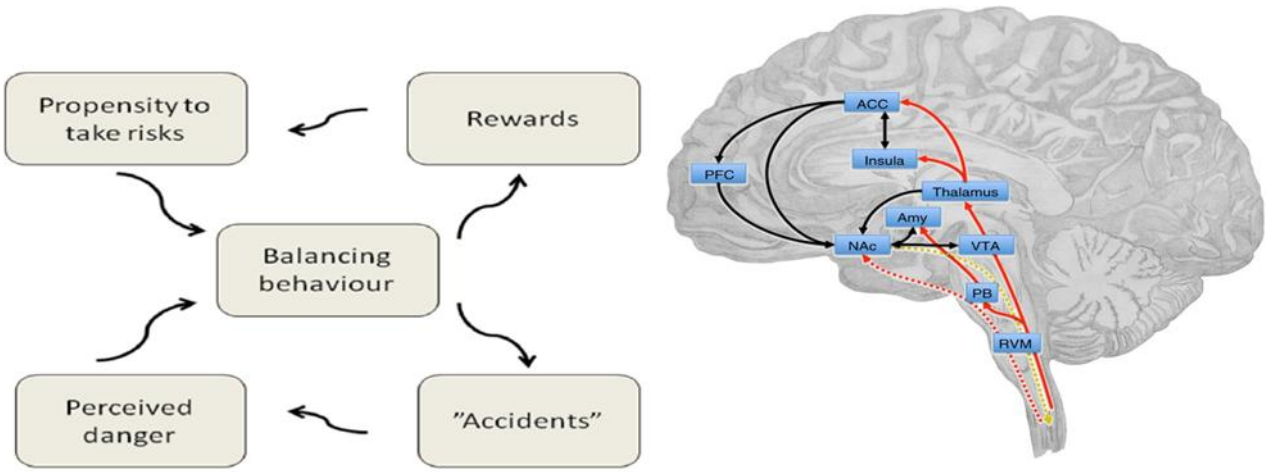

The Risk Thermostat Revisited:

A Predictor/ Corrector mechanism for Homeostatic Risk Regulation

Abstract

John Adams' Risk Thermostat model proposed that human behaviour regulates experienced risk rather than minimising objective hazard, accounting for phenomena such as behavioural adaptation and, under some conditions, risk compensation (Adams, 1995; Wilde, 1994). While influential, the model remained largely metaphorical, offering little account of the mechanisms by which risk is sensed, regulated, or stabilised. Here we propose an updated formulation grounded in contemporary neuroscience and predictive processing. We argue that risk regulation can be modelled as a homeostatic control process operating over precision-weighted expectations of harm and reward, rather than as a direct response to hazard magnitude. Within this framework, anxiety functions as a low-threshold anticipatory signal associated with uncertainty about future harm (Grupe & Nitschke, 2013), pain as a high-threshold signal associated with realised or imminent bodily threat (Apkarian et al., 2009; Wiech, 2016), and reward anticipation as a countervailing signal encoding positive prediction error about the value of action policies (Schultz, 1997; Schultz, 2016). This reinterpretation does not claim a single unified neural mechanism for "risk" but offers a control-theoretic lens through which disparate findings in pain, anxiety, motivation, and safety behaviour can be coherently related.

Key words - *Risk thermostat; Predictive processing; Homeostatic control; Anxiety and pain; Risk compensation; Safety behaviour*

Introduction

Risk regulation is a central feature of human behaviour across domains including transport safety, occupational health, sport, finance, and everyday decision-making. Classical safety engineering approaches have often assumed that reducing objective hazard will proportionally reduce risky behaviour. In contrast, Adams argued that individuals tend to regulate their behaviour to maintain a preferred level of *perceived* risk, a principle illustrated by the Risk Thermostat metaphor (Adams, 1995).

The Risk Thermostat has proven useful in explaining behavioural adaptation to safety measures and the sometimes disappointing impact of technical interventions. However, it was

intentionally underspecified at the mechanistic level. Risk was treated as subjective and learned, but the processes by which it is sensed, compared, and regulated were left open.

