The test of truth: An experimental investigation of the norm of assertion

John Turri *

Philosophy Department, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L3G1, Canada

ARTICLE INFO

ABSTRACT

Assertion is fundamental to our lives as social and cognitive beings. Philosophers have recently built an impressive case that the norm of assertion is *factive*. That is, you should make an assertion only if it is *true*. Thus far the case for a factive norm of assertion been based on observational data. This paper adds experimental evidence in favor of a factive norm from six studies. In these studies, an assertion's truth value dramatically affects whether people think it should be made. Whereas nearly everyone agreed that a true assertion supported by good evidence should be made, most people judged that a false assertion supported by good evidence should not be made. The studies also suggest that people are consciously aware of criteria that guide their evaluation of assertions. Evidence is also presented that some intuitive support for a non-factive norm of assertion comes from a surprising tendency people have to misdescribe cases of blameless rule-breaking as cases where no rule is broken.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Assertion is fundamental to our lives as social and cognitive beings. By asserting we share information, coordinate behavior, and advance collective inquiry. Because assertion is so important, it is unsurprising that it is governed by cognitive or epistemic norms. Some assertions should be made and some should not. But what distinguishes the ones that should be made from the ones that should not? In short, what is the norm of assertion? Should any false assertions be made? Or should we make an assertion only if it is true? In short, is the norm of assertion *factive*?

Over the past decade, in philosophy there has been an explosion of interest in the norms of assertion, all aimed at answering the question, what is the norm of assertion? Experimental cognitive science has not directly addressed the question, but some have indicated that they think the answer is straightforward. For example, consider this recent passage by two leading cognitive scientists:

> [L]anguage is not only conventional in the sense of involving socially shared arbitrary sound-meaning assignments; language is a rule-governed, normatively structured practice. Given the conventional rules that constitute a language, there are appropriate and inappropriate, right and wrong uses of the linguistic devices in making speech acts (*with assertions, for example, being governed by norms of truth*) that apply to all speakers, including the self, in an agent-general manner. (Rakoczy & Tomasello, 2009: 205; emphasis added)

Rakoczy and Tomasello then report a fascinating experiment designed to test whether children as young as 2 years old can 'differentially track kinds of mistakes' for different types of speech act, in particular commands and assertions. When it comes to assertion, the researchers treat *falsity* as 'inappropriate', 'wrong', and meriting 'normative intervention' even by a child's lights (209–10). In

* Tel.: +1 519 888 4567x32449; fax: +1 519 746 3097.
E-mail address: john.turri@gmail.com

short, even a 2-year-old can be expected to recognize that assertion has a factive norm.

Rakoczy, Tomasello and perhaps many others think it's obvious that assertion's norm is factive, and, as we shall see momentarily, an impressive case has been built in support of a factive norm. Nevertheless, many claim to find a factive norm counterintuitive and propose alternative accounts. They argue that truth is too strong of a standard and that the relevant standard is instead what you *believe* to be true or what you have *good evidence* for. Consider Grice's famous Maxims of Quality, whose 'Supermaxim' is, 'Try to make your contribution one that is true,' and whose sub-maxims are, '1. Do not say what you believe to be false. 2. Do not say that for which you lack adequate evidence' (Grice, 1989: 27). Here the emphasis is on avoiding mendacity and on adequate evidence for the truth of an assertion, not truth *per se*. Wilson and Sperber (2012: 49, 47) go one step further than Grice and contend that our linguistic practices are governed by 'no maxim of truthfulness at all.' In light of such disagreement, it's no trivial question whether assertion's norm is factive.

Although there has been no experimental work in this area, empirical considerations have shaped the debate over assertion's norm, albeit at a fairly high level of abstraction. If assertion is indeed a rule-governed activity—as virtually everyone agrees that it is—then we should look to the behavior of competent language users as a guide to what the rules are. Of course, assertion does not come with a rule book any more than natural languages come with grammar guides. Nor do we expect competent language users to have an explicit, articulable theory about the rules of assertion any more than we expect them to have such a theory about the rules of syntax. But we can gain evidence about the rules of assertion by looking for distinctive patterns in two related areas: the normal give-and-take surrounding assertion and competent speakers' normative intuitions about assertions. It is an empirical matter whether such patterns exist and, if so, what they are like.

Consider the primary case made for the leading theory in this area, *the knowledge account of assertion*, which says that knowledge is the norm of assertion. Since knowledge requires truth, knowledge is *ipso facto* a factive norm. The best and most widely recognized evidence for the knowledge account is *an explanatory argument from linguistic patterns* (Turri, 2010b, 2011a; Williamson, 2000). I cannot canvass all the relevant evidence here, but in order to give the reader an idea of why the knowledge account has gained so much support, I will briefly discuss the most suggestive data points.

First, one way of prompting someone to make an assertion is to ask, 'What time is it?'. But an equally effective, and practically interchangeable, prompt is to ask, 'Do you know what time it is?'. Competent speakers respond to the two questions similarly. But why would that be? Proponents of the knowledge account explain it as follows. Because knowledge is the norm of assertion, my question 'Do you know what time it is?' enables you to infer that I want you to make the relevant assertion and, thus, functions as *an indirect request* for you to make the assertion. This is similar to how my question 'Can you pass the salt?' can function as an indirect request for you to pass the salt.

Second, when you're asked a question, even if the question has nothing to do with you or what you know, it is normally okay to respond by saying, 'Sorry, I don't know.' Suppose you're asked, 'What is the conversion rate from liters to quarts?', and you respond, 'Sorry, I don't know.' Normally, your response will be judged perfectly acceptable. But you and your epistemic state are irrelevant to the content of the question, so why is that response any more acceptable than, say, 'Sorry, Paris is the capital of France' would be? Proponents of the knowledge account explain it as follows. By saying 'Sorry, I don't know', you're informing the questioner that you lack the appropriate normative standing to answer her question, which is surely relevant in the context.

Third, when you make an assertion, even if the content of your assertion has nothing to do with you or what you know, it is normally appropriate to ask you, 'How do you know that?'. What explains the default propriety of this response? Proponents of the knowledge account explain it as follows. By making an assertion, you represent yourself as satisfying the norm of assertion; and knowledge is the norm; so the question is appropriate because it asks whether you're accurately representing yourself.

Fourth, asking 'How do you know?' is understood as implicitly challenging my authority to make an assertion. More aggressive than asking 'How do you know?' is 'Do you really know that?'. More aggressive yet is 'You don't know that!' or 'You don't know what you're talking about'. What explains this range of aggressiveness? Proponents of the knowledge account explain it as follows. 'How do you know?' implicitly challenges my authority to assert Q, by asking me to demonstrate that I do in fact have it; 'Do you really know that?' explicitly challenges my authority, by questioning whether I have it; and 'You don't know that!' explicitly rejects my authority. Explicitly questioning someone's authority is more aggressive than implicitly questioning it, and explicitly rejecting someone's authority is more aggressive than explicitly questioning it.

