**The Error of God, Revisited**

Ryan McKay\*a,b, Robert M. Ross a,b, Katherine O'Lone a,b & Charles Efferson\*c

a Department of Psychology, Royal Holloway, University of London, Egham, Surrey TW20 0EX, United Kingdom

b ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, NSW 2109, Australia

c Department of Economics, University of Zurich, Blümlisalpstrasse 10, CH-8006 Zurich, Switzerland

\*Joint Corresponding Authors:

Email: ryantmckay@mac.com; charles.efferson@econ.uzh.ch

Phone: +44 1784 41 4993; +41 44 634 36 65

**Acknowledgment:** This research was supported by John Templeton Foundation grant no. 52257.

If belief in God is an error, it may at least be an adaptive one.

~ *Johnson, “The error of God: Error management theory, religion, and the evolution of cooperation” (2009, p. 170)*

We enjoyed Dominic Johnson’s rich, lucid and thought-provoking new book, in which he provides a detailed exposition of his *supernatural punishment theory.* Johnson argues that the development of two unique cognitive abilities – theory of mind and complex language – upended “the entire cost-benefit calculus of selfishness” (2015, p. 146). As a result, social transgressions such as adultery were more likely to be detected than before, and the consequences of detection could be considerably graver (transgressions could now be reported to absent third parties, exacerbating reputational damage for the transgressor). According to Johnson:

The heightened probability and costs of exposure for social transgressions… may have favored the evolution of traits that suppress, rather than promote, selfish behavior… My proposal is that supernatural punishment was just such a trait. This almighty deterrent helped to steer us away from the real-world punishment that could be so costly to Darwinian fitness. (2015, p. 168.)

The claim that humans have an evolved propensity to believe in supernatural punishers is both influential and controversial. In this commentary we play devil’s advocate, focusing on a particular component of Johnson’s argument. Let us assume that Johnson is right that the emergence of mentalising and linguistic capacities increased the “probability and costs of exposure for social transgressions”. Why would unjustified beliefs in the existence of supernatural punishers be an adaptive response to this new scenario? Here Johnson invokes *error management theory* (e.g., Haselton & Buss, 2000; Haselton & Nettle, 2006; Johnson, 2009):

[W]hen we set the sensitivity of a smoke alarm, it is better to err on the side of caution because the costs of a false alarm are negligible (however annoying it may be at the time), whereas the costs of being burned to death in a fire are great indeed. One should thus expect the smoke alarm to go off a bit too often. Whenever the true probability of some event—such as the detection of snakes, fires, or selfish behavior—is uncertain, a biased decision rule can be better than an unbiased one if we want to avoid the more dangerous error. (2015, p. 169.)

We agree with Johnson that erring on the side of caution is an excellent strategy for longevity and reproductive success.[[1]](#footnote-1) Consider base jumping. For a base jumper there are at least two relevant “errors”: opening one’s chute earlier than necessary and opening one’s chute later than necessary. As the latter error can typically only be made once, an unbiased decision rule here (pulling the chute with the aim of making both types of error equally frequently, so as to minimise overall errors) will tend to result in death by the second jump. In a sense, the word “error” in this context is misleading. According to the Oxford dictionary, error is “the state or condition of being wrong in conduct or judgement” (Simpson & Weiner, 1989), but minimising errors in this context would be poor judgement indeed. Thus even extreme sports enthusiasts should err on the side of caution. The goal should not be to minimise errors, but to minimise *costs*.

Does “erring on the side of caution” require erroneous beliefs? Johnson thinks so. In considering the errors that one might make in deciding whether to cheat or not, he writes:

If false negative errors (assuming stealth and getting caught) are more costly than false positive errors (assuming detection and missing a reward), then *only exaggerated* estimates of the probability of detection—such as a belief that supernatural agents are observing your behavior all the time—will help you to avoid the worst of the two errors. (2015, p. 169; italics in original.)

Johnson is explicit here that *only* exaggerated estimates – erroneous beliefs – help to avoid costly errors. This strong claim seems easy to refute. After all, most people with unbiased beliefs about the probability of costly errors will still act so as to avoid those errors. To see this, consider a different example. Imagine you are handed a gun with six chambers, and informed that two of the chambers are loaded with bullets. Your task is to pull the trigger, and you can choose to point the gun either at your head or at a nearby sand bag. You will be paid a dollar if you pull the trigger while pointing the weapon at your head. In this scenario there are two relevant errors: 1) Dry firing (pulling the trigger on an empty chamber) at the sand bag, and missing out on a dollar; [[2]](#footnote-2) and 2) Discharging a live round at your head, and missing out on the rest of your life.

