Available online at www.sciencedirect.com

ScienceDirect

Cognitive Psychology 55 (2007) 232–257

Cognitive Psychology

www.elsevier.com/locate/cogpsych

Simplicity and probability in causal explanation

Tania Lombrozo [*]

Department of Psychology, University of California, Berkeley, 3210 Tolman Hall, Berkeley, CA 94720, USA

Accepted 29 September 2006
Available online 9 November 2006

Abstract

What makes some explanations better than others? This paper explores the roles of simplicity and probability in evaluating competing causal explanations. Four experiments investigate the hypothesis that simpler explanations are judged both better and more likely to be true. In all experiments, simplicity is quantified as the number of causes invoked in an explanation, with fewer causes corresponding to a simpler explanation. Experiment 1 confirms that all else being equal, both simpler and more probable explanations are preferred. Experiments 2 and 3 examine how explanations are evaluated when simplicity and probability compete. The data suggest that simpler explanations are assigned a higher prior probability, with the consequence that disproportionate probabilistic evidence is required before a complex explanation will be favored over a simpler alternative. Moreover, committing to a simple but unlikely explanation can lead to systematic overestimation of the prevalence of the cause invoked in the simple explanation. Finally, Experiment 4 finds that the preference for simpler explanations can be overcome when probability information unambiguously supports a complex explanation over a simpler alternative. Collectively, these findings suggest that simplicity is used as a basis for evaluating explanations and for assigning prior probabilities when unambiguous probability information is absent. More broadly, evaluating explanations may operate as a mechanism for generating estimates of subjective probability.

Keywords: Causal explanation; Simplicity; Inference to the best explanation; Subjective probability

[*] Fax: +1 510 642 5293.
E-mail address: lombrozo@berkeley.edu

0010-0285/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2006.09.006

1. Introduction

In everyday life as in science, data are inevitably consistent with multiple explanations. Does Mercury trace epicycles around the earth or follow an elliptical orbit around the sun? Is Hamlet's behavior due to love-sickness, insanity, or a sinister plot to avenge his father's death? Because the true state of the world is underdetermined, selecting the best explanation requires more than consistency with data. Sherlock Holmes, a master of underdetermined inference, advised that to evaluate explanations we "balance probabilities and choose the most likely" (Doyle, 1986b, p. 30). In the spirit of the rational detective, people may likewise evaluate explanations by comparing probabilities and choosing the most likely. But unfortunately for Holmes and the rest of us, explanations rarely come equipped with probabilities—even in fiction. Evaluating explanations therefore requires either a mechanism for generating probabilities or a non-probabilistic basis for selecting the best explanation.

Several scientists and scholars have advocated simplicity as a basis for evaluating explanations. In what has come to be known as Occam's Razor, William of Occam suggested that the number of entities invoked in an explanation should not be multiplied beyond necessity (Baker, 2004). Sir Isaac Newton described a similar maxim in the *Principia*, writing that "we are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances" (Newton, 1953/1686). These endorsements of simplicity illustrate a strategy that philosophers call "inference to the best explanation" (Harman, 1965; Lipton, 2002; Peirce, 1998): when multiple explanations are possible, choose the one that (if true) would best explain the evidence at hand. If simpler explanations are better explanations, then (all else being equal) one should select the simplest explanation.

Although simplicity can be evaluated in the absence of information about probability, simplicity and probability are intimately related. Newton advocated simplicity precisely because he believed simpler explanations were more probable, an assumption that stemmed from his belief that "nature is pleased with simplicity, and affects not the pomp of superfluous causes" (Newton, 1953/1686). Formal analyses of simplicity in philosophy, statistics and computer science likewise suggest that simpler explanations should be accorded higher probability, but where Newton turned to metaphysics, contemporary scholars consider the properties of probabilistic inference (e.g. Forster, 2000). In particular, complex hypotheses run the risk of fitting aspects of the data that result from noise or idiosyncratic properties of the data points that happened to be sampled. As a result, complex hypotheses may fit observed data very closely, but generalize to novel data more poorly than simpler alternatives. Formal metrics for simplicity, such as Minimum Description Length (Rissanen, 1978), Bayesian Occam's Razor (Jeffreys & Berger, 1992) and the Akaika information criterion (Forster, 2000), address this problem by assigning simpler hypotheses a higher prior probability—the probability assigned to a hypothesis before data has been observed. Once data is observed these probabilities are updated, so while simplicity and probability may correspond in the absence of data, complex hypothesis can be deemed more probable than simple alternatives as data accumulates.

Recent work in psychology supports the psychological reality of a preference for simplicity, as well as a role for simplicity in probabilistic inference. Chater (1996), for example, advocates a simplicity metric known as Kolmogorov Complexity, according to which simplicity is equivalent to being producible by a short program for a universal Turing machine (Li & Vitanyi, 1997). He shows that adopting this notion of simplicity implies a correspondence between simplicity and probability, and argues that this correspondence may hold for perception (but see van der Helm, 2000). Bayesian models of category learning (Feldman, 2000; Griffiths, Christian, & Kalish, 2006) and causal induction (Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2006) likewise provide support for the claim that simpler hypotheses are accorded a higher probability. More broadly, simplicity has been invoked in explanations of object perception (Leeuwenberg & Boselie, 1988; Pomerantz & Kubovy, 1986), concept learning (Shepard, Hovland, & Jenkins, 1961), cognitive development (Andrews & Halford, 2002), and perceived phrase structure in language and music (Bod, 2002). Chater and Vitanyi (2003) advance the more ambitious proposal that simplicity may be an organizing principle for cognitive science.

Despite widespread interest in simplicity, few studies have examined the role of simplicity in the context that originally motivated Occam and Newton: the evaluation of explanations. In particular, there is only indirect support for the claim that simpler explanations are preferred and that this preference guides probabilistic judgments. The findings from Feldman (2000), Griffiths et al. (2006), and Lu et al. (2006) suggest that human judgments conform to the predictions of a Bayesian model with prior probabilities that favor simpler hypotheses, but in these tasks participants are not asked to judge explanations. Instead they make inferences that involve the evaluation of hypotheses that may be represented only implicitly. Moreover, the formal metrics for simplicity used in these models are difficult to apply in the messy, real-word contexts that characterize everyday cognition. These studies leave open whether simplicity guides judgments when explanations are explicitly compared, and how simplicity and probability are related in such judgments.

Previous work on simplicity in explanation confirms that both simple and probable explanations are valued (Einhorn & Hogarth, 1986; Thagard, 1989), but fails to address the relationship between simplicity and probability explored by philosophers and statisticians. Read and Marcus-Newhall (1993), for example, examined whether people prefer explanations that involve fewer propositions, a prediction of Thagard's Theory of Explanatory Coherence (Thagard, 1989). Participants were given multiple pieces of data (e.g. that Cheryl has felt tired, gained weight, and had an upset stomach), and asked to evaluate a simple explanation (that Cheryl is pregnant) and a more complex one (that Cheryl has mononucleosis, has stopped exercising, and has a virus). Whether asked to judge the explanations' probability or "goodness," participants rated the simple explanation significantly higher than the complex alternative. Although this finding suggests that simpler explanations are favored, it cannot rule out the possibility that participants were responding purely on the basis of probability. On the assumption that having mononucleosis, stopping exercise, having a virus, and being pregnant are approximately equally likely, the conjunction of the first three cannot exceed the probability of the fourth. Indeed, pregnancy could be much less likely, and still have greater probability than the conjunction of *all three* of the preceding explanations.[1]

To examine whether simpler explanations are preferred independently from probabilistic assumptions like those that may have operated in the Read and Marcus-Newhall (1993) experiment, simplicity and probability information would need to be unconfounded by

[1] Read and Marcus-Newhall (1993) happened to collect the relevant data for examining this hypothesis in testing a related hypothesis about explanatory breadth. Specifically, they collected the subjective probabilities assigned to each of the individual hypotheses (e.g. having mononucleosis). For all stimulus sets, the probability for the complex explanation (obtained by multiplying the probabilities of the three components) was in fact lower than the probability assigned by participants to the simple explanation.

providing participants with independent evidence for the probability of each explanation. Lagnado (1994) did precisely this. Participants indicated their preference for one of two explanations for a patient's symptoms: that the patient had a single disease D_1 (which can cause both symptoms) or that the patient had both diseases D_2 and D_3 (which can each cause one of the symptoms). Participants were also provided with the probability of contracting disease D_1 or both diseases D_2 and D_3. Lagnado (1994) found that most participants preferred the simpler explanation when it was more probable, found the explanations equally good when described as equally probable, and preferred the complex explanation when told that the probability of contracting both D_2 and D_3 was greater than the probability of contracting D_1. While these data suggest that simplicity is not valued once an explanation's probability is provided, Lagnado acknowledges a serious methodological problem that prevents strong conclusions from being drawn. Participants were 20 computer science graduate students who saw the medical diagnosis problem with probability information as a within-subjects manipulation. In other words, each participant was asked to consider the same problem when the simpler hypothesis was more probable, when the hypotheses were equally probable, and when the complex hypothesis was more probable. This procedure likely made the probability manipulation transparent, and encouraged a probability-savvy population to respond on the basis of probability.

