BRILL

HOBBES STUDIES 28 (2015) 132-148

Hobbes STUDIES
brill.com/hobs

Two Psychological Defenses of Hobbes's Claim Against the "Fool"

Gregory J. Robson
University of Arizona, Tucson, Arizona UNITED STATES
grobson@email.arizona.edu

Abstract

A striking feature of Thomas Hobbes's account of political obligation is his discussion of the Fool, who thinks it reasonable to adopt a policy of selective, self-interested covenant breaking. Surprisingly, scholars have paid little attention to the potential of a psychological defense of Hobbes's controversial claim that the Fool behaves irrationally. In this paper, I first describe Hobbes's account of the Fool and argue that the kind of Fool most worth considering is the covert, long-term Fool. Then I advance and critically assess two psychological arguments according to which the Fool's policy of self-interested covenant breaking is prudentially irrational. The first argument holds that, taken together, the deep guilt from early-stage covenant breaking, the cumulative guilt from continued covenant breaking, and the high statistical risk of detection during high-volume covenant breaking (which increases greatly when one is desensitized to guilt) render the Fool's policy irrational. The second argument holds that the Fool's policy is irrational because it puts him at risk of adopting a psychologically intolerable view of his fellow covenanters and, specifically, the extent to which they can be trusted.

Keywords

Hobbes's fool – covenants – justice – guilt – Leviathan

A striking feature of Thomas Hobbes's account of political obligation is his discussion of the Fool, who thinks it reasonable to adopt a policy of selective, self-interested covenant breaking. Hobbes must show that the Fool is irrational in order to uphold the rationality of keeping one's promises, including the promise to submit with others to the authority of a sovereign. Hobbes argues that the Fool's policy is unwise since, if caught, the Fool will lose the cooperation

 | DOI 10.1163/18750257-02802003

of others and risk both being excluded from societies for defense and returning to the perilous state of nature. Hobbes's reply to the Fool has left many commentators unconvinced, primarily because it is not clear that all covenant breakers can reasonably expect to be caught over the long-term.[1]

Scholars such as Kinch Hoekstra and A. P. Martinich have advanced interesting arguments defending Hobbes's claim that the Fool is irrational.[2] Yet these arguments are not only controversial[3] but also focus almost exclusively on the risks the Fool takes in terms of (a) his personal security or (b) the benefits he derives from social cooperation.[4] The Fool's risks go beyond personal security and lack of social benefits, however. The Fool also risks developing an intolerable psychology if he is over-burdened by guilt from violating covenants or by a terribly bleak view of the extent to which his fellow covenanters can be trusted. Surprisingly, scholars have paid little attention to the potential of a defense that emphasizes these risks.

In this paper, I first describe Hobbes's account of the Fool and argue that the kind of Fool most worth considering is the covert, long-term Fool. Then I advance and critically assess two arguments that rely on empirical claims about human psychology to defend Hobbes's controversial conclusion about the Fool. The Guilt Argument holds that, taken together, the deep guilt from early-stage covenant breaking, the cumulative guilt from continued covenant breaking, and the high statistical risk of detection during high-volume covenant breaking (which increases greatly when one is desensitized to guilt) render the Fool's policy irrational. The Interpersonal Interactions Argument holds that

1 Zaitchik expresses a fairly common view in the scholarship: "The problems with this [Hobbes's] reply are so astoundingly obvious that one must wonder how Hobbes dared to give it" (A. Zaitchik, "Hobbes's Reply to the Fool: The Problem of Consent and Obligation," *Political Theory*, 10, 2 (May 1982), 246). See also G. Kavka, "Right Reason and Natural Law in Hobbes's Ethics," *The Monist*, 66, 1 (Jan. 1983), 128.

2 According to Hoekstra, Hobbes primarily has in mind the "explicit" Fool, and this Fool is plausibly unwise. According to Martinich, Hobbes thinks that the Fool must justify his actions via propositions that are necessarily true if true at all, and the Fool is truly foolish because he cannot do so. See K. Hoekstra, "Hobbes and the Foole," *Political Theory*, 25, 5 (Oct. 1997), esp. 623–629; and A. P. Martinich, *Hobbes* (New York: Routledge, 2005), 103–4.

3 Against Hoekstra, Peter Hayes argues that Hobbes is not targeting the "explicit" Fool and "the only fool worth arguing against is the silent one." Moreover, Doug Jesseph argues that the propositional standard described by Martinich (see fn. 2) implausibly requires a rational agent to be certain of decisional outcomes *ex ante*. Jesseph therefore argues that Hobbes must have supported a less stringent standard of rationality. See P. Hayes, "Hobbes's Silent Fool: A Response to Hoekstra," *Political Theory*, 27, 2 (Apr. 1999), 225–26; and D. Jesseph, review of *Hobbes*, by A. P. Martinich, *Notre Dame Philosophical Reviews*, 7 June 2006 [http://ndpr.nd.edu/news/25043-hobbes/, accessed 15 Sep. 2012].

