Explanatory Considerations Guide Pursuit

Patricia Mirabile (patricia.mirabile@sorbonne-universite.fr)
Sciences Normes Décision, Sorbonne Université
Paris, France

Tania Lombrozo (lombrozo@princeton.edu)
Department of Psychology, Princeton University
Princeton, NJ, USA

Abstract

Evidence is typically consistent with more than one hypothesis. How do we decide which hypothesis to pursue (e.g., to subject to further consideration and testing)? Research has shown that *explanatory considerations* play an important role in learning and inference: we tend to seek and favor hypotheses that offer good explanations for the evidence we invoke them to explain. Here we report three studies testing the proposal that explanatory considerations similarly inform decisions concerning pursuit. We find that ratings of explanatory goodness predict pursuit (though to a lesser extent than they predict belief), and that these effects hold after adjusting for subjective probability. These findings contribute to a growing body of work suggesting an important role for explanatory considerations in shaping inquiry.

Keywords: explanation; pursuit; abduction; active learning

From belief to pursuit

"Faced with tracks in the snow of a certain peculiar shape," writes (Lipton, 2003), "I infer that a person on snowshoes has recently passed this way." This form of inference, familiar from both science and everyday life, is known as *inference to the best explanation* (IBE). Harman (1965) explains that in drawing this inference to an explanatory hypothesis, one infers, from the premise that a given hypothesis would provide a better explanation for the evidence than would any other hypothesis, to the conclusion that the given hypothesis is true (Harman, 1965).

Recent work in the cognitive science of explanation has confirmed and helped characterize the role of IBE in human cognition: both children and adults tend to prefer some explanations over others, and these preferences affect which hypotheses they favor (Lombrozo, 2016; for factors that might affect these preferences, see Colombo, Bucher, & Sprenger, 2017). For example, in Douven and Mirabile (2018), participants read about two possible explanations for six realistic events: a target "best" explanation (that had on average received higher quality ratings in a previous experiment) and an alternative explanation (that had received lower ratings). One group of participants was asked to rate the explanatory quality of both explanations (how well each explained the event) and a second group of participants rated the probability of each explanation, additionally indicating whether they accepted the target explanation as the true one. They found that the mean goodness ratings of the first group were better predictors of the acceptance rate of the target explanation by the second group than the probability ratings of that same group.

Findings like these suggest that explanatory considerations play an important role in guiding belief – indeed, in some cases a stronger role than that played by probability (see also Douven & Schupbach, 2015). But they also raise an important puzzle that has been a perennial challenge for advocates of IBE: why treat explanatory considerations as a good guide to what is true? After all, the world may not be simple, elegant, or otherwise conform to a good explanation. This challenge is especially acute when explanatory considerations diverge from probabilistic considerations (see van Fraassen, 1989).

One possibility is that the practice of favoring hypotheses that offer better explanations is a good epistemic policy in the sense that it has positive epistemic consequences, even if it doesnt *directly* result in an inference to a hypothesis that is more likely to be true. Along these lines, Wilkenfeld and Lombrozo (2015) introduce the idea of "Explaining for the Best Inference," whereby the practice of explaining (and of seeking *good* explanations) might improve our epistemic standing through a suite of downstream cognitive effects. Indeed, seeking and evaluating explanations facilitates the discovery of subtle patterns (e.g., Williams & Lombrozo, 2010; Walker & Lombrozo, 2017), even when the generated explanations are inaccurate (Walker, Lombrozo, Legare, & Gopnik, 2014). Explaining also encourages processes such as comparison (Edwards, Williams, Gentner, & Lombrozo, 2019), abstraction (Walker & Lombrozo, 2017), and metacognitive calibration (Rozenblit & Keil, 2002), which can be beneficial even if the agent fails to make an inference to a true explanation.

In the current paper, we turn our attention to the idea of IBE as an effective epistemic policy that could guide learners over time. Rather than focusing exclusively on the role of explanatory considerations in making an inference to (or evaluating the probability of) a given hypothesis at a given time, we consider whether and how explanatory considerations affect the decision to *pursue* one hypothesis over another – that is, to subject a hypothesis to further consideration or testing.