Fifth, assertions of the form 'Q, but I don't know that/ whether Q' strike us as inconsistent. But their *content* is perfectly consistent, so why do they seem inconsistent? Proponents of the knowledge account explain it as follows. Knowledge is the norm of assertion, so in order to properly assert a conjunction of the form, 'Q, but I don't know that/ whether Q', you must know each conjunct. But your knowing the first conjunct ('Q') would falsify the second conjunct ('I don't know whether Q'), in which case you could not possibly know the conjunction. And by asserting the conjunction, you represent yourself as knowing it (because you represent yourself as satisfying the norm of assertion). So by asserting the conjunction, you represent yourself as knowing something that you could not possibly know, which explains the inconsistency. In a word, what you assert is inconsistent with how you represent yourself.

In sum, the overall case for the knowledge account is impressive, based on data gleaned from social observation and introspection. These data have acted as a leash, fruitfully constraining as well as inspiring the theoretical debate. But it's quite a long leash (Turri, 2013). It would be useful to add to our data set relevant observations tempered by proper controls. This paper accomplishes that

by testing whether people's judgments about cases better fit a factive or non-factive norm of assertion. The results support the view that the norm is factive.

Before proceeding, I should say a few words about the sort of normativity at stake in the debate over assertion's norm, and a word about how the various proposals have been framed. I begin by characterizing the sort of normativity at stake and contrasting it with other sorts of normativity.

For the most part, when theorists ask what the norm of assertion is, they have something very specific in mind. They want to know what rule or rules help constitute the practice of assertion, make that form of activity possible. They are interested in *constitutive normativity* (Brandom, 1983; Rescorla, 2009; Searle, 1969, 1979). Constitutive norms or rules are internal to and distinctive of a practice. Consider chess as an example. Chess is the game that it is precisely because it has the rules that it does: the rules help constitute the game of chess, and distinguish it from other board games, such as checkers. On the current approach, something similar is true of assertion: assertion is the speech act that it is precisely because it has the rule that it does. When someone performs a particular speech act and that speech act is an assertion, what makes it an assertion rather than something else, such as a guess, conjecture or guarantee, all of which have the same 'direction of fit' as assertion (Anscombe, 1957; Turri, 2010a)? One answer is that assertion is *individuated normatively* (Williamson, 1996). Assertion is the unique speech act that has a certain standard of correctness, more stringent than that of guessing, less stringent than that of guaranteeing. This standard of correctness is the norm of assertion.

Constitutive normativity differs from moral and prudential normativity. For example, a particular move might be allowed by the rules of chess even if making that move would be immoral or imprudent because, say, your opponent is a child who will be devastated by the ensuing crushing defeat, or because your opponent is a short-tempered mobster who will break your fingers if you checkmate him. Similarly, the constitutive norm of assertion might allow assertions which are nevertheless forbidden by other norms. For example, it might be illegal in a certain country to assert that the emperor has no clothes, just as it might be illegal or imprudent to checkmate the emperor in a game of chess. But the norms that prohibit these assertions are not norms *of assertion*; they are not norms that help constitute the practice of assertion. Rather, they are norms that happen to pertain to particular assertions.

For these reasons, those who favor a non-factive norm of assertion are not content to point out that sometimes we should, in *some* sense of 'should', say something false, or that sometimes we should, all things considered, say something false. To object on such grounds is merely to confuse a more generic sense of 'should' with the 'should' associated with the norm of assertion. Such confusion is akin to noting that in some circumstances you should move your rook diagonally—say, on pain of torture, or because it will drive your opponent to distraction—and concluding that therefore the rules of chess allow the rook to move diagonally.

The literature on assertion's norm suffers from some terminological inconsistency. What is the right way to express assertion's constitutive norm? Some say, 'You should make an assertion only if …'; others say 'You ought to make an assertion only if …'; others say, 'You may make an assertion only if …'; and others say, 'You must: make an assertion only if …'. (See Turri, 2013 for more on the variety.) I'm not going to resolve this inconsistency here. Instead, I will simply opt for the 'should' formulation. And when probing laypeople I will stick to asking about what a speaker 'should' say or whether anything 'incorrect' has been done. It is legitimate to question whether different terminology would lead to different results. But choices must be made in order to get the project off the ground and I have chosen to start here. I welcome and encourage further work that makes different choices.

1.1. A synopsis of competing accounts

Theorists have defended two main factive accounts of assertion's norm:

The truth account you should assert Q only if Q is true (Weiner, 2005)

The knowledge account you should assert Q only if you know Q (Unger, 1975; Williamson, 2000, Hawthorne, 2004, Schaffer, 2008, Turri, 2011a, Benton, 2011)

The knowledge account is a factive account because knowledge requires truth. (There has recently been some controversy over this. See Hazlett, 2010, 2012; Turri, 2011b, and Buckwalter ms.) I have already explained the main argument for the knowledge account. For the remainder of this paper, I will simply ignore the differences between the two specific factive accounts, focusing instead on the factive genus.

Theorists have defended two main non-factive accounts of assertion's norm:

The belief account you should assert Q only if you believe Q. (Bach, 2008; Bach & Harnish, 1979)

The justification account you should assert Q only if believing Q is justified (or: reasonable) for you. (Douven, 2006; Lackey, 2007)

Belief is not factive, because you can believe something false. And virtually everyone agrees that justification is not factive either, because it seems obvious that you can be justified in believing something false. That is, *fallibilism* about justification is almost universally accepted (Turri and Friedman, 2013; Sutton, 2007 dissents).

Justification is fallible only if it is possible to have a justified false belief. To be clear, this is only a necessary condition, not a sufficient condition. To avoid near-total skepticism about justification, something *much* more than the *mere possibility* is needed. Pairing the justification account with an infallibilist conception of justification or an overly demanding fallibilist conception would completely undermine the primary motivation for a justification account.

1.2. Two counts against factive accounts: intuition and argument

Two objections have led many to reject a factive account and in its place endorse a justification account. Researchers who follow this path endorse a non-skeptical, fallibilist conception of justification. (Note: although the justification account and truth account are logically compatible, so too is it logically consistent to reject one and accept the other. Researchers who find the following two objections persuasive do just that.)

The first objection is simply that there are straightforward counterexamples to factive accounts. The counterexamples follow a recipe: a protagonist has good evidence for believing Q, believes Q on that basis, thereby gaining a justified belief that Q, which he then gives voice to by asserting Q. Now simply add to the story that, despite the good evidence, Q is false and, thus, the assertion is false. Taken to the extreme, the protagonist is a radically deceived brain-in-a-vat somehow able to make assertions, perhaps by wiggling a parietal lobe. (Compare Douven, 2006: 477, and Lackey, 2007: 607.) Factive accounts entail that the protagonist's assertion is incorrect. Opponents of factive accounts report having a strong intuition that such justified or reasonable false assertions are not incorrect *qua* assertions.[1] Call this the *non-factivity intuition* and theorists who endorse it *non-factivists*.

The second objection is a simple argument about justified false assertions. I will understand the argument to go like this: a speaker who makes a justified false assertion is not thereby properly criticizable or blameworthy; so the speaker has probably not broken a norm of assertion; so assertion probably does not have a factive norm (Lackey, 2007: 603, 597; Douven, 2006: 476–7; Hill & Schechter, 2007: 109).

The following experiments address both ways of resisting factive accounts: the non-factivity intuition and the argument from blamelessness to correct behavior.

2. Experiment 1A

Do competent speakers generally share the non-factivity intuition? Or does their assessment reflect an implicit commitment to the opposite *factivity intuition*, to wit, that a false assertion is *ipso facto* an assertion that should not be made? This section reports a very simple initial experiment to help answer these questions.

