frac{P(I,Z|H_N) + P(I,-Z|H_N)}{P(I,Z|H_W) + P(I,-Z|H_W)} \tag{A.3}$$

We now assume that I is conditionally independent of H_i, given the state of Z. That is, given that Z is stipulated to be either true or false, our ignorance I has no bearing on whether H_1 or H_2 is the correct hypothesis (and vice versa). This seems intuitive, since I is typically relevant only insofar as it helps us to determine whether or not Z is true.[3] This assumption allows us to rewrite the likelihood term for I as:

$$\frac{P(I|H_N)}{P(I|H_W)} = \frac{P(I|Z) \cdot P(Z|H_N) + P(I|-Z) \cdot P(-Z|H_N)}{P(I|Z) \cdot P(Z|H_W) + P(I|-Z) \cdot P(-Z|H_W)} \tag{A.4}$$

By Bayes' theorem:

$$P(I|Z) = \frac{P(Z|I) \cdot P(I)}{P(Z)} \tag{A.5}$$

Inserting this expression into Eq. (A.4) along with the corresponding expression for $-Z$, we find that:

$$\frac{P(I|H_N)}{P(I|H_W)} = \frac{\frac{P(Z|I) \cdot P(I)}{P(Z)} \cdot P(Z|H_N) + \frac{P(-Z|I) \cdot P(I)}{P(-Z)} \cdot P(-Z|H_N)}{\frac{P(Z|I) \cdot P(I)}{P(Z)} \cdot P(Z|H_W) + \frac{P(-Z|I) \cdot P(I)}{P(-Z)} \cdot P(-Z|H_W)} \tag{A.6}$$

Substituting $f^{+Z} = P(Z|I)/P(Z)$ and $f^{-Z} = P(-Z|I)/P(-Z)$ and replacing this likelihood term into Eq. (A.2), we derive the final result, Eq. (5) from the main text:

$$\frac{P(H_N|X,I)}{P(H_W|X,I)} = \frac{P(H_N)}{P(H_W)} \cdot \frac{P(X|H_N)}{P(X|H_W)} \cdot \frac{P(Z|H_N) \cdot f^{+Z} + P(-Z|H_N) \cdot f^{-Z}}{P(Z|H_W) \cdot f^{+Z} + P(-Z|H_W) \cdot f^{-Z}}$$

[3] There are some cases where this assumption might not hold, particularly if one hypothesis implies that the evidence is more likely to be absent than the other (e.g., if one suspect but not another is capable of tampering with the evidence). That said, these sorts of cases reflect a somewhat different causal structure from that considered here, in that the ignorance is itself *evidence* favoring one hypothesis over the other.

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cogpsych.2016.06.004.

