Deceptive and Counter-Deceptive Machines
Papers from the AAAI 2015 Fall Symposium

Can Accomplices to Fraud Will Themselves to Innocence, and Thereby Dodge Counter-Fraud Machines?

Selmer Bringsjord
Deep Detection LLC
Troy NY 12180 USA

Alexander Bringsjord
Deep Detection LLC
Troy NY 12180

1 Introduction

Most readers will be at least vaguely familiar with the tragic-in-so-many-ways case of Ponzi schemer Bernard "Bernie" Madoff, now ensconced in prison, presumably for life. Teamed with others, we have rather elaborately analyzed and modeled Ponzi scheming using a highly expressive computational logic we refer to as a *cognitive calculus*; in particular, we have showed, formally, that had Madoff been somewhat more meta-cognitive, he would have had a much better chance of evading detection (Bringsjord et al. 2015).[1] We suspect that on the other hand most readers will be *un*familiar with the fate of Madoff's business associates. In particular, we bring to your attention that, as reported recently in the *Wall Street Journal* (Sterngold 2015), Madoff's accountant David Friehling recently received an eyebrow-raising light sentence: home detention, and no prison time whatsoever. Why is Madoff locked up for life, whereas his accountant won't be locked up at all? The reason, as explained in (Sterngold 2015), is two-fold: Friehling coöperated with prosecutors, and — and this is key for the present paper — apparently his failure to study the documents and reports that he affirmed secured his agnosticism with respect to the proposition ϕ that Madoff was a fraud. In turn, this agnosticism blocked blame. This brief paper explores, briefly, the consequeces of this second reason — not for Friehling, whom we assume for the record to have received an entirely appropriate sentence, but rather for an arbitrary possible accomplice a assumed for ease of exposition to be an accountant to Ponzi schemer p, within a context that includes another human agent c, and a financial detective d. We assume that d is an artificial agent in the classic sense of a computing machine that takes in percepts and performs actions as its output (Russell and Norvig 2009). We specifically assume that d is armed with a formal definition of fraud, and as d receives percepts about the cognitive states of human agents in the environment, d seeks continuously to determine whether or not any of these agentt have cognitive states that satisfy the definition. We assume that d somehow receives all available percepts. We let $\alpha, alpha', \ldots$ range over arbitrary agents.

The sequel's sequence is: First (§2) we explain why it's provably the case that agnosticism on the part of potential accomplice a with respet to fraudster b entails that that potential accomplice is *not* him/herself guilty of fraud (under specific assumptions that we make clear). We next (§3) provide a recipe for an accomplice to a fraudster to dodge certain counter-fraud machines, by a sheer act of will (under — once again — specific assumptions that we make clear). Then (§4) we pose the question "Can one control by an act of will whether one believes?", and answer with a qualified Yes that lead, in §5, to the position that belief should be "degree-ified." A short section regarding our next research steps wraps up the paper.

2 Agnosticism Implies No Fraud?

Again, apparently Friehling benefited in significant part from the fact that apparently he didn't know that Madoff was a Ponzi schemer. The judge is quoted in (Sterngold 2015) as noting that Friehling contended he was unaware of Madoff's fraud because he (Friehling) didn't review Madoff's documents in detail. Presumably, put a bit more precisely, what the judge presumably meant, couched in terms of our imaginary a and p, is that a, with respect to the proposition that p is a frauster, was agnostic. Moreover, a would have provided no signs to observer c to indicate otherwise.

Given the technical literature on nature of fraud, the judge is here being quite reasonable. For it turns out that on what is hitherto apparently the only rigorous account of fraud in the literature, (Firozabadi, Tan, and Lee 1999), agnosticism *does* preclude fraud. We explain this now, in the form of an argument.

To commence the argument, note that two general conditions are in (Firozabadi, Tan, and Lee 1999) held to be essential for a to be guilty of fraud, a "deception" condition and a "violation" condition. We focus now on the deception condition, which is disjunctive in structure. Following (Firozabadi, Tan, and Lee 1999) directly, but slightly modifying their notation to make it closer to our customary usage [e.g. see (Bringsjord and Govindarajulu 2013)], we use **B** as a belief operator, and lower-case Greek letters to refer to states-of-affairs (represented as formulae). In addition, following (Firozabadi, Tan, and Lee 1999) directly, $\mathbf{E}(c, \psi)$ means that agent c brings it about that ψ. Here then are the