2.1. Participants, materials and procedure

Sixty-six participants (36 female, aged 18–60 years, *M* = 32 years, *SD* = 12 years) were recruited using Amazon Mechanical Turk and paid $0.20 for approximately 2–3 min of their time. Participants were located throughout the United States and 97% listed English as a native language. Participants were not allowed to take the survey

more than once. I excluded data from 5 participants who failed a comprehension question designed to ensure that they were reading the scenario carefully enough. But including these participants in the analysis yielded the same pattern of results described below. Participants were not allowed to re-take any survey and participants who had taken previous similar surveys were excluded by their AMT Worker ID. The procedure for recruiting and compensating participants was the same for each of the experiments reported below.

Participants were randomly assigned to one of two conditions: True and False. Participants read the following story (alternate text in brackets and separated by a slash):

> Maria is a watch collector. She owns so many watches that she cannot keep track of them all by memory alone. So she maintains a detailed inventory of them. She keeps the inventory up to date. Maria knows that the inventory is not perfect, but it is extremely accurate[2]. Today Maria is having guests over for dinner. Soon after dinner is served, one of her guests asks, "Maria, do you have a 1990 Rolex Submariner in your watch collection?" Maria consults her inventory. It says that she does have a 1990 Rolex Submariner in her collection. [But this is one of those rare cases where the inventory is wrong: she does not have one/And this is just another case where the inventory is exactly right: she does have one].

In both the False and True conditions, Maria is justified in thinking that she has a 1990 Rolex in her collection. She is justified for the same reason in each condition: because of an entry in the inventory that she knows to be extremely reliable. On most standard accounts of justification, Maria's justification is equally strong in both conditions. She has fallible but nevertheless very good justification for thinking that the watch is in her collection. Any difference between the two conditions cannot plausibly be explained by saying that Maria had better justification in one condition or the other.

Following the story, participants answered two comprehension questions, followed by a question about what Maria should say (the test question). These were the questions, with response options in brackets:

1. Is there a 1990 Rolex Submariner in Maria's collection? [Yes/No]
2. If Maria tells her guest that she has a 1990 Rolex Submariner in her collection, she will be saying something _____. [True/False]
3. Should Maria tell her guest that she has a 1990 Rolex Submariner in her collection? [Yes/No]

The questions were always asked in the same order. Response options were presented in random order. The questions appeared on a single screen and the story remained at the top of the screen throughout. Participants could change their answers while they remained on the screen. After testing, participants filled out a brief demographic ques-

---

[1] 'Qua assertion' because, as explained earlier, everyone agrees that they might be incorrect according to other rules that aren't constitutive of assertion, such as legal, prudential or moral rules.

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

tionnaire. These same procedures were also used in each of the experiments reported below.

2.2. Results and discussion

Response to the test question—about whether Maria should tell her guest that the watch is in her collection—will tell us whether participants share the non-factivity intuition. If they share the non-factivity intuition, then participant response probably will not differ between False and True conditions. The mere fact that the assertion would be false should not change their estimation of whether Maria should make it. By contrast, if they share the factivity intuition, then they will probably answer that Maria should not make the assertion in the False condition more than in the True condition.

There was a significant effect of condition: over 80% of participants in the False condition responded that Maria should not make the statement compared to only 3% in the True condition, Fisher's exact test, $p < .001$, all tests two-tailed unless otherwise noted. Moreover, participants in True responded that Maria should make the statement at rates far greater than would be expected by chance, just as participants in False said that she should not make the statement at rates far greater than chance, binomial, both $ps < .001$. This pattern is what we would expect if participants shared the factivity intuition but it's not what we would expect if they shared the non-factivity intuition.

It might be objected that although Maria is well justified in thinking that she has a 1990 Rolex in her collection, the comprehension questions focused only on truth-related matters, neglecting whether she was justified. Perhaps this led participants to ignore the fact that Maria is justified, which explains why participants failed to respond as non-factive accounts would predict. The next experiment addresses this concern.

3. Experiment 1B

What would happen if, before asking participants to evaluate whether Maria should make the assertion, we asked them only one question about Maria's justification, or only one question about what's true in the case, or no question at all about either of those things? Investigating this will affect how we should interpret the results from Experiment 1A. If changing or eliminating the question results in participants answering that Maria should make the false but well justified assertion, then that will provide some relief for non-factivists. But if it does not result in that, then it will provide them only more grief. So which will it be: relief or grief?

3.1. Participants, materials and procedure

Ninety-two new participants (26 female, aged 18–65 years, $M = 27$ years, $SD = 9.5$ years, 98% native English speakers) were randomly assigned to one of three conditions: No Question, Reason Question, and Truth Question.

All participants read the False version of the story in Experiment 1A. Conditions differed in how participants were questioned. Following the story, in the Reason Question condition, participants were asked,

It is _____ for Maria to think that she has a 1990 Rolex Submariner in her collection. [reasonable/unreasonable]

followed by the test question,

Should Maria tell her guest that she has a 1990 Rolex Submariner in her collection? [Yes/No]

In the Truth Question condition, participants were asked,

Is there a 1990 Rolex Submariner in Maria's collection? [Yes/No] followed by the test question. Participants in No Question were asked *only* the test question.

Participants in all conditions were then taken to a separate screen (they could not go back and change previous answers at this point), where they were asked,

When you were deciding whether Maria should tell her guest that she has a 1990 Rolex Submariner in her collection, which factor seemed more important to you? [Maria's evidence about what was in her collection/The truth about what was in her collection]

The purpose of asking this question is to see whether participants are able to reflectively identify the criteria guiding their evaluation of an assertion. Previous research on *moral* evaluation has found that participants are sometimes very unreliable at identifying the criteria that guide their moral assessments (Cushman, Young, & Hauser, 2006).

3.2. Results and discussion

We want to know whether the earlier results favoring factive accounts were at least partly due to the way participants were questioned. If they were, then there probably will be an effect of condition on participant response in the present experiment. If they were not, then there probably will not be an effect of condition.

There was no effect of condition on participant response, $\chi^2(2) = 3$, $p = .223$. Moreover, participants in each condition answered that Maria should not make the assertion at rates exceeding what we would expect by chance: 68% in Reason Question, 80% in No Question, and 86% in Truth Question, binomial, all $ps \leqslant .05$. It is especially noteworthy that participants in Reason Question answered 'No' at rates greater than chance. Even when asked *only* about Maria's justification, participants still think that she should not make the assertion. This is difficult to explain if assertion's norm is not factive but easily explained if it is factive.

Participant assessment of what mattered more in their evaluation of the case, evidence or truth, did not differ from their answer to whether Maria should make the false but well justified assertion, McNemar's test, $p = .508$. That is, participants who answered that Maria should make the assertion strongly tended to identify evidence as mattering more, and participants who answered that Maria should not make the assertion strongly tended to identify truth as mattering more. Among those who said that Maria should make the assertion, 85% said that evidence mattered more. Among those who said that Maria should not make the assertion, 92% said that truth mattered more.

Overall, 75% of participants said that truth mattered more, which far exceeds what could be expected by chance, binomial, $p < .001$. These results suggest that participants are able to reliably report the criteria guiding their evaluation of assertions, at least on the most basic way of soliciting such reports.