Over the past two decades, neuroscience has increasingly framed perception, action, and affect as forms of active inference under uncertainty rather than stimulus–response mappings (Friston, 2010; Clark, 2013). Pain, anxiety, and reward are now widely understood as context-sensitive states shaped by expectations, learning, and control demands rather than as direct readouts of sensory input (Barrett, 2017). This shift creates an opportunity to revisit the Risk Thermostat as a *biologically plausible control model*, without claiming that the brain literally computes "risk" as a single variable.

Risk as a Regulated Variable

In physiological homeostasis, regulated variables are maintained within tolerable bounds rather than eliminated entirely. Temperature, glucose concentration, and blood pressure fluctuate, but are constrained by feedback control. Adams' central claim—that humans regulate *felt* risk rather than objective hazard—implicitly places risk in this category.

From a neurocomputational perspective, *experienced risk* can be modelled as an emergent quantity reflecting expectations about harm under uncertainty, weighted by confidence (precision) in those expectations (Friston et al., 2016). Importantly, this does **not** imply that risk is identical to prediction error. Rather, prediction error provides one class of signals that inform whether current or anticipated states violate tolerable bounds of safety or value.

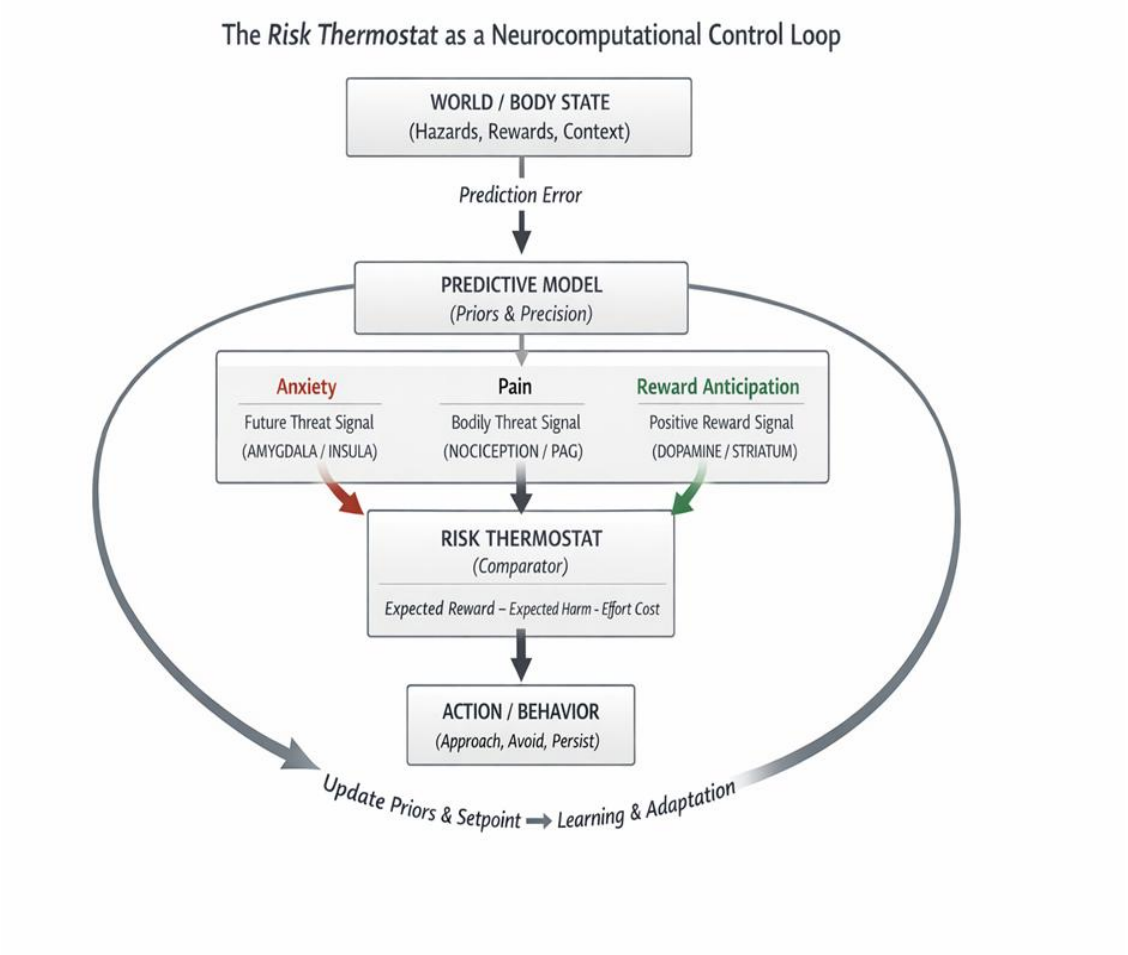

Figure 1 – The risk thermostat revisited

This framing preserves Adams' behavioural insight while avoiding the category error of equating a sociotechnical construct ("risk") with a single neural signal.

The Aversive Control Arm: Anxiety and Pain

Within this control-theoretic framing, anxiety and pain can be treated as distinct but related aversive signals contributing to risk regulation.

Anxiety is best characterised as a *future-oriented* state associated with uncertainty and anticipation of potential harm. Neurobiological accounts consistently emphasise its link to uncertainty, volatility, and increased sensitivity to threat-related prediction error (Grupe & Nitschke, 2013; Paulus & Stein, 2010). Anxiety often arises in the absence of current injury and appears to bias behaviour toward caution and monitoring rather than immediate withdrawal.

Pain, by contrast, is typically *present-oriented* and closely tied to bodily integrity. Contemporary accounts emphasise that pain perception is not a direct measure of tissue damage, but is strongly modulated by expectations, context, and meaning (Apkarian et al., 2009; Wiech, 2016). Nevertheless, pain tends to function as a higher-threshold signal, demanding immediate protective or corrective action.