References

Bartlett, F. C. (1932). *Remembering: An experimental and social study*. Cambridge, UK: Cambridge University Press.
Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-Stay, Lose-Sample: A simple sequential algorithm for approximating Bayesian inference. *Cognitive Psychology, 74*, 35–65.
Bonawitz, E. B., & Lombrozo, T. (2012). Occam's rattle: Children's use of simplicity and probability to constrain inference. *Developmental Psychology, 48*, 1156–1164.
Bourne, L. E. (1970). Knowing and using concepts. *Psychological Review, 77*, 546–556.
Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3–5.
Danks, D. (2014). *Unifying the mind: Cognitive representations as graphical models*. Cambridge, MA: MIT Press.
Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*, 274–290.
German, T. P., & Johnson, S. C. (2002). Function and the origins of the design stance. *Journal of Cognition and Development, 3*, 279–300.
Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684–704.
Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7*, 145–171.
Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 334–384.
Griffiths, T. L., Vul, E., & Sanborn, A. N. (2012). Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science, 21*, 263–268.
Johnson, S. G. B., Jin, A., & Keil, F. C. (2014). Simplicity and goodness-of-fit in explanation: The case of intuitive curve-fitting. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the Cognitive Science Society* (pp. 701–706). Austin, TX: Cognitive Science Society.
Johnson, S. G. B., Johnston, A. M., Toig, A. E., & Keil, F. C. (2014). Explanatory scope informs causal strength inferences. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the Cognitive Science Society* (pp. 2453–2458). Austin, TX: Cognitive Science Society.
Johnson, S. G. B., Kim, H., & Keil, F. C. (2016). Explanatory biases in social categorization. In *Proceedings of the 38th annual conference of the Cognitive Science Society*. .
Johnson, S. G. B., Merchant, T., & Keil, F. C. (2015a). Argument scope in inductive reasoning: Evidence for an abductive account of induction. 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*. 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., Rajeev-Kumar, G., & Keil, F. C. (2014). Inferred evidence in latent scope explanations. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the Cognitive Science Society* (pp. 707–712). Austin, TX: Cognitive Science Society.
Johnson, S. G. B., Rajeev-Kumar, G., & Keil, F. C. (2015d). Belief utility as an explanatory virtue. 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. 1009–1014). Austin, TX: Cognitive Science Society.
Johnston, A. M., Johnson, S. G. B., Koven, M. L., & Keil, F. C. (2016). Little Bayesians or little Einsteins? Probability and explanatory virtue in children's inferences (in preparation).
Johnston, A. M., Johnson, S. G. B., Koven, M. L., & Keil, F. C. (2015). Probabilistic versus heuristic accounts of explanation in children: Evidence from a latent scope bias. 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*. Austin, TX: Cognitive Science Society.
Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.
Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430–454.
Kanizsa, G. (1976). Subjective contours. *Scientific American, 234*, 48–52.
Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28*, 107–128.
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.
Kim, N. S., & Keil, F. C. (2003). From symptoms to causes: Diversity effects in diagnostic reasoning. *Memory & Cognition, 31*, 155–165.
Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin, 110*, 499–519.
Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature, 427*, 244–247.
Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 585–589.
Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology, 55*, 232–257.

Please cite this article in press as: Johnson, S. G. B., et al. Sense-making under ignorance. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.06.004

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99*, 167–204.

Lombrozo, T. (2012). Explanation and abductive inference. In K. J. Holyoak & R. G. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 260–276). Oxford, UK: Oxford University Press.

Lombrozo, T., & Rehder, B. (2012). Functions in biological kind classification. *Cognitive Psychology, 65*, 457–485.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review, 115*, 955–984.

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

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge, MA: MIT Press.

Michotte, A., Thinès, G., & Crabbé, G. (1964). Les complements amodaux des structures perceptives. In *Studia psychologica*. Leuven, Belgium: Publications Universitaires de Louvain.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289–316.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, UK: Oxford University Press.

Osherson, D., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*, 185–200.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco, CA: Morgan Kaufmann.

Pearl, J. (2000). *Causality: Models, reasoning, and inference* : . Cambridge University Press.

Peirce, C. S. (1997). In P. A. Turrisi (Ed.), *Pragmatism as a principle and method of right thinking: The 1903 Harvard lectures on pragmatism*. Albany, NY: State University of New York Press (Original work published 1903).

Popper, K. (1959). *The logic of scientific discovery*. London, UK: Routledge (Original work published 1934).

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26*, 303–343.

Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology, 65*, 429–447.

Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology, 50*, 264–314.

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.

Samarapungavan, A. (1992). Children's judgments in theory choice tasks: Scientific rationality in childhood. *Cognition, 45*, 1–32.

Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review, 69*, 379–399.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*, 261–267.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Sussman, A. B., Khemlani, S. S., & Oppenheimer, D. M. (2014). Latent scope bias in categorization. *Journal of Experimental Social Psychology, 52*, 1–8.

Tenenbaum, J. B., & Griffiths, T. L. (2001b). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*, 629–640.

Tenenbaum, J. B., & Griffiths, T. L. (2001a). The rational basis of representativeness. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd annual conference of the Cognitive Science Society* (pp. 1036–1041). Mahwah, NJ: Erlbaum.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Von Helmholtz, H. (2005). *Treatise on physiological optics* (Vol. III) Mineola NY: Dover (Original work published 1867).

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*, 245–272.

Please cite this article in press as: Johnson, S. G. B., et al. Sense-making under ignorance. *Cognitive Psychology* (2016), http://dx.doi.org/10.1016/j.cogpsych.2016.06.004