One concern about both of the experiments up until now is that some participants might be answering the test question negatively even though they do not think that the norm of assertion is factive. For instance, perhaps some participants think that the norm of assertion is non-factive but they also think that a false statement by Maria might have negative consequences. Furthermore, some of these participants might reasonably think that avoiding the negative consequences trumps whatever the norm of assertion permits. Accentuating this concern is the fact that the test question asks whether Maria should 'tell her guest' that she has the watch in her collection. It's a legitimate worry that some participants answered the test question negatively because they think that hosts, *qua* hosts, must tell their guests only truths. This is consistent with their thinking that assertors, *qua* assertors, are not required to assert only truths. The next experiment addresses these concerns.

4. Experiment 2

This experiment aims to address the concerns mentioned at the end of the previous section by doing two things. First, it varies how serious are the implied consequences of a false assertion by Maria. Second, the test question does not ask whether Maria 'should tell her guest' something but instead whether she 'should say' it in response to people in different relationships to her. If the concerns raised at the end of the previous section are valid, then as the potential negative consequences of a false assertion rise, participants probably will be more likely to say that Maria should not make the assertion. By contrast, if participant response is keyed to the norms of assertion specifically—rather than, say, the norms of prudent action—then participants probably will not be more likely to do that.

4.1. Participants, materials and procedure

One hundred and five new participants (36 female, aged 18–63 years, $M = 31.18$ years, $SD = 10.01$ years, 99% native English speakers) were randomly assigned to one of three conditions: Low, Medium, and High.

Participants in each condition read a variation on the False version of the story from Experiment 1A. Maria is no longer hosting guests for dinner but instead is simply asked a question by people in different capacities and for different purposes. In story for the Low condition, Maria's neighbor asks her a question out of idle curiosity, so the implied cost of speaking falsely is low. In the story for Medium, an insurance appraiser asks Maria a question in order to assess her collection's value, so the implied cost of speaking falsely is higher but not intimidating. In the story for High, a federal prosecutor asks Maria a question

as part of an investigation, so the implied cost of speaking falsely is high.

After reading the story, participants in all conditions were asked two questions. First, they were asked the test question:

> Should Maria say that she does have a 1990 Rolex Submariner in her collection? [Yes/No]

Then they were taken to a separate screen (they could not go back and change their previous answer), where they were asked,

> When you were deciding whether Maria should tell her guest that she has a 1990 Rolex Submariner in her collection, which factor seemed more important to you? [Maria's evidence about what was in her collection/ The truth about what was in her collection]

4.2. Results and discussion

We're primarily interested in whether there is an effect of condition on response to the test question. If there is, then it complicates the interpretation of participant response to the test question. More specifically, to the extent that participants answer negatively more frequently as the perceived costs of speaking falsely rise, it complicates the interpretation of earlier results favoring factive accounts. By contrast, if there is no effect of perceived cost of error on response to the test question, then that lends support to my earlier interpretation of the findings in Experiments 1A and 1B. In line with that, I made two predictions about the results for the present experiment. First, there would be no effect of condition on response to the test question. Second, participants would once again tend to say that Maria should not make the false but well justified assertion.

Both predictions were true. First, there was no effect of condition on response to the test question, $\chi^2(2) = .655$, $p = .721$. Higher implied costs of speaking falsely did not lead participants to more frequently answer that Maria should not speak falsely. Second, overall 70% of participants answered that Maria should not make the false but well justified assertion, which far exceeds what could be expected by chance, binomial, $p < .001$, and replicates the result from Experiment 1B, binomial, $p = .095$, test proportion = .77. What's more, participants answered this way in each of the three conditions: 71% in Low, 65% in Medium, and 74% in High, binomial, $ps \leqslant .049$, one-tailed. (A one-tailed test was used here because, based on previous results, the prediction was directional.)

Once again, participant assessment of what mattered more in their evaluation of the case, evidence or truth, did not differ from their answer to whether Maria should make the false but well justified assertion, McNemar's test, $p = .774$. Among those who said that Maria should make the assertion, 83% said that evidence mattered more. Among those who said that Maria should not make the assertion, 91% said that truth mattered more. Overall, 71% of participants said that truth mattered more, which far exceeds what could be expected by chance, binomial, $p < .001$. As we saw in Experiment 1B, participants seemingly are able to identify the criteria guiding their evaluation.

Table 1
Percentage of participants in Experiments 1A, 1B, and 2 who answered that Maria should not make a well justified but false assertion.

	Should not
Experiment 1A	
True	3
False	80
Experiment 1B	
Reason Question	68
No Question	80
Truth Question	86
Experiment 2	
Low	71
Medium	65
High	74

Altogether the results from our first three experiments (see Table 1) provide evidence for at least three conclusions. First, the non-factivity intuition is not widely shared. Second, by contrast, the factivity intuition is widely shared. It appears that roughly 3/4 of people respond as a factive account would predict, whereas roughly 1/4 of people respond consistently with the non-factivity intuition. An assertion's truth-value significantly affects how most people evaluate it. Third, people's evaluation of assertion is, to a surprising extent, insensitive to the implied costs of speaking falsely. Whether Maria's neighbor asks out of idle curiosity or a federal prosecutor asks in the course of an investigation, participants were equally likely to answer that Maria should not make a false but well justified assertion.

These results could be strengthened by observing similar patterns using a different dependent measure. All the experiments thus far collected responses on a forced-choice dichotomous scale. But perhaps this method of questioning masks important differences that will be revealed if participants are given greater flexibility to express themselves. The next experiment pursues this possibility.

5. Experiment 3

This experiment explores the robustness of the main findings from the first three experiments. It does this by using a different method of questioning to detect similar patterns in participant response.

5.1. Participants, materials and procedure

One hundred new participants (25 female, aged 18–64 years, $M = 29.35$ years, $SD = 10.31$ years, 97% native English speakers) were randomly assigned to one of four conditions in a 2 (Cost: Low/High) × 2 (Truth-value: True/False) between-subjects design. Participants in each condition read one story.

The Truth-value factor varies whether the inventory correctly lists that Maria owns the watch. The Cost factor varies the implied costs of a false statement from Maria. In the story for Low True, Maria's neighbor asks her a question 'just out of idle curiosity,' Maria consults her inventory, and the inventory accurately lists that she owns the watch. In the story for High True, a federal prosecutor asks

Maria a question 'for purposes of an investigation,' and the inventory accurately lists that she owns the watch. The stories for the Low False and High False conditions are the same except that the inventory is wrong when it lists that she owns the watch. Otherwise, the storyline was the same as in previous experiments reported above.

After reading the story, participants were asked to rate their agreement or disagreement with the following statement: Maria should say that she does have a 1990 Rolex Submariner in her collection. Responses were collected on 7-point scale, anchored at Strongly Disagree through Strongly Agree. Responses were coded from −3 (Strongly Disagree) through +3 (Strongly Agree), creating a midpoint of 0 (Neutral). (Participants never saw numerical values, only the qualitative descriptions.)