Treating anxiety and pain as two operating regimes of aversive control—rather than as categorically separate systems—provides a parsimonious explanation for their frequent interaction, including the amplification of pain by anxiety and the persistence of aversive states after tissue healing.

The Appetitive Control Arm: Anticipation of Reward

Risk regulation cannot be understood solely in terms of avoidance. Organisms routinely accept danger, effort, and discomfort in pursuit of valued outcomes. This requires a countervailing signal that justifies exposure to risk.

Dopaminergic systems are widely interpreted as signalling *reward prediction error*—outcomes that are better or worse than expected—rather than pleasure itself (Schultz, 1997; Schultz, 2016). In control-theoretic terms, such signals increase the expected value of certain action policies, thereby permitting risk-taking when anticipated benefits outweigh anticipated costs.

Within the Risk Thermostat framework, reward anticipation raises tolerance for uncertainty and potential harm, not by suppressing aversive signals, but by re-weighting their influence on action selection. This helps explain why identical hazards may be experienced as unacceptable in some contexts and actively sought in others.

Arbitration, Setpoints, and Learning

For a thermostat to function, competing signals must be compared relative to a reference level. In neural terms, this corresponds to integrative processes that combine expected reward, expected harm, and effort costs when selecting actions. Models such as the Expected Value of Control framework identify candidate mechanisms for such arbitration without claiming a single dedicated "risk module" (Shenhav et al., 2013).

Risk tolerance is not fixed. It is shaped by development, learning, social norms, prior outcomes, and health status (Adams, 1995; Slovic, 2016). Safe experiences can raise tolerance, while unexpected harm can lower it. This slow adaptation mirrors the adjustment of setpoints seen in other homeostatic systems.

Risk Compensation: Conditional, Not Universal

A frequent criticism of risk thermostat models is that empirical evidence for risk compensation is mixed. This criticism is well-founded. Early claims of near-complete behavioural offset (e.g., Peltzman, 1975) have not been uniformly supported, and later analyses emphasise that compensation depends on conditions such as visibility of safety measures, feedback salience, incentives, and skill (Hedlund, 2000).