Pursuing explanations

Pursuing hypotheses is an important part of any search for explanations: doctors order medical tests before establishing a diagnosis, detectives interrogate suspects and verify alibis, and scientists collect evidence to assess their theories. Decisions about pursuit are especially critical in science: not only must we justify to academic peers, funding agencies, and sometimes the general public why a given hypothesis is worthy of pursuit, but this very investigation can also serve as a "criterion of demarcation" between scientific and non-scientific endeavors. According to Popper (2005), a hallmark of science is the generation of theories that can be submitted to a method of critical testing: scientific theories should make predictions that can be falsified by empirical evidence.

However, both in principle and due to time and resource limitations, we are unable to investigate all hypotheses, even all *good* hypotheses: we must instead decide which hypotheses are worth pursuing, and which hypotheses are worth pursuing first. Nyrup (2015) argues that the justification of pursuit is the most legitimate use of IBE – more legitimate even that the justification of belief. His core idea is that a hypothesis that offers a good explanation has higher "epistemic value" if true than its salient competitors, and that this justifies giving it priority when deciding which hypotheses to pursue first.

Formal analyses additionally support the idea that a policy of favoring better explanations could pay off downstream, even if it does not lead to an accurate inference right away. Specifically, Kelly (2007) introduces a formal notion of simplicity, and contends that simple hypotheses should be preferred because adopting simple rather than more complex hypotheses will reduce to a minimum the number of necessary reversals of opinion before arriving at the true hypothesis, and therefore allow us to converge to the truth more quickly. Douven (2016) shows that under certain conditions, artificial agents using update rules that favor better explanations (defined according to a particular measure of "explanatory power") converge faster on the truth than artificial agents with probabilistic (Bayesian) update rules. Evaluating the goodness of an explanation might therefore be a key consideration when deciding whether to pursue it.

In the present research, we report three studies designed to address the following four questions. First (Q1), are people more likely to pursue one hypothesis over another to the extent it offers a good explanation for the data? Second (Q2), is this evaluation partially comparative, such that the explanatory goodness of alternatives will also matter, with a given hypothesis more likely to be pursued to the extent its alternative offers a poor explanation? Third (Q3), does explanatory goodness have an effect on pursuit that is not reducible to the effects of subjective probability on pursuit? Based on the findings from Douven and Mirabile (2018) concerning belief, we expect positive answers to these questions. However, we also expect pursuit and belief to diverge, given their differential costs (in terms of both requisite resources, and the consequences of getting things right vs. wrong). This prompts our final question (Q4): Does explanatory goodness differentially affect pursuit versus belief?

Study 1

In Study 1, we address Q1 - Q4 using materials adapted from Douven and Mirabile (2018). In a within-subjects design, participants were shown six vignettes that each described a disruptive event. They were presented with two possible hypotheses that might explain the event, and asked to rate the goodness and probability of each hypothesis. They also indicated which hypothesis they would recommend investigating first ("pursuit"), and which hypothesis they were more inclined to believe ("belief"). This allowed us to examine the link between perceived explanatory goodness and pursuit, as well as its relationship to probability and belief.

Method

Participants Participants were 72 adults recruited from Amazon Mechanical Turk (33 female, 39 male, ages 20-69, M = 35). Participation was restricted to MTurk workers with unique IP addresses in the United States who had completed at least 1000 HITs with a minimum approval of 99%. An additional 35 participants completed the study, but were excluded from analyses for failing one or more attention checks (described below).

Materials Six vignettes were lightly adapted from the stimuli used by Douven and Mirabile (2018). In these vignettes, experts (scientists, detectives, doctors) are attempting to explain a disruptive event (e.g., the flooding of a village, a murder, or a patient's symptoms), and they have generated two possible explanatory hypotheses, where these hypotheses are independent and are not jointly exhaustive. For instance, in one vignette, participants read about a womans murder, where one hypothesis is that the murder was committed by her jealous husband, and another hypothesis is that the murder was committed by a coworker trying to prevent her from sharing incriminating evidence. Based on the ratings of explanatory goodness provided by participants in Experiment 1 of Douven and Mirabile (2018), one of the hypotheses was classified as offering what we expected to be perceived as the *best explanation*, and the other as offering the *second best explanation*. These designations were used in analyses, but were not presented to participants.