The current paper explores the roles of simplicity and probability in explanation evaluation, using a task similar to Read and Marcus-Newhall (1993) and Lagnado (1994). Four experiments explore whether simpler explanations are preferred when participants are explicitly engaged in evaluating explanations, as well as the relationship between simplicity and probability. Experiment 1 confirms that, all else being equal, people prefer more probable explanations, as well as explanations that are simpler in the sense of invoking fewer causes. Experiment 2 examines how information about simplicity and probability trade-off when a complex explanation has more probabilistic evidence than a simple alternative. In particular, the experiment examines whether simplicity informs judgments by elevating the prior probability assigned to simpler explanations. Experiment 3 replicates Experiment 2 in a different format, and also examines whether simplicity influences an explicitly probabilistic judgment: the perceived frequency of a cause invoked in explanation. Finally, Experiment 4 examines whether the tendency to privilege simpler explanations persists when probabilistic evidence leaves no room to doubt that a complex explanation is most probable. The pattern of findings is used to support the claim that simpler explanations are preferred, and that this preference is used as a basis for probabilistic judgments.

2. Experiment 1: Simplicity and probability as explanatory virtues

Experiment 1 was designed to confirm the intuition that simplicity and probability are indeed explanatory virtues. Following Newton, simplicity was defined in terms of the number of causes invoked in an explanation, but this metric is appealing for reasons beyond Newton's endorsement. Simplicity so defined is easy to quantify, and number comparisons are straightforward to evaluate. In addition, an abundance of evidence supports the idea that causal considerations are psychologically salient, especially in explanation (Lombrozo & Carey, 2006; Lombrozo, 2006). People clearly distinguish between causes and effects (e.g. Waldmann, 1996) and are sensitive to causal order for tasks like categorization (Ahn, Kim, Lassaline, & Dennis, 2000; Rehder, 2003a, 2003b). People are also sensitive to the presence of multiple possible causes, for example controlling for confounds in causal induction tasks (Spellman, 1996a, 1996b; Spellman, Price, & Logan, 2001). For these reasons, number of causes is a psychologically plausible metric for simplicity of potential relevance to other cognitive tasks.

Experiment 1 thus examines whether explanations involving fewer causes are deemed more satisfying in the absence of probability information, and whether explanations involving causes that are more probable are deemed more satisfying in the absence of a simplicity difference. Satisfaction was chosen as the relevant dimension to emphasize to participants that the judgment was a subjective, psychological one. Phrases such as "most likely" were avoided, as they might prejudge the issue of what criterion to apply in evaluating explanations. Experiment 1 additionally explores whether participants justify a preference for simpler explanations by appeal to simplicity, probability, or something else.

2.1. Methods

2.1.1. Participants

Forty-eight undergraduates and summer school students (44% male; mean age = 21, *SD* = 3) from an elite university participated in exchange for course credit, payment, or a small gift.

2.1.2. Materials

Experimental materials consisted of a two-page questionnaire. On the first page, participants read about a fictional alien planet, along with an alien's symptoms and possible causes. They were then asked to select the most satisfying explanation for the alien's symptoms, and permitted to choose among various possibilities. The information provided varied as a function of condition. In the *simplicity* condition, participants read a scenario like the following:

> There is a population of 750 aliens that lives on planet Zorg. You are a doctor trying to understand an alien's medical problem. The alien, Treda, has two symptoms: Treda's *minttels are sore* and Treda has developed *purple spots.*
>
> *Tritchet's syndrome* always causes both *sore minttels* and *purple spots.*
>
> *Morad's disease* always causes *sore minttels*, but the disease never causes *purple spots.*
>
> When an alien has a *Humel infection*, that alien will always develop *purple spots*, but the infection will never cause *sore minttels.*
>
> Nothing else is known to cause an alien's *minttels to be sore* or the development of *purple spots.*
>
> What do you think is the most satisfying explanation for the symptoms that Treda is exhibiting?

(A) Treda the alien has *Tritchet's syndrome.*
(B) Treda the alien has *Morad's disease.*
(C) Treda the alien has a *Humel infection.*
(D) Treda the alien has *Tritchet's syndrome* and *Morad's disease.*
(E) Treda the alien has *Tritchet's syndrome* and a *Humel infection.*
(F) Treda the alien has *Morad's disease* and a *Humel infection.*

The symptoms could be explained by appeal to a single disease (in this case, Tritchet's syndrome) or two diseases (e.g. Morad's disease and a Humel infection). Thus the target explanations (A vs. F) varied in simplicity, but no probability information was provided. Note that D and E also account for both symptoms, but subsume A which is itself sufficient.

In the *probability* condition, participants read a scenario like the following:

> There is a population of 750 aliens that lives on planet Zorg. You are a doctor trying to understand an alien's medical problem. The alien, Treda, has two symptoms: Treda's *minttels are sore* and Treda has developed *purple spots.*
>
> *Tritchet's syndrome* always causes both *sore minttels* and *purple spots.* You know that *Tritchet's syndrome* is present in about 50 of the aliens on Zorg.
>
> *Morad's disease* also always causes both *sore minttels* and *purple spots. Morad's disease* is present in about 73 of the aliens on Zorg.
>
> Nothing else is known to cause an alien's *minttels to be sore* or the development of *purple spots.*
>
> What do you think is the most satisfying explanation for the symptoms that Treda is exhibiting?
>
> (A) Treda the alien has *Morad's disease.*
> (B) Treda the alien has *Tritchet's syndrome.*
> (C) Treda the alien has *Tritchet's syndrome* and *Morad's disease.*

Either disease could account for the symptoms, but one was more probable than the other. Thus the level of simplicity was constant across the target explanations (A vs. B), but probability varied. On the second page of the questionnaire, participants in both conditions were asked to justify their explanation choice from page one.

In constructing the alien medical scenarios, four different sets of symptoms were employed, all involving one recognizable symptom (e.g. purple spots) and one "blank" symptom (e.g. sore minttels) to reduce the extent to which participant's could employ prior beliefs in judging whether the symptoms might have a common etiology.

2.1.3. Design and procedure

Participants were randomly assigned to the *simplicity* condition or the *probability* condition. The order in which diseases were presented, the name of the disease accounting for both symptoms, and the specific symptom pair mentioned were counterbalanced. The candidate answers to the why-question were presented in one of several random orders.

2.2. Results

When asked to choose between a one- or two-cause explanation for a set of symptoms, the overwhelming majority of participants (96%) selected the simpler, one-cause explanation. This preference was significantly different from choosing one of the six available options randomly (Binomial test, $p < .01$) or choosing between the one- and two-cause explanations randomly (Binomial test, $p < .01$). When asked to explain a set of symptoms by appeal to a more or less common disease, 92% of participants chose the more probable explanation. This preference was significantly different from choosing one of the three available options randomly (Binomial test, $p < .01$) or choosing between the more and less probable disease explanations randomly (Binomial test, $p < .01$).

To analyze participants' justifications for their explanation choices, the responses were coded into one of four categories. Justifications were coded in the *simple* category if simplicity was directly or indirectly invoked. Participants most frequently invoked simplicity directly, but some participants noted that the one-cause explanation was "prettier" or "less complex." Justifications were coded in the *sufficient* category if participants suggested that the single disease was sufficient to account for both symptoms, without explicitly invoking simplicity or a related concept. These responses were likely motivated by the idea that if a simple explanation is sufficient a more complex one is unnecessary, but they were distinguished from *simple* justification because the reliance on simplicity was implicit. Justifications were coded as *probable* if the participant wrote that the chosen explanation seemed more likely to be true. Finally, explanations were coded as *other* if they failed to conform to one of the previous categories. Two coders coded all 48 justifications, with an inter-coder agreement of 96%. Disagreements were resolved by discussion. Only four justifications were categorized as *other*.