Most reasonable people in this situation would “err” on the side of caution and point the gun at the sand bag. Such behaviour would not require “erroneous” beliefs about the probability of the costly error (e.g., believing the chamber has five live rounds rather than two). Indeed, even someone who slightly *underestimated* the probability of a bullet in this scenario (e.g., believing the chamber has just a single live round) would surely still err on the side of caution.

To return to the domain of cheating, exaggerated estimates of the probability of detection are not the *only* way to avoid costly errors. As with the gun example, individuals with estimates biased in the opposite direction – who are overly sanguine about their chances of avoiding detection – may still refrain from cheating if they appreciate how costly it is to be caught. But are exaggerated estimates the *best* way to avoid costly errors? Here is Johnson again:

The best solution to avoiding detection… is a mechanism that overestimates the true probability that detection will occur—*exaggerated* estimates outperform *accurate* estimates, because the latter will engender more [costly[[3]](#footnote-3)] mistakes. (2015, p. 169; italics in original.)

We acknowledge that someone whose estimate of the likelihood that cheating will be detected is so exaggerated that they *never* cheat will avoid at least as many costly errors as someone with an unbiased estimate. Indeed, that person will *never* make the costly error of cheating and being caught (McKay & Efferson, 2010). So we acknowledge that exaggerated estimates may be the best strategy for minimising costly errors. However, just as adaptive decision-making is not about minimising errors, it is also not about minimising *costly* errors. Rather, adaptive decision-making is about minimising *overall costs*. One should err on the side of caution, certainly – it is best to make fewer costly errors than errors that are less costly – but ensuring that one errs *completely* on the side of caution is not the best way to minimise overall costs.

In any given situation, the decision about whether or not to transgress depends on one’s beliefs (e.g., beliefs about the probability of detection). There is a level of belief at which the optimal behaviour (assuming throughout that the agent wishes to maximise their expected payoff) switches from transgressing to not transgressing, and unbiased (Bayesian) beliefs lead one to switch behaviour at the right point. Unbiased beliefs ensure that one never chooses the option with the lower expected payoff.  Biased beliefs, however (e.g., the exaggerated estimates that Johnson describes), ensure that one will, with some positive probability, choose sub-optimally. In the long run this strategy cannot but fail to minimise overall costs.

To illustrate this, assume two possibilities exist:

A: an agent is watching you

B: no-one is watching you

If A is true, the best behaviour is to refrain from transgressing. If B is true, the best behaviour is to transgress (e.g., cheat, steal). Further assume that if A is true and you choose to transgress, the error cost is relatively large (e.g., punishment, exile, death). In contrast, if B is true and you refrain from transgressing, the error cost is relatively small (a missed opportunity). Figure 1 depicts this scenario. On the horizontal axis is the Bayesian posterior belief that A is true. The Bayesian posterior represents the theoretically justifiable belief after integrating all available evidence in a fashion consistent with Bayes' rule (Bayes, 1763). The solid line shows the expected value of not transgressing as a function of the Bayesian posterior. The dashed line shows the expected value of transgressing. At *p\_c*, the two behaviours have the same expected value and a Bayesian expected value maximiser is indifferent between them. For any posterior greater than *p\_c*, refraining from transgressing is advantageous (maximises expected value). For any posterior less than *p\_c*, transgressing is advantageous.

­– – Insert Figure 1 about here – –

Because of the cost asymmetry, the range of posterior beliefs for which it is advantageous to refrain from transgressing is much larger than the range for which transgressing is advantageous. We can say, in this precise sense, that an expected value maximiser with Bayesian beliefs will be heavily biased in favor of not transgressing. Simply put, if our Bayesian has at least a weak belief that an agent is watching, she will refrain from transgressing in order to avoid the large error costs associated with being caught transgressing. This logic does not require a cognitive bias. Indeed, because someone with Bayesian beliefs has integrated all available evidence in the optimal way, we are explicitly considering a decision maker *without* a cognitive bias.