4 For a review of attempts to vindicate Hobbes's judgment that the Fool acts irrationally, see M. LeBuffe, "Hobbes's Reply to the Fool," *Philosophy Compass*, 2 (2007), 31–45.

the Fool's policy is irrational because it puts him at risk of adopting a pessimistic, psychologically intolerable view of his fellow covenanters' degree of trustworthiness. These two responses not only are mutually compatible but also harmonize with other arguments for Hobbes's position, suggesting the promise of a multi-pronged defense of Hobbes's claim against the Fool.[5]

1 Hobbes's Reply to the Fool

Hobbes describes the Fool in *Leviathan*:

> The Foole hath sayd in his heart, there is no such thing as Justice; and sometimes also with his tongue; seriously alleging, that every mans conservation, and contentment, being committed to his own care, there could be no reason, why every man might not do what he thought conduced thereunto: and therefore also to make, or not make; keep, or not keep Covenants, was not against Reason, when it conduced to ones benefit.[6]

The Fool, says Hobbes, denies in his heart or by his words that Reason requires strict compliance with valid covenants—i.e., acting justly.[7] By "covenants," Hobbes has in mind contractual arrangements in which at least one party agrees to perform in the future.[8] Rather than acting as if bound *in foro interno* (internally, in the Fool's mind) by Hobbes's third law of nature (which requires strict compliance with valid covenants), the Fool adopts a self-interested, case-by-case approach to keeping covenants.[9] "The central substantive issue that Hobbes and the Fool disagree over," writes Gregory Kavka, "is whether it is rational for an agent to violate core moral rules when doing so promises to

5 Several scholars hold that Hobbes's discussion of the Fool is one of the most important passages in *Leviathan*. See, e.g., Kavka, "Right Reason and Natural Law in Hobbes's Ethics," 127; and Zaitchik, "Hobbes's Reply to the Fool: The Problem of Consent and Obligation," 245.

6 T. Hobbes, *Leviathan*, ed. R. Tuck (Cambridge: Cambridge University Press, 2005), XV, 72. Citations of *Leviathan* are to chapter and section.

7 See R. Hobbes's, "Hobbes's unReasonable Fool," *Southern Journal of Philosophy*, 30, 2 (1992), 95.

8 The Fool's claim is meant to apply to covenants in and out of the commonwealth whose nature, duration, and elaborateness may vary considerably. For example, A agrees to purchase a good from B at a certain price or in exchange for A's subsequent provision of a specified service.

9 The kind of covenanting at issue is a mutual, good-faith transferring of right that obligates the future performance of one or all parties. It consists in a targeted laying-down of one's rights relative to the party with whom one is covenanting. Reasonable fear that another party will not perform invalidates a covenant. See Hobbes, *Leviathan*, XIV, 64–66.

benefit her."[10] Hobbes uses the example of the Fool to pose the question whether it is reasonable for a party A not to abide by the terms of an agreement with a party B (or, two or more parties) if one or both of the following conditions obtain. (1) A knows that B, the counterparty, has already performed, or that a common power will compel B's performance. (2) A thinks that A can break the covenant without being found out, such that B, who has not performed, will still perform. Hobbes aims to show that the Fool acts unreasonably by enacting a policy of breaking covenants in cases where the net benefit seems high. But Hobbes also believes that the Fool, so long as he is secure, cannot reasonably violate *any* of the laws of nature (which remain binding after the initial covenant to institute sovereign rule) or the laws established by the sovereign. The rationality of "foolish" behavior thus concerns the general question whether it is rational to act unjustly or immorally whenever doing so advances one's own interests (thus, I use terms like "covenant breaking" herein to refer to injustice and immorality broadly construed).

In a crucial part of his reply to the Fool, Hobbes addresses the Fool's willingness to break covenants regarding confederations formed for mutual self-defense. Hobbes's argument runs as follows. A subject S in a state of nature, and thus in a state of war, forms a confederation with others on the expectation that they will defend the confederation (and therefore S) from external threats. Yet if S "declares he thinks it reason[able] to deceive those that help him," says Hobbes, S "can in reason expect no other means of safety, than what can be had from his own single power."[11] According to Hobbes, S can then be received into and retained by a society for peace and defense only on account of the error of those receiving and retaining S. But precisely this fact should give S pause and occasion to re-think his policy. For "a man cannot reasonably reckon upon" the errors of others who receive and retain him "as the means of his security."[12] And if the society leaves behind or expels S – neither of which S can reliably predict – S will be at risk of being harmed or even killed.[13] Since a policy of self-interested covenant breaking is apt to thrust S into grave conflict, such a policy, says Hobbes, is unmistakably foolish.

Kavka summarizes the most compelling objection to Hobbes's reply to the Fool: "[I]f the probability of discovery and punishment is small enough and

10 Kavka, "The Rationality of Rule-Following: Hobbes's Dispute with the Fool," *Law and Philosophy*, 14, 1 (Feb. 1995), 9. I have replaced Kavka's term "Foole" with "Fool."