Among the 96% of participants in the simplicity condition who chose the simpler explanation, 17% justified their choice by appeal to simplicity, 39% by appeal to sufficiency, and 39% by appeal to probability. Those in the latter category generally noted that it seemed more likely that the alien would have one rather than two diseases. In the probability condition, all participants who chose the more probable explanation justified their choice by appeal to probability.

2.3. Discussion

Experiment 1 confirmed the intuitive predictions that in the absence of probability information people prefer simpler explanations, and in the absence of a simplicity difference people prefer more probable explanations. Participants' justifications also shed light on the basis for their preference: 44 of 48 participants justified their choice of explanation by appeal to either probability or simplicity (as reflected in a *probable*, *simple*, or *sufficient* justification). These patterns of justification indicate that simplicity and probability both influence an explanation's "goodness," but cannot establish whether simplicity leads to better explanations because simpler explanations are judged to be more probable, or because simplicity is an end in itself. Many participants were explicit in justifying the choice of a simpler explanation by appeal to probability, suggesting that simplicity may have been used as a cue to probability. However, an alternative possibility suggested in Lagnado (1994) is that participants assumed the diseases have comparable baserates, in which case the single disease explanation is genuinely more likely.[2] Participants in Experi-

[2] This assumes that the probability of contracting a disease is less than .5; otherwise, contracting two would be more likely than contracting one. This is a reasonable assumption about diseases, but might not hold in other cases.

ment 2 were provided with the baserate for each disease, which allows this possibility to be evaluated.

3. Experiment 2: Simplicity versus probability

In Experiment 1, participants exhibited a preference for simplicity in the absence of probability information. To examine whether the preference results from participants' inferring equal baserates for the three diseases, participants in Experiment 2 were provided with the baserate of each disease. Specifically, participants saw stimuli like those from the *simplicity* condition in Experiment 1, but with disease baserate information presented as in the *probability* condition. Using these baserates, participants could in principle compare the probability that an alien with both symptoms has a single disease that causes both symptoms (call it D_1) to the probability that the alien has two diseases that each cause one symptom (call them D_2 and D_3). In some cases, the complex explanation will be more likely.

Examining how participants trade-off simplicity and probability can also distinguish approaches to explanation evaluation. Non-probabilistic approaches like Thagard (1989, 2000) and Read and Marcus-Newhall (1993) predict that participants will always select the simpler explanation, regardless of the disease baserates. In contrast, probabilistic approaches like Lagnado (1994) predict sensitivity to baserates, with the simpler explanation selected only when it is in fact more probable. A final possibility is that participants will respond to both simplicity and probability. This is what might be expected if simpler explanations are accorded a higher prior probability.

3.1. Methods

3.1.1. Participants

Participants were 144 undergraduates and summer school students from an elite university (60% female; mean age = 22, *SD* = 5) who completed the study in exchange for course credit, payment, or a small gift. An additional three participants were replaced for leaving questions blank, and four participants were replaced for reasons explained in the Results section.

3.1.2. Materials

Experimental materials consisted of a two-page questionnaire. On the first page, participants read a scenario like that in the *simplicity* condition from Experiment 1, but with the baserate for each disease provided. For example, after being told "Tritchet's syndrome always causes both sore minttels and purple spots," participants were also told "You know that Tritchet's syndrome is present in about 50 of the aliens on Zorg." The baserate for each disease varied, with a total of eight sets of baserates employed. Assuming the two diseases causing a single symptom are probabilistically independent,[3] the probability of having two diseases can easily be computed by multiplying the

[3] An additional experiment was conducted to verify that participants believe contracting two diseases is probabilistically independent. Twenty undergraduates from an elite university responded to the following question: "Suppose there are two diseases, D_1 and D_2. Do you think someone who has D_1 is more or less likely to have D_2 than someone who does not have D_1? Circle one: More/Less." Participants were evenly split, with 50% selecting "more" and 50% selecting "less."

Table 1
Disease frequencies from Experiment 2

$P(D_1):P(D_2 \text{ and } D_3)$	D_1	D_2	D_3
15:1	50	50	50
1:1	50	197	190
9:10	50	195	214
4:5	50	225	210
2:3	50	250	220
1:2	50	268	280
1:3	50	330	340
1:10	50	610	620

For each probability ratio, the corresponding number of aliens with each disease (from a population of 750) is indicated. Each probability ratio corresponds to a single baserate condition.

corresponding probabilities (e.g. 225/750 * 210/750).[4] Based on this value, the eight baserate conditions can be characterized in terms of the corresponding probability ratios for having the disease causing both symptoms compared to having the remaining two diseases. The baserates and probability ratios are presented in Table 1. There were three different sets of symptoms employed in the alien scenario, generated as in Experiment 1.

On the second page of the questionnaire, participants were asked to explain why they chose the explanation they did on the first page. In addition, they were asked to complete a math problem that required computing a joint probability to ensure that participants knew how to compute joint probabilities.

3.1.3. Design and procedure

Participants were randomly assigned to *baserate* condition, making baserate a between-subjects factor with 18 participants per condition. The order in which diseases were presented, the name of the disease accounting for both symptoms, and the specific symptom pairs mentioned were counterbalanced. The candidate answers to the why-question were presented in one of several random orders.

3.2. Results

3.2.1. Explanation choices

To characterize how participants traded off simplicity and probability, the percentage of participants selecting the single disease corresponding to the simpler explanation, call it D_1, was examined as a function of baserate condition (see Fig. 1). However, such an analysis treats all non-D_1 explanations equivalently, when in fact they are not. In addition to the most reasonable two-cause alternative (D_2 and D_3), there are two one-cause explanations that fail to account for the symptoms (D_2 alone or D_3 alone) and two two-cause explanations that overdetermine the symptoms (D_1 and D_2, D_1 and D_3). Some participants chose

[4] Note that as computed, this probability includes cases in which the alien may also have the third disease. Probabilities were computed in this way because they correspond to the natural interpretation of the explanation choices (i.e. “Tritchet’s syndrome” means “Tritchet’s syndrome and no commitment to the other diseases,” not “Tritchet’s syndrome and definitely not the other diseases”) as confirmed by participants’ interpretations, but also because this provides a more conservative test of the influence of simplicity.

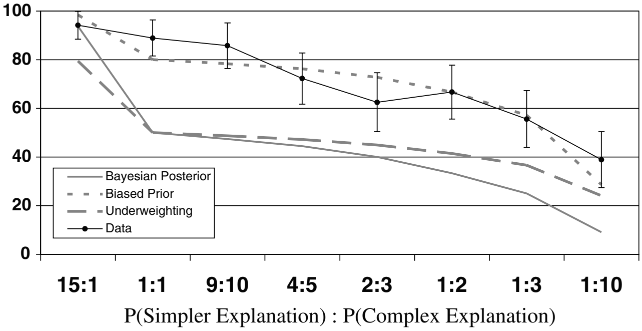

Fig. 1. Predictions and data from Experiment 2. The *x*-axis corresponds to the probability ratio of the simpler explanation to the more complex alternative. The *y*-axis corresponds to the percentage of participants selecting the simpler explanation. The data are indicated with black circles; the solid gray line presents the values that would be expected if the percentage of participants selecting the simpler explanation corresponded to the Bayesian posterior probability for the simpler explanation at the corresponding probability ratio ("Bayesian posterior"). Also illustrated are predictions corresponding to two ways in which the Bayesian calculation might be biased: by involving a prior probability favoring simpler explanations ("biased prior") or by under-weighting the relevance of the provided probability information ("underweighting").

these explanations (four total), but it was clear from their justifications that they intended either D_1 or D_2 and D_3. For example, one participant selected D_2 and justified the choice by claiming that only D_2 accounts for both symptoms. In no case in this or the following experiments did a participant select one of these alternatives and justify the choice in a way that suggested they intended their selection. While this in itself is an interesting finding, participants who did not select D_1 or D_2 and D_3 were replaced in this and the subsequent experiments to simplify analysis. Thus for the data presented, the participants not selecting D_1 are coextensive with those selecting D_2 and D_3.

Overall, 71% of participants chose the simpler explanation, and the remaining 29% the complex alternative. However, a 2 × 8 χ^2 test of independence revealed a significant effect of *baserate* condition on explanation choice ($\chi^2(7) = 25$, $p < .01$). As illustrated in Fig. 1, increasingly fewer participants selected the simpler explanation as the corresponding disease (D_1) became less probable than the two-cause alternative (D_2 and D_3). This suggests that probability indeed informed participants' choices, but was not immune to the simplicity difference. Even when D_2 and D_3 was 10 times more likely than D_1, over a third of participants judged D_1 more satisfying.

