The evolved architecture of hazard management: Risk detection reasoning and the motivational computation of threat magnitudes

DOI: 10.1017/S0140525X06009538

John Tooby[a] and Leda Cosmides[b]

[a]*Department of Anthropology, Center for Evolutionary Psychology, University of California, Santa Barbara, CA 93106-3210;* [b]*Department of Psychology, Center for Evolutionary Psychology, University of California, Santa Barbara, CA 93106.*

tooby@anth.ucsb.edu **http://www.psych.ucsb.edu/research/cep/**

cosmides@psych.ucsb.edu

http://www.psych.ucsb.edu/research/cep/

Abstract: The architecture of the hazard management system underlying precautionary behavior makes functional sense, given the adaptive computational problems it evolved to solve. Many seeming infelicities in its outputs, such as behavior with "apparent lack of rational

motivation" or disproportionality, are susceptibilities that derive from the sheer computational difficulty posed by the problem of cost-effectively deploying countermeasures to rare, harmful threats.

Boyer & Lienard's (B&L's) landmark work represents a decisive advance in our understanding of the evolutionary psychology of ritual behavior, viewed as a byproduct of adaptations for avoiding danger. We strongly endorse the view that there is an evolved, species-typical suite of neurocomputational adaptations designed to deal effectively with dangers and the deployment of countermeasures – what we have called *hazard management* or *precaution systems* (Fiddick et al. 2000). We also agree that the themes of ancestrally recurrent danger (e.g., contagion, danger to offspring) that pervade obsessive-compulsive disorder (OCD) ideation, together with the hyperactivation of precautionary checking subroutines, indicate that OCD results from breakdowns in these evolved systems (Cosmides & Tooby 1999). In particular, we have been pursuing the hypothesis that there is an evolved domain-specific inferential specialization designed to reason about whether an appropriate precaution has been taken, conditioned on the presence of the danger it protects against. A growing body of evidence suggests that such a *risk detection specialization* exists (somewhat parallel to the cheater detection system) and is cognitively distinct (Fiddick et al. 2000), neuropsychologically dissociable (e.g., from reasoning about social contracts; Stone et al. 2002), and involves distinct patterns of neural activation, as judged by neuroimaging findings (Ermer et al., in press).

This risk detection reasoning (and attentional) subsystem appears to use cognitive primitives at the level of *hazard*$_i$ *(present/absent), countermeasure for i* (*in effect, not in effect*), and it draws attention to conditions in which a danger may be present but its appropriate precaution may not have been taken. We believe that when this checking subroutine produces the inference that a specific hazard$_i$ might be present in the absence of its specifically associated countermeasure, this output potentiates the regulatory circuitry governing motivations to take associated safeguards. We suspect that this same system, to accomplish its detection function for evolutionarily prepared dangers, accesses what B&L call the *Potential Hazard Repertoire*, and the *Evolutionary Precaution Repertoire*. We therefore view these as evolutionarily prepared subsets of two more encompassing repertoires that include *all represented hazards*, and *all represented countermeasures*, respectively. That is, the risk detection subsystem not only functionally links "innate" countermeasures (e.g., washing) to ancestral hazards (e.g., disease exposure), but also links evolutionarily novel countermeasures (e.g., backing up) to evolutionary novel dangers (e.g., hard disk crashes).

The adaptive computational problems posed by reasoning, however, are dwarfed by the magnitude of the design problems posed by the task of computing valuations in a fitness enhancing way (Tooby et al. 2004). The selective intensity of an adaptive problem is a function of the frequency of the selective event, multiplied by the magnitude of its fitness consequences. Events that (1) happen frequently over the lifespan, (2) where outcomes follow rapidly, and (3) where outcomes can be readily assayed for their value allow the evolution of feedback systems that shape and weight actions reliably (allowing, e.g., motor skill acquisition). In contrast, a selectively significant set of detrimental events (*threats*) will be sufficiently harmful (e.g., disease, predation, ambush, social disgrace) to make it potentially cost-effective to take countermeasures even when their incidence is low. Because their incidence is low, however, sampling error and a paucity of observations will make uncertainty great, accurate ascertainment of their true probabilities difficult, and the change in their probability associated with a given countermeasure or predictive cue even harder to determine. Indeed, a successful precaution may preclude a harmful event from happening, so that it is never observed by an individual. How can the observer tell whether a threat has disappeared, whether the precaution remains necessary, or whether good luck was responsible for the apparent disappearance of a threat that remains real? The rarer an event is, the more its true probability will be hidden from an observer among a large range of possible values.