Procedure Each participant received all six vignettes, with the order of the two hypotheses in each vignette randomized across participants. The study consisted of three phases: goodness and probability ratings, belief and pursuit decisions, and distraction questions, which doubled as attention checks. The distraction questions always appeared between the other two phases, which appeared first or last (randomized across participants).

In the goodness and probability ratings phase, participants received all six vignettes, and for each rated the two corresponding hypotheses on *explanatory goodness* and *probability*, with order randomized across participants. For explanatory goodness, participants were asked: "How good do you think each of these hypotheses is as an explanation for why *[the* *event occurred]*?", with the corresponding event specified in the stimuli participants saw. Responses were collected on a continuous scale from 0 to 100, where 0 meant that an explanation was very bad, 50 meant that an explanation was neither good nor bad, and 100 meant that an explanation was very good. For probability, participants were asked: "How likely do you think each of these hypotheses is?" Responses were collected on a continuous scale from 0 to 100, where 0 meant that the hypothesis had 0% probability of being true, 50 meant that the hypothesis was equally likely to be true or not true, and 100 meant that the hypothesis had 100% probability of being true. We also included an attention check in which participants were instructed to select zero on the two continuous scales.

In the pursuit and belief decisions phase, participants received all six vignettes, and for each rated the two corresponding hypotheses for pursuit and belief, with order randomized across participants. For the pursuit decisions, participants were told: "The [experts] only have enough resources to investigate and test *one of the two* hypotheses before deciding on an explanation. They could also decide to save their resources and not investigate or test either of the two hypotheses. What do you think they should do?" Participants could select either hypothesis or indicate that they didn't think either of the hypotheses should be investigated. For the belief judgment, participants were asked: "Which of the two hypotheses are you more inclined to believe is the true explanation of why *[the event occurred]*?", (again, the corresponding event was specified in the stimuli participants saw). Participants could select either hypothesis or indicate that they were not inclined to believe either of the hypotheses.

The distraction phase consisted of two questions that doubled as attention checks. Participants read a list of words and, depending on a randomly assigned condition, copied into a text box the first word from that list that referred to an animal, a fruit, or a season. Participants also counted the number of animals in a picture.

After completing these three phases of the study, participants provided demographic information.

Results & Discussion

To examine whether and how explanatory considerations affect pursuit (Q1 and Q2), we fit a logistic binomial mixed-effects model (Q1/Q2 model) predicting participants probability of deciding to pursue the (antecedently defined) best explanation, as opposed to the second best explanation or neither explanation. Our choice of model and dependent variable allowed us to parallel the analyses in Douven and Mirabile (2018), where acceptance of the target "best" explanation was used as a dichotomous dependent variable. Explanatory goodness ratings for the best explanation and for the second best explanation were both centered on 50 and included as fixed effects. Vignette was included as a group-level random effect.

This model found a positive coefficient for the goodness of the best explanation ($p < .001$), with a 5.2% increase in the odds of choosing to pursue an explanation for each one-point increase in goodness. It also found a negative coefficient for the goodness of the second best explanation ($p < .001$), with a 5.2% *decrease* in the odds of choosing to pursue the best explanation for each one-point increase in the goodness of the second best explanation. These results provide a positive answer to Q1: explanatory considerations did predict pursuit. They also provide an answer to Q2: while the goodness of the better hypothesis mattered, the goodness of the alternative mattered as well.