A logistic regression analysis was conducted to examine the hypothesis that simplicity informs judgments by establishing the prior probability assigned to explanations. The natural log of the probability ratio was used as a predictor for the percentage of participants choosing the simpler explanation in each *baserate* condition, as this choice results in a straightforward interpretation of the regression parameters. To understand why, it helps to consider how these parameters relate to the computations that would be performed by an idealized Bayesian agent. In this task, the ideal agent's data would result in a slope parameter of 1 and a constant of 0, as indicated in the "Bayesian Posterior" curve on Fig. 1. A non-ideal agent could have a bias in favor of simplicity at either of two stages in the inference process, each corresponding to a parameter of the logistic function. A slope significantly less than 1, illustrated with the dashed line, would suggest that the agent underweights the importance of probability: as evidence in favor of D_2 and D_3 accumulates, the agent fails to reduce the probability of choosing D_1 accordingly. In contrast, a constant significantly greater than zero, illustrated by the dotted line, reflects a bias at the level of the prior probability. The non-ideal agent could overweight, underweight, or appropriately weight probability information, but starts out with disproportionate confidence that D_1 is true.

The regression analysis resulted in a constant significantly different from zero ($\beta = -1.306$, $SE = .248$, $p < .01$), but a slope not significantly different from 1 ($\beta = .795$, $SE = .217$, $p = .8$). The slope does not rule out a correspondence to that of the ideal agent, suggesting that as a group, participants incorporate probability information appropriately: it was not over- or under-weighted. To understand the constant, which does deviate from the ideal agent, the preference for simpler explanations can be translated into a prior probability, thus considering simplicity and probability on the same metric. As a group, participants judged the simpler explanation more likely than the complex alternative by a factor of about 4 (2.3 to 6, .95 confidence interval), and this belief influenced what would be the prior probability in a Bayesian computation. When the probability ratio was 1:2, the percentage of participants choosing the simpler explanation corresponded to the ideal Bayesian's posterior probability for D_1 at a frequency of 1:(2/4), and so on for the other values. This behavior reflects a prior probability of about 79% (69–86%, .95 confidence interval). As a result, participants required disproportionate evidence in favor of the complex explanation before it rivaled the simpler alternative.

3.2.2. Explanation choice justifications

Justifications for participants' explanation choices were coded into one of five categories. The *simple*, *sufficient*, and *probable* categories were coded as in Experiment 1. An additional category, *misunderstood*, was added for this experiment. Justifications were coded as *misunderstood* if they suggested that the participant made an explanation choice based on some misunderstanding of the information they were provided.[5] In particular, some participants misunderstood statements about a disease *not causing* a symptom to mean that it *precluded* the appearance of that symptom by another disease. For example, one participant justified the choice of the simpler explanation by stating that "the information given negated the possibility of either Humel or Tritchet's syndrome." As before, explanations were coded as *other* if they failed to conform to one of the previous categories. Two coders coded all 144 justifications, with an inter-coder agreement of 94%. Disagreements were resolved by discussion.

Overall, 15% of participants justified their explanation choice by appeal to simplicity, 13% by appeal to sufficiency, and 51% by appeal to probability. Nine percent provided *other* justifications, and 12% potentially misunderstood. To analyze justifications as a function of explanation choice and *baserate* condition, the data were collapsed across neighboring baserates to increase the number of participants in otherwise small,

[5] There were a total of 17 participants classified as *misunderstood*. Because this number is non-negligible, the logistic regression analysis detailed above was repeated with these participants excluded. The resulting slope ($\beta = .915$, $SE = .246$) and constant ($\beta = -1.164$, $SE = .261$) parameters were not significantly different from those in the analysis already reported. These values correspond to use of probability information that is slightly closer to that of an ideal agent with a prior probability of 76% (66–84%, .95 confidence interval) for the simpler explanation.

and hence highly variable, categories. *Simple*, *sufficient*, and *misunderstood* justifications were only generated by participants who chose the simpler explanation, and their frequency did not vary as a function of baserate condition (χ^2-tests of independence, $p > .2$). In contrast, *probable* justifications were invoked for both the simpler and more complex explanation, and the number of *probable* justifications did change significantly as a function of *baserate* condition and explanation choice (χ^2 tests of independence, $p < .01$). Specifically, probability was invoked increasingly rarely for the simple explanation as it became less likely, while justifications that invoked probability became increasingly frequent for the two-cause explanation as it became more likely. Interestingly, the total number of appeals to probability did not change as a function of *baserate* condition when the data were collapsed across explanation choice. In other words, for each pair of baserates, about half of participants (15–20 out of 36) appealed to probability, but the explanation they used probability to support changed as a function of the probability ratio. This suggests that for these participants (51% total), probability was explicitly the relevant criterion for explanation evaluation, but simplicity informed their evaluation of probability.

3.2.3. Explanation choice and math ability

A majority of participants (69%) correctly answered the math problem. This percentage of correct choices was significantly different from chance responding ($p < .01$), and did not differ significantly across *baserate* conditions ($\chi^2(3) = .614, p = .893$). There was a small but non-significant correlation between answering the math problem correctly and choosing the two-cause explanation on the disease explanation problem ($r = .094, p = .263$); the correlation was comparable when only considering cases in which the two-cause explanation was more probable ($r = .15, p = .122$). This suggests that a failure to use probability information for the explanation problem is not the result of mathematical ignorance: participants who thought computing the joint probability was relevant would have known how to do so.

3.3. Discussion

Experiment 2 revealed that as a group, participants consider probability when choosing among alternative explanations, but are also influenced by simplicity. More precisely, participants incorporate probability information appropriately, but do so over prior probabilities that favor the simpler explanation over the alternatives by a factor of about four. As a result, disproportionate evidence in favor of the complex explanation was required for a majority of participants to select it over the simpler alternative. These findings suggest that a preference for simpler explanations is not due only to assumptions about the baserates of diseases, but rather to the higher prior probability assigned to simpler explanations. The influence of both simplicity and probability on explanation judgments departs from the predictions and findings of Thagard's Theory of Explamatory Coherence (Thagard, 1989) as well as probabilistic approaches like Lagnado (1994).

Conceptually this task is almost identical to that reported in Lagnado (1994) and described in the introduction. However, Lagnado found that none of his participants preferred a simple explanation when it was less probable than a complex alternative. The methodological differences between Lagnado's task and the one reported here may account for the greater tendency among his participants to respond on the basis of probability. In Lagnado's task each participant saw several examples with different baserates, with the consequence that the probability manipulation was relatively transparent. In the present experiment each participant saw a single scenario with a single set of baserates, so the probability manipulation was more opaque. And unlike Lagnado (1994), participants were not given the joint probability of D_2 and D_3, nor explicitly told that the conjunction of D_2 and D_3 was more likely than D_1 in the conditions in which it was. This information was omitted for two reasons. First, explicitly indicating these probabilities could establish task demands encouraging responses in terms of probability. Rather than having participants try to infer whether the experiment was about simplicity or probability, the goal was to have them use the information they found natural in informing their judgment. Second, joint probabilities are often absent in real-world decision-making contexts. Uncertain or incorrect joint probability estimates may be one reason why simplicity is used as a method for evaluating explanations in the first place (this hypothesis is explored in Experiment 4). Thus the current methods may provide a more naturalistic picture of people's reliance on simplicity and probability in explanation evaluation.

4. Experiment 3: Probabilistic consequences of simplicity

The results from Experiments 2 support the idea that simplicity and probability jointly influence explanation evaluation. In particular, simplicity may influence the prior probability that an explanation is true, with probability information in the form of empirical frequencies influencing the posterior probability of the explanations under consideration. Participants' justifications for their explanation choices also provide tentative support for the idea that simplicity is used as a cue to probability. For almost every probability ratio, a subset of participants who chose the simple explanation justified their choice by appeal to probability.

Experiment 3 replicates Experiment 2, but introduces an important methodological modification and an explicitly probabilistic judgment. Rather than presenting participants with numbers on paper—a format that lends itself to being approached as a math problem—participants "experience" the frequencies of various diseases by seeing many individual cases. Because participants are never presented with a numerical value corresponding to the frequency of each disease, they can also be asked to estimate the frequency of each disease after making an explanation choice. If people think simpler explanations are more likely to be true, they may systematically misremember the frequencies of causes to accommodate this belief.