5.2. Results and discussion

The main findings from the first three experiments were: the non-factivity intuition is not widely shared, the factivity intuition is widely shared, and raising the implied costs of speaking falsely does not affect people's evaluation of assertions. The previous experiments collected responses from participants using dichotomous scales. The present experiment tests whether we also observe those outcomes when participants have greater flexibility to express themselves. If we do not observe similar outcomes in this experiment, then it complicates my interpretation of earlier findings. If we do observe similar outcomes, then it supports my interpretation of earlier findings.

In line with the earlier findings, I made two predictions. First, there would be an effect of Truth-value on participant response. More specifically, participants would be inclined to agree with the statement in True conditions but inclined to disagree in the False conditions. Second, there would be no effect of Cost on participant response. That is, response in Low True would not differ from response in High True, and response in Low False would not differ from response in High False.

Both predictions were true. A two-way analysis of variance (ANOVA) revealed that there was an overall effect of condition on response to the test question, $F(3) = 26.12$, $p < .001$, $\eta_p^2 = .449$. There was a main effect of Truth-value, $F(1) = 77.75$, $p < .001$, $\eta_p^2 = .447$, but there was no effect of Cost, $F(1) = .37$, $p = .547$, and no interaction, $F(1) = .233$, $p = .630$. The mode in Low False was 'Disagree'; the mode in High False was 'Disagree'; the mode in Low True was 'Strongly Agree'; the mode in High True was 'Strongly Agree'. Pairwise comparisons using independent-samples t-tests revealed that participant response in Low False ($M = −.48$, $SD = 1.92$) did not differ from response in High False ($M = −.84$, $SD = 1.99$), $t(48) = .651$, $p = .518$; similarly, response in Low True ($M = 2.28$, $SD = 1.31$) did not differ significantly from response in High True ($M = 2.24$, $SD = 1.27$), $t(48) = .110$, $p = .913$. Because Cost entered into no main or interaction effects, I collapsed across that factor to gain statistical power. Participant response in True conditions was significantly above the midpoint ($M = 2.26$, $SD = 1.28$), $t(49) = 12.34$, $p < .001$, whereas response in False conditions was significantly below the midpoint ($M = −.66$, $SD = 1.94$), $t(49) = −2.4$, $p = .020$.

These results provide additional support for the three main findings from previous experiments. Once again, the truth-value of a proposition significantly affected people's evaluation of whether it should be asserted. Participants strongly agreed that a well justified *true* assertion should be made, but they tended to disagree that a well justified *false* assertion should be made. Moreover, once again, participant response was not affected by the implied cost of speaking falsely. Taken together these results corroborate earlier findings that participant response is keyed to the 'should' of assertoric normativity (as opposed to prudential normativity), that the non-factivity intuition is not widely shared, and that the factivity intuition is widely shared.

6. Experiment 4

One concern about all the experiments up until now is that I have only tested people's intuitions about assertions of *positive* claims (i.e. claims that something *is* the case). I have not tested people's intuitions about assertions of *negative* claims (i.e. claims that something is *not* the case). But if the norm of assertion is factive, then it is factive regardless of whether the assertion is positive or negative. Thus the results so far could be strengthened by investigating whether an assertion's *content* (positive/negative) affects people's intuitions about whether it should be made.

Another concern about the experiments thus far is that I have not given participants enough flexibility *of the right sort* to register what they think Maria should say. They were forced to either answer a dichotomous question about whether Maria should make a certain assertion, or to rate their agreement with the claim that Maria should make the assertion. Narrowly focusing on one particular "should" question or statement might prevent participants from meaningfully expressing their view of Maria's situation. For instance, even though participants definitely tend to say that Maria "should not" make a justified false assertion, we still do not know what they think she should say in such a situation. The results so far could be strengthened by allowing participants to directly select from multiple options what Maria should say.

This experiment addresses both of these concerns. On the one hand, it tests whether the effect of truth on assertability is similar for positive and negative claims. On the other hand, it gives participants much more flexibility to select what Maria should say.

6.1. Participants, materials and procedure

One hundred sixteen new participants (33 female, aged 18–62 years, $M = 28.45$ years, $SD = 8.73$ years, 96% native English speakers) were randomly assigned to one of four conditions in a 2 (Content: Positive/Negative) × 2 (Truth-value: True/False) between-subjects design.

Participants in each condition read one story. The basic story was the same as the Low cost story from Experiment 3 (Maria's neighbor asking her about the watch out of curiosity). The Content factor varies whether the inventory lists that Maria does/does not own the watch. The Truth-

value factor varies whether the inventory is correct or incorrect.

After reading the story, participants were asked a single question: How should Maria answer her neighbor's question? Participants were given seven options, displayed left-to-right across their screen. The options varied whether and how strongly Maria should claim or deny that the watch is in her collection. In conditions where the inventory says that Maria does have the watch (Positive conditions), participants chose from the following options, left-to-right:

"I definitely do have one."

"I do have one."

"I probably do have one."

"It's unclear."

"I probably don't have one."

"I don't have one."

"I definitely don't have one."

In conditions where the inventory says that Maria does not have the watch (Negative conditions), participants chose from the same options in reverse order, beginning with "I definitely don't have one" on the left, through "I definitely do have one" on the right.

Responses were coded from left-to-right, +3 (Maria should say that things are *definitely* as the inventory indicates) through −3 (Maria should say that things are *definitely not* as the inventory indicates), with a neutral midpoint of 0 ("It's unclear"). (Participants never saw numerical values, only the qualitative descriptions.) Thus coding for specific responses was reversed in Positive and Negative conditions (except for the neutral midpoint, which was the same in both), creating a consistent measure of how strongly Maria should assert in accordance with, or against, the evidence provided by the inventory.

6.2. Results and discussion

We want to learn two things from this study. First, were earlier results favoring factive accounts at least partly due to the fact that the stimuli featured only positive assertions, or will participants will view negative assertions similarly? If participants view negative assertions similarly, then there probably will not be a main effect of Content on response to the test question. Second, how will participants respond when they are empowered to more flexibly express themselves? Of special interest is what participants think Maria should say when she has a justified false belief about the contents of her collection.

A two-way ANOVA revealed that there was an overall effect of condition on response to the test question, $F(3) = 27.64$, $p < .001$, $\eta_p^2 = .425$. There was a main effect of Truth-value, $F(1) = 82.12$, $p < .001$, $\eta_p^2 = .423$, but there was no effect of Content, $F(1) = .762$, $p = .385$, and no interaction, $F(1) = .041$, $p = .841$. Pairwise comparisons using independent-samples t-tests revealed that the mean response in True Positive ($M = 2.0$, $SD = 1.3$) was significantly higher than in False Positive ($M = -.28$, $SD = 1.71$),

$t(56) = 5.69$, $p < .001$; similarly, the mean response in True Negative ($M = 2.28$, $SD = .649$) was significantly higher than in False Negative ($M = −.10$, $SD = 1.61$), $t(56) = 7.38$, $p < .001$. The mode in both True Positive and True Negative was +2, indicating that Maria should flat-out assert the true and well justified claim. The mode in both False Positive and False Negative was +1, indicating that Maria should say that the justified false claim is *probably* true. Fig. 1 visualizes the distribution of responses in the four conditions.

