

Thinking About The Liar, Fast and Slow

ROBERT BARNARD, JOSEPH ULATOWSKI, AND JONATHAN WEINBERG

ABSTRACT

The liar paradox is widely conceived as a problem for logic and semantics. On the basis of empirical studies presented here, we suggest that there is an underappreciated psychological dimension to the liar paradox and related problems, conceived as a problem for human thinkers. Specific findings suggest that how one interprets the liar sentence and similar paradoxes can vary in relation to one's capacity for logical and reflective thought, acceptance of certain logical principles, and degree of philosophical training, but also as a function of factors such as religious belief, gender, and whether the problem is treated as theoretical or practical. Though preliminary, these findings suggest that one reason the liar paradox resists a final resolution is that it engages both aspects described by so-called dual process accounts of human cognition.

The development of logic as a scientific discipline has brought with it an immense, purely verbal, manipulatory system of discriminations. ... Many people regard it as something next to mysterious that formal logic, from a strictly causal point of view, has attained its present form. One of the great tasks of scientists trained both in logic and in the behavioral or psychosocial sciences is to make this development understandable. A "natural history" of logic must be written just as one might write the natural history of horses or of tulips.

--Arne Næss, "A Study of Or"

The story of the liar paradox begins, according to tradition, with Epimenides the Cretan, who said, "All Cretans are liars." If all Cretans are liars and Epimenides is a Cretan, then Epimenides does not tell the truth when he says, "All Cretans are liars." In turn, if what Epimenides says is a lie, but at the same time says what is true, then he is a

liar who speaks truly--which, frankly, does not make much sense at all! It is difficult to not be engaged by the liar paradox once one has command of what the problem is.

Most work on the liar paradox has treated it as a special kind of logical or formal problem. Our aim here is to suggest that an understanding of the liar paradox can be advanced by taking a look not directly at the problem itself, but rather at how people think about the problem. There is, we maintain, an underappreciated psychological dimension to the liar paradox that may inform our work on the logical and formal dimensions of the problem. While the experience of confronting a simple contradiction is psychologically unremarkable, the experience of thinking through a version of the liar paradox requires one to appreciate inferential transitions and aporetic disorientation (cf. Rescher 2001). Armchair speculation alone cannot reveal the particular contours of how different people think about the liar paradox, but experimental investigation can, albeit indirectly.

2. Psychologising the Liar without Psychologism?

Our approach is not “psychologistic.” We are in no way arguing for a direct inference from “the folk think in such-and-such a way, when confronting the liar paradox” to “such-and-such a way is the normatively correct way to think about the liar paradox.” Nonetheless, there are at least three philosophical reasons to be interested in questions about the psychology of the liar.

First, philosophical work on the liar has not been devoid of psychological speculation, and, moreover, the philosophical community has not speculated as to how we might expect people think about the paradox. Maybe, our pre-theoretic and informal psychological resources are simply not up to the task of wrangling with such paradoxes. Roy Cook gives voice to such a view when he writes recently that “paradoxes often demonstrate, or at least suggest, that our most basic intuitions and platitudes regarding some of our most basic concepts -- including truth, collection, logic, knowledge, and belief -- are faulty in some sense or another” (Cook 2013, 1). When he suggests further that attempts to solve such paradoxes may require “entirely new approaches to these

concepts” (*ibid.*, 2), Cook seems to be begging off the study of these pre-theoretic platitudes at least partly because it would have to utilize tools and techniques with which the logician is relatively unfamiliar. We likewise see Barwise and Etchemendy (1987) as exemplars of this *folk incoherence* view:

The significance of a paradox is never the paradox itself, but what it is a symptom of. For a paradox demonstrates that our understanding of some basic concept or cluster of concepts is crucially flawed, that the concepts break down in limiting cases. And although the limiting cases may strike us as odd or unlikely, or even amusing, the flaw itself is a feature of the concepts, not the limiting cases that bring it to the fore. If the concepts are important ones, this is no laughing matter. ... [I]f our ordinary concept of truth is somehow incoherent, as the paradox suggests, this raises the question of whether the same incoherence infects the mathematical and scientific discourse that presupposes the intuitive notion. (Barwise and Etchemendy 1987, 4f)

This picture of the cognitive inadequacy of the folk strikes us as a plausible armchair psychological hypothesis -- yet, only a hypothesis. There could be other, rival plausible hypotheses.

For example, to frame his discussion of how he thinks someone might work through the liar paradox, Maudlin imagines a hypothetical survey and then articulates an intuitive model for how one might informally evaluate the liar sentence, “This sentence is false.”¹ He writes:

One imagines a multiple choice questionnaire, with the following curious entry:

* This sentence is false.

¹ For the remainder of the paper, we will call this sentence: “LIAR.” We will call it such even to its face. That’s just how we roll.

The starred sentence is:

- a) True
- b) False
- c) All of the above
- d) None of the above.

There follows a bit of informal reasoning. If a) is the right answer, then the sentence is true. Since it says it is false, if it is true it must really be false.

Contradiction. Ergo, a) is not the right answer. If b) is the right answer, then the sentence is false. But it says it is false, so then it would be true. Contradiction. Ergo, b) is not the right answer. If both a) and b) are wrong, surely c) is. That leaves d).

The starred sentence is neither true nor false. (Maudlin 2004, 1)

He then goes on to offer an account of how the process of “informal reasoning” described here interacts with wider issues involving language and logic:

The conclusion of this little argument is somewhat surprising, if one has taken it for granted that every grammatical declarative sentence is either true or false. But the reasoning looks solid, and the sentence is a bit peculiar anyhow, so the best advice would seem to be to accept the conclusion: some sentences are neither true nor false. This conclusion will have consequences when one tries to formulate an explicit theory of truth, or an explicit semantics for a language. But it is not so hard after all to cook up a semantics with truth-value gaps, or with more than the two classical truth values, in which sentences like the starred sentence fail to be either true or false. From this point of view, there is nothing deeper in the puzzle than a pathological sentence for which provision must be made. (Ibid.)

To be clear, Maudlin is not at all claiming that everyone, or even most people, will go through these steps in their mind when they confront LIAR and its ilk. His model of

LIAR-directed thought is not meant to model specific processes of human cognition; rather, we take him to be presenting a story about what pre-theoretic, informal resources are available to ordinary cognisers, in confronting such paradoxes, and their adequacy to arrive at a view about truth-value gaps. As opposed to Barwise and Etchemendy's model, Maudlin's account is an exemplification of a *folk coherence* model of LIAR-directed thought.

On this "Maudlin model", the conclusion that some sentences are neither true nor false follows fairly simply from any initial attempt to assign LIAR a truth value. The hard *philosophical* work for him comes next in presenting a theory of semantics and truth that can accommodate that conclusion, but it takes almost no rigorous philosophical reasoning to get to the conclusion in the first place. Nonetheless, if the story is a simple one, it is an empirically committed one as well. The picture of ordinary cognition about LIAR develops in a way that would have us conforming to some familiar logical principles, e.g. contradiction avoidance, but not to others, e.g. bivalence. We can thus ask both whether his account of ordinary cognition is accurate in extension -- in terms of whether people generally do have some resources that will bring them to the conclusion that there are truth-value gaps -- and also whether it is accurate in terms of the particular resources appealed to -- in terms of what logical principles people generally can access.²

So, we have our first motivation for commencing empirical work on the psychology of the liar paradox. While philosophers have appealed to different accounts of what our non-specialist resources are in this area, the accounts themselves are speculative at best, and as a group are not obviously consistent with each other (in a way that even dialetheists

² We might compare Maudlin's insistence on formal analysis with Magdziak's (2000, 2004) work emphasising a less rigorous alternative response to the liar paradox, but whose outcome is the same, i.e., "neither true nor false" is the appropriate response to the liar. While formal definitions of truth are symptomatic of the liar paradox, Magdziak considers partial and conditional definitions of truth which seem to be operative in colloquial and scientific discourse and tend to avoid alethic paradoxes. The trouble is that these less rigorous definitions cannot be generalised to cover the more rigorous formal discourse, and a philosophical resolution of the liar paradox is not forthcoming.

should be potentially embarrassed). Though the philosophers we have cited here are fairly recent, the question of folk coherence versus folk incoherence -- and attempts to answer that question in an empirically-grounded way -- have a longer history than that, and we take ourselves to be following in the footsteps of the Norwegian empirical semanticist Arne Næss. In 1938, after having performed numerous experimental studies on the ordinary person's conception of truth (cf. Næss 1938a, 1938b), Næss reported that people do not typically agree with contradictory or inconsistent expressions. Just as philosophers abhor inconsistency, so too do ordinary people.