Past research suggests that if simpler explanations are preferred and judged more probable, simplicity could influence the interpretation or representation of probability information. For example, Chapman (1967) and Chapman and Chapman (1967) found that participants reported illusory correlations between diagnostic signs and psychiatric conditions (e.g. reporting sexual content in Rorschachs and being homosexual), but only for sign-condition pairs that conformed to prior beliefs and hence could be explained. A variety of related studies find that prior beliefs can influence the evaluation of evidence (e.g. Chinn & Brewer, 1993; Fugelsang & Thompson, 2003; Koehler, 1993), suggesting that simplicity should be no exception: if people believe simpler explanations are better or more probable, then they may misinterpret or misremember probability information in a way that favors simpler explanations.

4.1. Methods

4.1.1. Participants

Participants were 108 undergraduate and summer school students from an elite university (56% female, mean age = 22, $SD = 7$) who completed the study in exchange for course credit or a small gift. Fourteen additional participants were excluded: two for failing to follow directions (they circled more than one answer when a single one was requested), two due to experimental error (they were presented with the wrong frequency information for their condition), and 10 for misunderstanding a crucial part of the information. Misunderstandings were identified by participants' explanation choice justifications as in Experiment 2, with a single judge coding the justification as *misunderstood* being sufficient grounds for exclusion. Participants who potentially misunderstood the task were eliminated because it is unclear whether the basis for their explanation preference was truly simplicity.

4.1.2. Materials

Thirty-six distinct computer presentations were created using Microsoft PowerPoint. Each presentation began with the following introduction and instructions:

> Welcome to planet Zorg! You are a doctor coordinating the research efforts of a small team investigating three diseases found among the aliens of planet Zorg. These diseases are Pilt's disease, Stemmel's Disease, and Brom's Disease.

Participants were then guided through information about which diseases cause which symptoms, as in Experiment 2. They were additionally given a response sheet that included this information, so remembering the relations between diseases and symptoms was not necessary. After this information, they read the following paragraph, designed to encourage participants to attend to the upcoming frequency information and suggest that the sampling was representative of the entire population:

> In addition to this information about symptoms, your research team has developed a test for the presence of each medical problem. The tests will help establish the prevalence of each disease in the population. Specifically, you hope to determine which diseases are rare and which are common, so knowing their frequencies is important.

This was followed by a description of how to identify a positive test for the first disease (one of: a red dot on an alien's forehead rather than a blue one; purple smoke coming out of a test-tube as opposed to no smoke; yellow eyes when a light was flashed versus black eyes). Participants then saw 10 screens, containing a total of 75 aliens (4–11 per screen), some of which were indicated as having the disease being tested according to the diagnostic sign previously introduced. Each screen was presented for two seconds with a one-second gap between screens. At the conclusion of the 10 screens, the diagnostic test for the second disease was introduced, and the process of providing 75 examples repeated for that disease. This was repeated for the third disease as well. Each disease involved a different diagnostic sign.

After the presentation of frequency information, participants read about Treda: "Treda the alien has two symptoms: smelly skin and purple plickets. What do you think is the most satisfying explanation for Treda's symptoms?" They were asked to circle one of six options: each disease individually and each pairwise combination. They were then asked to justify their choice. In the final part of the experiment, they were asked to estimate the percentage of the Zorg population with each of the three diseases.

4.1.3. Design and procedures

Participants were assigned randomly to one of the 36 computer presentations, with a total of three seeing each presentation in individual sessions. There were four *baserate* conditions with 27 participants each. The *baserate* conditions corresponded to the same frequency information as in Experiment 2 (out of 75 rather than 750 aliens) at the following ratios: 15:1, 9:10, 1:2, and 1:10. The order of the diseases, including the introduction with symptoms, the presentation of the frequencies, and the request for frequency estimates, was in a consistent order for each participant. There were a total of three orders, corresponding to a Latin square. The order in which the different diagnostic signs were employed was constant, so different diseases were assigned to each sign depending on the disease order. Because each disease required a different diagnostic test, and the tests were administered in different blocks, participants were not provided with any information about the co-occurrence of diseases. This parallels the information available in the previous questionnaire studies. Three different sets of symptoms were employed, as in Experiment 2. The *baserate* condition, order of disease presentations, and specific symptom pairs were counterbalanced. The explanation choices were presented in one of several random orders.

4.2. Results

4.2.1. Explanation choices

Overall, participants chose the simpler explanation 63% of the time, but the number selecting this option varied significantly as a function of the *baserate* condition ($\chi^2(3) = 7.26$, $p < .01$). Specifically, increasingly fewer participants chose the simpler explanation as it became less likely (see Fig. 2). But even when the simpler explanation was 10 times less likely than the two-cause alternative, 41% of participants continued to prefer it. Qualitatively these data replicate the findings from Experiment 2. There were no significant differences in the percentage of participants choosing the simpler explanation for three of the four corresponding *baserate* conditions in Experiment 2: 96 versus 94% at 15:1 ($\chi^2(1) = .09$, $p = .77$); 56 versus 67% at 1:2 ($\chi^2(1) = .56$, $p = .46$); and 41 versus 39% at 1:10 ($\chi^2(1) = .02$, $p = .90$). At the 9:10 probability ratio significantly fewer participants chose the

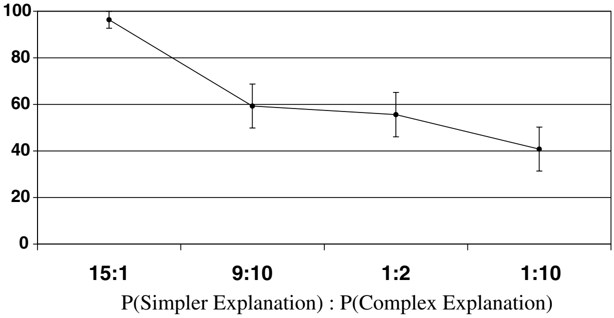

Fig. 2. Explanation choices from Experiment 3. The percentage of participants selecting the simpler explanation is indicated as a function of *baserate* condition.

simpler explanation in the computer task than in Experiment 2: 59 versus 89% ($\chi^2(1) = 4.6$, $p < .05$).

The logistic regression analysis detailed under the results section of Experiment 2 was repeated, resulting in a slope of .62 ($SE = .157$) and a constant of −.77 ($SE = .24$). The slope was not significantly different from the equivalent analysis in Experiment 2 ($p > .3$), but corresponds to a slight underweighting of probability information. However, the constant parameter was significantly smaller ($p < .05$) in this experiment. Whereas the estimate of people's prior probability for the simpler explanation was 79% in Experiment 2, the estimate for this experiment is 68% (57–78%, .95 confidence interval).[6]

There was no significant effect of the order in which the diseases were presented ($p > .1$), nor an interaction between disease order and *baserate* condition ($p > .3$).

4.2.2. Explanation choice justifications

Participant's justifications for their explanation choice were coded as in Experiment 2, with the exception that participants who potentially misunderstood the task were replaced (10 total). Two coders reviewed all 108 justifications with agreement of 96%. Disagreements were resolved by discussion. For analysis, values from neighboring probability ratios are combined to reduce noise among small and hence highly variable categories. Overall, 8% of participants provided *simple* justifications, 31% provided *sufficient* justifications, 56% provided *probable* justifications, and the remaining 5% provided *other* justifications. *Simple*, *sufficient*, and *other* justifications were almost always provided in support of the simpler explanation, while *probable* justifications were frequent for both the simple and complex explanations. Justification frequencies did not vary significantly as a function of *baserate* condition. However, as in Experiment 2, there was a significant interaction between *probable* justifications and *baserate* condition: at the 15:1 and 9:10 probability ratios most appeals to probability were in support of the simpler explanation (66% of *probable* justifications), but at the 1:2 and 1:10 probability ratios most appeals to probability were in support of the two-cause explanation (81% of *probable* justifications). Also as in Experiment 2, the total number of appeals to probability did not change as function of *baserate* condition when responses were collapsed across explanation choices—about 56% of participants justified their choice by appeal to probability in each *baserate* condition.

4.2.3. Frequency estimates

Because participants were provided with frequency information in the form of cases rather than summary values, participants could be asked to estimate disease frequencies to examine whether estimates vary as a function of frequency condition and explanation choice. In particular, will participants who choose a simple explanation overestimate the frequency of the disease invoked in that explanation? Average estimates for each disease are illustrated in Fig. 3. Each panel corresponds to the estimates for a single disease (D_1, D_2, or D_3), with the probability ratio along the x-axis and the average estimate for the percentage of the population with the corresponding disease along the y-axis. The actual values are indicated by solid gray. If participants were perfectly accurate in their assessments

[6] Note that the value obtained in this experiment was not significantly different from the estimate of 76% in Experiment 2 when participants who misunderstood the question were excluded.