11 Hobbes, *Leviathan*, XV, 73.

12 Hobbes, *Leviathan*, XV, 73.

13 The Fool will suffer under grave cultural deprivations as well, finding himself in a place where there is "no Knowledge of the face of the Earth; no account of Time; no Arts; no Letters" (Hobbes, *Leviathan*, XIII, 62).

the prospective gains of an offensive violation are large enough, it will be in accord with right reason to undertake the violation."[14] The key reason to find Hobbes's reply unconvincing is that many Fools can reasonably expect to benefit from selective, undetected covenant breaking. In what follows, I will buttress Hobbes's claim against the Fool by examining costs imposed by the Fool's psychology in order to show that far more Fools are irrational than Hobbes's own argument covers. As we shall see, two of these costs have no bearing on the Fool's risk of detection but nevertheless suggest his behavior is irrational.

First, however, it is important to get clear on the relevant notion of rationality. According to Hoekstra, most commentators take Hobbes to hold that the Fool's dictate to pursue his self-interest cannot diverge from his dictate to keep his covenants.[15] This view assumes it is always in the Fool's self-interest to uphold a policy of keeping covenants. It also implies that the Fool acts in a prudentially irrational way (thus, prudential rationality is the notion of rationality in which I am interested). The Fool's choice to adopt a policy of self-interested covenant breaking is irrational if, in terms of his wellbeing, the expected costs of his decision outweigh its expected benefits. It bears emphasis that Hobbes is focused not on whether the Fool might sometimes benefit by violating covenants, but on the more interesting question whether it is reasonable for him to pursue a *policy* of breaking covenants when non-compliance serves his self-interest.[16] On the preceding analysis, "reasonable" should be thought of as "prudentially rational."[17]

---

14 Kavka, "Right Reason and Natural Law in Hobbes's Ethics," 128.

15 David Gauthier, Jean Hampton, and the present author accept this common interpretation. Other interpreters (e.g., William E. Connolly), however, hold that Hobbes allows for such a divergence but, in such cases, the dictate to keep covenants overrides the dictate to secure one's self-interest. Still others (e.g., Kinch Hoekstra) attribute to Hobbes the pragmatic view that, in such cases, one should pursue one's self-interest rather than keep covenants. See D. Gauthier, *The Logic of Leviathan: The Moral and Political Theory of Thomas Hobbes* (Oxford: Clarendon Press, 1969), 8; J. Hampton, *Hobbes and the Social Contract Tradition* (Cambridge: Cambridge University Press, 1986), 65; and W. E. Connolly, *Political Theory and Modernity* (Oxford: Basil Blackwell, 1988), 26. Here I draw on Hoekstra, "Hobbes and the Foole," 621–22.

16 Other commentators agree that Hobbes equates rational decision-making with prudential decision-making. See, e.g., LeBuffe, "Hobbes's Reply to the Fool," 32. A fortuitous outcome from breaking a covenant does *not* show that a policy of breaking covenants is reasonable.

17 Most authors writing about Hobbes's Fool seem to use "rationally" and "reasonably" interchangeably, as I do herein, but it is not obvious that the terms are synonymous. See, e.g., W. M. Sibley, "The Rational Versus the Reasonable," *Philosophical Review*, 62, 4 (1953), 554–60.

II Overt and Covert Fools, Short-term and Long-term Fools

The kind of Fool whose rationality (or lack thereof) is most worth considering is the covert, long-term Fool. Hobbes emphasizes that the Fool's disregard of the third law of nature not only undermines the achievement and maintenance of social stability but also jeopardizes his own long-term wellbeing. To see how effective this response is, consider the case of the overt Fool and the covert Fool. (I propose this distinction in place of the usual distinction between the explicit Fool and the discreet Fool espoused by, e.g., Hayes and Hoekstra.[18]) Whereas the covert Fool secretly adopts a policy of violating covenants when it is in his self-interest, the overt Fool openly declares that he is, when it suits him, willing to (a) break covenants he has already formed or (b) form covenants but retain the option to break them. Hobbes seems to focus on covert rather than overt Fools, since it is strikingly imprudent to make such a declaration.[19] A strategy of open declaration would readily result in one's being cast out of society and – against the first law of nature, i.e., to seek peace – hurled on a life course marked by great conflict and strife.[20] Since stating publicly that one is willing to break covenants could only make covenant breaking less prudentially rational (for one would more easily be found out and cast out of society), a conclusion in favor of the covert Fool's irrationality will apply *a fortiori* to the overt Fool. A case for the covert Fool's irrationality is thus all the more important.