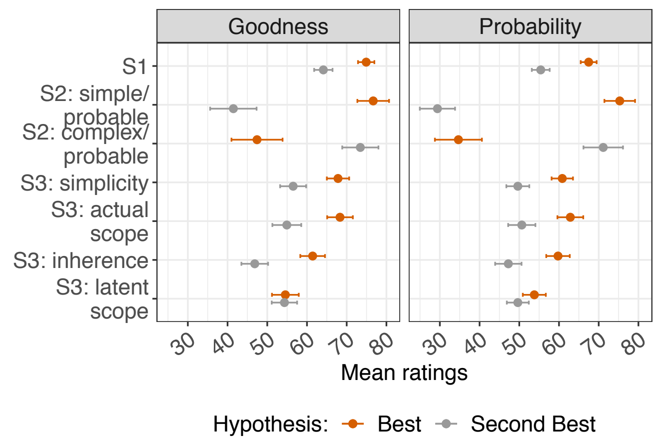

Figure 1: Mean ratings of goodness and of subjective probability for the two hypotheses for Study 1, Study 2 (Simple/Probable and Complex/Probable conditions) and Study 3 (Simplicity, Actual Scope, Latent Scope and Inherence conditions). Error bars represent 95% CI.

We next considered whether there were effects of explanatory considerations on pursuit that were not reducible to the effects of subjective probability on pursuit (Q3). To this end, we fit a logistic binomial mixed-effects model (Q3 model) predicting participants probability of deciding to pursue the best explanation, but in addition to the predictors included above, we also included a fixed effect for the probability assigned to the best explanation, and a fixed effect for the probability assigned to the second best explanation. Vignette was also included as a group-level random effect. There was a positive coefficient for the best explanation ($p < .001$), with a 4.8% increase in the odds of choosing to pursue the best explanation for each one-point increase in probability. There was also a negative coefficient for the second best explanation ($p < .001$), with a 4.9% *decrease* in the odds of choosing to pursue the best explanation for each one-point increase in the probability of its alternative. However, in this model, goodness ratings were not significant predictors. This suggests a negative answer to Q3: there was not evidence of effects of explanatory goodness on pursuit that were not reducible to the effects of subjective probability on pursuit. This result is potentially surprising in light of the findings from Douven and Mirabile (2018), which used essentially the same materials, but could be because goodness and probability were collected within-subjects and highly correlated: 0.73 for the best explanation, and 0.78 for the second best explanation. One aim of Study 2 is to more successfully tease apart goodness and probability ratings.

Finally, we evaluated whether explanatory goodness differentially predicted pursuit vs. belief (Q4). We fit a logistic regression mixed-effects model (Q4 model) predicting the probability of selecting the best explanation, with goodness rating for the best explanation, goodness rating for the second best explanation, and judgement type (belief vs. pursuit) as fixed effects, as well as interactions between judgement type and each goodness rating. We also included vignette as a group-level random effect. This model found that goodness ratings had a significant effect when predicting pursuit, and that this effect was significantly larger when predicting belief. A one-point increase in the goodness of the best explanation increased the odds of deciding to pursue by 5.1% ($p < .001$), and of deciding to believe by 11.5% ($p = 0.005$). On the other hand, a one-point increase in the goodness of the competing explanation *decreased* the odds of deciding to pursue by 5.0% ($p < .001$), and of deciding to believe by 7.3% ($p < .01$). Explanatory goodness thus had significant and differential effects on pursuit vs. belief, with the impact on belief larger than that on pursuit.

Study 2

Study 2 had two primary aims. First, we sought to revisit Q1-Q4 with materials that induced a weaker correlation between goodness and probability. Second, we sought to vary explanatory quality along a recognizable and objective dimension for which people's explanatory preferences have already been experimentally established: simplicity, defined as the number of unexplained causes invoked in an explanation (Pacer & Lombrozo, 2017). As in Study 1, participants were shown a vignette describing an unusual event with two possible explanatory hypotheses. Each hypothesis was either simple or complex, and described as either probable or improbable. By introducing simple/improbable and complex/probable hypotheses, we hoped to drive apart ratings of goodness and probability.

Method

Participants Participants in Study 2 were 135 adults recruited through Amazon Mechanical Turk as in Study 1 (56 female, 79 male, ages 19-72, M = 37). An additional 25 participants completed the study but were excluded from analyses for failing one or more attention checks.