These results further support a factive account. Once again, an assertion's truth-value significantly affected how people evaluate it, regardless of whether its content is positive or negative. And when given greater flexibility to indicate what should be said about a justified false proposition, very few participants said that the proposition should be flat-out asserted. In False conditions, the most common response was that Maria should assert that the justified false proposition is "probably" true, although surprising number of participants also said that Maria should assert *against* her evidence in order to speak the truth. By contrast, almost all participants said that a justified true proposition should be either asserted or emphatically asserted.

It's worth noting that, on the most natural interpretation, the assertion "I probably do (not) have one" is arguably true even when the flat-out assertion "I do (not) have one" is false, given what Maria knows about the inventory's reliability. In other words, Maria's evidence can make it true

that *probably she does (not) have the watch* even when it's false that *she does (not) have one*. Thus, in cases of justified false belief, the most common view is that Maria should make a *true* probabilistic assertion, though the bimodal distribution of responses suggests that this view is not uncontroversial. By contrast, in cases of justified true belief, nearly everyone agrees that she should make the true flat-out assertion, without any hedging or probabilistic qualification.

7. Experiment 5: rule-breaking and excuse validation

Recall that the second objection to factive accounts is a simple argument about justified false assertions: a speaker who makes a justified false assertion is not thereby properly criticizable (or: blameworthy); so the speaker has probably not broken a rule of assertion; so assertion probably does not have a factive norm. The argument assumes that blamelessness is a defeasibly good indication that no rule has been broken. (Its proponents are not committed to blamelessness being an indefeasible indication that no rule has been broken.) Granting that assumption for the sake of argument, this section reports an experiment that defeats its application to justified false assertions. The results arguably also discredit the assumption itself—not merely its application in the present case—but that is a more ambitious claim that must be developed elsewhere.
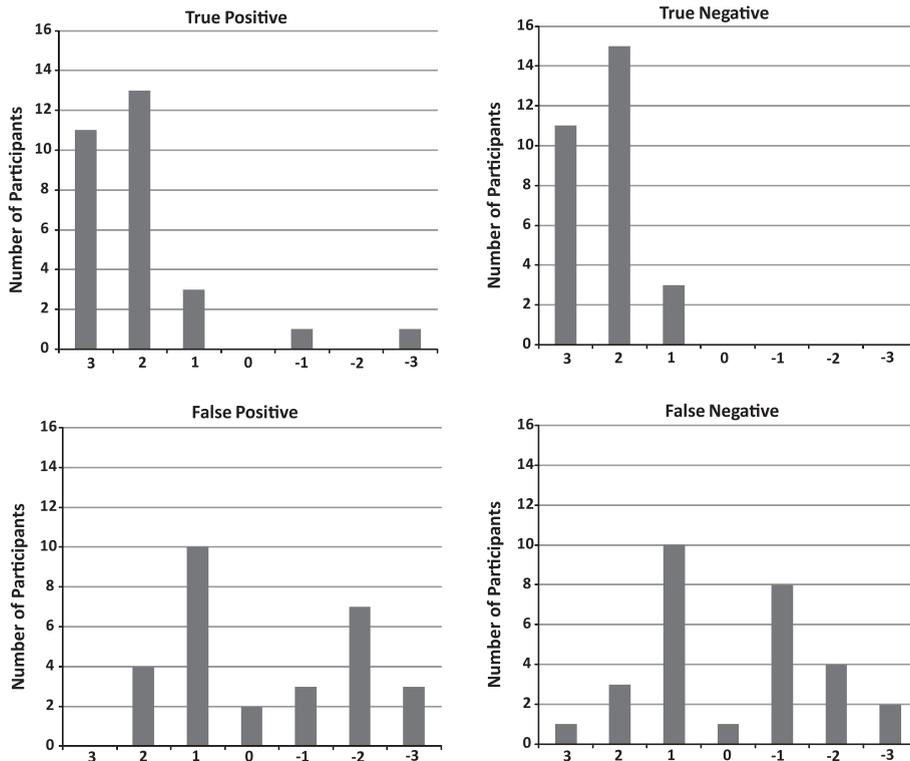

Fig. 1. Experiment 4: Distribution of responses across the four conditions: True Positive: the inventory correctly lists that Maria owns the watch; True Negative: the inventory correctly lists that Maria does not own the watch; False Positive: the inventory incorrectly lists that Maria owns the watch; False Negative: the inventory incorrectly lists that Maria does not own the watch. Values along the x-axis represent what Maria "should" say: 3 = the inventory is "definitely" right; 2 = the inventory is right; 1 = the inventory is "probably" right; 0 = it's "unclear"; -1 = the inventory is "probably" wrong; -2 = the inventory is wrong; -3 = the inventory is "definitely" wrong.

7.1. Excuse validation

Humans are sensitive to criticism for doing what is, by their lights, the right thing, even when it turns out to have been the wrong thing. Of course, we can be brought to see the error of our ways, but as we all know from experience, it can be challenging to admit to failure and wrongdoing, especially when someone criticizes or blames us.

We make allowance for this human sensibility in our dealings with others. We're often willing to excuse or even overlook what is in fact incorrect behavior if the transgressor had good evidence for thinking that their behavior was correct. If, from your perspective, you should do something, then it can easily seem harsh to criticize or blame you for doing it. And it seems even harsher yet to punish you for it.[3] A more natural and generous reaction is to try to help you improve your perspective so that it better matches what is in fact proper behavior. But sometimes even *that* much can seem like an imposition and we're willing to let further experience and reflection teach the transgressor his lesson in the fullness of time. What's more, sometimes we feel compelled to go so far as to insist that the transgressor "did nothing wrong." Such is the pressure we're prone to feel in the face of blameless wrongdoing. Call this the human tendency of *excuse validation*. (A related phenomenon, "blame validation," is discussed by Alicke, 1992 and Alicke & Rose, 2010.)

If excuse validation occurs, then it complicates the argument advanced by non-factivists earlier in Section 3. Non-factivists noted that a speaker is not properly criticizable for a false justified assertion, and concluded that a rule had not been broken. But we should be cautious about this inference. Our view on whether someone deserves criticism can affect whether we're willing to say, or perhaps even think, that a rule has been broken. If we want to reason from an intuition of blamelessness to the conclusion that no rule had been broken, then excuse validation is a factor that needs to be controlled for.

This section looks more carefully at the potential relationship between excuse validation and non-factivity intuitions about assertion. Although the motivation for non-factive accounts is real and, for some, irresistible, I suspect that it is not keyed directly to the norm of assertion. Instead it is keyed to our commitment to derivative principles of criticism for excusable violations, which fits a more general pattern. In short, the appeal of non-factivism is an expression of a tendency to excuse validation.

The relevance of distinguishing between correctness and blamelessness—or closely related distinctions—has not been lost on proponents of factive accounts (e.g. Williamson, 2000: 256, DeRose, 2002: 180). But earlier treatments have met with sustained criticism (e.g. Kvanvig, 2009; Lackey, 2007) and have not been developed in detail or with an eye toward empirical confirmation. Here I advance the discussion by identifying the phenomenon of excuse validation, experimentally establishing an approximate baseline frequency, and experimentally

establishing that participant response to cases of justified false assertions fits the observed pattern of excuse validation. This places a considerable burden on non-factivitsts who want to persist in arguing against factive accounts from premises about blamelessness.

7.2. Participants, materials and procedure

Sixty-five new participants (23 female, aged 18–54 years, *M* = 27 years, *SD* = 9.4 years, 97% native English speakers) were randomly assigned to one of two conditions: Driving and Asserting. I excluded data from twenty-seven additional participants who failed comprehension questions.