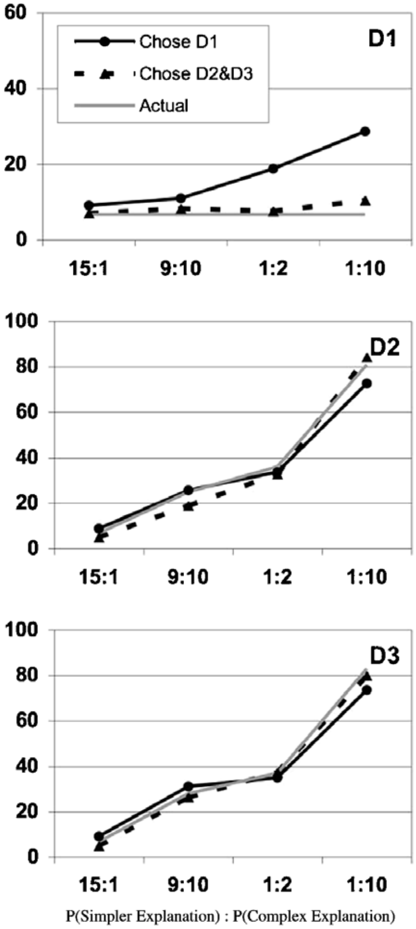

Fig. 3. Estimates of disease frequencies from Experiment 3. The average estimate for the percentage of the population with each disease is indicated. The actual values, corresponding to the data participants were shown, are indicated with solid gray lines. Participants' estimates are grouped according to whether they selected the simple, one-cause explanation or the complex, two-cause explanation as the most satisfying account of the alien's symptoms.

of frequencies, the data should fall along the solid lines. Indeed, on average, participants were remarkably accurate in estimating the disease frequencies.

A planned comparison of the estimate for D_1 as a function of explanation choice revealed the following. At the 15:1 and 9:10 probability ratios, average estimates of D_1 did not vary as a function of explanation choice. That is, on average, participants who chose the simple explanation generated estimates for the frequency of D_1 that did not vary from those of participants who chose the two-cause explanation. In contrast, participants who committed to the simple explanation at the 1:2 and 1:10 probability ratios estimated the frequency of D_1 to be greater on average than did participants who chose the two-cause explanation, though the variance in estimates was significantly greater in the former group. This comparison was significant at the 1:2 ratio (19 versus 8%, $t(14.6) = 2.18$, $p < .05$, two-tailed, equal variance not assumed), and suggestive at the 1:10 ratio (29% versus 10%,

$t(10.6) = 1.74$, $p = .11$, two-tailed, equal variance not assumed). When these data were analyzed in conjunction as a 2 by 2 ANOVA (baserate condition, 1:2 or 1:10, by explanation choice, one-cause or two-cause) with the estimate for D_1 as the dependent measure, there was a highly significant main effect of explanation choice ($F(1, 50) = 7.93$, $p < .01$) and no interaction between *baserate* condition and explanation choice ($F(1, 50) = .442$, $p = .51$). Thus participants who committed to the simple explanation when it was quite unlikely systematically overestimated the frequency of D_1. These same comparisons were not significant for estimates of D_2 and D_3 ($p > .2$), suggesting that the D_1 overestimation does not reflect a general handicap at estimating frequency information among participants who chose the simpler explanation.

The order in which diseases were presented did not influence frequency estimates ($p > .3$) nor interact with explanation choice or *baserate* condition ($p > .3$).

4.3. Discussion

Experiment 3 replicated the basic findings from Experiment 2 in a more ecologically valid context. When provided with probability information in the form of experienced frequencies, participants as a group were sensitive to probability information, but nonetheless had a general preference for the simpler explanation. As before, this bias principally manifested as a prior probability in favor of simpler explanations, not as a tendency to discount probability information. This preference also had explicitly probabilistic consequences, resulting in systematic overestimation of the frequency of the cause invoked in the simpler explanation. More precisely, participants who chose a simple explanation when it was unlikely to be true provided significantly higher estimates for the prevalence of the corresponding disease than did participants who chose the two-cause explanation in the same conditions (see Fig. 3).

What might account for this selective overestimation? One possibility is that some participants were simply bad at estimating frequencies, and for this reason chose to select an explanation on the basis of simplicity and provided inaccurate frequency estimates. This possibility seems highly unlikely given the selectivity and systematicity of the errors. Participants were generally accurate at probability ratios less favorable to the more complex explanation (15:1, 9:10), and also when estimating the frequency of diseases other than that invoked in the simplest explanation. Another possibility is that some participants confused the frequency of the disease in the simpler explanation, D_1, with either D_2 or D_3. Because D_2 and D_3 were more prevalent, such a mistake would result in inflated D_1 estimates. However, such a confusion would also result in systematic underestimation of the frequency of D_2 and D_3, which was not observed. More telling, only one participant provided a higher estimate for D_1 than either D_2 or D_3 at the 1:2 or 1:10 probability ratios.

Assuming the selective overestimation of D_1 was related to participants' preference for the simpler explanation, there is still a question of causal direction. It could be that participants overestimated D_1, and as a result chose the simpler explanation, or instead that they chose the simpler explanation, and as a result overestimated D_1. Once more, the selectivity and systematicity of the error provides a clue to the correct story. If participants inflated frequency estimates for a reason unrelated to the explanation task, such inflation would be expected across all conditions. The fact that inflated frequency estimates were restricted to the cases in which the simpler explanation was quite unlikely, namely the 1:2 and 1:10 probability ratios, speaks against this possibility. Thus participants who select an improbable, simple explanation most likely overestimated D_1 *as a result* of their explanatory commitment.

Synthesizing the findings from Experiments 1–3, the following picture emerges. Participants take both simplicity and probability into account when evaluating explanations, but do so in different ways. Simplicity influences the prior probability assigned to candidate explanations, with the simplest explanation receiving a higher prior probability. Probability information is used as a source of evidence to generate posterior probabilities, which are a function of both the prior and the evidence provided. If this Bayesian characterization of explanation evaluation is correct, one might expect the influence of simplicity on prior probability to diminish if participants have an alternative, probabilistic basis for generating priors. This possibility is examined in Experiment 4.

5. Experiment 4: Simplicity and probabilistic uncertainty

In Experiment 4, participants were asked to perform the same task as in Experiment 2, but any uncertainty associated with the probabilistic evidence was eliminated. This was accomplished by providing participants with the joint probabilities for all of the disease pairs. The rationale for this experiment is as follows. Participants may rely on simplicity to inform judgments when they lack strong probabilistic grounds for choosing an explanation. Experiments 2 and 3 required participants to compare explanations for which probabilities were not directly provided—in particular, comparing the probability of the simple and complex explanations required computing a joint probability. Simplicity may have informed the prior probability assigned to explanations by entering into this implicit joint probability calculation, with the result that the complex explanation seemed much less probable than the frequency information actually suggested. If this is the case, then participants may select the complex explanation when uncertainty about probability is removed, obviating the need to rely on simplicity as a basis for establishing prior probabilities.

Explicitly providing participants with the probability of each explanation provides a way to determine whether simplicity informs explanation evaluation when probabilistic uncertainty is eliminated, but the manipulation also presents a problem. The additional probability information may bias participants to respond on the basis of probability by changing the perceived task demands. As a result, a second condition is included in Experiment 4 for comparison. This condition is like the task in Experiment 2, but rather than being asked to select the most *satisfying* explanation, participants are asked to select the most *likely* explanation. Like providing additional probability information, requesting the most likely explanation should bias participants to respond on the basis of probability. But if the determinant of explanation choice is the strength of probabilistic evidence rather than task demands, then only providing joint probabilities should change responses from the patterns found in Experiment 2.

5.1. Methods

5.1.1. Participants

Participants were 48 undergraduates and summer school students from an elite university (58% female; mean age = 20, *SD* = 3) who completed the study in exchange for course credit, payment, or a small gift. One additional participant was excluded as a non-native English speaker who failed to understand the task.

5.1.2. Materials

Experimental materials consisted of a two-page questionnaire. In the *joint probability* condition, participants read a scenario like the one in Experiment 2 at the 2:3 probability ratio, with the exception that the following paragraph was added:

> Some of these aliens with one disease also have a second disease. For example, about 17 aliens have both *Brom's and Stemmel's disease*, about 15 have both *Brom's and Pilt's*, and about 73 have both *Stemmel's and Pilt's*.