Materials Two vignettes were created following the same structure as the stimuli used in Study 1. In these vignettes, scientists seek to explain an unusual event (either a change in the reproductive pattern of squirrels, or low crop yields in a given county), and they have generated two possibles hypotheses. One of these hypotheses was simple in the sense that it appealed to a single cause (exposure to one toxin, contamination by one pest), and the other was more complex in that it required the conjunction of two independent causes (two toxins, two pests). In addition, one of the hypotheses was described as being "quite probable" based on the data available to the scientists, and the other hypothesis was described as being "quite improbable."

Procedure The study had a between-subject design (2 vignettes x 2 probability conditions). Each participant received one vignette, with the order of the two presented hypotheses randomized across participants. Participants were randomly assigned to one of two probability conditions. In the simple/probable condition, the simple hypothesis was described as probable and the complex hypothesis as improbable. In the complex/probable condition, this pairing was reversed. In the main part of the study, participants responded to the same questions as in Study 1. They also responded to an attention check and completed one of the distraction tasks from Study 1 midway through the study.

Results & Discussion

First, we verified that Study 2 successfully reduced the high correlations between perceived goodness and probability observed in Study 1. We found correlations of 0.60 and 0.65 between goodness and probability in the simple/probable condition for the simple and complex explanations, respectively, and correlations of 0.58 and 0.76 in the complex/probable condition for the simple and complex explanations respectively. While these correlations remained strong, they were more modest than those in Study 1.

We next conducted the same analyses as those described in Study 1. To examine how explanatory considerations affect pursuit, we fit the Q1/Q2 model, but did not include vignette as a group-level random effect[1]. This model found a positive coefficient for the goodness of the simple explanation ($p < .001$), with a 9.8% increase in the odds of choosing to pursue an explanation for each one-point increase in goodness. It also found a negative coefficient for the goodness of the complex explanation ($p < .001$), with a 8.1% *decrease* in the odds of choosing to pursue the simple explanation for each one-point increase in the goodness of the complex explanation. These results again provide a positive answer to Q1: explanatory considerations did predict pursuit. They also provide an answer to Q2: while the goodness of the better hypothesis mattered, the goodness of the alternative mattered as well.

We next analyzed whether there were effects of explanatory considerations on pursuit that were not reducible to the effects of subjective probability on pursuit (Q3). There was a positive coefficient for the goodness of the simple explanation ($p < .001$), with a 7.9% increase in the odds of choosing to pursue the simple explanation for each one-point increase in goodness. There was also a negative coefficient for the goodness of the

[1] All analyses in Study 2 were first fit using mixed-effects models, with vignette as a group-level random effect. However, the regression analyses indicated a singular fit, so we fit the models again excluding the group-level random effect to ensure that the estimates were stable. Estimated coefficients in the fixed-effects and mixed-effects models were identical.

complex explanation ($p < .035$), with a 5.0% *decrease* in the odds of choosing to pursue the simple explanation for each one-point increase in the goodness of its alternative. However, in this model, subjective probability ratings were not significant predictors. This points to a positive answer to Q3: we found effects of explanatory goodness on pursuit that were not reducible to the effects of subjective probability. These findings could differ from those of Study 1 because goodness and probability were not as highly correlated as in Study 1, or because explanatory goodness was manipulated in the form of simplicity.

Finally, we evaluated whether explanatory goodness differentially predicted pursuit vs. belief by fitting the Q4 model. This model found that goodness ratings had a significant effect when predicting pursuit judgements, and that this effect was not significantly different when predicting belief. A one-point increase in the goodness of the simple explanation increased the odds of a participant deciding to pursue by 9.8% ($p < .001$), and a one-point increase in the goodness of the competing explanation *decreased* the odds of deciding to pursue by 8.1% ($p < .001$). Unlike Study 1, this suggests a negative answer to Q4.

Study 3

Studies 1-2 provided consistent answers to Q1 and Q2: participants were more likely to pursue one hypothesis over another to the extent they judged that hypothesis a good explanation, and its alternative a poor explanation. However, the answers to Q3 and Q4 were more variable across studies. In Study 3, we sought to revisit Q1-Q4 using a larger sample and more varied experimental materials.

