The Radicalism of Truth-insensitive Epistemology: Truth's Profound Effect on the Evaluation of Belief*

John Turri

john.turri@gmail.com

Abstract: Many philosophers claim that interesting forms of epistemic evaluation are insensitive to truth in a very specific way. Suppose that two possible agents believe the same proposition based on the same evidence. Either both are justified or neither is; either both have good evidence for holding the belief or neither does. This does not change if, on this particular occasion, it turns out that only one of the two agents has a true belief. Epitomizing this line of thought are thought experiments about radically deceived "brains in vats." It is widely and uncritically assumed that such a brain is equally justified as its normally embodied human "twin." This "parity" intuition is the heart of *truth-insensitive* theories of core epistemological properties such as justification and rationality. Rejecting the parity intuition is considered radical and revisionist. In this paper, I show that exactly the opposite is true. The parity intuition is idiosyncratic and widely rejected. A brain in a vat is not justified and has worse evidence than its normally embodied counterpart. On nearly every ordinary way of evaluating beliefs, a false belief is significantly inferior to a true belief. Of all the evaluations studied here, only blamelessness is truth-insensitive.

Keywords: norms; belief; truth; blame; knowledge; evidence

---

* This is the penultimate version of a paper forthcoming in *Philosophy and Phenomenological Research*. Please cite the final, published version if possible.

Introduction

Suppose that Walt is wondering whether a person with a certain name works at a particular company. Walt is given access to a list of the company's employees. The list is maintained by the company's human resources department. Just like the product of any human endeavor, the list is imperfect. But it is still highly reliable. Now Walt reads the list and it says that the company employs someone named "Gustavo." And this is just another occasion where the list is right. Does Walt's evidence justify him in believing that the company employs someone by that name? Now imagine a case exactly like the one just described except that it's Skyler with access to the list and this is one of those rare cases where the list is wrong. Does Skyler's evidence justify her in believing that the company employs someone by that name? Can the fact that the list is right in the one case but not the other create a difference in what Walt and Skyler are justified in believing? Can it affect what it's rational for them to believe? Can it affect what they should believe? What is reasonable for them to believe? What is the responsible thing for them to believe?

According to one dominant view in contemporary philosophy, the answer to these questions is a resounding "no." On this view, belief evaluation is relevantly *truth-insensitive*. One leading epistemologist writes as follows. Consider a "typical case" where there is "nothing odd" and "things are exactly as the person believes them to be." This would be Walt's case. Compare that to an "unusual case" where "the person has that very same evidence, but the proposition in question is nevertheless false." This would be Skyler's case. The "key thing to note" here is that in each case the person "has exactly the same reasons for believing exactly the same thing." Consequently, if the person has a good reason to believe the proposition in either case, then the person has a good reason in both cases. Their evidence and justification are equivalent across the cases,

despite the variation in truth. This way of thinking about evidence and justification is said to be "extremely plausible" (quotes from Feldman 2003, p. 29; see also Conee & Feldman 2004).

Focusing on more extravagant examples, consider a staple of contemporary epistemology, the notorious thought experiment involving a radically deceived "brain in a vat:"

> Could you not be floating in a tank while super-psychologists stimulate your brain electrochemically to produce exactly the same experiences you are now having, or even to produce the whole sequence of experiences you have had in your lifetime thus far? (Nozick 1981, p. 167)

Commenting on cases like this, another leading epistemologist says that such victims are "justified" because they "possess *good reason* for their beliefs — indeed, precisely the same reasons that we ourselves possess for our beliefs about the world." The super-psychologist's "machinations" cause the victims to have highly misleading evidence, "but this has no tendency to show that those reasons are not good ones." This verdict on the case, we're told, is "beyond question at an intuitive level" (BonJour 2003, pp. 185-186). It has "considerable intuitive" appeal (Wedgwood 2002, p. 350). "The intuition" here, we are told, is that you and your victimized counterpart in the super-psychologist's tank have "equally good evidence" for your respective beliefs (Russell 2001, p. 38; compare Cohen 1984, pp. 281-2). We have an "overpowering inclination to think" that the two of you have equally good evidence (Fumerton 2006, p. 93).

Some philosophers disagree with the truth-insensitive approach to belief evaluation just described. For example, one philosopher writes that belief evaluations that "abstract entirely away from external factors [are] uninteresting." They are uninteresting because they undermine the point of belief evaluations in the first place. We are social beings who need information to act

and plan our lives, so it is "essential" to human life that we have ways to identify good informants. Epistemic evaluations serve this purpose. This is why the evaluations are sensitive to "accuracy and etiology" (Greco 2013, p. 334; compare Craig 1990). On this approach, considerations of the reliability of evidence and what a belief is based on are partly constitutive of our evaluative practices.

The opposing view just described arguably falls short of claiming that, on a given occasion, a belief's truth value affects the quality of the agent's evidence, what the agent should believe, or what the agent is justified, rational, reasonable, or responsible in believing. (I will sometimes say "what the agent should believe" as short for the longer disjunctive formulation.) But at least one philosopher has endorsed the view that truth profoundly affects what an agent should believe. According to this philosopher, "one is permitted to believe" a proposition "only if one knows that" the proposition is true (Sutton 2007, p. 19; see also Littlejohn 2013). Moreover, knowledge requires truth, so "there are no false justified beliefs" (Sutton 2007, p. 1). This view has been labeled an "extraordinary" and "dissident doctrine" lacking in "intuitive credentials" (Conee 2007). It deviates from "the usual" and "traditional" approach to questions about belief evaluation, according to which there is no connection between what an agent should believe and truth (Chisholm 1989, p. 76).

But is it so extraordinary that truth can directly affect whether a belief is justified? Is it revolutionary to propose that false beliefs are not justified or should not be held? Some philosophers propose that such views are exactly the opposite of extraordinary or revisionary. The extraordinary and revisionist view, they suggest, is the one cultivated in narrow circles within "the professional philosophical community" (Sutton 2007, p. vii). They also propose that "we gain insight

and understanding" in epistemology by "reflecting on our thinking and practices" of belief evaluation (Greco 2010, p. 4). From this perspective, it is revisionary to *deny* that truth affects whether a belief is justified. And it is *not* revisionary to claim that false beliefs are not justified.