I cannot see what is meant by the claim that contradictory or inconsistent expressions "are permitted in the colloquial use" because no specific rules govern the "colloquial" use of true and its cognates. The claim of Tarski and others [that we dispense with intuitive adequacy because the ordinary person's conception is fundamentally flawed] is either materially scientific or it has no accurate meaning. As a materially scientific claim, we may devise an experiment with enough specification to make some meaningful observations about the use of language. I have succeeded in formulating such a control when it comes to the paradox of the liar. My results show that the words "true" and "false" (as well as a series of other words of similar merit) were not used by ordinary persons without limitation. People failed to affirm an expression considered a negation of itself. These results are of a statistical kind, and may eventually be invalidated through further materially scientific analysis. (Næss 1937/38, 383f [Ulatowski translation])³

It goes without saying that we likely hold inconsistent beliefs, though we may not be aware of them. Further, we might not acknowledge this fact even if it were clearly demonstrated to us. This is quite interesting since many philosophers believe that ordinary people, because of some lack of intellectual poise, would accept anything under any circumstances. That is not what Næss found (and certainly does not seem to be what we found, either).

³ Næss defends this view in a number of publications reprinted in Næss (2005).

The Næss questionnaires directly relevant to the study of whether non-philosophers accept inconsistencies or contradictions were never published, though he referred to them several times (cf. Næss 1938a, §6; 1981; 1992). Even Tarski himself cited the *now*-missing data (cf. Tarski 1944, 360). These so-called “Class-E” questionnaires asked respondents about their views on “verification,” “meaning,” “the semantic notion of truth,” and “the antinomy of the liar.”⁴ Later, Næss claimed:

According to Tarski and those following him, the *Umgangssprache* permits unlimited (*unbeschränkt*) use of the concept of truth. Propositions that negate themselves are permitted.

Such a hypothesis is empirical and we must ask, How is it testable? By what procedures? How is the metaphor of “permittance” eliminated? How are the rules of the *Umgangssprache* found?

The weight of the criticism of Tarski’s hypothesis is not that it is false, but that it is not made operational and therefore not tested.

A kind of test was made in 1936 and the result was negative. It made use of open questionnaires related to the antinomy of the liar. The persons speaking the *Umgangssprache* did not interpret any sentences in such a way that they negated themselves. The existence of a *rule* of the *Umgangssprache* that permits it was not in evidence, nor a *rule* that prohibits it. Rules may be *invented* that approximately picture the complex regularities of ordinary usage. In that case, there will be no rule of unlimited use of *true*. (Næss 2005, 70; cf. Næss 1981, 1992)

So, we have reports of Næss’ work seeming to indicate that ordinary people do not

⁴ That Næss never published these results is quite disappointing, indeed. We have to take whatever evidence we have of the Class-E questionnaires reported in Næss’ 1938a seriously.

embrace beliefs one might think are inconsistent, contradictory, or confused. Nonetheless, we only have these reports, and, moreover, one might argue that Næss' methods may be aligned more with ethnography than with experimental philosophy, and as such, his results would be somewhat limited in how much purchase they could give us on participants' aspects of cognition. Thus more, newer work is needed, and using different tools.

We have written thus far of "the folk" in a fairly undifferentiated way. But our further motivations for empirical work on the liar paradox involve two dimensions of complicating that too-simple picture of a uniform, universal set of cognitive resources and proclivities. One such dimension is that of demographic difference. Some work in experimental philosophy has indicated interpersonal differences in judgments in various philosophical domains, including variation by ethnicity (e.g., Machery et al. 2004) and personality type (e.g., Feltz and Cokely 2009). Although gender effects have proved elusive in the experimental philosophy literature, since there are at least some results indicating possible gender differences specifically in matters of truth (Barnard and Ulatowski 2013, Barnard and Ulatowski ms1), we took this opportunity to investigate whether gender correlated with different approaches to LIAR.

Surely other hypotheses of demographic differences are worthy of exploration in this regard, especially ethnicity (for see, e.g., Peng and Nisbett 1999). For this preliminary study, though, we only explored one further demographic variable. We conjectured that perhaps individuals' religious commitments could interact interestingly with how they would approach questions of paradox and contradiction, since many common religious doctrines endorse sets of propositions that are at least *prima facie* inconsistent (such as the Trinity). We wished to test the prediction as to whether self-reported theists were more tolerant of contradictions than self-reported non-theists.

Finally, if our first motivation to do empirical work on LIAR is to document what resources the folk in general do or do not bring to bear in engaging with it, our last motivation concerns the difference between the folk on the whole, and what might be termed more cognitively elite folks, either in terms of reflective ability or philosophical training. Clearly, trained philosophers have greater formal resources to wrestle with thorny

questions about, say, self-reference, or the theory of types. But while the folk may lack the wherewithal to address formal issues in and around the liar paradox, when we philosophers too step away from the areas of our technical training, it is an open empirical question whether our skills, training, or background as philosophers really give us better tools for untying that knot, compared with the folk. Even philosophers are prone to err in ways we like to imagine are only common among the folk. We do not deny that training and practice in formal methods may produce great differences in how well formal tasks can be performed, but we take it to be an open empirical question just where and to what extent philosophical training more generally leads to divergences from the cognition of non-philosophers. Although many philosophers have suggested sweeping and dramatic differences (cf. Ludwig 2007, Hales 2000, 2004, 2009, 2012; Williamson 2004, 2005, 2007, 2011, 2013), a fair number of studies have been conducted looking for such differences, and by and large they have shown philosophers' minds to work pretty much the same as those of non-philosophers (cf. Schulz et al. 2011, Machery 2009, Schwitzgebel and Cushman 2012, Schwitzgebel and Rust 2009).

And so a further hypothesis we were interested in considering is whether a person's tendency to think reflectively about versions of LIAR could produce systematic changes in people's judgments about it. Perhaps LIAR just seems to *strike* different people in different ways, and we may diverge in our quick, intuitive reactions. Other people generally find LIAR thought-provoking. It is not easy to discern from the armchair alone the extent to which people's thinking about LIAR operates in service of their initial reactions, or to what extent they are working through considerations that lead them to reflectively modify their evaluation. There are good theoretical psychological reasons to find both kinds of consequences of reflection possible here. A central part of the "dual process" theories of human cognition is that not only do we have distinct intuitive and reflective systems of judgment, they do not simply operate in parallel on separate domains. Rather, the reflective activity in "System 2" can serve as a check on the intuitive products of the cognitive heuristics in "System 1" (Evans 1984, Sloman 1996, Kahneman 2003), and yet the reflective system can also be subject to confirmation bias, and can just further

“lock in” whatever the agent’s initial take on a question may be (Mercier and Sperber 2011).⁵

We should consider one hypothesis whether we have an initial intuitive reaction that applies a classic truth-value to the liar sentence or liar-like paradox. For example, perhaps an intuitive matching bias (Evans and Lynch 1973; though, see also Evans 1998) may lead us to an initial response that LIAR is false, because, simply put, that's what it says. We see the word "false", and the mind finds it easiest to just use the word right before it as the answer.

Another candidate along similar lines would be the phenomenon of anchoring (Kahneman and Tversky 1972), in which an initially-provided value gets unconsciously recruited as an answer to a question, and it requires some mental effort to move away from that anchor. This can happen even in situations when one knows full well that there is no reason to expect the anchor in question to be a reliable guide, such as the last three digits of one's phone number anchoring people's estimates of when Attila the Hun departed Europe (Russo and Shoemaker 1989). It can require cognitive effort to move off of an initial estimate, and when we lack the will to do so, or when our cognitive resources are depleted, we are particularly susceptible to the influence of such anchors (Epley and Gilovich 2006). Moreover, there is also some evidence (e.g., Gilbert et al. 1993) that we tend to default to accepting presented claims as true, and that it takes at least a bit of further cognitive activity thereupon to reject them, and so some people may have an initial intuitive response ascribing truth to LIAR, especially, again, if they lack the will, interest, or cognitive resources to override such an initial reaction.