Specifically, we varied explanatory quality along four dimensions suggested by prior research to elicit reliable patterns of preferences in people's judgements. The first dimension was simplicity, defined in terms of the number of unexplained causes invoked in each explanation (e.g, explaining an illness with one toxin or the conjunction of two toxins). The second dimension was actual scope, defined as the number of observed effects explained (e.g., explaining all aspects of how a space shuttle had deviated from its trajectory or only some of them). The third dimension was latent scope, defined as the number of *unverified* effects predicted (e.g., one hypothesis predicts that prior to the volcano's irruption, the magma should have been relatively cool and the second predicts that a wider ranger of magma temperatures was possible–however, data on magma temperature prior to the irruption is not available). The fourth dimension was inherence, defined as an appeal to inherent/internal features versus extrinsic features (e.g., a flower's ability to wick off water is explained either by properties of its petals or by properties of the soil where it grows). Prior work has shown that with materials like those used here, people favor explanations that are simpler (Pacer & Lombrozo, 2017), broad in actual scope (Williams & Lombrozo, 2010), narrow in latent scope (Khemlani, Sussman, & Oppenheimer, 2010), and inherent (Cimpian & Salomon, 2014). While simplicity and actual scope are often defended as explanatory virtues, latent scope and inherence are typically assumed to reflect unwarranted biases. The procedure, materials, data collection plan, main predictions, and analyses for Study 3 were preregistered on the Open Science Framework platform prior to data collection and are available at `https://osf.io/6b58k/`.

Method

Participants Participants in Study 3 were 875 adults recruited from Amazon Mechanical Turk as in Studies 1-2 (446 female, 424 male, 2 non-binary/other and 2 who preferred not to respond, ages 18-87, M = 40). Following our preregistration, 1000 participants completed the study, with exclusions (N=125) based on failure to pass one or more attention check(s).

Materials Twelve vignettes were created following the same structure as the stimuli in Studies 1-2. In these vignettes, scientists generate two possible hypotheses to explain an unusual event. The two hypotheses in each vignette differed along a single dimension (simplicity, actual scope, latent scope, or inherence), with three vignettes targeting each dimension. The simplicity vignettes were similar to those in Study 2. In the actual scope vignettes, the best hypothesis explained all aspects of the explanandum, and the second best hypothesis explained only a subset. In the latent scope vignettes, the best hypothesis accounted for the *explanandum* without making unverified predictions, while the second best generated a prediction that it was not possible to verify. In the inherence vignettes, modified from Horne and Khemlani (2018), the best hypothesis invoked an inherent feature of the *explanandum*, and the second best invoked an extrinsic feature.

Procedure Each participant received one vignette, with the order of the two hypotheses randomized across participants. Aside from the fact that this study had a between-subjects design (4 dimensions of explanatory quality x 3 vignettes), the procedure was identical to that of Study 1.

Results & Discussion

To address Q1-Q4, we followed the analyses described in Studies 1-2[2]. We first fit the Q1/Q2 model. This model found a positive coefficient for the goodness of the best explanation ($p < .001$), with a 5.6% increase in the odds of choosing to pursue an explanation for each one-point increase in goodness.

[2]In our preregistered analyses, we planned to fit logistic binomial mixed-effects models that included as predictors the goodness and probability ratings of the best explanation and differences in goodness/probability ratings between the best and the second best explanation. However, upon analyzing the data, we found a high (>0.89) correlation between differences in goodness ratings and differences in probability ratings. Because high correlations between predictors in linear regressions can make the estimated coefficients unreliable, we replaced the difference predictors by the goodness and probability ratings of the second best explanation. The correlation between goodness and probability ratings ranged from >0.8 for the actual and latent scope virtues, to >0.82 for simplicity and >0.94 for inherence.