[1] Madoff's returns were simply preposterously consistent, and he should have believed that detectives would believe that such consistency was concocted.

two disjuncts for the deception condition:

D1 $\mathbf{B}(\alpha, \neg\psi) \wedge \mathbf{E}(\alpha, \mathbf{B}(\alpha', \psi))$
D2 $\neg\mathbf{B}(\alpha, \psi) \wedge \mathbf{E}(\alpha, \mathbf{B}(\alpha', \psi))$

The first step in our argument is to object to condition D2, in light of simple counter-examples: cases wherein D2 is satisfied, but where there is no deception. For instance, set ψ to the proposition τ ('They are no longer twins.') at the heart of the famous Twin Paradox arising from relativity.[2] For the counter-example, consider Jones, a well-intentioned teacher who very much wants to believe that relativity theory is true, but in his heart of hearts simply can't bring himself to believe something so disorientingly exotic. Recognizing that he nonetheless has an obligation to teach standing science (erected by a long line of geniuses in physics) to youths studying science, he seeks in his pedagogy to bring it about that for example τ is believed by these students, and succeeds. Is Jones engaging in deception? At least by our lights, and we assume by the lights of our readers, No. Hence we discard D2.

Note, though, that D1 is much more plausible. In this case, the deceptive agent outright believes that ψ is false, but nontheless attempts to bring it about that agent α' believes ψ.

In the next step, let's allow $\mathit{Fraudster}(\alpha)$ to denote that agent α is a fraudster. Given this, and a formal account of fraud that is based on D1 plus the addition of a violation condition, a guilty accomplice a to fraudster p would need to be such that

$$\mathbf{B}(a, \neg\neg\mathit{Fraudster}(p)) \wedge \mathbf{E}(a, \mathbf{B}(c, \neg\mathit{Fraudster}(p))$$

But then clearly the left conjunct here is blocked by agnosticism on the part of a. Our argument is thus complete.

3 Recipe for Dodging Counter-Fraud Machines

Now for the next phase of our investigation.

If belief is volitional, that is, if one can *decide* whether or not to believe a given proposition, there rises up a very easy and very peculiar route to dodging fraud when one is an accomplice a to fraudster p: viz., the agent a can just make sure he doesn't believe that p is a fraudster, *and* that he doesn't believe that p *isn't* a fraudster. Do that, and given the argument articulated in the previous section, one can will oneself to innocence! Hence, the imaginary counter-fraud artificial agent d is dodged. But, *is* belief volitional, as this dodge requires?

4 Is Belief Volitional?

Some philosophers have said that belief is *not* volitional. These thinkers have said that belief "happens to you" and is completely involuntary. They have defended this position in part by distinguishing between belief and *acceptance*. For example, (van Fraassen 1980) gives famous cases in which scientists purportedly *accept* certain propositions (parts of theories), but don't literally *believe* them. This distinction is used to do justice to such phenomena as that scientists certainly seem to sometimes in the course of their work weigh the evidence pro and con and then decide to "just go with" a given theory, and from there go on to work on the basis of that theory. Here the relevant philosophers say that these scientists *accept* theories, but they may or may not also *believe* them.

But clearly we *do* in fact sometimes decide to believe certain propositions, as we are reflecting upon the relevant evidence. Suppose that during the early afternoon you are sitting in a chair overlooking a lawn outisde a beautifully landscaped house somewhere in the suburbs of New York City, sipping some fine Douro wine, sitting with a friend. You friend says: "Do you believe that there is at least one cricket somewhere in this lawn, or in the groundcover fringe around it, before us?" Casting off the fact that this is an odd query, you start to think about it in earnest. You realize that before hearing this query, you had no belief either way about crickets and the lawn. Buy what about now? You realize that last night, sitting where you are sitting now, you heard crickets — at least one or two. Your friend is waiting for a response, and prods you: "Well, do you believe the proposition?" After thinking just a bit more, you decide to register your belief that there is in fact one cricket in the lawn-plus-fringe: "Yes," you reply, "I believe that there is at least one cricket therein — but I must say, the *degree* of my belief is very weak. I wouldn't wager anything substantive on the truth of the proposition."

This, we submit, is a perfectly coherent story.[3] We also maintain that it's a bit cooky to insist on some such thing as that you actually only *accept* the proposition, but don't believe it. Can you imagine saying to your friend, with a straight face, "Well, I don't *believe* that there's at least one cricket, but I do *accept* that there's at least one cricket"?