⁵ There has been some recent discussions recognising the diversity of processes occurring in both System 1 and System 2, which have suggested further refinements to the two (Frankish 2009, Stanovich 2009). Perhaps the range of processes involved in System 2 reasoning includes Type 3 processes involved in eliciting System 2 activity, e.g., resource allocation, conflict resolution, and the ultimate control of behavior, and mediating between the two systems (Evans 2009). We do not expect any such refinements to make trouble for our analysis here, however.

The “Maudlin model” of LIAR cognition discussed above fits well with that picture of reflection engaging with and correcting for initial intuitive responses, but there are other psychologically plausible models here, and we cannot tell just by introspecting which is true. Results like Cacioppo and Petty (1982) and Stanovich et al. (2013) indicate that confirmation bias can not only persist in the face of reflection, it can perhaps be intensified by it. High ability in reasoning can also be high ability in rationalizing. If the arguments of Mercier and Sperber (2011) are correct, then this myside bias even in reflective thought may be a central evolutionary design feature of our minds, and not just some unusual misfiring cylinder in the engine of cognition: our deliberative capacities have the function not so much of determining when we ourselves are wrong, but in persuading people -- both others and, for that matter, ourselves -- that we are right.⁶

To this end, we included a couple of different standard psychological scales for measuring our participants' interest in and skill at reflection. One instrument we used is the *Wason Selection Task* (Wason 1966), in which subjects are asked to consider a conditional, and one side of each of 4 cards, and they are then asked to say which of those cards would need to be turned over to see whether that conditional is true of it. In abstract versions of the task, subjects have a strong tendency to turn over the card representing the antecedent's being true, which is correct. Despite this, respondents tend to fail to turn over the card representing the consequent's being false, and also tend to turn over the card that represents the consequent's being true. A subset of subjects have tended to respond correctly, and, of these respondents, they have tended to be more thoughtful persons (Stanovich and West 1998). Moreover, in one version of the study conducted on a computer (Evans 1996), subjects were asked to use the mouse to hold their on-screen cursor over any card they were thinking about. It turned out that a strong predictor of whether a subject would correctly turn over the false-consequent card was whether or not the subject paused to think about that card at all. We therefore deployed the WST as one way of operationalising reflection on the part of our subjects, and we could look to see

⁶ See also Weinberg et al. (2013) for some recent experimental philosophy results suggestive of ways in which reflection may fail to overcome intuitive biases by introducing its own biases.

whether subjects who gave the logically normative responses on the WST, and those who did not, answered Liar-related questions in different ways.⁷

We also employed another instrument designed expressly to assess the tendency of participants to engage specifically in reflective thought. The *Cognitive Reflection Test* (Frederick 2005) is comprised of three questions that each have an obvious, intuitive, and incorrect answer, and a correct answer that can be recognised with the application of a fairly modest amount of reflection. Subjects are scored 0-3, based on how many of the questions they get correct.⁸ A high score requires both a willingness to second-guess one's initial responses, and also at least a modicum of mental wherewithal to suss it out. Importantly, scores on the CRT have shown to predict success at a great many cognitive tasks and decreased susceptibility to various cognitive biases, though not all of them (including myside biases, as noted earlier).⁹

The use of these measures is common in the extant literature, but we acknowledge two possible limitations to our approach related to their application in the case of the Liar. First, it is possible that a person might be willing to spend some time and cognitive effort on the WST and CRT tasks while not bothering to do so when confronted by the liar. At best, we are measuring our subjects' *general* disposition toward and capacity for reflective

⁷ There is a debate in the literature as to whether the response of turning over the true-antecedent card and the false-consequent card really is, in fact, the normatively correct response to the task. We suspect that readers interested in the liar paradox will not be tempted by that view, but those interested in it can look at, e.g., Oaksford and Chater (1994).

⁸ For example, one question reads: "If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes" The answer "100!" jumps right to mind, but it is not hard to see that the general rate of production here is that n machines can make n widgets in 5 minutes -- and so 5 is, in fact, the correct answer.

⁹ See also Pinillos et al. (2011) for an application of the CRT in an experimental philosophy context, to distinguish high-reflection from low-reflection subjects in a "Knobe effect" task.

thinking, but these instruments cannot determine whether any such capacity was brought to bear in the face of the deeply puzzling question of the liar paradox.¹⁰ Nonetheless, we take ourselves to have good reason to expect that WST and CRT scores, while certainly not perfect predictors of reflection on LIAR tasks, would generally correlate well enough with such reflection as to be a useful operationalisation.

Second, and more importantly, while these measures describe the extent to which some participants engaged, to a greater or lesser *degrees*, in reflective thought compared to other respondents, these measures do not tell us about the actual thinking processes employed by them. In particular, we cannot tell, e.g. the extent to which they did or did not explicitly consider or try to apply principles such as the principle of bivalence or the principle of non-contradiction. The results in aggregate do indicate whether responses tend to conform to such principles, and this is enough to warrant some cautious speculation.

3. Experimental Design and Methodology

Participants were recruited online and through *Amazon Mechanical-Turk* (“M-Turk”), and each respondent was compensated approximately \$1.50 (USD) for 10–15 minutes of

¹⁰ *Mutatis mutandis*, it is at least possible that some subjects were bored by the somewhat arbitrary-seeming tasks of WST and CRT, but whose attention -- and reflective powers -- were engaged by the deeply interesting question of LIAR.

their time.^{11,12} Our project attracted a geographically and demographically diverse group of 237 total participants aged 18–65 years. Of these, 162 (68.4%) self-identified as male and 74 (31.6%) self-identified as female. Further, 62 (26.2%) of the participants self-identified as having postgraduate training in philosophy (training beyond an undergraduate degree).¹³ Respondents were asked a series of questions, including a version of CRT and WST, as well as probing questions connected with the liar paradox (e.g., LIAR, the sentence-in-the-box paradox, Buridan’s Bridge paradox).¹⁴ Following on these questions, each person was asked to respond to a series of demographic questions regarding, for example, their political and religious affiliation, gender, age, education level attained, and socio-economic status.¹⁵

¹¹ Responses were collected through the University of Mississippi’s *Qualtrics* internet portal. Co-investigators sought and received IRB approval through the University of Mississippi (Approval #14x-163).

¹² M-Turk is populated by people who have access to the Internet and who are generally computer literate. We can understand why some might resist the use of this respondent pool, but we foresaw no reason not to employ the M-Turk population. Beyond that, we placed few restrictions on those who could respond to the survey questionnaire. Only respondents who self-identified as being a part of sensitive populations, e.g., children under the age of eighteen, prisoners, etc., were forbidden from submitting responses. Published studies of the M-Turk population suggest that the sample is in many ways superior to the standard social psychological sample of college students (Buhrmester et al. 2011, Paolacci et al. 2010).

¹³ To ensure that we would be able to compare philosophically trained participants to philosophical novices, in addition to using M-Turk, we solicited participants with philosophical training through announcements on philosophy oriented social networking sites, professional blogs, and listservs. That fewer females participated in the study may have been partly due to the gender imbalance in the sample from the philosophy profession.

¹⁴ For an overview of the paradox, see Ulatowski (2003).

¹⁵ No other personal information, such as name, address, email address, or phone number, was collected from respondents in order to preserve their anonymity.

We employed a “within-subjects” study design such that the same subjects were queried in all cases.¹⁶ On the working assumption that the ability to manifest reflective thinking would influence how one thinks about the liar, we hypothesized that how people evaluate the liar sentence would vary based upon their performance on the CRT and WST.

4. LIAR Experiment

All participants were presented with a randomized list of 17 claims (see Appendix A), and asked to sort them into one of four possible classifications: True (TRUE), False (FALSE), Both true and false (BOTH), Neither true nor false (NEITHER). Among the 17 claims listed in Appendix A are the following:

[LIAR]: “This sentence is false”

[CON]TRADICTION: “A sentence cannot be both true and false”

[BI]VALENCE: “Every sentence is either true or false”

Then, all participants received a CRT and WST before receiving a probe question, including the liar paradox or liar-like paradox.

As we saw earlier with Maudlin and with Barwise and Etchemendy, attempts to philosophically analyze problems like the liar paradox tend to make a wide range of assumptions, implicit or explicit, about the philosophizing subject. One immediate way in which the empirical investigation can be relevant to problems like LIAR is that we can recognize the meaningful effect that fairly common individual level differences might have.