Participants were then asked to choose the most satisfying explanation as in Experiments 2 and 3. In the *likely* condition, participants read a scenario like the one in Experiment 2 at the 2:3 probability ratio, with the exception that they were asked to select the most *likely* explanation rather than the most *satisfying* explanation. For both conditions, the second page of the questionnaire was identical to that in Experiment 2.

5.1.3. Design and procedure

Participants were randomly assigned to condition. The order in which diseases were presented and the specific symptoms mentioned were counterbalanced. The candidate answers to the why-question were presented in one of several random orders.

5.2. Results

5.2.1. Explanation choices

In the *joint probability* condition, only 17% of participants chose the simpler explanation. This percentage was significantly lower than the 61% at the corresponding probability ratio, 2:3, from Experiment 2 ($\chi^2(1) = 8.85$, $p < .01$). In the *likely* condition, 46% of participants chose the simpler explanation. This percentage was significantly higher than the 17% selection rate observed in the *joint probability* condition ($\chi^2(1) = 4.75$, $p < .05$), but not significantly different from the 61% at the corresponding probability ratio, 2:3, from Experiment 2 ($\chi^2(1) = 1.34$, $p > .2$).

5.2.2. Explanation justifications

Explanation choice justifications were coded as in Experiment 2 by two coders. Agreement was 88%, with disagreements resolved by discussion. Overall, the choice of the simpler explanation was justified 13% of the time by appeal to *simplicity*, 40% of the time by appeal to *sufficiency*, 33% of the time by appeal to *probability*, and 13% by appeal to other reasons. The choice of the complex explanation was justified 94% of the time by appeal to *probability* and 6% of the time by appeal to other reasons. Justifications did not vary significantly as a function of condition.

5.2.3. Explanation choice and math ability

Overall, 75% of participants answered the math problem correctly, and correct responses did not vary as a function of condition ($t(46) = -.656$, $p = .525$). As in Experiment 2, there was not a significant correlation between explanation choice and answering the math problem correctly in the *joint probability* condition ($r = .05$, $p = .83$). However, there was a significant relationship between explanation choice and answering the math problem correctly in the *likely* condition ($r = .51$, $p < .05$). Specifically, participants who selected the simpler explanation were more likely to answer the math problem incorrectly. This suggests that requesting participants to select the most *likely* explanation as opposed to the most *satisfying* did change task demands, leading participants who knew how to compute joint probabilities to answer on the basis of probability, and those who did not to answer on the basis of intuitions about probability. This also suggests that under some conditions, a preference for simpler explanations may result from probabilistic ignorance.

5.3. Discussion

Experiment 4 demonstrates that when ambiguity about probability information is reduced, a significantly greater number of participants are willing to override a simplicity preference and select a more probable two-cause explanation over a simpler alternative. Moreover, this result is not due merely to a change in task demands, as participants asked to select the most likely explanation did not exhibit a similar decrease in preference for the simpler explanation. These findings help elucidate the conditions under which simpler explanations are preferred. When probability evidence is ambiguous (as in Experiments 2 and 3), simpler explanations are assigned a higher prior probability, perhaps because complex explanations are penalized by the way in which joint probabilities are evaluated. But when probabilistic evidence is unambiguous, prior probabilities are effectively provided and the need to use simplicity is obviated.

6. General discussion

This paper began by considering the roles of simplicity and probability in choosing among candidate causal explanations. The data suggest that both simplicity and probability are relevant dimensions for assessing explanations, with simplicity increasing both "explanatory goodness" and perceived probability. Specifically, people may rely on simplicity as a basis for evaluating explanations when direct probability information is absent (Experiment 1) or opaque (Experiments 2 and 3 versus 4). The preference for simplicity manifests as a higher prior probability assigned to simple explanations, with the consequence that disproportionate evidence for a complex explanation is required before it will be favored over a simpler alternative (Experiments 2 and 3). And finally, committing to a simple but improbable explanation can inflate the perceived frequency of the cause invoked in the simple explanation (Experiment 3).

6.1. Relationship between explanation and probability

The findings reported above are consistent with two broad accounts of the role of simplicity in informing probabilistic judgments. First, it could be that simpler explanations are judged more probable in virtue of making for better explanations. On this view, simpler explanations receive elevated prior probabilities because they are judged more explanatory, not because of their simplicity per se. If this account is correct, then other explanatory virtues like consistency, scope, and fruitfulness may also lead to elevated prior probabilities. Alternatively, simplicity may enjoy probabilistic privilege for reasons not mediated by explanation. For example, if the world is believed to be simple, then simple explanations are more likely to be true not because they're more explanatory, but because they're more likely to describe the world.

In practice, these accounts are difficult to distinguish. For instance, it could be that simplicity is an explanatory virtue *because* simpler explanations are believed to be more probable. However, a variety of previous results indicate that explanatory considerations other than simplicity play a prominent role in probabilistic judgments (see Lombrozo, 2006, for review), suggesting that simplicity may inform probability because it is explanatory. One of Tversky and Kahneman's most remarkable demonstrations of the conjunction fallacy illustrates this nicely (Tversky & Kahneman, 1983). When asked to estimate the probability that the US would break off diplomatic relations with the Soviet Union, participants judged the event more probable if additionally told the break followed a Soviet invasion of Poland. Although the latter event included the former, and was therefore less probable, embedding the diplomatic break in a plausible causal story that explained the US decision made it seem more likely. Studies also show that generating possible explanations for an event or fact can increase its judged probability (Koehler, 1991, 1994), and that even when explanations are not solicited, a target statement is judged more probable when it can be explained in the same way as a preceding statement than when the explanation is different (Sloman, 1994, 1997). These effects extend to contexts with potentially serious consequences, like jury decisions, which have been shown to vary depending on whether witnesses are called randomly or in an order that corresponds to a causal narrative (Pennington & Hastie, 1988, 1992). If the mechanisms responsible for these findings are at work in the present studies, it seems likely that the influence of simplicity on probability is mediated by explanation.

6.2. Relationship to previous work

There has been no shortage psychological research on causal reasoning, and equally abundant philosophical work on simplicity, but the present findings differ importantly from these past contributions. In cognitive psychology, the emphasis has been on causal induction: tasks that require inferring the presence of an unknown causal relationship from contingency or mechanism information (e.g. Cheng, 1997; Griffiths & Tenenbaum, 2005; in computer science Halpern & Pearl, 2000). For example, ΔP, the difference between the probability of an effect in the presence and the absence of a cause, has been proposed as a measure of the strength of a causal relationship (Cheng & Novick, 1990). In contexts where more than one potential cause may be responsible for an effect, people appropriately conditionalize on the presence or absence of the possible confound instead of applying ΔP (Spellman, 1996b; Spellman et al., 2001). Such findings suggest that people are sensitive to number of causes in causal induction, though (to my knowledge) experiments only indirectly address the role of simplicity (Lu et al., 2006; Novick & Cheng, 2004), and participants are not asked to judge the quality of explanations.

The findings reported here also differ from discussions of simplicity in philosophy by involving causal tokens rather than types. Within philosophy, simplicity is generally invoked and justified in contexts of causal induction, like scientific theorizing (e.g. Sober, in press; but see Sober, 1991). In such cases, the alternatives differ in the number of types of entities or relations they posit in the world, rarely in the number of tokens. In contrast, deciding whether known causes are present is an inference over causal tokens. It is not clear that considerations of parsimony that operate over types should also hold for tokens. Whether the same considerations do operate in people's inferences over types and tokens is an open and interesting psychological question.

Finally, the literature in social psychology, and in particular in attribution theory, has explored causal judgments that resemble those in these tasks. For example, the role of multiple causes has been explored in the context of discounting. Several findings suggest that whether situational factors are deemed causally relevant to some behavior can influence the extent to which dispositional factors are considered, and vice versa (see McClure, 1998 for review; Morris & Larrick, 1995). In explaining specific outcomes, one or two causes may be preferred as a function of the extremity and valence of what's being explained (McClure, Lalljee, & Jaspars, 1991; McClure, Lalljee, Jaspars, & Abelson, 1989). Unlike the experiments here, this literature almost exclusively explores whether an additional cause is invoked once one cause is identified. The choice is between cause A or causes A and B, not between cause A and causes B and C. While simplicity is relevant to both choices, A is necessarily more probable than its conjunction with another cause. As a result, the tension between simplicity and probability explored in this paper never arises.