A more principled reason for rejecting the relevance of truth to justification begins by connecting justification to blame (Ginet 1975; Chisholm 1977; Alston 1988). Consider Walt and Skyler again, each of whom believes that the company employs someone named "Gustavo." They base their respective beliefs on what the reliable employee list says. In Walt's case the list is accurate on this occasion, whereas in Sklyer's case it isn't. Skyler is no more blameworthy than Walt is for believing that someone named "Gustavo" is employed. But to be justified in holding a belief just is to be blameless in holding it. So Skyler is no less justified than Walt is. Analogous points hold for the "brain in a vat" example. So truth does not affect whether a belief is justified. To put the argument more succinctly: blamelessness is insensitive to the actual truth value on a particular occasion, and justification just is blamelessness, so justification is similarly truth-insensitive.

This paper addresses two questions. First, and primarily, is a truth-sensitive conception of belief evaluation revisionary or does it accurately reflect our ordinary evaluative practices? Second, and subordinately, is the ordinary evaluation of belief equivalent or intimately tied to the assignment or denial of blame? I report three experiments to help answer these questions. Experiment 1 tests whether belief evaluations are truth-sensitive when using six different types of normative vocabulary: evidence, justification, should, rationality, reasonableness, and responsibility. All six forms of evaluation were extremely truth-sensitive. Experiments 2 and 3 test belief evaluations in "brain in a vat" scenarios. A normally embodied human's beliefs were judged much

more favorably than a brain in a vat's. Judgments of blame were truth-insensitive, but they did not meaningfully contribute to belief evaluations or the assessment of quality of evidence.

Experiment 1

Method

Participants. Nine hundred and sixty-one U.S. residents were tested (aged 18–72 years, mean age = 31 years; 95% reporting English as a native language; 349 female). Participants were recruited and tested online using Amazon Mechanical Turk and Qualtrics and compensated $0.35 for approximately 2 minutes of their time. Repeat participation was prevented.

Materials and Procedure. Participants were randomly assigned to one of twenty-four conditions in a 2 (Truth Value: true, false) × 2 (Cover Story: watch, bang) × 6 (Evaluation: evidence, justified, should, rational, responsible, reasonable) between-subjects design. Each participant read a single story about an agent who gets evidence for a certain proposition. The Truth Value factor manipulated whether the proposition is true or false. I did not expect an effect of Cover Story and included it as a robustness check. Here are the stories for the Watch and Bang conditions (Truth Value manipulations in brackets and separated by a slash):

> (Watch) Maria is a watch collector who owns over ten thousand watches. She cannot keep track of all her watches by memory alone, so her accountant maintains a detailed inventory of them. Maria knows that the inventory isn't perfect, but it is extremely accurate. Someone asks Maria, "Do you own a 1990 Rolex Submariner?" Maria consults the inventory and it says that she does have

one. In fact, she [does/does not] have one.

(Bang) Angelo is camping in a State park. Last night he watched the fire-works celebration over the lake. After a good night's sleep, he woke up in his tent shortly after sunrise. He goes to check the email on his smartphone when, all of a sudden, he hears two loud, sharp bangs ring out in the forest behind him. It is deer-hunting season. In fact, the two loud bangs [were/were not] gunshots.

After reading the story, participants answered a single test question. The Evaluation factor manipulated which test question was asked. Here are the different questions:

1. What does Maria's/Angelo's evidence justify her/him in believing?
2. What is Maria/Angelo justified in believing?
3. What should Maria/Angelo believe?
4. What is it rational for Maria/Angelo to believe?
5. What is the responsible thing for Maria/Angelo to believe?
6. What is it reasonable for Maria/Angelo to believe?

Responses were collected on a standard 6-point Likert scale anchored with the following left-to-right across the participants screen:

- "I definitely have one/Those definitely were gunshots" = 3
- "I have one/Those were gunshots" = 2
- "I probably have one/Those probably were gunshots" = 1
- "I probably do not have one/Those probably were not gunshots" = -1
- "I do not have one/Those were not gunshots" = -2
- "I definitely do not have one/Those definitely were not gunshots" = -3

Participants then advanced to a new screen and answered a comprehension question to ensure that they were reading the story carefully. Here are the comprehension questions for the two stories (options shown in brackets, rotated randomly):

- Who maintains the inventory? [Maria's accountant/Maria herself/Maria's collector]
- When Angelo heard the bangs, what was he going to do? [check his email/sleep in his tent/watch the celebration]

Participants filled out a brief demographic questionnaire after testing.

Results

Nearly 96% (919 of 961) of participants passed the comprehension question. I excluded from the analysis those who failed, but including them does not affect the pattern of results reported below.

An analysis of variance revealed that response to the test question was unaffected by participant age, participant gender, and Evaluation type. There was a large effect of True Value, $F(1, 844) = 293.70$, $p < .001$, $\eta_p^2 = .259$, all reported tests two-tailed. (See Fig. 1.) There was a marginally significant, extremely small, and unpredicted effect of Cover Story, $F(1, 844) = 3.86$, $p = .051$, $\eta_p^2 = .005$. This unpredicted and minuscule effect is interesting only insofar as it illustrates the robustness of the effect of truth value across multiple narrative contexts; I will ignore it in what follows.

Follow-up analysis revealed that mean response to the test question was significantly higher in true conditions ($M = 1.95$, $SD = 0.94$) than false conditions ($M = 0.39$, $SD = 1.64$), independent samples t-test, $t(715.92) = 17.64$, $p < .001$. The magnitude of the mean difference was ex-

tremely large, MD = 1.56, 95% CI [1.39, 1.73], Cohen's d = 1.31. Mean response in true conditions was significantly above the theoretical neutral midpoint (= 0), $t(464) = 44.97$, $p < .001$, MD = 1.95, 95% CI [1.87, 2.04], Cohen's d = 2.09 (extremely large effect size); it was above the neutral midpoint in false conditions, $t(453) = 5.09$, $p < .001$, MD = 0.39, 95% CI [0.24, 0.54], Cohen's d = 0.24 (small effect size). In true conditions, the modal response for each Evaluation type was for the agent to adopt the flat-out belief (i.e. "I have one" or "Those were gunshots"). In false conditions, the modal response for each Evaluation type was for the agent to adopt the probabilistic belief (i.e. "I probably have one" or "Those probably were gunshots"). (See Fig. 2.)