¹⁶ Questions were counterbalanced to detect carryover effects, but none was discovered. Our suspicion is that not finding an order effect may have been because the number of participants was too small. Though, that suspicion cannot be confirmed without running another study with more respondents.

INSERT FIGURE 1 ABOUT HERE

Figure 1. LIAR (N = 236) Response pattern

The distribution of people's responses to LIAR reflects in the following pattern: 5.0% TRUE, 56.4% NEITHER, 17.8% BOTH, and 20.8% FALSE (See *Figure 1*). Initially, two key points emerge from our data: (1) people's responses, despite having to choose among just the four we experimenters provided them, are richly varied and fragmented, and (2) the NEITHER response, which is one of the four choices offered, is the most common.

Determining whether the distribution of responses reflects explicit differences of logical or metalogical beliefs or arises from other kinds of individual level factors is something that interested us in our analysis. So, first, we considered whether the awareness of philosophical matters in people's responses to LIAR. Subjects were individuated according to a measure we deemed "philosophical awareness." "Philosophically aware" participants (according to our stipulation) were those who self-reported as having had more than 2 college level courses in philosophy ($N = 93$). We compared them with "philosophically unaware" respondents, those who self-reported having had fewer than 2 college level courses in philosophy ($N = 144$).

INSERT FIGURE 2 ABOUT HERE

Figure 2. Comparison of Philosophically Aware and Philosophical Unaware Responses to LIAR

While the observed response patterns between the philosophically aware and unaware appear to differ, the difference is not statistically significant ($\chi^2 = 5.559, ns$) (See *Figure*

2).¹⁷ This result makes sense, however, if philosophical training must reach a threshold level in order to have a measurable effect or if the philosophical training must be combined with some other factor. The liar paradox is a problem familiar to many philosophers, and one would naturally expect that a higher level of philosophical training ought to have a meaningful effect on how one responds to it.

To examine this possibility we considered whether the responses of people with high levels of philosophical training would differ from the responses of others. We compared responses provided by “philosophical experts” (graduate students and holders of a Ph.D. in philosophy), $N = 61$, with those participants who are non-expert (everyone else), $N = 175$.

INSERT FIGURE 3 ABOUT HERE

Figure 3. Comparison of Philosophical Experts and Non-Experts Responses to LIAR

In this case, the responses not only appear to differ but there was a statistically significant difference between the responses of experts and non-experts ($\chi^2 = 11.701$, $df = 3$, $p < .008$, Low-to-medium effect size, *Cramer's V* = .223) (See *Figure 3*).

Other categorical factors that might influence how one responds to LIAR, of course, include whether or not one is a reflective thinker. Philosophical expertise and reflective thought are often thought of as complementary, despite that exercising one's expertise and exercising reflective thought are not necessarily the same thing. To test whether reflective thinking makes a difference, we looked at response patterns to LIAR based upon whether

¹⁷ A Pearson χ^2 statistic enables researchers to investigate whether the response distributions of categorical variables, e.g., “what is your gender?”, “Do you have a Ph.D. in philosophy?”, differ from one another. It evaluates how likely it is that any observed difference between the sets arose by chance.

respondents gave a “passing” response to our modified WST¹⁸ and whether they scored a perfect ‘3’ on the CRT.

INSERT FIGURE 4 ABOUT HERE

Figure 4. Comparison of WST Passers (N = 52) and WST Failers (N = 184) to LIAR

In the WST case, the responses differ dramatically for the NEITHER and FALSE responses. Comparing WST Passers with WST failers on LIAR, we see that the difference between them was statistically significant ($\chi^2 = 16.084$, $df = 3$, $p < .001$, Low-to-Medium Effect, *Cramer's V* = .261) (See *Figure 4*).

Similar results cropped up when we compared the maximal score of “3” on CRT with those who scored lower than that.

INSERT FIGURE 5 ABOUT HERE

Figure 5. Comparison of CRT Failers (N = 114) and CRT Passers (N = 122)

Again, we see that the response patterns differ where the two most obvious differences are NEITHER and FALSE. For CRT Passers and CRT Failers on LIAR, the differences in responses were statistically significant ($\chi^2 = 9.214$, $df = 3$, $p < .027$, Low-to medium effect, *Cramer's V* = .198) (See *Figure 5*). These two results suggest a strong connection between the capacity for reflective thought and how people respond to LIAR.

Of these two tests, the CRT's scaled response allowed us to look at whether or not higher CRT scores would correlate with a specific response pattern to LIAR. Appealing to

¹⁸ Of respondents, only people who selected the minimum, ideal answer, i.e., turning over the card with the true antecedent, and the card with the false consequent, received a “passing” grade counting as “correct.” If the participant selected any other set of answers, they were counted as getting the WST “wrong.”

Maudlin’s model, and the fact that NEITHER was the most common response, we defined a new variable combining all non-NEITHER responses to LIAR (NEITHER responses, $N = 133$; Non-NEITHER responses, $N = 103$). We then compared the combined non-NEITHER response, what we will call below “OTHER” responses, with NEITHER responses.

INSERT FIGURE 6 ABOUT HERE

*Figure 6. Comparison of NEITHER responses with Other responses for LIAR
(by percentages of respondents)*

Here the difference between the distribution of NEITHER/OTHER responses varies significantly by CRT score category ($\chi^2 = 14.157$, $df = 3$, $p < .003$) (See *Figure 6*). There is an observable trend that CRT score and percentage of respondents answering NEITHER to LIAR are correlated (Kendall’s $\tau c = .214$, $est T = 3.172$, $p < .002$). Even when we control for philosophical expertise, this correlation remains significant (Non-experts: Kendall’s $\tau c = .164$, $est T = 2.040$, $p < .045$; Experts Kendall’s $\tau c = .253$, $est T = 2.113$, $p < .035$). That the correlation between CRT score and response patterns is present for philosophical experts and non-experts alike shows that the effect of philosophical education and the effect of reflective thinking are not the same effect.

We have reported experimental results above that show a significant number of people accept that a response to LIAR is NEITHER. Such a response seems to be a denial of the law of excluded middle, i.e., the logical principle that states some proposition is either true or false. Because we partly expected that response to be popular for LIAR, we decided also to ask participants whether they believed the law of non-contradiction, i.e., “A sentence cannot be both true and false,” is true or not-true. Interestingly, 68.9% of 90 participants who said that CON is true believed LIAR was NEITHER, while 48.8% of 146 respondents who agreed CON is not true responded that LIAR was NEITHER.

INSERT FIGURE 7 ABOUT HERE

Figure 7. Comparison of Respondents on the Law of Non-Contradiction

Again, as we have found in the other cases, the response patterns differ, especially for the response NEITHER. Similarly, and as one might expect, respondents who believed that CON was not true tended to respond BOTH to LIAR, and participants who believed that CON is true tended to agree less with BOTH to LIAR. The difference was statistically significant ($\chi^2 = 12.948$, $df = 3$, $p < .005$, Low-to medium effect, *Cramer's V* = .234) (See *Figure 7*). However, as will be important to our arguments later, both CON-endorsers and CON-rejecters shared the strong modal response of NEITHER.

We also queried respondents on the principle of bivalence, i.e., the view that every proposition is either true or false. 49.2% of 59 respondents who believed the principle of bivalence is true responded that LIAR is NEITHER, and 58.8% of 177 respondents who believed the principle of bivalence is not-true responded that LIAR is NEITHER.

INSERT FIGURE 8 ABOUT HERE

Figure 8. Comparison of Respondents on the Principle of Bivalence

Just as we have discovered with CON responses and others, the response patterns differ for BI. The difference between response patterns was statistically significant ($\chi^2 = 10.278$, $df = 3$, $p < .016$, Low-to medium effect, *Cramer's V* = .209) (See *Figure 8*) -- though, again, with many participants choosing NEITHER. In short, it seems that while some of the overall variation in responses may be explained in terms of differences in endorsement of general logical principles, conscious access to those principles themselves do not seem to be playing much of a role in determining the most general pattern of responses

Before proceeding we want to note two additional results. As we noted above, work by Barnard and Ulatowski (2013, ms1) found interesting differences in the trends of how men and women think about truth, and in particular, their data seem to indicate that women respondents are on average less likely to believe truth is correspondence to reality, and women respondents are more inclined on average than men to believe that some

degree of evidence is required for truth. In turn, we looked at whether there was a corresponding effect of gender on the response pattern for LIAR. We did find a significant difference between male and female response patterns to LIAR ($\chi^2 = 16.502$, $df = 3$, $p < .001$, Low-to medium effect, *Cramer's V* = .264) (See *Figure 9*). We did not make an explicit prediction in advance of what this difference might look like, but we can offer a post hoc interpretation of our finding. So far as we can tell from the data, women were on balance more willing than men to respond with FALSE to LIAR.