An argument upholding the Fool's irrationality should also focus on (what I call) the long-term Fool rather than the short-term Fool. Consider the case of a short-term Fool who reasonably expects to live only three more months. Is this "Fool" truly foolish to break a covenant that he considers detrimental to his wellbeing? Suppose that he can undetectably cheat on a contract, save money as a result, and put the money towards something he finds quite meaningful at the very end of his life. Such a course of action is *prima facie* reasonable, especially if this Fool has no other way to obtain the money, can keep his deceit to

18 Commentators usually distinguish the "explicit Fool" from the "discreet Fool" (or, silent fool). But I prefer the terms overt Fool and covert Fool. These terms pair more naturally as linguistic opposites than "explicit" and "discreet," and they better capture the public or open (i.e., overt) nature of the one kind of Fool, and the private or secret (i.e., covert) nature of the other kind of Fool.

19 Thus I agree – but for different reasons – with scholars who take Hobbes to focus on the covert fool (e.g., Hayes, who discusses the "silent fool") rather than the overt fool (e.g., Hoekstra, who discusses the "explicit fool"). See Hayes, "Hobbes's Silent Fool: A Response to Hoekstra," 225–26; and Hoekstra, "Hobbes and the Foole," 623–29.

20 See Hobbes, *Leviathan*, XIV, 64.

himself, and need only conceal his behavior for a few months. Far from appearing foolish, this "Fool" seems prudentially justified. Knowing that he has no long-term future, he reasonably sets out *ex ante* to break covenants for his own benefit.

Fortunately for the defender of Hobbes, Hobbes's conclusion about the Fool applies with greater force to the long-term Fool. Here, as we shall see, the Hobbesian has more attractive arguments at her disposal. The focal question of this paper is whether Hobbes – who implicitly focuses on the long-term Fool – is right that the covert, long-term Fool is genuinely foolish, not whether his own argument for that conclusion is tenable. I will offer and critically assess two arguments that (in Michael LeBuffe's terminology) count as "direct" defenses of Hobbes's conclusion, in that each suggests that the covert, long-term Fool misuses his prudential reasoning when trying to serve his self-interest.[21]

III The Guilt Argument

The two arguments – The Guilt Argument and The Interpersonal Interactions Argument – are partly motivated by Hobbes's discussion of "*Nosce teipsum*" ("read thyself," on Hobbes's translation) in the introduction to *Leviathan*. *Nosce teipsum*, says Hobbes, implies

> that for the similitude of the thoughts, and Passions of one man, to the thoughts, and Passions of another, whosoever looketh into himself, and considereth what he doth, when he does *think*, *opine*, *reason*, *hope*, *feare*, &c, and upon what grounds; he shall thereby read and know, what are the thoughts, and Passions of all other men, upon the like occasions.[22]

Here Hobbes is not only exhorting one to read (or know) oneself. He is also describing how one can gain access to others' psychologies by way of measured reflection on one's own psychological states and mental operations, including one's modes of reasoning. The two main arguments that I will soon develop (again, the guilt and interpersonal interactions arguments) are continuous

21 LeBuffe, "Hobbes's Reply to the Fool," 32. Indirect strategies, in contrast, hold that the Fool's misguided policy of self-interested covenant breaking owes to relevant background beliefs such as the Fool's atheism or inadequate appreciation of the third law of nature. For further discussion of the indirect approach, see LeBuffe, "Hobbes on the Origin of Obligation," *British Journal for the History of Philosophy*, 11, 1 (2003), 15–39.

22 Hobbes, *Leviathan*, The Introduction, 2 (italics original).

with this insight. They are motivated by the idea that one's capacity to reason to the "thoughts, and Passions of all other men, upon the like occasions" can serve as a powerful constraint upon one's willingness to break covenants validly entered into with others.

The Basic Guilt Argument

In a multidisciplinary study on guilt aversion, Chang et al. show that "the anticipation of guilt can motivate cooperative behavior" such that "people often choose to cooperate when they can better serve their interests by acting selfishly."[23] My first pro-Hobbes argument – the Guilt Argument (GA) – concerns the ways in which one's covenant breaking both affects and is affected by one's feelings of guilt (if any) after breaking covenants. This argument provides useful resources for sustaining the claim – arguably inadequately defended by Hobbes himself[24] – that the Fool is truly foolish.[25] As Hobbes does, I will argue that a policy of self-interested covenant breaking (again, "covenant breaking" in this paper concerns the broader scope of immorality and injustice) can be irrational due to its unacceptably bad consequences. The GA will suggest that one should not enact a policy of covert, self-interested covenant breaking because it effects a sense of guilt that is prohibitively "costly" in the psychological sense.

We turn now to two sub-arguments: the "diminishing marginal costs argument" and the "desensitization argument." The DMCA holds that (a) initial covenant breaking can yield high levels of per-covenant guilt and (b) later-stage covenant breaking increases cumulative guilt.[26] The desensitization argument shows that, if one adopts a Foolish policy (even despite the psychological cost of experiencing this guilt), one's desensitization to guilt will eventually lead

23 L. J. Chang, M. Dufwenberg, A. G. Sanfey, and A. Smith, "Triangulating the Neural, Psychological, and Economic Bases of Guilt Aversion," *Neuron*, 70, 3 (12 May 2011), 560.

24 Recall that Hobbes arguably overlooks the ability of many "Fools" to adopt covenant-breaking dispositions and reasonably expect never to get caught.