6.3. Beyond number of causes

The findings reported above suggest that simpler explanations are preferred and judged more likely, but only a limited notion of simplicity is explored. For example, do people continue to prefer explanations involving fewer causes when all explanations involve multiple causes? Are interacting causes treated differently from independent causes? How are simplicity differences reconciled with differences in fit between the explanation and the data, as in curve fitting? Simplicity can be quantified in a number of ways, presenting the possibility that people are sensitive to a number of different metrics, perhaps relying on some more than others as a function of context. While "number of causes" is a psychologically plausible metric, another promising approach is to begin with computationally well-defined notions of simplicity, such as Minimum Description Length (Rissanen, 1978) or Bayesian Occam's Razor (Jeffreys & Berger, 1992), and search for psychological analogues.

The present research could also be extended by exploring the role of simplicity in different cognitive tasks. While simplicity plays an important role in judgments of explanation satisfaction, the effects may be different in tasks framed in terms of causal inference, or in which inferring a causal explanation is just one step in a chain of reasoning to, for example, determine which medication is appropriate for a patient, or what penalty appropriate to a crime. In addition to influencing the evaluation of explanations, simplicity may play a role in the generation of potential explanation. For example, people may be more likely to recognize that a single known cause can account for an effect, failing to recognize that a conjunction of known causes can do the same. In the four experiments reviewed above, participants were provided with candidate explanations, so effects of simplicity necessarily entered in the evaluation of explanations. In real-world contexts, where people are generally responsible for both generating and evaluating possible explanations, the influence of simplicity may be much more pronounced.

7. Conclusions

The present findings reveal the psychological reality of two related ideas from statistics and philosophy. First, simplicity plays a privileged role in assigning prior probabilities, as suggested by the literatures on formal metrics of simplicity in model selection. Second, considerations of how well something explains guide inference, as proposed by the literature on Inference to the Best Explanation. More broadly, evaluating explanations may serve as a mechanism for generating subjective probabilities.

Acknowledgments

This research was supported by a National Defense Science and Engineering Graduate Fellowship and a National Science Foundation Graduate Fellowship. I thank Susan Carey and Tom Griffiths for helpful discussion and comments on earlier drafts of the paper, and Jessica Jane Rutstein for help with the experiments. I also thank Nick Chater, David Lagnado, Barbara Spellman, and an anonymous reviewer for detailed and insightful comments. Selected results were presented at the Annual Summer Interdisciplinary Conference in 2004 and at the 26th Annual Conference of the Cognitive Science Society in 2004.

References

Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41*, 361–416.

Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology, 45*, 153–219.

Baker, A. (2004). Simplicity. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy (winter 2004 edition)*, URL = <http://plato.stanford.edu/archives/win2004/entries/simplicity/>.

Bod, R. (2002). A unified model of structural organization in language and music. *Journal of Artificial Intelligence Research, 17*, 289–308.

Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior, 6*, 151–155.

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology, 6*, 193–204.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review, 103*, 566–581.

Chater, N., & Vitanyi, P. (2003). Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Science, 7*(1), 19–22.

Cheng, P. W. (1997). From covariation to causation: a causal power theory. *Psychological Review, 104*, 367–405.

Cheng, P., & Novick, L. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58*(4), 545–567.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: a theoretical framework and implications for science instruction. *Review of Educational Research, 63*, 1–49.

Doyle, A. C. (1986). Sherlock Holmes: The complete novels and stories(Vol. 2). New York: Bantam Press.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99*, 3–19.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature, 407*, 630–633.

Forster, M. R. (2000). Key concepts in model selection—performance and generalizability. *Journal of Mathematical Psychology, 44*, 205–231.

Fugelsang, J. A., & Thompson, V. A. (2003). A dual-process model of belief and evidence interactions in causal reasoning. *Memory & Cognition, 31*, 800–815.

Griffiths, T. L., Christian, B. R., & Kalish, M. L. (2006). Revealing priors on category structures through iterated learning. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the cognitive science society* (pp. 1394–1399). Mahwah, NJ: Erlbaum.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 354–384.

Halpern, J. Y. & Pearl, J. (2000). *Causes and explanations: A structural-model approach. Technical Report R-266, UCLA Cognitive Systems Lab.* Los Angeles, CA.

Harman, G. (1965). The inference to the best explanation. *Philosophical Review, 74*, 88–95.

Jeffreys, W. H., & Berger, J. O. (1992). Ockham's razor and Bayesian analysis. *American Scientist, 80*, 64–72.

Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin, 110*, 499–519.

Koehler, D. J. (1994). Hypothesis generation and confidence in judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 461–469.

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes, 56*, 28–55.

Lagnado, D. (1994). *The psychology of explanation: A Bayesian approach*. Masters Thesis. Schools of Psychology and Computer Science, University of Birmingham.

Leeuwenberg, E., & Boselie, F. (1988). Against the likelihood principle in visual form perception. *Psychological Review, 95*(4), 485–491.

Li, M., & Vitanyi, P. M. B. (1997). *An introduction to Kolmogorov complexity and its applications*. New York, NY: Springer-Verlag.

Lipton, P. (2002). *Inference to the best explanation*. New York, NY: Routledge.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences, 10*(10), 464–470.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99*(2), 167–204.

Lu, H., Yuille, A., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2006). Modeling causal learning using Bayesian generic priors on generative and preventive powers. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the cognitive science society* (pp. 519–524). Mahwah, NJ: Erlbaum.

McClure, J. (1998). Discounting causes of behavior: Are two causes better than one? *Journal of Personality and Social Psychology, 74*, 7–20.

McClure, J., Lalljee, M., & Jaspars, J. (1991). Explanations of extreme and moderate events. *Journal of Research in Personality, 25*, 146–166.

McClure, J., Lalljee, M., Jaspars, J., & Abelson, R. P. (1989). Conjunctive explanations of success and failure: the effect of different types of causes. *Journal of Personality and Social Psychology, 56*, 19–26.

Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another: a normative analysis of discounting in causal attribution. *Psychological Review, 102*(2), 331–355.

Newton, I. (1953/1686). *Philosophiae naturalis principia mathematica*. Reprinted in H. Thayer (Ed.) Newton's Philosophy of Nature. New York: Hafner.

Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review, 111*, 455–485.

Peirce, C. S. (1998). *The essential Peirce: Selected philosophical writings, 1893–1913*. Bloomington, IN: Indiana University Press.

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: the effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 521–533.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: tests of the story model for jury decision making. *Journal of Personality and Social Psychology, 62*, 189–206.

Pomerantz, J. R., & Kubovy, M. (1986). Theoretical approaches to perceptual organization. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. II). New York, NY: John Wiley & Sons, Inc.

Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: a parallel distributed processing account. *Journal of Personality and Social Psychology, 65*(3), 429–447.

Rehder, B. (2003a). Categorization as causal reasoning. *Cognitive Science, 27*, 709–748.

Rehder, B. (2003b). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1141–1159.

Rissanen, J. (1978). Modeling by the shortest data description. *Automatica, 14*, 465–471.

Shepard, R., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied, 75*, 1–42.

Sloman, S. A. (1994). When explanations compete: the role of explanatory coherence on judgments of likelihood. *Cognition, 52*, 1–21.

Sloman, S. A. (1997). Explanatory coherence and the induction of properties. *Thinking and Reasoning, 3*, 81–110.

Sober, E. (1991). *Reconstructing the past: Parsimony, evolution, and inference*. Cambridge, MA: Bradford Books.

Sober, E. (in press). Parsimony. In S. Sarkar & J. Pfeifer (Eds.), *The Philosophy of science: An encyclopedia* (Vol. 2). New York, NY: Routledge.

Spellman, B. A. (1996a). Conditionalizing causality. In D. R. Shanks & K. Holyoak (Eds.), *Causal learning* (pp. 167–206). San Diego, CA: Academic Press.

Spellman, B. A. (1996b). Acting as intuitive scientists: contingency judgments are made while controlling for alternative potential causes. *Psychological Science, 7*, 337–342.

Spellman, B. A., Price, C. M., & Logan, J. M. (2001). How two causes are different from one: the use of (un)conditional information in Simpson's paradox. *Memory & Cognition, 29*, 193–208.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12*, 435–467.

Thagard, P. (2000). Probabilistic networks and explanatory coherence. *Cognitive Science Quarterly, 1*, 93–116.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Review, 90*, 293–315.

van der Helm, P. A. (2000). Simplicity versus likelihood in visual perception: from surprisals to precisals. *Psychological Bulletin, 126*, 770–800.

Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks & K. Holyoak (Eds.), *Causal learning* (pp. 47–88). San Diego, CA: Academic Press.