Given the Barnard and Ulatowski results, the male participants in our current study might be comparatively reticent for any claim to count as true or false unless it, or its negation, corresponds with the way the world is, and LIAR, since it lacks such a truthmaker, would seem a poorer candidate for either one of the traditional truth values. Moreover, since no evidence plays a role in deciding the truth value of LIAR, women might be somewhat more willing to offer a rejection of it as FALSE. We would note, however, that this difference shows up against a background of NEITHER still being the seemingly preferred response for both groups. Our claim is not that women and men deploy the notion of truth in *radically* different ways, let alone that they have anything like different concepts of truth, but only that these findings suggest some intriguing differences around the edges regarding how men and women confront the paradox.

INSERT FIGURE 9 ABOUT HERE

Figure 9. Comparison of Male v. Female Respondents

Moreover, also, as mentioned above, we hypothesized that since the theological commitments of some religious traditions affirm claims that can be interpreted as contradictory, e.g., that in the holy trinity 1=3, theists might have an openness to contradiction which in turn might have an effect on responses to LIAR. Here, too, we found a significant difference, between theist and non-theist participants ($\chi^2 = 10.033$, $df =$

3, $p < .018$, Low-to medium effect, *Cramer's V* = .206) (See *Figure 10*).¹⁹

INSERT FIGURE 10 ABOUT HERE

Figure 10. Comparison of Theist and Non-Theist

These individual level differences are interesting to contemplate, but their overall importance in understanding how the folk think about the liar requires a more rigorous evaluation.

5. What Liars Beneath: Towards sorting out the contributions of these various factors

What story do our data tell about what ordinary cognitive resources do or do not get deployed when confronting LIAR, and by whom? First, our data suggest that further reflection is leading respondents to have a different *intuitive* response. Our contention is that reflection, when an ordinary person confronts the liar paradox, should be understood on a model of reorganizing the available information in a manner that *shifts* their intuitive responses, and not on a model of recruiting general information into our conscious deliberations, which is then used to *override* an initial intuitive response. Consider the Evans (1996) WST study: on that account, the key job of reflection in performing that task seems to be to just get System 1 paying attention to the "~Q" case at all -- but, once it does, it is easily and intuitively seen to be a card that needs to be turned over. In general, WST studies do not seem to report any residual intuitive sense in their subjects that that card should not be turned over; this can be contrasted with the psychologically persistent

¹⁹ We collected fine-grained responses from participants allowing them to self-identify by broad religious tradition, e.g., Protestant Christian, Jewish, Muslim, etc. These responses were combined so that people who self-identified in one of these religious traditions were counted as theists, and those who answered "none" or "other" were coded as non-theist. (There were only a small handful of "other"s.)

intuitive pull of, e.g., the conjunction fallacy in the face of explicit reflective correction (See Kahneman and Tversky 1996).

Our reasons for preferring the “shifting” model over the “overriding” model, in this area, are the results described in Figures 7 and 8: our participants displayed a variety of attitudes towards basic logical principles, and while their doing so produced interesting effects on their patterns of responses, the principles themselves do not seem to be driving the NEITHER response. How could they, when they are not even endorsed by groups of participants who have that response? Our data cannot rule out, and perhaps they even suggest, that *some* participants are relying consciously on those principles. But it can only be a minority of them at best.

We thus arrive at the following account of the general difference between those who gave the NEITHER response and the smaller but not insubstantial number of our participants who gave such answers as FALSE. We conjecture that the latter group are primarily participants who simply failed to reflect on the target sentence. When confronted with a sentence that says of itself that it is false, these participants apparently resisted the urge -- or, more likely, failed even to have any impetus -- to reconsider the truth value of the sentence, and simply chose TRUE or FALSE in effortless accordance with the deliverances of their System 1. As the data seem to suggest, the pragmatic response to the truth value of LIAR is FALSE because that is exactly what the sentence says it is -- FALSE. Such folks judge the truth value of the sentence efficiently without necessarily any accompanying awareness of the source of that judgment.

In contrast, participants who tended to say NEITHER in response to LIAR are more reflective. Upon reflection, these respondents *see* that flat-footed responses, TRUE or FALSE, lead to an unwieldy result: *if LIAR is false, then it is TRUE* and *if LIAR is true, then it is FALSE*. Sometimes explicitly but more often implicitly, such participants take it that neither TRUE nor FALSE, *sans phrase*, is a stable response to LIAR, so they choose NEITHER.

One might still ask: *Why do so comparatively few subjects, both reflective and not, say BOTH?* The “Maudlin model” is helpful here. Remember, Maudlin suggested that once the person comprehends the paradoxicality of TRUE and FALSE, BOTH is no longer a

live option. He writes, “If both [TRUE] and [FALSE] are wrong, surely [BOTH] is” (Maudlin 2004, 1). Our theory is that reflective subjects intuit that neither TRUE nor FALSE are legitimate options, and then can draw the easy inference along those lines Maudlin suggests. For respondents who are relying only upon the deliverances of their intuitive System 1, BOTH is not a live option just as NEITHER is not a live option -- their reason seems to be: *that’s not what the sentence says of itself*. The more complicated answers simply fail to present themselves as intuitive candidates, if reflection is not brought to bear.

To discover which interpretation is getting closer to what the data report is to figure out which of the factors matter most in driving participants’ responses to LIAR. In order to determine which of the factors that generate distinct responses to LIAR are most salient, we had to perform a logistic regression analysis. The regression model included factors that were earlier found to yield significant differences in response patterns: respondent gender and theism, CRT score, WST performance, and acceptance of bivalence and non-contradiction. Significant results indicate that a factor has an effect on response patterns for LIAR after controlling for all other factors in the model.²⁰ The β -coefficient reports how many standard deviations a dependent variable will change per standard deviation increase in the predictor variable.

We performed two multinomial logistic regression analyses. One compared NEITHER with FALSE (*Table 1*), and the other compared NEITHER with BOTH (*Table 2*).

	β -coefficient (SE)	LOWER	Odds Ratio	UPPER
CON TRUE	0.445 (.451)	.651	1.576	5.177
BI TRUE	-1.112 (.448)*	.137	.329	.791
WST PASS	-1.1614 (.823)*	.040	.199	.998

²⁰ The model was not hierarchical or analyzed stepwise, we had no reason to order the supposed predictors.

THEIST	0.600 (.382)	.863	1.823	3.851
PHIL EXPERTISE	-0.399 (.605)	.205	.671	2.195
GENDER	0.819 (.389)*	1.059	2.269	4.858
CRT = 0	-0.376 (.525)	.246	.686	1.919
CRT = 1	-0.755 (.597)	.146	.470	1.514
CRT = 2	-0.346 (.477)	.278	.708	1.803

*Table 1. A Summary of Logistic Regressions Results for NEITHER vs. FALSE
(95% confidence interval; * = $p < .05$)*

When we originally compared responses from people, e.g., who answered that CON is TRUE or who self-reported as theist, with those from people in the complement class we discovered significant differences. But those particular differences do not remain significant when the effect from other factors in the model are considered at the same time. This is not the case for acceptance of BI, gender, and WST performance. These three factors have an effect that is sufficiently robust to remain significant when we control for the effect of all the other factors. Plainly put, these factors seem to drive whether a respondent answers NEITHER (as opposed to FALSE). A similar regression analysis was performed looking at what factors would influence respondents to respond NEITHER rather than BOTH. In this case, the only factor that remained significant was

respondent gender.²¹

	β -coefficient (SE)	LOWER	Odds Ratio	UPPER
CON TRUE	-1.023 (.614)	.108	.360	1.198
BI TRUE	-1.039 (.592)	.111	.354	1.130
WST PASS	-1.392 (.932)	.040	.249	1.545
THEIST	0.852 (.467)	.938	2.343	5.854
PHIL EXPERTISE	-0.253 (.718)	.190	.776	3.173
GENDER	1.035 (.485)*	1.088	2.851	7.282
CRT = 0	-2.630 (.778)	.125	.573	2.630
CRT = 1	0.785 (.652)	.611	2.191	7.860
CRT = 2	0.409 (.561)	.501	1.505	4.520

*Table 2. A Summary of Logistic Regressions Results for NEITHER vs. BOTH
(95% confidence interval; * = $p < .05$)*

Given that the Liar sentence says of itself that it is false, we treated that response as the main intuitive default view. Our regression analysis was designed to see what factors in the model could help explain why participants answered something other than “False.”