25 Zaitchik, too, offers a defense that is different from Hobbes's own, specifically a nongenetic Rawlsian account. See Zaitchik, "Hobbes's Reply to the Fool: The Problem of Consent and Obligation," esp. 259–63. For another Rawlsian approach, see Kavka, "Right Reason and Natural Law in Hobbes's Ethics," esp. 128–30.

26 According to Lisa Lindsey, guilt is a negative affective state (presumably like fear or nervousness) that people can anticipate and want to avoid. It makes sense, then, that people would want to keep covenants in order to avoid feeling guilty. See L. Massi, "Anticipated Guilt as Behavioral Motivation: An Examination of Appeals to Help Unknown Others Through Bone Marrow Donation," *Human Communication Research*, 31, 4 (Oct. 2005), 453–81.

one to break many more covenants, significantly increasing the statistical risk of detection (and, therefore, the prudential risk of being a Fool). The GA combines these sub-arguments to show that, for more persons than Hobbes's argument covers, a Foolish policy is irrational because it is psychologically untenable.

The Diminishing Marginal Costs Argument (DMCA)

The DMCA, which applies the relevant concept from economics to the human psyche, emphasizes that early-stage covenant breaking is psychologically very costly. It also holds that, while the initial marginal costs are very high, the costs decrease significantly as one breaks more and more covenants, eventually becoming so small that they no longer constrain one's actions psychologically. Further, the DMCA provides resources that help demonstrate the prudential irrationality of high-volume covenant breaking, even if the guilt-costs per covenant broken are low. The subsequently discussed argument from desensitization will lend further support to this third claim.

The first part of the DMCA concerns the high initial costs of covenant breaking in terms of the guilt it produces. (By "guilt" I mean roughly the inner discomfort that one feels from acting against one's moral judgment.[27]) Consider a *pure record* of covenanting, to wit, a history of covenanting in which a person has never broken a single covenant. This person would occupy an early position on the diminishing marginal costs curve and be guilt-free such that he would find covenant breaking virtually intolerable. The first time he breaks a covenant, he would, on this argument, experience significant guilt and possibly even suffer a psychologically traumatic "loss of innocence."[28] Now if the psychological pain of initial covenant breaking is strong enough, then it will *always* be pragmatically irrational for one to break one's first covenant. It might be rational for one to break covenants down the road, after one is already acclimated to being a scoundrel. Yet, on this argument, the Fool can rationally break a covenant *only if he has first broken another covenant irrationally*. A perfectly rational agent with a strong enough sensitivity to guilt would therefore never break a covenant, let alone adopt a policy of breaking covenants. Such an agent would never become a Fool!

Now even if the irrationality of first-time covenant breaking is interesting chiefly as a speculative possibility, the concept of the DMC curve still helps us

27 Here I mean "acting" in a broad sense that includes choosing to maintain undesirable mental states that one can avoid.

28 Consider the high degree of guilt that a person who has never cheated another would feel from secretly defrauding a trusting friend of his livelihood.

to see that the costs of the first few covenant-breakings will be high even if they become progressively lower per covenant. Looking *prospectively* at adopting a policy of covenant breaking, a rational agent with minimal guilt with respect to covenant breaking will recognize the high degree of guilt that he could incur from actually implementing such a policy. Were he to pursue such a policy he would knowingly risk experiencing overwhelming guilt, a cost he may indeed be unwilling to bear.[29]

This argument does not, however, entitle the Hobbesian to claim that *no one* can ever rationally adopt or continue to pursue a Foolish policy. Since many people already feel guilty about some of their past actions, the points above apply only to a subset of covenant breakers.[30] The potential guilt-cost of breaking a given covenant may be low and therefore a weak deterrent for people who already have significant guilt from immoral or unjust activities (more on this later). And in highly experienced Fools, covenant breaking might not produce any (psychologically uncomfortable) guilt that gives them reason not to continue their self-interested covenant breaking. More needs to be said, then, about why people who feel little guilt per instance of covenant breaking cannot rationally adopt or continue to pursue a "foolish" policy.

The second part of the DMCA offers a partial solution to this problem. Although each additional unit on the curve (each additional instance of covenant breaking) costs less (produces less guilt) than the previous one, the overall guilt-burden reflects all of those individual guilt-costs. Accordingly it will be cold comfort to the potential Fool to realize that his fourth instance of covenant breaking would add less guilt than his third, when he knows that the fourth would exacerbate his already high degree of guilt from the first, second, and third instances. The concept of *cumulative guilt*, then, helps to bolster the Guilt Argument. Nonetheless, we must also address another important cohort of covenant breakers. These "fools" have become *so* accustomed to breaking important promises that they have managed to (or have convinced themselves they have) become "numb" even to past guilt from covenant breaking. Is it irrational even for these people to enact covenant breaking policies?