It's also worth considering the proportion of participants who recommended at least flat-out belief. Participants recommend at least flat-out belief when they recommend believing that the target proposition is either true (i.e. "I have one" or "Those were gunshots") or definitely true (i.e. "I definitely have one" or "Those definitely were gunshots"). In true conditions, 72% of participants made flat-out recommendations, whereas only 28% did in false conditions. (See Fig. 1.) Being assigned to a true rather than false condition increased the odds of a flat-out recommendation by over 770% (or a factor of 7.71), binomial logistic regression, $p < .001$, 95% CI for the odds ratio [5.75, 10.35].

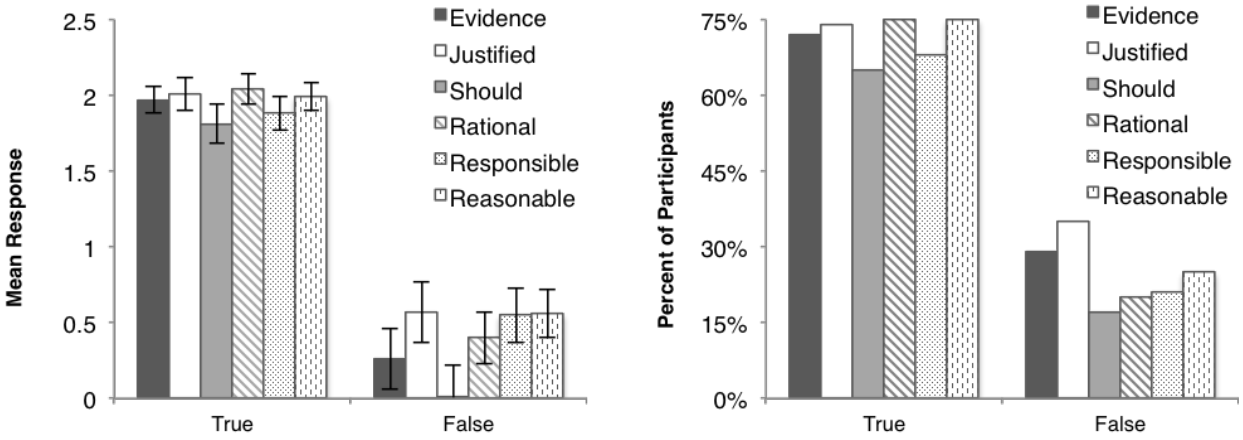

Fig. 1. Experiment 1. Left panel: mean response to the test question in True and False conditions across six Evaluation types; scales ran -3 to +3; error bars +/- one SEM. Right panel: percent of participants offering flat-out recommendations for belief.

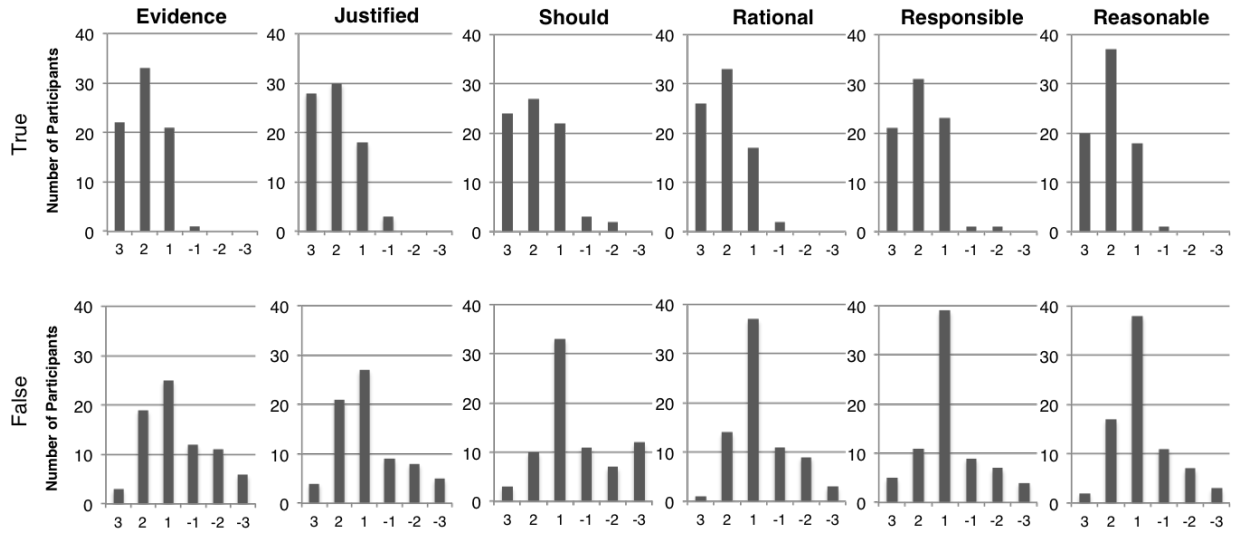

Fig. 2. Experiment 1. Distribution of responses to the test question across twelve conditions (collapsing across Cover Story). 3 = definitely true; 2 = true; 1 = probably true; -1 = probably false; -2 = false; -3 = definitely false.

Discussion

In this experiment, the ordinary evaluation of belief and evidence was affected by the truth value of the relevant proposition. The size of truth's effect was extremely large. I observed this same basic pattern across different narrative contexts and when using different normative vocabulary to prompt evaluations.

Experiment 2

Proponents of truth-insensitive epistemology often claim thought experiments involving radically deceived agents, such as "brains in vats," provide powerful intuitive support for their view. This experiment investigates how people evaluate the beliefs of a brain in a vat. It also tests whether their evaluations are predicted by judgments about blame.

Method

Participants. Two hundred and ten new participants were tested (aged 18 to 58 years, mean age = 29 years; 98% reporting English as a native language; 54 female). Participants were recruited and tested the same way as in Experiment 1, except that this time they were compensated $0.40 for closer to three minutes of their time. I offered higher compensation because participants in this experiment were being asked to read more text and answer more questions.

Materials and Procedure. Participants were randomly assigned to one of three conditions — Normal, False, and Brain — in a between-subjects design. Each participant read a single story. Here are the stories for the three conditions

(Normal/False) Victor is a healthy human adult sitting on his patio in a fine neighborhood. Victor is currently enjoying a variety of perfectly vivid sensory experiences, thanks to a team of scientists who helped save his life with a supercomputer that detected a heart condition. Victor was unaware that scientists could do that, just as he was unaware that he had a bad heart condition in the first place. Everything seems perfectly normal to him now. The scientists monitor him regularly. ¶[1] As Victor sits there on his patio, it seems as though a reddish four-legged animal with pointy ears and bushy tail is walking through a nearby flowerbed. To Victor, it seems like he is looking at a fox. His experiences seem entirely natural. [And/But] things [are exactly/are not] as they seem to Victor: as he sits there on his patio, he [is/is not] looking at a fox.