²¹ These regression results should not be overlooked, since the first results show how WST Passers, BI truthers, or being female increases the log odds of choosing NEITHER over FALSE (cf. *Table 1*) and the second results only show being female increases the log odds of choosing NEITHER over BOTH (cf. *Table 2.*). The effect of commitment to bivalence and logical acumen operationalized by WST is what we would, in some sense, hope to find. The gender result is perplexing, for it suggests that there may be a real, yet unexplained, effect of gender on how one responds to LIAR. One might try to explain the effect in terms of the fact that the philosopher sample was heavily male. But if that were the key factor, one would expect that philosophical expertise would also end up as a significant predictor. According to our analysis, it is not. (Statistical aside: Because we employed SPSS to produce the logistic regression data, we ought to remind the reader that instead of t ratios, it calculates using the “Wald statistic,” which is a test statistic distributed as a χ^2 where $df = 1$ and its value equals the square of a t ratio.)

When all of the other factors are controlled for, people who reject bivalence are about three times more likely to respond NEITHER than they are likely to respond FALSE. When all of the other factors are controlled for, those who can pass the modified WST test (i.e. understand classical logical implication) are about five times more likely to respond NEITHER than they are likely to respond FALSE. When all of the other factors are controlled for, men are roughly 2.25 times more likely to respond NEITHER than women are likely to respond NEITHER.²²

Our work here is at best suggestive, and in particular should not be taken as in any sense definitive in the absence of further exploration and, especially, replication. And much more work is required to determine to what extent reflection is efficacious because we bring logical principles to bear, or instead because it gets our intuitive mental logic capacities working properly on the problem -- and indeed, whether there are interesting interpersonal differences in the answer to that question.

6. Sentence-in-the-Box Experiment results

One experiment showing that reflective people tend to choose the NEITHER response for LIAR might not suffice for the explanation we gave of the responses to LIAR in Section 5. Some respondents were randomly assigned the sentence-in-the-box case. Participants receiving the sentence-in-the-box case (hereafter “SB”) were presented with the following:

INSERT FIGURE 11 ABOUT HERE

Figure 11. The Sentence-in-the-Box Example

And they were asked to choose whether the proposition is: TRUE, FALSE, BOTH, or NEITHER. Of the 60 study participants who received SB, 58.3% responded NEITHER, 20.0% responded FALSE, 16.7% responded BOTH, and 5.0% responded TRUE (*Figure*

²² Regression analysis suggests that although responses to LNC and BI play a role in predicting responses to the LIAR, none of these factors has a sufficiently large R^2 value to ground the claim that they drive the response to LIAR.

12.).

INSERT FIGURE 12 ABOUT HERE

Figure 12. SB Response Pattern

If we compare the raw descriptive data of responses of LIAR (cf. *Figure 1.*) with the responses of SB (*Figure 12.*), we notice a striking similarity between them: no univocal response, but with a strong modal endorsement of NEITHER. Indeed, the overall proportions of the four answers are highly similar across the two tasks. Once we saw how closely the two sets of data overlapped, we decided to subject the results of SB to analyses similar to the ones we ran for LIAR.

What we found for SB was largely consistent with our findings for LIAR. For philosophical expertise, the differences were statistically significant ($\chi^2 = 8.469$, $df = 3$, $p < .037$, Medium effect, *Cramer's V* = .376) (See *Figure 13.*).

INSERT FIGURE 13 ABOUT HERE

Figure 13. Comparison of Experts (N = 18) and Non-Experts (N = 42) for SB.

For WST Passers and Failers, the difference was statistically significant ($\chi^2 = 11.854$, $df = 3$, $p < .008$, Medium-to-large effect, *Cramer's V* = .444) (See *Figure 14.*)

INSERT FIGURE 14 ABOUT HERE

Figure 14. Comparison of WST Passers (N = 52) and WST Failers (N = 184) for SB.

For CRT Passers and Failers, the difference was statistically significant ($\chi^2 = 8.042$, $df = 3$, $p < .045$, Medium effect, *Cramer's V* = .366) (See *Figure 15.*)

INSERT FIGURE 15 ABOUT HERE

Figure 15. Comparison of CRT Passers (N = 122) and CRT Failers (N = 114) for SB.

When we use respondents' actual score on the CRT to assess the role of reflective thinking in a more fine-grained way, the difference was statistically significant for SB, just as in LIAR ($\chi^2 = 24.752$, $df = 9$, $p < .003$). There is an observable trend that CRT score and percentage of respondents answering NEITHER to SB are correlated (Kendall's $\tau c = .259$, $est T = 2.865$, $p < .004$) (Figure 16.). When we control for philosophical expertise, the correlation remains significant for experts (Experts: Kendall's $\tau c = .48$, $est T = 2.306$, $p < .02$).

INSERT FIGURE 16 ABOUT HERE

*Figure 16. Comparison of NEITHER responses with Other responses for SB
(by percentages of respondents)*

The collected data on SB resemble the LIAR data, which seems to show that we have a conceptual replication of the effect concerning reflectivity and liar-like semantic paradoxes. Primarily, the finding just is that respondents whose intuitive responses tend to reveal a certain amount of reflection agree with the response NEITHER for SB, while those participants who responded with FALSE seem to reveal that their intuitive responses were driven by pragmatic, automatic, and associative responses. The stark difference between low-CRT and high-CRT scoring subjects is plainly visible in Figure 15.

These dramatic differences by CRT score lead us to an important cautionary note about how to interpret what our results seem to be saying about the population on the whole. By design, our sample recruited a higher proportion of philosophers and of high-CRT people than is representative of the population on the whole. (For example, Frederick (2005) reports that about 60% of his college-aged participants got a score of CRT-0 or CRT-1; in contrast ***) Although NEITHER is still a more common answer

than any other single answer in the low-CRT groups, FALSE is a *very* close second behind it in that part of our sample.

Of course, we should not close the case just yet. Even putting aside the fact that we are only reporting one study here, there are also a great many variations on the liar paradox, and perhaps there is an iteration of it that yields dissimilar results, i.e., data that show people tend not to reflect in some response to some instances of semantic paradox. We believed early on that people's responses to LIAR might differ greatly from their responses to other paradoxes, largely because people might not have an overwhelming reaction to LIAR. After all, think of how freshmen in an Introductory Philosophy course often react to the notion that the principle of moral or epistemic relativism, i.e., "there are no objective moral / epistemic principles," is self-defeating. Sometimes novice thinkers just are not able to grasp the problem. It is too difficult for the person to step back and get a look at the principle for what it is saying. This led us to our third and final experiment, exploring one further way of presenting the liar paradox

7. "Theoretical" Buridan's Bridge Experiment Results

While there is good statistical evidence that people look at LIAR and SB in similar ways, when less abstract formulations of the paradox are employed the example might elicit different intuitive responses.

We take it that both LIAR and the SB are abstract because the sentences refer to themselves, and some people, particularly those who employ language without being cognizant of distinct ways of manufacturing some forms of ostensive definition, might not see the sentence for doing just that. People fail to see how the proposition refers to itself.

Other versions of the liar and liar-like paradoxes are not so abstract. One example is the paradox of Buridan's bridge. Even though it is an example of a semantic liar-like paradox, in order to understand it one need not abstract away from what some proposition says to comprehend the aporetic reasoning the paradox evokes. One merely needs to hear the story and understand the narrative to comprehend that what one of the

characters appearing in the vignette seems to land us in an aporetic situation. Since the paradox requires the reader not to employ an abstract form of reasoning, including the recognition of how a sentence refers to itself, we have distinguished it from LIAR and SB.

Some participants were randomly assigned to receive the following version of the Buridan's bridge paradox.

Suppose that Jones stands guard at a footbridge crossing a river. It's the only foot bridge for many miles. Smith approaches the bridge and asks Jones whether he may cross. Jones declines. Smith pleads with Jones to allow him to cross the bridge because he doesn't want to walk to the next foot bridge many miles from this one.