Participants read one story about either a driver who blamelessly breaks the speed limit, or about a speaker who blamelessly makes a false statement. To accentuate their blamelessness, each protagonist's rule-breaking is saliently caused by the actions of another person in the story. After the story, participants were asked four comprehension questions, then two test questions: whether the protagonist is criticizable for their behavior, and whether there is a sense in which the protagonist's behavior is incorrect.

Here are the story and questions for the Driving condition:

> Doreen just had her car serviced and is driving home from the mechanic's shop. She wants to get home without unnecessary delay, but she does not want to break the speed limit. The speed limit is 55 miles per hour, so she looks down to see how fast she is going. The speedometer says that she is going 55 miles per hour. But the speedometer is wrong, since the mechanic disrupted a setting on the speedometer without anyone noticing. As a result, Doreen is actually driving 60 miles per hour.
> D1. How fast does Doreen think she is driving? [50/55/60]
> D2. Is it reasonable for Doreen to think that she is driving that fast? [Yes/No]
> D3. How fast is Doreen driving? [50/55/60]
> D4. Is Doreen's speed above the limit? [Yes/No]
> D5. Should Doreen be criticized for driving that speed? [Yes/No]
> D6. Is there a sense in which it is incorrect for Doreen to drive that speed? [Yes/No]

Here are the story and questions for the Asserting condition.

> Robert just purchased an 1804 US silver dollar and is placing it in a display in his library. He is having dinner guests over tonight, and he wants to tell them about the silver dollar. He sets the silver dollar in its display, closes the library doors behind him, and goes to greet his guests. Soon after, as Robert serves his guests dinner, he tells them, "There is an 1804 US silver dollar in my library." But between the time Robert left his library and the time he served dinner, a coin thief

---

[3] Actually, criticism and blame are mild forms of punishment. But 'punish' connotes something much more severe, beyond mere verbal punishment.

quickly slipped through his library window, stole the silver dollar, and escaped without anyone noticing. As a result, Robert's coin is not in the library.

A1. Where does Robert think his coin is located? [in the library/in the thief's possession]

A2. Is Robert's belief about the coin's location reasonable? [Yes/No]

A3. Where is Robert's coin? [in the library/in the thief's possession]

A4. Is Robert's statement false? [Yes/No]

A5. Should Robert be criticized for making that statement? [Yes/No]

A6. Is there a sense in which it is incorrect for Robert to make that statement? [Yes/No]

7.3. Results and discussion

The purpose of the Driving condition is to establish an approximate base rate for excuse validation. That is, responses in this condition help set a baseline for how frequently we can expect participants to react to what obviously is *real but blameless objective rule-breaking* by not merely excusing it, but by going so far as to *deny* that the transgressor behaved in incorrectly in any sense.

It is a plain fact of the case that Doreen is breaking the speed limit: she is driving 60 m.p.h. in a 55 m.p.h zone. Indeed, it was a comprehension question whether Doreen was driving above the limit (i.e. question D4). Only participants who answered 'Yes' were included in the analysis. But it is also obvious that Doreen lacks evidence that she is breaking the speed limit, and indeed that she has excellent evidence that her driving speed is legal. So of course she should not be criticized, much less ticketed, for breaking the speed limit. Her objective rule-breaking is blameless and fully excusable. Or so it can easily seem.

In light of all that, I predicted that many participants would engage in excuse validation by answering 'No' to question D6—that is, by denying that Doreen's driving speed is incorrect in any sense. And this is exactly what we observe. Only 49% answered that there was a sense in which it was incorrect for Doreen to drive that fast, which does not differ significantly from what could be expected by chance, binomial, $p > .99$. This is a very surprising result, especially given that only participants who correctly answered comprehension question D4 ('Is Doreen's speed above the limit?') were included in the analysis. This is no doubt related to the fact that 97% answered that Doreen should not be criticized for driving that speed.

I predicted that we would see a similar response pattern in the Asserting condition. Again, this is exactly what we observe. Participants responded to the two test questions very similarly in both Asserting and Driving conditions (Fisher's, both ps > .95, two-tailed). In the Asserting condition, 96% said that Robert should not be criticized, and only 46% answered that there was a sense in which it was incorrect for him to make the statement.

What could explain this striking similarity? The best explanation is that protagonists in both stories engage in blameless and excusable rule-breaking, and that a predictable proportion of participants in both conditions react to this by engaging in excuse validation: they literally deny

that any transgression has occurred, even though they recognize—and had just agreed!—that it has. Given the similarities of structure and content between Driving and Asserting, it must then be that *making a false statement* breaks a relevant and salient rule in Asserting just as *driving over the speed limit* breaks a relevant and salient rule in Driving. In other words, a factive norm of assertion is part of the best explanation of the observed similarities in response patterns.

I want to be clear about how I'm marshaling this data into an argument against non-factive accounts. Excuse validation occurs when people detect that a certain action breaks a rule, note that it was broken blamelessly, and then deny that the transgressor broke the rule. People tend to engage in excuse validation at a certain rate, but this rate cannot be gleaned from the armchair. Participant response in Driving helps to establish an approximate base rate for excuse validation experimentally. The response pattern in Asserting does not differ from the pattern in Driving.[4] This similarity is striking and cries out for explanation. One good explanation for this similarity is that participants in Asserting are also engaging in excuse validation at roughly that same rate. But in order for them to do that, *they must think that a relevant rule has been broken*. Given the structure of Asserting and its isomorphism to Driving, the only plausible candidate for such a rule is a factive norm of assertion. And this is further evidence in favor of a factive norm of assertion.

8. Conclusion

I accomplished two main things in this paper. First, I adopted an experimental approach to investigating the norms of assertion, placing the discussion on surer empirical ground and hopefully inspiring further work in an area ripe for experimental investigation. Second, I strengthened the case for a factive norm of assertion.

Results from six experiments support several conclusions. First, the non-factivity intuition is not widely shared (Experiments 1A, 1B, 2, 3, and 4). That is, most people do not judge that well justified but false assertions should be made. Second, the factivity intuition is widely shared (Experiments 1A, 3 and 4). That is, most people judge that false assertions should not be made, even when they are strongly supported by the available evidence. Third, people's evaluation of the assertions is keyed to assertoric normativity rather than, say, prudential normativity (Experiments 2 and 3). Fourth, people can reliably identify the criteria guiding their evaluation of assertions (Experiments 1B and 2). Similar results were observed across several experiments that questioned participants differently, including adding or eliminating comprehension questions, collecting responses on different scales, and changing the available options. Overall, a factive norm of assertion well explains some rather striking patterns in people's reaction to cases. Moreover, intuitions that appear to conflict with a factive norm are arguably due to a general tendency to say

---

[4] In subsequent work, my lab has replicated this result multiple times. We have replicated it not only with stories involving driving, but with a range of other activities too, including baking, farming, and chess (Turri & Blouw, in press).

that blameless transgressors did nothing incorrect (Experiment 5).

Up until now, inquiry into the norms of assertion has been broadly observational though not experimental. But an experimental approach is warranted because competing theories about the norm of assertion generate testable predictions, given two plausible and widely shared assumptions.