(Brain) Victor is a healthy human brain sitting in a vat of fluid in a fine laboratory. Victor is currently enjoying a variety of perfectly vivid sensory experiences, thanks to a team of scientists creating them through a supercomputer that is hooked up to Victor. Victor was unaware that scientists could do that, just as he was unaware that his body died and he was put in a vat in the first place. Everything seems perfectly normal to him now. The scientists monitor him regularly. ¶ As Victor sits there in his vat, it seems as though a reddish four-legged animal with pointy ears and bushy tail is walking through a nearby flowerbed. To Victor, it seems like he is looking at a fox. His experiences seem entirely natural. But things are not as they seem to Victor: as he sits there in his vat, he is not looking

[1] Indicates a paragraph break on the participant's screen.

at a fox.

After reading the story, participants answered a series of questions.

1. Victor _____ believe that he is looking at a fox. (should/only thinks he should)
2. Victor _____ for thinking that he is looking at a fox. (has good evidence/only thinks he has good evidence)
3. Victor _____ that he is looking at a fox. (knows/only thinks he knows)
4. Could Victor be properly blamed for thinking that he is looking at a fox? (Yes/No)
5. Victor _____ looking at a fox. (was/was not)
6. Victor was _____. (on a patio/in a laboratory)

Each of the first four questions was answered on a new screen while the story remained at the top of the screen. The final two questions were comprehension questions; they were answered from memory together on a new screen. All response options were rotated randomly.

Results

Approximately 80% (165 of 210) of participants passed the comprehension questions. I eliminated from the analysis forty-five participants who failed a comprehension question. Including them does not change the pattern of results reported below.

Assignment to condition affected response to the should question, $\chi^2(2, 165) = 53.64$, $p < .001$, Cramer's V = .570 (large effect size), the evidence question, $\chi^2(2, 165) = 40.21$, $p < .001$, Cramer's V = .494 (large effect size), and the knowledge question, $\chi^2(2, 165) = 42.85$, $p < .001$, Cramer's V = .510 (large effect size). But assignment to condition did not affect response to the blame question, $\chi^2(2, 165) = 0.83$, $p = .659$.

Follow-up pairwise comparisons revealed that participants in Normal were significantly more likely than participants in False to answer that the agent should believe the proposition (91% to 10%), $\chi^2(1, 107) = 48.34$, $p < .001$, Cramer's V = .672 (large effect size), that the agent has good evidence (79% to 29%), $\chi^2(1, 107) = 26.9$, $p < .001$, Cramer's V = .501 (large effect size), and that the agent knows the proposition (68% to 13%), $\chi^2(1, 107) = 31.21$, $p < .001$ Cramer's V = .540 (large effect size); but there was no difference in how frequently participants answered that the agent could be blamed for holding the belief (22% to 16%), $\chi^2(1, 107) = 0.82$, $p = .462$.

Participants in Normal also were more likely than participants in Brain to answer that the agent should believe the proposition (91% to 39%), $\chi^2(1, 120) = 34.20$, $p < .001$, Cramer's V = .534 (large effect size), that the agent has good evidence (79% to 28%), $\chi^2(1, 120) = 31.95$, $p < .001$, Cramer's V = .516 (large effect size), and that the agent knows the proposition (67% to 21%), $\chi^2(1, 120) = 26.81$, $p < .001$, Cramer's V = .473 (large effect size); but there was no difference in how frequently participants answered that the agent could be blamed for holding the belief (22% to 19%), $\chi^2(1, 120) = 0.24$, $p = .659$.

Participants in False and Brain did not differ in their response to the should question, $\chi^2(1, 103) = 2.65$, $p = .139$, the evidence question, $\chi^2(1, 103) = 0.02$, $p = 1$, the knowledge question, $\chi^2(1, 103) = 0.95$, $p = .435$, or the blame question, $\chi^2(1, 103) = 0.20$, $p = .795$.

The percentage of participants who answered that the agent should believe the proposition was above chance rates in Normal, $\chi^2(1, 62) = 40.32$, $p < .001$, below chance rates in False, $\chi^2(1, 45) = 11.76$, $p < .001$, and suggestively but non-significantly below chance rates in Brain, $\chi^2(1, 58) = 2.48$, $p = .11$. The percentage of participants who answered that the agent has good evi-

dence was above chance rates in Normal, $\chi^2(1, 62) = 20.90$, $p < .001$, and below chance rates in False, $\chi^2(1, 45) = 8.02$, $p < .005$, and in Brain, $\chi^2(1, 58) = 11.66$, $p < .001$. The percentage of participants who answered that the agent knows the proposition was above chance rates in Normal, $\chi^2(1, 62) = 7.81$, $p = .005$, and below chance rates in False, $\chi^2(1, 45) = 24.20$, $p < .001$, and in Brain, $\chi^2(1, 58) = 19.93$, $p < .001$. The percentage of participants who answered that the agent could be blamed for holding the belief was below chance rates in Normal, $\chi^2(1, 62) = 18.65$, $p < .001$, in False, $\chi^2(1, 45) = 21.36$, $p < .001$, and in Brain, $\chi^2(1, 58) = 22.35$, $p < .001$.

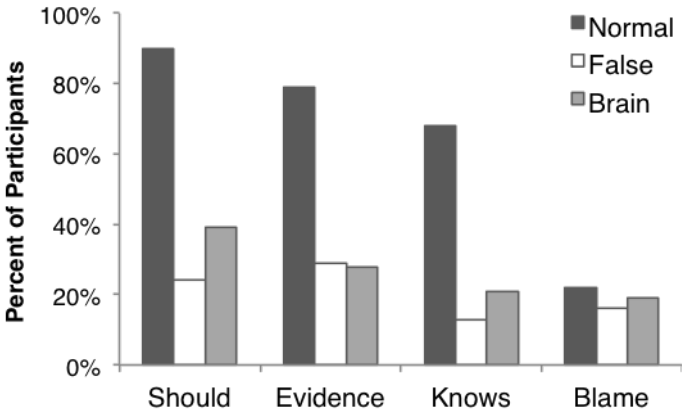

Fig. 3. Experiment 2. Percent of participants across three conditions who answered that the agent should believe the proposition, has good evidence for the proposition, knows the proposition, and could be blamed for believing the proposition.