Jones says, "if the first sentence you utter is true, then I will allow you to cross the bridge. But, if the first sentence you utter is false, then I will throw you into the water."

Smith responds, "you will throw me into the water."

If Jones throws Smith into the water, then the first sentence Smith uttered was true and Jones should have let him pass. If Jones lets Smith pass, then the first sentence Smith uttered was false and Jones should have thrown him into the water.²³

Participants were asked to evaluate whether

²³ All participants were asked to read the vignette and then answer what Jones should do. We call it the "practical" version because it asks participants to choose what Jones should do, rather than tell us whether the sentence "you will throw me into the water" is true or false. Respondents were able to choose one of the following answers in the "practical" case: Jones should throw Smith into the water, Jones should allow Smith to pass, or Jones should not do anything. We have left aside our findings from the "practical" case for this article because we preferred to focus upon the experiment that elicited response to liar and liar-like semantic paradoxes.

THROW: You will throw me into the water

is TRUE, BOTH, NEITHER, or FALSE. 114 subjects were presented with this item. The general distribution looked like *Figure 17*.

INSERT FIGURE 17 ABOUT HERE

Figure 17. Descriptive Data for Buridan's Bridge THROW case.

This overall response pattern for THROW is observably different from the response pattern for LIAR. We note, for instance, that the modal response to THROW is BOTH. At a statistical level the response patterns are significantly different. This was determined using a likelihood ratio analysis comparing LIAR and THROW ($\lambda = 19.058$, $df = 9$, $p < .025$, Small-to-medium effect, *Cramer's V* = .213).²⁴

At a minimum, we do now have some reason to think that this presentation of the liar leads to a different pattern of responses. We are inclined towards the conjecture that the ability to confront the paradox abstractly is diminished by the practical factors involved in how the paradox is presented. But the best we can say at present is that our data are consistent with such a conjecture; further investigation of practical presentations of the liar are required.

8. Conclusion

With our results now before us, let us conclude by revisiting our empirical hypotheses that motivated our study, as initially discussed in section 2.

First, regarding the disagreement among philosophers regarding the coherence or incoherence of folk cognitive resources to address the liar, we do take ourselves to have

²⁴ For this comparison, the Pearson's χ^2 test results do not quite reach the threshold of significance. ($\chi^2 = 15.446$, $df = 9$, $p = .079$, *ns*). We suspected that this was a reflection of the smaller sample size. The likelihood ratio test is more sensitive; it is equivalent to the standard Pearson's χ^2 test for large data sets and is preferred for smaller sample sizes.

some evidence for a particular psychological model of how people engage the liar paradox. It appears that quick, associative responses to theoretical forms of the paradox, such as LIAR and SB, tend to be simply “true” or “false”, mostly “false”. But fast and frugal responses are distinct from more reflective intuitive responses, and that reflection, especially if aided by some philosophical training, generally leads people to endorse a “neither” response. For practical presentations of the paradoxes, such as the Buridan’s Bridge paradox, respondents have a greater tendency to stay with that fast, pre-reflective response. In all, it seems that initial and pre-reflective folk responses are perhaps not well-suited to engage with the paradox. The further resources that are put into play by a process of reflection are able to track the paradoxical nature of sentences like LIAR, and on the whole, something like the “Maudlin model” seems to be consistent with our findings. Interestingly, it seems that the relevance of reflection may be less a matter of bringing explicit beliefs about logical principles to bear, and more a matter of shifting the deliverances of our intuitive systems themselves. The psychological element of Barwise and Etchemendy’s “breakdown” model is perhaps disconfirmed, in that under reflection, subjects do not fragment into a range of responses as if at random; rather, the NEITHER response becomes more and more attractive, apparently, the more reflective and the more philosophically-trained a person is.

Our second set of hypotheses, regarding some interpersonal variation in responses to the paradoxes, also received some confirmation: within our sample, female subjects did give a somewhat different pattern of responses than male subjects, and likewise for self-identified theists and self-identified non-theists. As noted above, such results would need to be replicated and extended before taken as in any way definitive. Those two variables are surely confounded with other demographic variables as well, such as personality types, political orientation, and socio-economic status, and this initial study has only controlled for a few such possible confounds.

Regarding the question of philosophical expertise, the basic news here seems to be basically encouraging for philosophers: greater reflection and greater philosophical training were both factors that shifted subjects’ pattern of responses away from TRUE or FALSE and towards NEITHER. Proponents of a truth-value gap approach will perhaps

find some evidential ammunition in that result, and we generally incline that way ourselves. Nonetheless, we think it important to issue two big caveats in that regard. First, there is still at least *some* variability in the even the most expert sample, with, e.g., 15% or so answering FALSE (though our sample apparently contained vanishingly few expert philosophers willing to offer the dialethic answer BOTH.) Second, the fact that philosophical training and cognitive reflection produce a convergence in our subjects' answers, certainly does not entail that they have converged on the truth.

In general, it is important to distinguish between a descriptive psychological model, and any normative implications one might take to follow from that model. (This is an important part of why, as we emphasized at the start, we were not at all recommending using x-phi surveys as any sort of replacement for formal and philosophical work on the paradox itself, and, in this way, we are assuming Næss' methodological mantle (cf. Næss 1938a, §97).) In this case, at least *prima facie*, our descriptive results are congruent with a particular normative analysis: since greater reflection or expertise lead to an increased tendency towards the "neither" response, that is legitimately some reason to think that that response is more likely to be the *correct* one. Although we are basically sympathetic to that normative inference, we would emphasize that it is indeed an inference, and as such, it needs to be licensed by some further assumptions about reflection and philosophical training. Those assumptions themselves need to be subjected to further empirical scrutiny. We do have a bit of evidence suggesting that there *may* be some differences in the way that different people do come to grapple with the paradox, including potentially gender or religious background. And should such differences turn out to be robust, this could undermine our sense that these data perhaps provide evidence for the normative psychological claims. Nonetheless, we would emphasize that the normative import of our findings are conjectural, at best, pending further research. Despite their preliminary nature, we hope that our findings serve to highlight the potential value of further inquiry into the psychological dimension of the liar paradox and similar problems. Research of the kind we discuss here can help to shed light old problems and lead to an awareness of new areas of potential research.

Professor
University of Mississippi
Department of Philosophy & Religion
100 Bryant Hall
University, MS 38677 U.S.A.
rwbjr@olemiss.edu

Lecturer
University of Waikato
Philosophy Programme
Postal Bag 3105
Hamilton 3240
New Zealand
julatows@waikato.ac.nz

Associate Professor
University of Arizona
Tucson, AZ 85721 U.S.A.
jmweinberg@email.arizona.edu

References

- Barwise, J. and J. Etchemendy. 1987. *The Liar*. Palo Alto, CA: CSLI Publications.
- Barnard, R. and J. Ulatowski. 2013. Truth, Correspondence, and Gender. *Review of Philosophy and Psychology* 4: 621-638.
- Barnard, R. and J. Ulatowski. ms1. Just the Facts? Evidence, Gender, and the Ordinary View of Truth. University of Waikato, unpublished manuscript.
- Buhrmester, M., T. Kwang, and S. Gosling. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6.1: 3-5.
- Cacioppo, J. and R. Petty. 1982. The Need for Cognition. *Journal of Personality and Social Psychology* 42.1: 116-131.
- Cook, R. 2013. *Paradoxes*. Wiley.
- Epley, N. and T. Gilovich. 2006. The Anchoring-and-Adjustment Heuristic: Why the Adjustments are Insufficient. *Psychological Science* 17.4: 311-318.
- Evans, J. St.B.T. 1984. Heuristic and Analytic Processes in Reasoning. *British Journal of Psychology*. 75: 451-468.
- Evans, J. St.B.T. 1996. Deciding Before You Think: Relevance and Reasoning in the Selection Task. *British Journal of Psychology*. 87: 223-240.
- Evans, J. St.B.T. 1998. Matching Bias in Conditional Reasoning: Do We Understand It After 25 Years? *Thinking and Reasoning* 4: 45-82.
- Evans, J. St.B.T. 2009. How Many Dual-Process Theories Do We Need? One, Two, or Many? In *In Two Minds: Dual Processes and Beyond*, edited by Jonathan St. B.T. Evans and Keith Frankish, pp. 33-54. Oxford: Oxford University Press.
- Evans, J. St.B.T. and J.S. Lynch. 1973. Matching Bias in the Selection Task. *British Journal of Psychology* 64: 391-397.
- Evans, J. St.B.T. and D.E. Over. 1996. *Rationality and Reasoning*. Hove, England: Psychology Press.
- Frankish, K. 2009. Systems and Levels: Dual-System Theories and the Personal-Subpersonal Distinction. In *In Two Minds: Dual Processes and Beyond*, edited by Jonathan St. B.T. Evans and Keith Frankish, pp. 89-107. Oxford: Oxford University