It also found a negative coefficient for the goodness of the second best explanation ($p < .001$), with a 4.1% *decrease* in the odds of choosing to pursue the best explanation for each one-point increase in the goodness of the second best explanation. These results again provide a positive answer to Q1: explanatory considerations did predict pursuit. They also provide an answer to Q2: while the goodness of the better hypothesis mattered, the goodness of alternatives mattered as well.

We next analyzed whether there were effects of explanatory considerations on pursuit that were not reducible to the effects of subjective probability by fitting the Q3 model. There was a positive coefficient for the goodness of the best explanation ($p < .001$), with a 2.7% increase in the odds of choosing to pursue the best explanation for each one-point increase in goodness, and a positive coefficient for the subjective probability of the best explanation ($p < .001$), with a 3.4% increase in the odds of choosing to pursue the best explanation for each one-point increase in probability. There was also a negative coefficient for the goodness of the second best explanation ($p = .0026$), with a 1.8% *decrease* in the odds of choosing to pursue the best explanation for each one-point increase in the goodness of its alternative, and a negative coefficient for the subjective probability of the second best explanation ($p < .001$), with a 2.9% *decrease* in the odds of choosing to pursue the best explanation for each one-point increase in the goodness of its alternative. Like study 2, this provided a positive answer to Q3: the effect of explanatory considerations held even when the effect of probability judgements was also taken into account.

Next, we evaluated whether explanatory goodness differentially predicted pursuit vs. belief by fitting the Q4 model. This model found that goodness ratings had a significant effect when predicting pursuit judgements, and that this effect was significantly larger when predicting belief: a one-point increase in the goodness of the best explanation increased the odds of a participant deciding to pursue by 5.6% ($p < .001$) and of deciding to believe by 13.6% ($p < .001$). On the other hand, a one-point increase in the goodness of the competing explanation *decreased* the odds of deciding to pursue by 4.1% ($p < .001$), and of deciding to believe by 10.4% ($p < .01$). As in Study 1, explanatory goodness thus had significant and differential effects on pursuit vs. belief, with a larger impact on belief.

Finally, we repeated the three analyses just described for each of the four sets of vignettes corresponding to each virtue. These analyses revealed the same patterns of answers to Q1-Q2 as in the full data set, but some departures for Q3 and Q4. Specifically, we found a negative answer to Q3 for simplicity and actual scope, and a negative answer to Q4 for latent scope.

General Discussion

Across three studies, we find evidence that explanatory considerations affect pursuit: participants were more disposed to pursue a hypothesis to the extent it offered a good explanation, and to the extent its competitor offered a poor explanation. In Studies 2-3, we also found that the effect of explanatory considerations on pursuit were not reducible to the effects of subjective probability on pursuit. Finally, in Studies 1 and 3, we found that explanatory goodness had a larger impact on belief than on pursuit. Discrepancies across the three studies could have resulted from the high correlations between ratings of goodness and of subjective probability, but it is notable that Study 3–which had the largest sample–found positive answers to all four of our guiding questions. However, it is important to note that these results raise open questions about the direction of a potential causal relationship between explanatory considerations and pursuit, and indeed they do not rule out the possibility that pursuit decisions might be causing judgements of explanatory goodness, rather than the reverse.

Why might explanatory considerations affect pursuit? As suggested in the introduction, pursuing good explanations could facilitate learning (Lombrozo, 2016), have higher expected epistemic value (Nyrup, 2015), or provide a more efficient route to the truth (Kelly, 2007; Douven & Schupbach, 2015). The pursuit of good explanations might therefore improve our overall epistemic standing (Wilkenfeld & Lombrozo, 2015), even if the true hypothesis is not the most explanatory. If these ideas are correct, they provide a justification for IBE that side-steps many of the traditional worries concerning its application to belief.

Interestingly, however, the impact of explanatory goodness on pursuit was smaller than that on belief. In a context where unjustified pursuit is more costly (given limited resources) than erroneous belief, participants might be more reluctant to recommend pursuit on the basis of explanatory considerations alone. Moreover, decisions to pursue might be more sensitive to pragmatic considerations that compete with explanatory goodness, or to the goal of reducing uncertainty by maximizing expected information gain.