To further explore the relationship between belief evaluation, truth, evidence, knowledge, and blame, I conducted a binary logistic regression analysis. (See Table 1.) The model contained answer to the should question as the outcome, and it contained assignment to condition and answers to the evidence, knowledge, and blame questions as predictors. Assignment to condition

and answers to the evidence and knowledge questions all made unique significant contributions to predicting the outcome. This means that, even when controlling for the judgment that the agent has good evidence, both condition (truth value) and knowledge judgments significantly predicted whether participants thought that the agent should believe the proposition. Blame judgments were not predictive — indeed, they were utterly inconsequential. Similarly, regression analyses showed that blame judgments did not predict evidence judgments (p = .474) or knowledge judgments (p = .171) either.

Table 1. Experiment 2. Logistic regression predicting the judgment that the agent should believe the proposition. Reference class for Condition: Brain. Reference class for Evidence and Knowledge: the "only thinks" contrasts. Reference class for Blame: the agent cannot be properly blamed.

Predictor	B	S.E.	Wald	df	*p*	Odds Ratio	95% C.I. for Odds Ratio	
							LLCI	ULCI
Condition								
False	-0.77	0.49	2.46	1	.117	0.46	0.18	1.21
Normal	1.70	0.57	8.99	1	.003	5.46	1.80	16.56
Evidence	1.44	0.46	10.07	1	.002	4.24	1.74	10.34
Knowledge	1.51	0.53	8.15	1	.004	4.53	1.61	12.77
Blame	-0.10	0.57	0.03	1	.857	0.90	0.30	2.73
Constant	-1.13	0.35	10.51	1	.001	0.32		

Discussion

The results replicate the basic finding from Experiment 1. Ordinary evaluations of belief were deeply affected by the truth value of the relevant proposition. I observed this same basic pattern using a completely new story based on a famous thought experiment from the literature about a

"brain in a vat." Moreover, regression analysis revealed the same basic pattern even when controlling for the influence of evidence on the evaluations of belief. Knowledge judgments significantly affected evaluations of belief, even when controlling for the influence of truth and evidence.

I also observed three noteworthy null results. First, blame judgments had no effect on the evaluation of belief or evidence. Second, although truth significantly affected evaluations of belief, evidence and knowledge, it did not affect evaluations of blame. Third, a fairly routine case of false belief was judged no differently from a case of radical deception in terms of what the agent should believe, how good his evidence is, and whether he knows.

Experiment 3

Some philosophers might be suspicious of the results from Experiment 2 because participants were not asked to make comparative judgments about the quality of evidence in the two cases. More specifically, many philosophers report having a powerful *comparative* intuition about a normally embodied human and his brain-in-a-vat counterpart, to wit, that they have equally good evidence. But in Experiment 2 no participant evaluated both the normal human and the brain in a vat. This experiment addresses the suspicion by having participants make comparative judgments of quality of evidence.

Method

Participants. Ninety-one new participants were tested (aged 19 to 74 years, mean age = 34 years; 98% reporting English as a native language; 37 female). Participants were recruited and

tested the same way as in earlier experiments, except that this time they were compensated $0.50 for three to four minutes of their time. I offered higher compensation because participants in this experiment were being asked to comprehend a longer and more complicated passage.

Materials and Procedure. Participants were randomly assigned to one of three conditions — Human First, Brain First, and Plain — in a between-subjects design. Participants in all conditions read the same story about two characters, Harvey and Louis. The story was closely based on the story used in Experiment 3 about Victor. Instead of having one group of people read about a normal human and another group read about a brain in a vat, this time each participant read a single story about both characters. Harvey is the healthy human and Louis is the brain in a vat. Here are the two blocks of text that made up the story:

> (Harvey) Harvey is a healthy human adult sitting on his patio in a fine neighborhood. Harvey is currently enjoying a variety of perfectly vivid sensory experiences, thanks to a team of scientists who helped save his life with a supercomputer that detected a heart condition. Harvey was unaware that scientists could do that, just as he was unaware that he had a bad heart condition in the first place. Everything seems perfectly normal to him now. The scientists monitor him regularly. ¶ As Harvey sits there on his patio, it seems as though a reddish four-legged animal with pointy ears and bushy tail is walking through a nearby flowerbed. To Harvey, it seems like he is looking at a fox. His experiences seem entirely natural. And things are exactly as they seem to Harvey: as he sits there on his patio, he is looking at a fox.
>
> (Louis) Louis is a healthy human brain sitting in a vat of fluid in a fine labo-

> ratory. Louis is currently enjoying a variety of perfectly vivid sensory experiences, thanks to a team of scientists creating them through a supercomputer that is hooked up to Louis. Louis was unaware that scientists could to that, just as he was unaware that his body died and he was put in a vat in the first place. Everything seems perfectly normal to him now. The scientists monitor him regularly. ¶ As Louis sits there in his vat, it seems as though a reddish four-legged animal with pointy ears and bushy tail is walking through a nearby flowerbed. To Louis, it seems like he is looking at a fox. His experiences seem entirely natural. But things are not as they seem to Louis: as he sits there in his vat, he is not looking at a fox.

I included three conditions to guard against possible order effects and priming effects caused by asking comprehension questions. Participants in the Human First condition read the Harvey block first and the Louis block second. Then they answered four comprehension questions and, on a separate screen, the key test question. The only difference for the Brain First condition was that participants read the Louis block first and the Harvey block second. Participants in the Plain condition were presented with the two blocks of text in randomized order and then answered only the test question.

These were the comprehension questions:

1. Is Harvey looking at a fox? (Yes/No)
2. Is Louis looking at a fox? (Yes/No)
3. Who is sitting on a patio? (Harvey/Louis/neither/both)
4. Who is sitting in a vat? (Harvey/Louis/neither/both)

The “Yes/No” and “Harvey/Louis” options were always rotated randomly. To avoid confusing

participants, the "neither/both" options for questions 3 and 4 always appeared after the "Harvey/Louis" options. This was the test question:

5. Who has better evidence for thinking that he's looking at a fox?

Responses to the test question were collected on a standard 7-point Likert scale anchored with "Definitely Harvey" (= 3), "Harvey," "Probably Harvey," "Neither" (= 0), "Probably Louis," "Louis," "Definitely Louis" (= -3).