What would a cognitive bias accomplish? To see this, imagine an expected value maximiser with a cognitive bias, $δ$, in favor of A. For any Bayesian posterior, *p*, our cognitively biased decision maker believes that A is true with probability p + $δ$. In other words, this person has an exaggerated estimate of A. As Figure 1 shows, for any Bayesian posterior greater than *p\_c*, this bias has no behavioural consequences. Both a Bayesian and a cognitively biased decision maker will make the truly optimal choice and refrain from transgressing. Analogously, for any Bayesian posterior less than *p\_b* = *p\_c* - $δ$, the cognitive bias has no behavioural consequences. Both types of cognition lead to transgressing, the truly optimal choice. Critically, however, for any Bayesian posterior between *p\_b* and *p\_c*, the Bayesian will transgress, which is the optimal behaviour, while the cognitively biased individual will make the sub-optimal choice (i.e., refrain from transgressing). Altogether, the Bayesian expected value maximiser always chooses the best behaviour, which is simply another way of saying that Bayesian beliefs integrate all available evidence in the optimal way. The cognitively biased expected value maximiser sometimes chooses the best behaviour, but not always. This is simply another way of saying that, when considering a single decision-making domain without constraints on the cognitive processes available, the notion of an adaptive cognitive bias makes little sense. If cognition is biased, it must by definition occasionally lead to sub-optimal choices.

To summarise, Johnson claims that exaggerated estimates of the probability that transgressions will be detected are a) the only; and b) the *best*, way to avoid the costly error of transgressing and being caught. We have shown that the first of these claims is incorrect, and that the second is beside the point. The first claim is incorrect because individuals with unbiased beliefs about the probability of costly errors will still try to avoid those errors (indeed, even individuals who *under*estimate the probability of costly errors may still try to avoid those errors). The second claim is beside the point because adaptive decision-making is not about minimising errors, or even about minimising costly errors – it is about minimising overall costs. And in that respect, unbiased estimates outperform exaggerated estimates.

Johnson’s *supernatural punishment theory* is a bold and important thesis, and his new book cements his status as one of the key theorists in the field. But no one is completely without error. He may be right that the evolutionary advent of language and theory of mind increased the “probability and costs of exposure for social transgressions” (2015, p. 168). He may also be right that belief in punitive supernatural agents inhibits social transgressions. In our opinion, however, the “error of Johnson” is the assumption, based on error management theory, that exaggerated estimates of the probability of detection – erroneous beliefs, including beliefs in supernatural punishment – are the most adaptive response to the novel evolutionary scenario he describes. As we have shown, the most adaptive cognitive architecture in this scenario is the architecture of an agent who maximises expected value under Bayesian beliefs.

**Figure caption.**

**Figure 1.** Expected value of transgressing (dashed line) and not transgressing (solid line) as a function of the Bayesian posterior belief that A is true (i.e., that an agent is watching you). At *p\_c* a Bayesian expected value maximiser is indifferent between transgressing and not transgressing. For any posterior greater than *p\_c*, transgressing is the sub-optimal behaviour. For any posterior less than *p\_c*, transgressing is the optimal behaviour.

**References**

Bayes, T. R. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London, 53*, 370–418.

Haselton, M. G. & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology, 78*(1), 81–91.

Haselton, M. G. & Nettle, D. (2006). The paranoid optimist: an integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review, 10*(1), 47–66.

Johnson, D. D. P. (2009). The error of God: Error management theory, religion, and the evolution of cooperation. In S.A. Levin (Ed.), *Games, Groups, and the Global Good*, Springer Series in Game Theory (pp. 169-80). Berlin: Springer Physica-Verlag.

Johnson, D. (2015). *God is watching you: how the fear of God makes us human*. Oxford: Oxford University Press.

McKay, R. & Efferson, C. (2010). The subtleties of error management. *Evolution & Human Behavior, 31,* 309-319.

Simpson, J. & Weiner, E. (1989). error. *Oxford English Dictionary*. Oxford: Oxford University Press. Retrieved from <http://www.oxforddictionaries.com/definition/english/error>

1. To be clear, “erring on the side of caution” means acting so as to make more costly errors at a lower rate than one makes less costly errors. [↑](#footnote-ref-1)
2. Assume there are no other relevant costs, e.g., wounded pride. [↑](#footnote-ref-2)
3. We assume that by “engender more mistakes” Johnson means, “engender more costly mistakes”. Accurate estimates will engender more costly mistakes than exaggerated estimates, but fewer mistakes overall. [↑](#footnote-ref-3)