Indeed, because many threats are produced by design by antagonistically co-evolving organisms, selection often makes threats maximally unpredictable to their victims. Yet the motivational system, in order to allocate effort among possible precautionary actions and other fitness-promoting activities, must compute and assign to each threat (given a set of cues) a value, as an approximate function of its expected cost and its expected probability. That is, for each represented threat, the system must compute a regulatory variable: a *threat index*. It also needs to compute values for cues that predict changes in the probability of the threat, as well as values for the effectiveness of countermeasures. (The categorization of threats ought not to be just a function of their "objective" external resemblance to each other, but more importantly of how similar their countermeasures are: That is, "pollution" is an evolved idea not because it represents a single kind of threat – it does not – but because the environmental threats it lumps together can be attenuated using the same kinds of countermeasures.) Because of their different evidentiary bases, threat indices cannot be computed in the same way in which positive payoffs driving reward-seeking decisions can be. Accordingly, the precautionary system is adaptively designed to produce feelings of "compulsion" (action motivated in a way that is divorced from any proximate goal or confirmatory payoff – unlike, say, foraging). Normally, precautionary compulsions should be trumped whenever situations invite alternative actions whose payoffs exceed the threat index. Although what might be called *optimal defense theory* has some powerful analogies with optimal foraging theory, it also has disanalogies which would have selected for the hazard management system to become a computationally differentiated part of the motivational architecture.

A well-engineered system should supplement observations of the incidence of rare costly events and countermeasures with other sources of information. These include (1) correlated cues to conditions of heightened threat (Neyman-Pearsonian decision theory suggests that the system ought to be biased to overinterpret the diagnosticity of candidate predictors, as in post-traumatic stress disorder [PTSD]); (2) non-frequentist causal models of countermeasures (e.g., physical barriers to threats); (3) decoupled imaginative simulations (Tooby & Cosmides 1990) and quasi-counterfactual representations such as "near misses" (when dysregulated, these recalibrations constitute obsessions); (4) possible transgenerational epigenetic reweighting (see Tooby et al. 2003); (5) genetic inheritance (the heritable personality dimension psychoneuroticism may exist as an adaptation to allow local and transgenerational recalibration of threat indices through genetic or epigenetic reweighting (see Tooby et al. 2003); and (6) social sources of information.

The social dimension especially illuminates collective ritual behavior. Observations gathered by multiple conspecifics provide more accurate estimates of actual threat magnitudes – the adaptationist rationale for circuits that reset threat indices partly based on observed fear reactions in others (Cook & Mineka 1987). Moreover, the high uncertainty hovering over incidences and countermeasure effectiveness leaves the hazard system susceptible to error, volatile reweighting, individual differences, and social entrainment (including manipulation). Seeing others devote considerable effort to a collective ritual presented as a countermeasure advertises their threat indices, inducing observers to reweight. Finally, because of human improvisational intelligence (Tooby & DeVore 1987), we think there is a proper domain for some precautionary ritual behavior, where it functions as preparation for complexly managed, instrumental activity in dangerous and unpredictable environments whose negotiation necessitates high levels of skill acquisition, rapid reaction time, and organized material readiness. Aspects of military training, seamanship, *katas*, mountain climbing – even medicine concoction and some cooking – exemplify aspects of functional precautionary ritual behavior. This minor caveat aside, B&L have powerfully illuminated underlying commonalities in ritual behavior.