- Press.
- Frederick, S. 2005. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives* 19.4: 25–42.
- Gilbert, D., R. Tatarodi, and P. Malone. 1993. You Can't Not Believe Everything You Read. *Journal of Personality and Social Psychology* 65.2: 221–233.
- Hales, S. 2000. The Problem of Intuition. *American Philosophical Quarterly* xx:135–147.
- Hales, S. 2004. Intuition, Revelation, and Relativism. *International Journal of Philosophical Studies* 12.3: 271–295.
- Hales, S. 2009. Moral Relativism and Evolutionary Psychology. *Synthese* 166.2: 431–447.
- Hales, S. 2012. The faculty of intuition. *Analytic Philosophy* 53(2): 180–207.
- Kahneman, D. 2003. Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review* 93.3: 1449–1475.
- Kahneman, D. and A. Tversky. 1972. Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology* 3: 430–454.
- Kahneman, D. and A. Tversky. 1996. On the Reality of Cognitive Illusions. *Psychological Review*. 103: 582–591.
- Ludwig, K. 2007. The Epistemology of Thought Experiments: First Person Versus Third Person Approaches. *Midwest Studies in Philosophy* 31.1: 128–159.
- Machery, E. 2009. *Doing Without Concepts*. Oxford: Oxford University Press.
- Machery, E., R. Mallon, S. Nichols, and S. Stich. 2004. Semantics, cross-cultural style. *Cognition* 92.3: B1–B12.
- Magdziak, M. 2000. Na antynomii semantycznych w języku potocznym. *Kwartalnik Filozoficzny* 28(4): 137–142.
- Magdziak, M. 2004. Prawda w języku potocznym i paradoks kłamcy. *Kwartalnik Filozoficzny* 32(2): 133–143.
- Maudlin, T. 2004. *Truth and Paradox: Solving the Riddles*. Oxford: Oxford University Press.
- Mercier, H. and D. Sperber. 2011. Why Do Humans Reason? Arguments for an Argumentative Theory. *Behavioral and Brain Sciences* 34: 57–111.
- Næss, A. 1937/1938. Zum Vortrag von Kokoszynska über Einheitswissenschaft. *Erkenntnis*

7: 382–384.

- Næss, A. 1938a. 'Truth' as Conceived By Those Who Are Not Professional Philosophers. Oslo: Jacob Dybwad Universitetsforlaget.
- Næss, A. 1938b. Common-Sense and Truth. *Theoria* 4.1: 39–58. [Reprinted in Næss 2005]
- Næss, A. 1961. A Study of *Or. Synthese* 13: 49–60. [Reprinted in Næss 2005]
- Næss, A. 1981. The empirical semantics of key terms, phrases and sentences: Empirical semantics applied to nonprofessional language. In *Philosophy and Grammar*, pp. 135–154, edited by S. Kanger and S. Ohman. Dordrecht: D. Reidel Publishing Company. [Reprinted in Næss 2005]
- Næss, A. 1992. How Can the Empirical Movement Be Promoted Today? A Discussion of the Empiricism of Otto Neurath and Rudolf Carnap. In *From an Empirical Point of View: The Empirical Turn in Logic*, pp. 197–255, edited by E. M. Barth, J. van Dormael, and F. Vandamme. Gent, Belgium: Communication and Cognition. [Reprinted in Næss 2005]
- Næss, A. 2005. *Selected Works of Arne Næss, vol. 8: Common Sense, Knowledge, and Truth*, edited by H. Glasser and A. Drengson. Dordrecht: Kluwer Academic Publishers.
- Paolacci, G., J. Chandler, and P.G. Iyengar. 2010. Running Experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5.5: 411–419.
- Peng, K. and R.E. Nisbett. 1999. Culture, Dialectics, and Reasoning about Contradiction. *American Psychologist* 54.9: 741–754.
- Pinillos, A., N. Smith, G.S. Nair, C. Mun, and P. Marchetto. 2011. Philosophy's New Challenge: Experiments and Intentional Action. *Mind and Language* 26.1: 115–139.
- Rescher, N. 2001. *Paradoxes: Their Roots, Range, and Causes*. LaSalle, IL: Open Court Press.
- Russo, J. and P. Shoemaker. 1989. *Decision Traps: Ten Barriers to Brilliant Decision Making and How to Overcome Them*. New York: Simon and Schuster.
- Schwitzgebel, E. and F. Cushman. 2012. Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers. *Mind and Language* 27.2: 135–153.
- Schwitzgebel, E. and J. Rust. 2009. The Moral Behaviour of Ethicists: Peer Opinion. *Mind*

- 118: 1043–1059.
- Slooman, S.A. 1996. The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin* 119: 3–22.
- Stanovich, K. 2009. Distinguishing the Reflective, Algorithmic, and Autonomous Minds: Is It Time for a Tri-Process Theory? In *In Two Minds: Dual Processes and Beyond*, edited by Jonathan St. B.T. Evans and Keith Frankish, pp. 55–88. Oxford: Oxford University Press.
- Stanovich, K. and R. West. 1998. Cognitive Ability and Variation in Selection Task Performance. *Thinking and Reasoning* 4: 193–230.
- Stanovich, K. and R. West. 2000. Individual Differences in Reasoning: Implications for the Rationality Debate. *Behavioral and Brain Sciences* 23: 645–665.
- Stanovich, K. and R. West. 2003. The Rationality Debate as a Progressive Research Program. *Behavioral and Brain Sciences* 26: 531–534.
- Stanovich, K., R.F. West, and M.E. Toplak. 2013. Myside Bias, Rational Thinking, and Intelligence. *Current Directions in Psychological Science* 22: 259–264.
- Tarski, A. 1944. The Semantic Conception of Truth and the Foundations of Semantics. *Philosophy and Phenomenological Research* 4: 341–376.
- Ulatowski, J. 2003. A Conscientious Resolution of the Action Paradox on Buridan's Bridge. *Southwest Philosophical Studies* 25: 97–106.
- Wason, P.C. 1966. Reasoning. In *New Horizons in Psychology*, edited by B. Foss. Harmondsworth, England: Penguin: 135–151.
- Weinberg, J., J. Alexander, C. Gonnerman, and S. Reuter. 2013. Restrictionism and Reflection: Challenge Deflected, or Simply Redirected. *The Monist* 95.2: 200–222.
- Williamson, T. 2004. Philosophical 'Intuitions' and Scepticism about Judgment. *Dialectica* 58.1: 109–153.
- Williamson, T. 2005. Armchair Philosophy, Metaphysical Modality and Counterfactual Thinking. *Proceedings of the Aristotelian Society* 128: 1–23.
- Williamson, T. 2007. *The Philosophy of Philosophy*. London: Wiley-Blackwell.
- Williamson, T. 2011. Philosophical Expertise and the Burden of Proof. *Metaphilosophy* 42.3: 215–229.

Williamson, T. 2013. Review of Joshua Alexander, *Experimental Philosophy: An Introduction*. *Philosophy* 88.3: 467–474.

APPENDIX A:

List of 17 Claims

1. The pope wears mitres.
2. The square root of -1 is e .
3. Marge Simpson perms her hair.
4. Aluminum is composed of bauxite.
5. William H. Taft was President of the United States from 1909-1913.
6. The Earth is flat.
7. A sentence cannot be both true and false.
8. There are 13 items in a baker's dozen.
9. Every sentence is either true or false.
10. $54 + 87 = 141$
11. God has three letters.
12. Both Aldous Huxley and C.S. Lewis died on November 22, 1963, the same day
President John F. Kennedy was assassinated.
13. The empire state building is in New Jersey.
14. Mississippi is hot.
15. A student says, "all students are liars."
16. Harpo Studios Chicago is owned by Oprah Winfrey.
17. Ted Williams batted an average of .406 in the 1941 baseball season.