The upshot is that for those of a logicist persuasion when in comes to AI [e.g. see (Bringsjord 2008)], instead of adding a new fundamental operator to epistemic logics to represent 'accepts' but not 'believes' and 'knows,' it makes more sense to allow that belief comes in different levels of *strength*, and that volitional "acceptance" is really volitional "low-level belief. So the scientist who decides to accept a theory is deciding to belief that that theory holds, but at a low level of belief.

5 Toward Degree-ified Belief

But if this is the position to be adopted, and we do adopt it, cognitive calculi, and the epistemic logics within them that are able to do justice to the position, must be based

[2] Details are unimportant. It suffices to note that two twins on Earth are together at time t in our conventional dating system, and then one of them travels by rocketship into space, returns, and the two are reunited at some time well after t in our dating system. Are the twins still twins? If being the same age is a requirement for being twins, the answer is No: the one who stayed is older.

[3] No doubt there are many supporting parables to be pondered. E.g., don't "leaps" of faith entail decisions to believe? Also worth considering is that if we interview scientists and ask if they believe a theory they have been characterized [per (van Fraassen 1980)] as merely accepting, they will almost certainly say they believe it — but maybe not strongly.

not just on an operator **B**, but on a *range* of such operators. This means, more specifically, following the tiered strength-factor approach of (Chisholm 1977), which has been followed in some prior work on computational logic (Bringsjord et al. 2008), is that instead of just **B**, we shall need to include $\mathbf{B}^1$ for belief that some proposition is more probable than not (which covers the cricket case), $\mathbf{B}^2$ for belief that some proposition is beyond reasonable doubt, $\mathbf{B}^3$ for belief that some proposition is evident, and $\mathbf{B}^4$ for belief that some proposition is indubitable.

Given this machinery, it will not be possible to dodge d. The reason is that while a can avoid believing to a high degree that p is a fraudster, there is no way to avoid a low level of belief. That is, the following would not be blocked:

$$\mathbf{B}^1(a, \neg\neg Fraudster(p)) \wedge \mathbf{E}(a, \mathbf{B}^k(c, \neg Fraudster(p))$$

6 Next Steps

We are working on a "degreeified" formal definition of fraud, and a cognitive calculus that correspondingly provides degrees of belief (and, for reasons to be explained, knowledge as well). Given that calculus, we then will continue to work toward a counter-fraud artifiical detective that uses it. At the symposium, we will report on our progress on these fronts.

References

Bringsjord, S., and Govindarajulu, N. S. 2013. Toward a Modern Geography of Minds, Machines, and Math. In M¨ller, V. C., ed., *Philosophy and Theory of Artificial Intelligence*, volume 5 of *Studies in Applied Philosophy, Epistemology and Rational Ethics*. New York, NY: Springer. 151–165.

Bringsjord, S.; Taylor, J.; Shilliday, A.; Clark, M.; and Arkoudas, K. 2008. Slate: An Argument-Centered Intelligent Assistant to Human Reasoners. In Grasso, F.; Green, N.; Kibble, R.; and Reed, C., eds., *Proceedings of the 8th International Workshop on Computational Models of Natural Argument (CMNA 8)*, 1–10. Patras, Greece: University of Patras.

Bringsjord, S.; Govindarajulu, N. S.; Licato, J.; Sen, A.; Johnson, J.; Bringsjord, A.; and Taylor, J. 2015. On Logicist Agent-Based Economics. In *Proceedings of Artificial Economics 2015 (AE 2015)*. Porto, Portugal: University of Porto.

Bringsjord, S. 2008. The Logicist Manifesto: At Long Last Let Logic-Based AI Become a Field Unto Itself. *Journal of Applied Logic* 6(4):502–525.

Chisholm, R. 1977. *Theory of Knowledge 2nd ed*. Englewood Cliffs, NJ: Prentice-Hall.

Firozabadi, B. S.; Tan, Y.-H.; and Lee, R. 1999. Formal Definitions of Fraud. In McNamara, P., and Prakken, H., eds., *Norms, Logics, and Information Systems*. Amsterdam, The Netherlands: IOS Press. 275–287.

Russell, S., and Norvig, P. 2009. *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice Hall. Third edition.

Sterngold, J. 2015. Madoff Accountant Gets Light Sentence. *The Wall Street Journal*. In "Markets" section of iPad edition, on the the 29th of May.

van Fraassen, B. 1980. *The Scientific Image*. Oxford, UK: Oxford University Press.