Results

Participants in the Human First and Brain First conditions answered 95% of the comprehension questions correctly, indicating that understood the details of the story. There was no effect of condition on response to the test question (Human First, M = 1.80, SD = 1.19; Brain First, M = 1.48, SD = 1.15; Plain, M = 1.83, SD = 1.37), univariate analysis of variance, $F(2, 88) = 0.75$, $p = .478$, $\eta_p^2 = .017$. Mean response to the test question (M = 1.70, SD = 1.23) was significantly above the neutral midpoint, one sample t-test, $t(90) = 13.17$, MD = 1.70, 95% CI [1.45, 1.96], Cohen's d = 1.38 (very large effect size). Modal response was "Definitely Harvey." Overall, nearly 80% (72 of 91) of participants agreed or were inclined to agree that Harvey's evidence was better. Less than 20% of participants (18 of 91) thought that Harvey's and Louis's evidence was equally good. One participant answered that Louis's evidence was better.

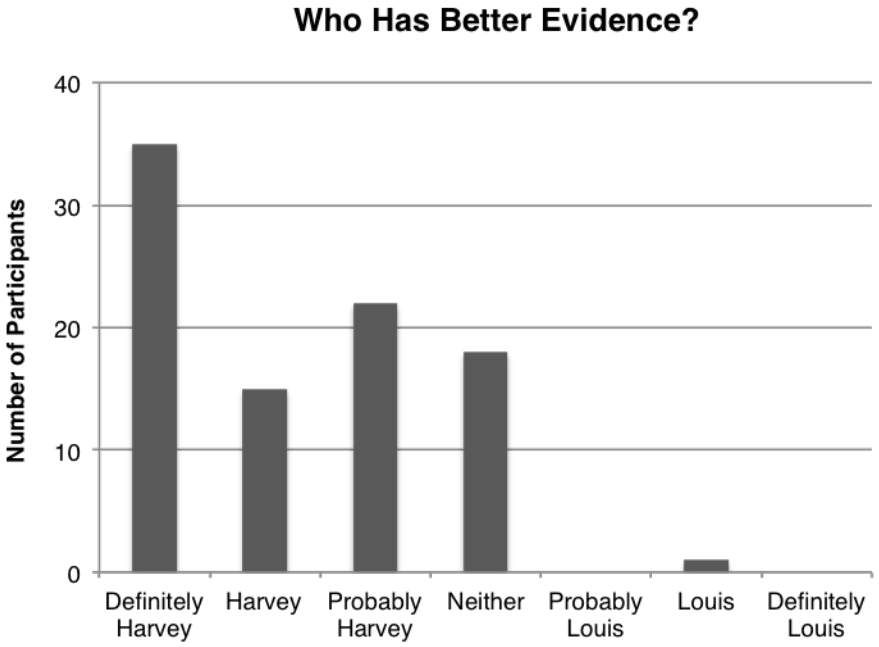

Fig. 4. Experiment 3. Number of participants selecting the various responses to the question, "Who has better evidence" for thinking that the relevant proposition is true? Harvey: a normally embodied human. Louis: a human "brain in a vat."

Discussion

The results replicate and generalize a main finding from Experiment 2. In Experiment 2, participants in a between-subjects design rated a normal human's evidence significantly better than a brain in a vat's evidence. In this experiment, participants did the same thing in a within-subjects design. This strongly supports the conclusion that, on the ordinary way of evaluating belief and evidence, a brain in a vat's evidence is inferior.

Conclusion

Many philosophers have emphatically endorsed a *truth-insensitivity hypothesis*: certain core, philosophically important evaluative properties of belief are insensitive to whether it is true. These allegedly truth-insensitive properties include whether a belief is justified, well supported by the agent's evidence, rational, reasonable, and responsible. The truth-insensitivity hypothesis is said to be "extremely plausible," "beyond question," and supported by our "intuition" — by our "overpowering inclination" to think that it's true (Feldman 2003, p. 29; BonJour 2003, p. 186; Russell 2001, p. 38; Fumerton 2006, p. 93). In short, the hypothesis is intuitively compelling and pretheoretically appealing. To deny the truth-insensitivity hypothesis is labelled radical, dissident, and counterintuitive (e.g. Conee 2007). But some philosophers have resisted these strong claims. They have suggested that the truth-insensitivity hypothesis itself is counterintuitive and violates commonsense principles of belief evaluation (e.g. Sutton 2007). Some have also provided principled accounts of why we should expect the truth-insensitivity hypothesis to violate commonsense principles (e.g. Greco 2010).

This paper investigated whether the truth-insensitivity hypothesis is counterintuitive and revisionary. The results were absolutely clear: ordinary belief evaluation is deeply truth-sensitive. Across three experiments, I observed a consistent pattern whereby belief and evidence were judged much better when the proposition in question was true rather than false. This was true in mundane cases and also in extraordinary thought experiments about "brains in vats" and their normally embodied counterparts. It was true for evaluations elicited with a wide range of normative vocabulary, including justification, evidence, rationality, reasonableness, responsibility, and what an agent should believe. It was true when people evaluated the brain in a vat and the normal

human separately and when they judged them simultaneously.

I also found strong initial evidence against a principled argument offered in support of the truth-insensitivity hypothesis. The argument has two key premises. First, whether a person can be blamed for holding a belief based on certain evidence is insensitive to the belief's truth value on this particular occasion. Second, for a belief to be justified just is for it to be blamelessly held. From these two premises the truth-insensitivity hypothesis follows for the property of justification. I found evidence that commonsense endorses the argument's first premise: blame judgements were truth-insensitive. However, I also found that blame judgments didn't predict evaluations of belief or quality of evidence. Indeed, blame judgments were the only evaluative judgments for which I found any evidence of truth-insensitivity.

There are at least two ways to explain the overall results. On the one hand, proponents of the truth-insensitivity hypothesis might be relying on idiosyncratic intuitions about the evaluation of belief and evidence, both when arguing against competing theories and when discussing fanciful thought experiments about envatted brains and Cartesian demons (e.g., Lehrer & Cohen 1983; BonJour 2002; Feldman 2003). On the other hand, they might be relying on perfectly normal intuitions about *blamelessness*, which they either misunderstand or misdescribe using vocabulary that, as it turns out, actually expresses highly truth-sensitive forms of evaluation.