Several limitations are worth noting. First, participants reasoned about relatively abstract and unfamiliar material. Second, participants did not pursue explanations themselves (e.g., through further consideration or evidence gathering). Future work could investigate decisions to pursue (vs. believe) with more realistic materials, and testing a richer set of pursuit-relevant behaviors. It would also be fruitful to investigate whether the role of explanatory considerations changes as a function of the relevant consideration (as we began to explore in Study 3), in different environments (e.g., with different cost structures), and as a function of the learners goals (e.g., to achieve truth vs. avoid error).

More ambitiously, future research should investigate how pursuit and belief are integrated into a broader model of truth-seeking behavior that involves explanation generation, pursuit, the collection of evidence, hypothesis revision, and ultimately belief. Our findings suggest that explanatory considerations affect this process at two important stages, belief and pursuit, but leave open how they shape everyday and scientific inquiry more broadly.

Acknowledgments

We wish to thank the Concepts and Cognition Lab and Igor Douven for valuable input, as well as the SND research center and Sorbonne Université for funding Patricia Mirabile's stay as a visiting scholar at Princeton University.

References

Cimpian, A., & Salomon, E. (2014, May). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences*, *37*(5), 461–480.

Colombo, M., Bucher, L., & Sprenger, J. (2017, Sep). Determinants of judgments of explanatory power: Credibility, generality, and statistical relevance. *Frontiers in Psychology*, *8*.

Douven, I. (2016). Inference to the best explanation: What is it? and why should we care. In K. McCain & T. Poston (Eds.), *Best explanations: New essays on inference to the best explanation.* Oxford: Oxford University Press.

Douven, I., & Mirabile, P. (2018, Nov). Best, second-best, and good-enough explanations: How they matter to reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(11), 1792–1813.

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

Edwards, B. J., Williams, J. J., Gentner, D., & Lombrozo, T. (2019). Explanation recruits comparison in a category-learning task. *Cognition*, *185*, 21–38.

Harman, G. H. (1965). The inference to the best explanation. *The Philosophical Review*, *74*(1), 88–95.

Horne, Z., & Khemlani, S. (2018, July). Conceptual constraints on generating explanations. In C. Kalish, M. Rau, J. Zhu, & T. Rogers (Eds.), *Cogsci 2018* (pp. 1815–1820).

Kelly, K. T. (2007). How simplicity helps you find the truth without pointing at it. *Induction, Algorithmic Learning Theory, and Philosophy*, 111–143.

Khemlani, S. S., Sussman, A. B., & Oppenheimer, D. M. (2010, Nov). Harry potter and the sorcerer's scope: latent scope biases in explanatory reasoning. *Memory & Cognition*, *39*(3), 527–535.

Lipton, P. (2003). *Inference to the best explanation*. Routledge.

Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *TICS*, *20*(10), 748–759.

Nyrup, R. (2015, Dec). How explanatory reasoning justifies pursuit: A peircean view of ibe. *Philosophy of Science*, *82*(5), 749–760.

Pacer, M., & Lombrozo, T. (2017). Ockham's razor cuts to the root: Simplicity in causal explanation. *Journal of Experimental Psychology: General*, *146*(12), 1761.

Popper, K. (2005). *The logic of scientific discovery*. Routledge.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive science*, *26*(5), 521–562.

van Fraassen, B. C. (1989). *Laws and symmetry*. Oxford University Press.

Walker, C. M., & Lombrozo, T. (2017). Explaining the moral of the story. *Cognition*, *167*, 266–281. (Moral Learning)

Walker, C. M., Lombrozo, T., Legare, C. H., & Gopnik, A. (2014). Explaining prompts children to privilege inductively rich properties. *Cognition*, *133*(2), 343–357.

Wilkenfeld, D. A., & Lombrozo, T. (2015, Oct). Inference to the best explanation (ibe) versus explaining for the best inference (ebi). *Science & Education*, *24*(9-10), 1059–1077.

Williams, J. J., & Lombrozo, T. (2010). The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science*, *34*(5), 776–806.