29 Chang et al. suggest that the experience of guilt may be a significant expected cost that people account for when deciding whether to deceive others. Indeed, "a guilt-aversion mechanism underlies decisions to cooperate," and people "may even experience a preview of their future guilt at the time of the decision … ultimately motivat[ing] them to cooperate" (Chang et al., "Triangulating the Neural, Psychological, and Economic Bases of Guilt Aversion," 561, 566).

30 An interesting question is whether all people with a sense of guilt must have acted irrationally before developing it.

The Argument from Desensitization

The second part of the GA – the argument from desensitization – may be able to fill precisely this gap. Surely Fools who knowingly break more and more covenants could conceivably suppress or overcome their cumulative guilt, even to the point of becoming *generally desensitized* to covenant-related guilt (such guilt, if any, no longer being action-guiding). At some point a Fool may cease to feel very guilty about his past covenant breaking such that future violations may cause him little or no additional guilt. (We commonly call persons desensitized to the claims of justice and morality "sociopaths," whether they are born that way or become desensitized over time.) Now this line of argument may seem to suggest that intransigent Fools are therefore not irrational on psychological grounds. But here the Hobbesian can avail herself of an interesting countermove: she can point to a heightened risk of detection *precisely on account of this numbness*.

As such a Fool breaks more and more covenants and experiences less and less guilt per covenant broken, the Fool's simple, statistical risk of detection will increase significantly, and his ability consistently to keep covenants may become much diminished as well. Consider Sissela Bok's observation that the typical liar's distorted conscience leaves him particularly ill equipped to prescind from the urge to lie.[31] Similarly, as the covenant-breaker becomes less and less sensitive to covenant-related guilt, he will become increasingly willing to break covenants. And even if (in a given case) lack of guilt and chance of detection do not track one another organically, the statistical risk of detection will clearly be far higher for the person who has broken many important covenants such that he has become insensitive to the normal guilt-constraint. (This claim about the Fool's increased probability of detection is properly part of the GA insofar as his minimal degree of guilt is responsible for the increase.) The Hobbesian could then argue that this Fool will be less likely or less able to cover his tracks safely and, against reason, more likely to be cast out of society into the perilous state of nature. Hobbes, of course, thinks this risk is rarely if ever worth taking.

The Combined Guilt Argument

The Hobbesian may now be able to use the DMCA *in tandem* with the argument from desensitization to defend Hobbes's claim against the Fool. The combined argument runs as follows. A potential Fool would need to consider prospectively the substantial, possibly unmanageable level of guilt per instance that could result from secretly breaking one's first few major covenants, as well

31 S. Bok, *Lying: Moral Choice in Public and Private Life*, 2nd ed. (New York: Vintage Books, 1999), 94–95.

as the cumulative guilt-burden that would accrue if he continued in this vein. This would quite possibly deter him from becoming a Fool. But, even if it did not, he would still have to contend with the possibility of eventually breaking so many covenants as to become numb to the cumulative guilt, psychologically warped, and desensitized to added guilt from further covenant breaking. Most people would be hard-pressed to choose to become psychologically warped, desensitized scoundrels. And worse still, those who covertly break so many covenants as to become desensitized to the normally accompanying guilt also place themselves in a high-risk category for being found out. (This puts in play Hobbes's original argument about being cast out of society.) Only a genuine Fool would put himself in such a position!

III.ii The Combined Guilt Argument Assessed

Can the foregoing, combined GA ground Hobbes's claim against the long-term covert Fool? Consider two of its controversial assumptions:

(1) Activities other than covenanting will not so dull prospective covenanters to guilt as to make, say, first-time covenant breaking (injustice, immorality) quite tolerable for them.
(2) High-volume covenant breaking significantly increases the probability that one will get caught.

Let's begin with (1). Although pursuing a policy of covenant breaking would force many Fools to bear prohibitively high psychological costs, certain potential Fools may already have substantial guilt – even from apparently amoral mishaps (e.g., forgetting to help a friend after agreeing to do so) – before breaking even a single covenant. If a "Fool" already has a strong sense of guilt, and guilt becomes easier to bear per unit the guiltier one becomes, then initial covenant breaking might be far less psychologically costly for him than it is for many other Fools.[32] Also, depending on the particular Fool, any worry about adding to one's cumulative guilt might prove impotent as a deterrent. Moreover, like (1), (2) is true for many but perhaps not all Fools. High-volume covenant breaking presumably enables some Fools to gain valuable experience. Any resultant decrease in a Fool's likelihood of being caught might offset her higher statistical chance of being caught from breaking more covenants. But

32 For example, one's desire to be moral in activities other than covenanting (e.g., supererogatory giving) can give one a sense of guilt if one fails to live up to one's expectations.

then (2) would not be true for her. All of which appears to suggest that the GA holds for many but not all Fools.