The present framework accommodates this variability. If risk regulation depends on experienced uncertainty and control rather than hazard per se, then compensation should occur only when interventions alter subjective risk in a way that invites behavioural adjustment. Where interventions reduce harm without changing experience, compensation may be weak or absent.

Limitations and Competing Accounts

This account is explicitly a *model*, not a claim of neural identity. It does not deny the relevance of alternative frameworks such as reinforcement learning, prospect theory, or the somatic marker hypothesis. Rather, it offers a control-theoretic perspective that can coexist with these approaches.

Key limitations include the difficulty of operationalising "experienced risk" in experimental settings and the risk of over-generalising from laboratory paradigms to real-world sociotechnical systems. Empirical tests of the model would require manipulations that independently vary objective hazard, perceived controllability, and reward structure while measuring behavioural and physiological responses.

Conclusion

John Adams' Risk Thermostat captured a durable behavioural insight: humans regulate risk as it is experienced, not as it is statistically defined. Contemporary neuroscience does not replace this insight but helps to situate it within a broader understanding of biological control.

By treating anxiety, pain, and reward anticipation as interacting control signals rather than isolated phenomena, the Risk Thermostat can be reframed as a homeostatic regulation of action under uncertainty. This reframing does not claim finality or universality but provides a coherent bridge between safety science and modern neurocognitive theory.

David Slater

References

Adams, J. (1995) *Risk*. London: UCL Press.

Apkarian, A.V., Baliki, M.N. and Geha, P.Y. (2009) 'Towards a theory of chronic pain', *Progress in Neurobiology*, 87(2), pp. 81–97.

Barrett, L.F. (2017) *How Emotions Are Made: The Secret Life of the Brain*. Boston, MA: Houghton Mifflin Harcourt.

Berridge, K.C. and Robinson, T.E. (1998) 'What is the role of dopamine in reward? Hedonic impact, reward learning, or incentive salience?', *Brain Research Reviews*, 28(3), pp. 309–369.

Clark, A. (2013) ‘Whatever next? Predictive brains, situated agents, and the future of cognitive science’, *Behavioural and Brain Sciences*, 36(3), pp. 181–204.

Friston, K. (2010) ‘The free-energy principle: A unified brain theory?’, *Nature Reviews Neuroscience*, 11(2), pp. 127–138.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O’Doherty, J. and Pezzulo, G. (2016) ‘Active inference and learning’, *Neuroscience & Biobehavioural Reviews*, 68, pp. 862–879.

Grupe, D.W. and Nitschke, J.B. (2013) ‘Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective’, *Nature Reviews Neuroscience*, 14(7), pp. 488–501.

Hedlund, J. (2000) ‘Risky business: Safety regulations, risk compensation, and individual behavior’, *Injury Prevention*, 6(2), pp. 82–90.

Paulus, M.P. and Stein, M.B. (2010) ‘Interoception in anxiety and depression’, *Brain Structure and Function*, 214(5–6), pp. 451–463.

Peltzman, S. (1975) ‘The effects of automobile safety regulation’, *Journal of Political Economy*, 83(4), pp. 677–725.

Schultz, W. (1997) ‘Dopamine neurons and their role in reward mechanisms’, *Journal of Neurophysiology*, 80(1), pp. 1–27.

Schultz, W. (2016) ‘Dopamine reward prediction error signalling: A two-component response’, *Nature Reviews Neuroscience*, 17(3), pp. 183–195.

Shenhav, A., Botvinick, M.M. and Cohen, J.D. (2013) ‘The expected value of control: An integrative theory of anterior cingulate cortex function’, *Neuron*, 79(2), pp. 217–240.

Slovic, P. (2016) *The Perception of Risk*. London: Routledge.

Wiech, K. (2016) ‘Deconstructing the sensation of pain: The influence of cognitive processes on pain perception’, *Science*, 354(6312), pp. 584–587.

Wilde, G.J.S. (1994) *Target Risk*. Toronto: PDE Publications.