The first assumption is that the normative intuitions of skilled practitioners are a source of evidence about what the practice's rules are. This assumption is arguably shared by those who appeal to competent speakers' intuitions to support theories of syntax (e.g. Chomsky, 1965), experimental investigations into the adequacy of various theories of the relationship between semantics and pragmatics, and other forms of psycholinguistic experimentation (e.g. Gibbs & Moise, 1997, and the essays in Noveck and Sperber, 2004). Other things being equal, we should expect skilled practitioners to reliably identify what should and should not be done, what must and must not be done according to the rules of the practice.

Importantly, this is not to say that skilled practitioners have an explicit, articulable theory about what the rules are, or that they can reliably identify the actual rule(s) of the practice when asked to select them from a set competing proposals. The assumption is not that skilled practitioners tend to be good theorists; rather, it's that patterns in their concrete, first-order intuitions and judgments manifest their skill in applying the relevant rules (Chomsky, 1957).[5] In short, their intuitive reaction to cases tends to manifest their competence, resulting in detectable patterns. We can then use these patterns when theorizing about the practice's rules. The patterns will make some proposals much less likely than others, given what we expect from skilled performance.

The second assumption is that assertion is a social practice that competent speakers are skilled at, which is utterly uncontroversial. Given these two assumptions, the normative intuitions of competent speakers are a valuable source of evidence about the norms of assertion. Other things being equal, if that group strongly tends to judge that assertions of a certain sort should not be made, then we should conclude that those assertions probably violate the norm of assertion. Similar remarks apply when investigating the norms of other speech acts, such as questioning, commanding, guaranteeing, and promising. The present discussion can thus serve as a model for investigating the norms of other speech acts.

All the studies reported here were conducted on adults in the United States. This is a noteworthy limitation and future research could profitably investigate these matters cross-culturally and developmentally. First, is there cross-cultural variation in the evaluation of assertions and informants? If members of other cultures respond significantly differently to the same experimental stimuli, then that would raise questions about how to interpret the results reported here, and indeed about whether there is a universally shared, normatively constituted practice of assertion. Second, are there developmental differences in the evaluation of assertions and informants? Do young children and adolescents tend to view these matters differently from adults? If so, at what point do the differences appear and disappear?

In the process of investigating the norms of assertion, we encountered a surprising and potentially important pattern in people's normative judgments about blameless rule-breaking: excuse validation. Experiment 5 suggests that many people are surprisingly unwilling to identify blameless rule-breaking as rule-breaking. This raises a host of interesting questions about the nature and function of rules in human society, as well as the relationship between rules, blame and punishment. Further work in this area could investigate the causes and limits of excuse validation. For example, if blamelessly breaking the rule has seriously negative social consequences, then perhaps people will be more likely to correctly identify it as rule-breaking.

Acknowledgments

---

[5] Chomsky (1957: p. 97, n. 5) cautions against 'replac[ing] an operational test of behavior … by an informant's judgment about his behavior. The operational tests for linguistics notions may require the informant to respond, but not to express his opinion about his behavior, his judgments about synonymy, about phonetic distinctness, etc. The informant's opinions [about these things] may be based on all sorts of irrelevant factors.' Postulating a set of rules to explain the observed linguistic behavior is the linguist's job.

References

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology, 63*(3), 368–378.

Alicke, M., & Rose, D. (2010). Culpable control or moral concepts? *Behavioral and Brain Sciences, 33*(04), 330–331.

Anscombe, G. E. M. (1957). *Intention*. Oxford: Blackwell.

Bach, K. (2008). Applying pragmatics to epistemology. *Philosophical Issues, 18*(1), 68–88.

Bach, K., & Harnish, R. M. (1979). *Linguistic communication and speech acts*. Cambridge, Mass: MIT Press.

Benton, M. A. (2011). Two more for the knowledge account of assertion. *Analysis, 71*(4), 684–687. http://dx.doi.org/10.1093/analys/anr085.

Brandom, R. (1983). Asserting. *Nous, 17*(4), 637–650.

Buckwalter, Wesley (in preparation). Factive verbs and protagonist projection. University of Waterloo.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment. *Psychological Science, 17*(12), 1082–1089.

DeRose, K. (2002). Assertion, knowledge, and context. *The Philosophical Review, 111*(2), 167–203.

Douven, I. (2006). Assertion, knowledge, and rational credibility. *Philosophical Review, 115*(4), 449–485.

Gibbs, R. W., & Moise, J. F. (1997). Pragmatics in understanding what is said. *Cognition, 62*, 1–24.

Grice, H. P. H. (1989). *Studies in the way of words.* Cambridge: Harvard University Press.

Hazlett, A. (2010). The myth of factive verbs. *Philosophy and Phenomenological Research, 80*(3), 497–522.

Hazlett, A. (2012). Factive presupposition and the truth condition on knowledge. *Acta Analytica, 27*(4), 461–478.

Hill, C., & Schechter, J. (2007). Hawthorne's lottery puzzle and the nature of belief. *Philosophical Issues, 17*, 102–122.

Kvanvig, J. (2009). Assertion, knowledge, and lotteries. In D. Pritchard & P. Greenough (Eds.), *Williamson on knowledge* (pp. 140–160). Oxford: Oxford University Press.

Lackey, J. (2007). Norms of Assertion. *Noûs, 41*(4), 594–626.

Noveck, I. A., & Sperber, D. (2004). *Experimental pragmatics.* Palgrave Macmillan.

Rakoczy, H., & Tomasello, M. (2009). Done wrong or said wrong? Young children understand the normative directions of fit of different speech acts. *Cognition, 113*(2), 205–212.

Rescorla, M. (2009). Assertion and its constitutive norms. *Philosophy and Phenomenological Research, 79*(1), 98–130.

Schaffer, J. (2008). Knowledge in the image of assertion. *Philosophical Issues, 18*(1), 1–19.

Searle, J. R. (1969). *Speech acts.* Cambridge: Cambridge University Press.

Searle, J. R. (1979). *Expression and meaning.* Cambridge: Cambridge University Press.

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

Turri, J. & Blouw, P. (in press). Excuse validation.

Turri, J. (2010a). Epistemic invariantism and speech act contextualism. *Philosophical Review, 119*(1), 77–95.

Turri, J. (2010b). Prompting challenges. *Analysis, 70*(3), 456–462.

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

Turri, J. (2011b). Mythology of the factive. *Logos & Episteme, 2*(1), 143–152.

Turri, J. (2013). Knowledge and suberogatory assertion. *Philosophical Studies.* http://dx.doi.org/10.1007/s11098-013-0112-z.

Turri, J., & Friedman, O. (2013). Winners and losers in the folk epistemology of lotteries. In: Beebe, J.R. (Ed.), Advances in experimental epistemology. Continuum.

Unger, P. (1975). *Ignorance: a case for skepticism.* Oxford: Oxford University Press.

Weiner, M. (2005). Must we know what we say? *Philosophical Review, 114*(2), 227–251.

Williamson, T. (1996). Knowing and asserting. *Philosophical Review, 105*(4), 489–523.

Williamson, T. (2000). *Knowledge and its limits.* Oxford: Oxford University Press.

Wilson, D., & Sperber, D. (2012). *Meaning and relevance.* Cambridge: Cambridge University Press.