Indeed, given the substantial variation in possible human psychological responses to pursuing a "foolish" trajectory, it is reasonable to expect Fools to feel guilt (if at all) in different degrees, at different times, and in different respects. Some Fools, for example, will become completely desensitized; others will become only mildly desensitized or desensitized only during particular activities or in particular relationships—e.g. only in business or certain business relationships, or only in friendship or certain friendships. Since humans respond in psychologically diverse ways to their own instances of covenant breaking, it may seem that the GA applies only to a subset of Fools because it depends too heavily, one might argue, on controversial assumptions about what a human would do or what her psychology would be like at a given time. There is more to this story, however. Prior to pursuing a "foolish" policy, no Fool can know whether or to what extent acting immorally or unjustly will prove psychologically detrimental to her. Faced with such epistemic uncertainty, *any* person deliberating about whether to become a Fool ought to consider the real risk that she will become intolerably guilty, or, if she becomes desensitized to covenant-related guilt, that she will rashly expose herself to danger by breaking too many covenants. A rational deliberator would take such a risk seriously, perhaps treating it as a decisive reason not to act foolishly.

IV The Interpersonal Interactions Argument

We turn now to the second psychological strategy for vindicating Hobbes's judgment against the Fool. This strategy relies on (what I call) the Interpersonal Interactions Argument (IIA) and is not weighed down by assumptions (1) and (2) above. The IIA holds that (a) a policy of self-interested covenant breaking will leave the Fool with a bleak and intolerable view of the extent to which other covenanters can be trusted, and (b) the Fool misuses his prudential reasoning faculties by charting a life course in which he risks incurring the psychological costs associated with (a). The desire to avoid such costs gives the Fool good reason not to pursue a policy of self-interested covenant breaking.

According to the IIA, the Fool will learn by personal experience how easy it can be to break covenants. And he will ask himself: "If I can break covenants fairly easily without others being aware of it, couldn't they have broken covenants *with me* without *my* knowing it?" Answering this unsettling question in the affirmative, the Fool will come to view others within his social network suspiciously. Many of his social interactions will then become colored by a fear

that his friends' and confidants' warm outward dispositions towards him do not reflect their deeper motivation to benefit themselves by deceiving him. Dispensing with his former view of other persons as honest covenanters, this covert Fool will begin constantly "looking over his shoulder" for fear of being exploited by apparently friendly but actually dishonest covenanters. More generally, he will be haunted by the thought that his society has been and will continue to be populated by covert Fools who are just as deceitful as he is.

Adopting a long-term policy of covenant breaking would then be irrational for this Fool insofar as it would make his everyday life unbearable by leaving him with a psychologically intolerable view of those around him – business partners, friends, etc. – as potentially treacherous, self-interested covenanters. Such a policy would make him feel victimized by their duplicity. Worse still, it would leave him with the unsettling thought that, while being victimized, he often would not even know it. The more success this Fool has at avoiding detection, the bleaker his psychosocial picture of the world will become. A Fool who violates only a few covenants successfully may attribute his success mainly to luck. But if he manages to break numerous covenants without detection, he will then have systematic evidence that others with similar covenant breaking competency can break covenants with relative ease. Such a Fool would have good reason to worry that others are breaking covenants with him, and would experience a loss of basic trust in them. Importantly, here we have a defense of the irrationality of covenant breaking that is untethered to the question whether non-compliance with valid covenants increases one's chances of returning to Hobbes's terrible state of nature. Even a Fool who adopts a highly risk-averse policy in order to avoid detection may find such a policy too psychologically unsettling to be rational.

IV.ii The IIA Assessed

Of the philosophical objections that may be raised against the IIA, one seems particularly worthy of discussion. In a society already characterized by rampant covenant breaking, the Fool would see himself as an individual in a classic prisoner's dilemma.[33] Knowing what he himself is likely to do, the Fool would have good reason to expect non-compliance by most others. So he

33 Hampton, Kavka, LeBuffe, and Palumbo all suggest that the Fool can be seen as an actor in a prisoner's dilemma in which the sovereign can threaten to punish violators, incentivizing people to comply with covenants. See Hampton, *Hobbes and the Social Contract Tradition*, esp. 132–37; Kavka, *Hobbesian Moral and Political Theory* (Princeton: Princeton University Press, 1986), esp. 137–56; LeBuffe, "Hobbes's Reply to the Fool," esp. 33–35; and A. Palumbo, "Playing

would break covenants to ensure that others do not take advantage of him unilaterally. The question arises whether, if forced to decide between the two, the Fool would be better off (a) breaking covenants in order to preempt others from imposing costs on him by breaking covenants with him; or (b) not breaking covenants in order to avoid the negative (psychological) effects he anticipates from his own covenant breaking. If (a) is better for the Fool's wellbeing than (b), then enacting a policy of self-interested covenant breaking might be the more rational option for him in this highly dysfunctional society. Even if it will engender in many Fools seemingly intolerable, paranoid psyches, such a policy may be necessary for their peace and wellbeing, to the small extent either is possible in such a society.