Either way, the present findings have implications for epistemological theorizing moving forward. Epistemology, perhaps more than any other branch of analytic philosophy, prides itself on respecting patterns in ordinary thought and talk (e.g. Locke 1690, book 4.11; Reid 1764; Moore 1959; Austin 1956; Goldman 1979; Goldman 1993; Dretske 1981; Cohen 1988; Cohen 2013; DeRose 1995; DeRose 2009; Fantl & McGrath 2002; Fantl & McGrath 2009; Fantl & Mc-

Grath 2013; Hawthorne & Stanley 2008; Greco 2010; Turri 2011; Turri 2013; Turri 2014). But recent work has uncovered several critical junctures where core debates in contemporary epistemology have stalled and even degenerated because of mistaken claims about commonsense. For example, for decades philosophers have repeatedly asserted that commonsense epistemology is committed to an epistemic closure principle (e.g. Stroud 1984; Vogel 1990; Hawthorne 2004). These assertions have been made and granted without any evidence. Even those who reject epistemic closure have assumed that this was a demerit that must be outweighed by theoretical benefits (e.g. Dretske 1970; Dretske 2013; Nozick 1981). But when the matter was studied experimentally, it turned out that commonsense actually rejects closure (Turri 2015a; Turri 2015b).

To take another example, for decades philosophers have assumed that knowledge must be reliably produced (e.g. Goldman 1976; Goldman 2012; Dretske 1981; Sosa 1991; Sosa 2007; Greco 2010). This assumption was supported by claiming that commonsense epistemology is a form of "proto-reliabilism." But when the matter was studied experimentally, it turned out that commonsense implicitly rejects reliabilism, allowing for even highly unreliably produced knowledge (Turri under review a).

To take a third example, for decades philosophers have discussed "fake barn" cases. Philosophers have claimed that people are strongly inclined to deny knowledge in such cases (e.g. Goldman 1976; Pritchard 2005). A theory which treats "fake barn" cases differently is rejected out of hand, on the grounds that it is prohibitively counterintuitive (e.g. Pritchard 2008; Pritchard 2009). Those accused of countenancing a verdict of knowledge spend considerable time rebutting the accusation (e.g., Greco 2010, pp. 76-80). Even those who suggest that, ultimately, we should count them as instances of knowledge feel compelled to justify this verdict,

often at great length (e.g. Sosa 2007, lectures 2 and 5). But when the matter was studied experimentally, it turned out that "fake barn" cases were overwhelmingly judged to be cases of knowledge (Colaco, Buckwalter, Stich & Machery 2014; Turri, Buckwalter & Blouw 2014; Turri under review).

The present paper tells a similar story about the hypothesis of truth-insensitivity. For decades claims about "brains in vats" and other radically deceived agents have been treated with great seriousness and respect by professional epistemologists. An acceptable theory of knowledge, it has long been assumed, must either explain why a brain in a vat is equally justified (rational, reasonable, etc.) as a normally embodied counterpart, or explain why this "parity" intuition is so compelling despite being false. Even philosophers who ultimately want to deny that a brain in a vat is equally justified expend considerable time and energy trying to accommodate or diagnose the parity intuition (e.g. Goldman 1986; Goldman 1988; Sosa 1991; Sosa 2003; Comesaña 2002; see Littlejohn 2009 for illuminating discussion). But this is all a huge mistake. Decades of debate have been based on a myth. The parity intuition deserves no deference. It is idiosyncratic and goes against the central tendencies in commonsense epistemology. The only detectable "overpowering inclination" is to reject the parity intuition and the idiosyncratic forms of truth-insensitive epistemology that go along with it.

Proponents of truth-sensitive epistemology have proposed an explanation for why belief evaluations are truth-sensitive. The explanation appeals to the social function of belief evaluations, which is to flag good sources of information for purposes of planning and reasoning (Greco 2010; see also Craig 1990). Given the evaluations' purpose, we should not expect them to be truth-insensitive. This is a suggestive and promising starting-point to be elaborated and de-

fended further in future work (for related discussion and findings regarding the evaluation of decision-making, see Turri in press).

Of course, proponents of truth-insensitive epistemology might propose new arguments to think that belief evaluation, beyond blamelessness, *should* be insensitive to truth. Such attempts are to be welcomed and should be judged fairly on their merits. But lazy appeals to intuition and armchair generalizations about psychological tendencies should be excluded from serious discussion and, indeed, actively discouraged. Epistemologists have already wasted too much time developing idle explanations of spurious "facts," journals have devoted too much space to discussions predicated on myth, and too many students over several generations have had their own intuitions and commonsense marginalized and disregarded. Future attempts to support truth-insensitive evaluations should begin by acknowledging the revisionist's burden, which for so long has been mistakenly thought to rest with the opposition.

Acknowledgments — For helpful feedback, I thank Wesley Buckwalter, Clayton Littlejohn, David Rose, and Angelo Turri. This research was supported by the Social Sciences and Humanities Research Council of Canada and an Early Researcher Award from the Ontario Ministry of Economic Development and Innovation.

References

Alston, W. P. (1988). The deontological concept of epistemic justification. Philosophical Perspectives, 2, 257–299.

Austin, J. L. (1956). A plea for excuses. Proceedings of the Aristotelian Society, 57, 1–30.

BonJour, L. (2002). *Epistemology: classic problems and contemporary responses*. Rowman & Littlefield.

BonJour, L., & Sosa, E. (2003). *Epistemic justification: Internalism vs. externalism, foundations vs. virtues*. Malden, Mass.: Blackwell.

Chisholm, R. (1989). Theory of Knowledge (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Chisholm, R. (1989). *Theory of knowledge* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Cohen, S. (1984). Justification and truth. Philosophical Studies, 46(3), 279–295. doi:10.1007/BF00372907

Cohen, S. (1988). How to be a Fallibilist. Philosophical Perspectives, 2, 91–123.

Cohen, S. (2013). Contextualism defended. In M. Steup, J. Turri, & E. Sosa (Eds.), *Contemporary debates in epistemology* (2nd ed., pp. 69–75). Malden, Mass.: Wiley-Blackwell.

Colaco, D., Buckwalter, W., Stich, S., & Machery, E. (2014). Epistemic intuitions in fake-barn thought experiments. Episteme, 11(02), 199–212. doi:10.1017/epi.2014.7

Comesaña, J. (2002). The diagonal and the demon. Philosophical Studies, 110(2), 249–266.

Conee, E. (2007). Review of Jonathan Sutton, *Without Justification*. Notre Dame Philosophical Reviews, 12. Retrieved from http://ndpr.nd.edu/review.cfm?id=11803

Conee, E., & Feldman, R. (2004). *Evidentialism: essays in epistemology*. Oxford: Oxford University Press.

Craig, E. (1990). *Knowledge and the state of nature: an essay on conceptual synthesis*. Oxford: Oxford University Press.

DeRose, K. (1995). Solving the skeptical problem. The Philosophical Review, 104(1), 1–52.

DeRose, K. (2009). *The case for contextualism*. Oxford: Oxford University Press.

Dretske, F. I. (1970). Epistemic operators. The Journal of Philosophy, 67(24), 1007–1023.

Dretske, F. I. (1981). *Knowledge and the flow of information*. Cambridge, Mass.: MIT Press.

Dretske, F. I. (2013). The case against closure. In M. Steup, J. Turri, & E. Sosa (Eds.), Contemporary debates in epistemology (2nd ed.). Wiley-Blackwell.

Fantl, J., & McGrath, M. (2002). Evidence, pragmatics, and justification. The Philosophical Review, 111(1), 67–94.

Fantl, J., & McGrath, M. (2009). *Knowledge in an uncertain world*. Oxford University Press.

Fantl, J., & McGrath, M. (2013). Practical factors affect whether you know. In M. Steup, J. Turri, & E. Sosa (Eds.), *Contemporary debates in epistemology* (2nd ed., pp. 84–95). Wiley-Blackwell.

Feldman, R. (2003). *Epistemology*. Upper Saddle River, NJ: Prentice Hall.

Fumerton, R. (2006). *Epistemology*. Malden, MA: Blackwell.

Ginet, C. (1975). *Knowledge, perception, and memory*. Dordrecht: D. Reidel.

Goldman, A. I. (1976). Discrimination and perceptual knowledge. Journal of Philosophy, 73(20), 771–791.

Goldman, A. I. (1979). What is justified belief? In G. Pappas (Ed.), *Justification and knowledge*. Dordrecht: Reidel.

Goldman, A. I. (1986). *Epistemology and cognition*. Harvard University Press.

Goldman, A. I. (1988). Strong and weak justification. Philosophical Perspectives, 2, 51–69.

Goldman, A. I. (1993). Epistemic folkways and scientific epistemology. Philosophical Issues, 3, 271–285.

Goldman, A. I. (2012). Reliabilism and contemporary epistemology. Oxford: Oxford University

Press.

Greco, J. (2010). *Achieving knowledge: a virtue-theoretic account of epistemic normativity*. Cambridge: Cambridge University Press.

Greco, J. (2013). Justification is not internal. In M. Steup, J. Turri, & E. Sosa (Eds.), *Contemporary debates in epistemology* (2nd ed.). Malden, Mass.: Wiley-Blackwell.

Hawthorne, J. (2004). *Knowledge and lotteries*. Oxford: Oxford University Press.

Hawthorne, J., & Stanley, J. (2008). Knowledge and action. Journal of Philosophy, 105(10), 571.

Lehrer, K., & Cohen, S. (1983). Justification, truth, and coherence. Synthese, 55(2), 191–207.

Littlejohn, C. (2009). The new evil demon problem. Internet Encyclopedia of Philosophy.

Littlejohn, C. (2013). The Russellian retreat. Proceedings of the Aristotelian Society, 113(3), 293–320. doi:10.1111/j.1467-9264.2013.00356.x

Locke, J. (1975). *An essay concerning human understanding*. (P. H. Nidditch, Ed.). Oxford: Clarendon Press.

Moore, G. E. (1959). *Philosophical papers*. New York: Collier Books.

Nozick, R. (1981). *Philosophical explanations*. Cambridge, Mass: Harvard University Press.

Pritchard, D. (2005). *Epistemic luck*. New York : Oxford University Press.

Pritchard, D. (2008). Greco on knowledge: virtues, contexts, achievements. The Philosophical Quarterly, 58(437-447), 1–11.

Pritchard, D. (2009). Apt performance and epistemic value. Philosophical Studies, 143(3), 407–416. doi:10.1007/s11098-009-9340-7

Reid, T. (1997). *An inquiry into the human mind on the principles of common sense*. (D. R. Brookes, Ed.). University Park, Penn.: Pennsylvania State University Press.

Russell, B. (2001). Epistemic and moral duty. In M. Steup (Ed.), *Knowledge, truth, and duty: essays on epistemic justification, responsibility, and virtue* (pp. 34–48). Oxford: Oxford University Press.

Sosa, E. (1991). Knowledge in Perspective. Cambridge: Cambridge University Press.

Sosa, E. (2007). *A virtue epistemology: apt belief and reflective knowledge, volume I*. Oxford University Press.

Stroud, B. (1984). *The significance of philosophical skepticism*. Oxford: Clarendon Press.

Sutton, J. (2007). *Without justification*. MIT Press.

Turri, J. (2011). The express knowledge account of assertion. Australasian Journal of Philosophy, 89(1), 37–45. doi:10.1080/00048401003660333

Turri, J. (2013). The test of truth: An experimental investigation of the norm of assertion. Cognition, 129(2), 279–291. doi:10.1016/j.cognition.2013.06.012

Turri, J. (2014). Knowledge and suberogatory assertion. Philosophical Studies, 167(3), 557–567. doi:10.1007/s11098-013-0112-z

Turri, J. (2015a). An open and shut case: epistemic closure in the manifest image. Philosophers' Imprint, 15(2), 1–18.

Turri, J. (2015b). Skeptical appeal: the source-content bias. Cognitive Science, 39(2), 307–324. doi:10.1111/cogs.12153

Turri, J. (in press). Evidence of factive norms of belief and decision. Synthese.

Turri, J., Buckwalter, W., & Blouw, P. (2014). Knowledge and luck. Psychonomic Bulletin & Review. doi:10.3758/s13423-014-0683-5

Vogel, J. (1990). Are there counterexamples to the closure principle? In M. D. Roth & G. Ross

(Eds.), Philosophical Studies Series (Vol. 48, pp. 13–27). Springer. doi:10.1007/978-94-009-1942-6_2

Wedgwood, R. (2002). Internalism explained. Philosophy and Phenomenological Research, 65(2), 349–369.