One possible response to this objection draws on work in contemporary decision theory. Suppose that a person lives in a society where he knows that most people are a lot like him. The person and his fellow residents are not psychological duplicates, but they are fairly similar in respect of their willingness to break covenants.[34] Interestingly, if this person keeps his covenants he will then have *evidence* that others will keep their promises in the future. For it will be more epistemically likely that other people will keep *their* covenants if he can reasonably assume that others will act as he does. And conversely, breaking covenants will give him evidence that others will break covenants with him in the future. Suppose one thinks that one should act so as to maximize the epistemic probability of achieving one's goals.[35] This view of practical reason, if correct, has important implications for the Fool.[36] Insofar as the person's goals depend on the future cooperation of others, he has reason not to break covenants—even if no one will ever find out and he is not overburdened by guilt from past covenants![37]

Hobbes. The Theory of Games and Hobbesian Political Theory," *UEA Papers in Philosophy*, New Series 8 (1996), 1–29. Here I draw primarily on LeBuffe's illuminating article.

34 For discussion of prisoner's dilemmas with psychological twins, see P. Weirich, "Causal Decision Theory," *Stanford Encyclopedia of Philosophy*, 5 Oct. 2012 [http://plato.stanford.edu/entries/decision-causal/, accessed 20 Apr. 2015].

35 The standard "one-boxer" about Newcomb's problem in decision theory calls for maximizing expected utility rather than following the dominance strategy espoused by two-boxers. See R. Nozick, "Newcomb's Problem and Two Principles of Choice," in N. Rescher (ed.), *Essays in Honor of Carl G. Hempel* (Dordrecht: Reidel, 1969), 114–46.

36 Of course, the force of this response depends on whether one-boxing generates evidence in a legitimate way.

37 One advantage of this one-boxer version of the IIA is that all it requires – besides, of course, a commitment to one-boxing – is basic similarity in the Fools' willingness to break covenants.

Overall, the IIA is akin to the GA in terms of its value for the Hobbesian. Whether or not one subscribes to the aforementioned view of practical reason, the IIA unveils an additional, psychological cost of covenant breaking that has so far been neglected in the literature. This argument demonstrates that a policy of self-interested covenant breaking is irrational for many more Fools than Hobbes's argument covers because of its potential to be psychologically disastrous for the violator. Despite the highly diverse ways in which covenant breakers can respond psychologically to their own immoral behavior, it bears reemphasis that no prospective Fool can be certain *ex ante* of how severe his own psychological response will be. (To Hobbes's claim that no violator can be sure of whether he will be detected and expelled from the commonwealth, we can add that no violator can reliably predict whether his actions will ultimately take a grave psychological toll on him.)

A key implication of the IIA is that the Fool risks negating many of the benefits made possible for him by his covenanting to install a sovereign who will resolve the prisoner's dilemma in the state of nature and provide conditions under which members of the commonwealth can reasonably trust each other. In embarking on a course of action whose psychological consequences he cannot predict, the Fool risks, in the most extreme case, losing critical benefits and comforts associated with life outside of the state of nature. For if the Fool can no longer trust anyone, he can no longer live a peaceful, productive, and commodious life in the commonwealth. Furthermore, since no prospective Fool (however psychologically healthy he may be at the time) can know in advance how he will respond to his own covenant breaking, every Fool – even if he adopts a highly risk-averse policy in order to avoid detection – may find the potential psychological cost of such a policy too great to incur rationally. Interestingly, this risk applies to the Fool *even if* he is never detected and expelled from society. So the Fool's prudentially irrational behavior is a function not only of his risk, noted by Hobbes, of being expelled from society, but also of the kind of person he risks becoming (and in particular, the kind of psychology he risks developing) while continuing to live in the commonwealth. Having successfully escaped the miserable state of nature, the Fool now risks rendering his own life miserable in the commonwealth itself.

V Conclusion: The Two Strategies in Broader Perspective

The case of Hobbes's "Fool" raises the question whether it can be rationally defensible to adopt a policy of self-interested covenant breaking. After distinguishing covert and overt Fools and short-term and long-term Fools, I offered

two psychological defenses of Hobbes's view that the Fool is truly foolish. Both the Guilt Argument and the Interpersonal Interactions Argument show that Foolish behavior can be significantly more costly than Hobbes's argument suggests. Moreover, these strategies may be jointly deployed to cover a broader array of cases than Hobbes's argument covers—and, of course, the more costs associated with covenant breaking, the less prudentially rational it becomes. Interestingly, the strategies may also be compatible with the aforementioned accounts by Hoekstra, Kavka, and Martinich.[38] If all Fools need not be irrational for the same reason (e.g., the guilt from acting foolishly) or the same kind of reason (e.g., the psychological costs of covenant breaking), then a package argument of this sort could further strengthen Hobbes's case against the Fool. Building on the analysis in this paper, such an argument would account for *both* the grave risk of returning to the state of nature and the substantial psychological costs associated with pursuing a policy of self-interested covenant breaking.

38 As discussed in the footnotes above, these approaches include (a) interpretive accounts that focus on either the covert or overt Fool or a certain standard of rationality, and (b) game theoretic accounts that consider the rationality of adopting a policy of self-interested covenant breaking.