ChatGPT

Consequence Minimization: Formal Model and Cross-Disciplinary Review

Introduction

Consequence Minimization (CM) refers to the decision principle of prioritizing the avoidance of catastrophic or absorbing negative outcomes before pursuing maximization of gains. In essence, an agent guided by CM is "safety-first": ensuring survival or a baseline outcome is secured with high priority, and only then seeking additional benefits. This report develops a formal modeling framework for CM and reviews interdisciplinary literature – from behavioral economics and cognitive neuroscience to psychology, game theory, and evolutionary biology – that illuminates this principle. We also outline empirical paradigms (e.g. "termination" tasks, stress inductions) to test CM's predictions, emphasizing how CM can be distinguished from related concepts like loss aversion, regret minimization, and conventional risk aversion. The goal is a comprehensive understanding of CM: its theoretical formulation, its relationship to known decision strategies, and its grounding in biological and behavioral research.

Formal Modeling Framework for CM Agents

Core Assumptions: In the CM framework, the agent holds a *baseline threshold* for an acceptable outcome (e.g. a minimum wealth, survival probability, or utility level). Outcomes below this threshold are treated as **catastrophic** (perhaps leading to ruin or irreversible loss). The agent's preferences are *lexicographic* or *threshold-weighted* – avoiding threshold breach dominates other considerations. Formally, one can define a lexicographic utility: given actions *A* and *B* with outcome distributions, *A* is preferred to *B* **iff** the probability of falling below the threshold under *A* is less than that under *B*; only if these catastrophic risks are equal does the agent then compare expected gains or other criteria. This captures the "avoid ruin at all costs" ethos of CM. Equivalently, the agent *minimizes the probability of disaster* (or some risk measure of severe loss) first and *maximizes expected gain second*. This modeling idea has precedent in finance as **Roy's safety-first criterion** (Roy, 1952), which selects portfolios that minimize the probability of returns falling below a disaster level. In decision theory terms, CM resembles a *maximin* or *minimax* rule focusing on worst-case outcomes: the agent "minimizes the possible loss for a worst-case scenario" (ensuring the minimum payoff is as high as possible). Unlike standard maximin, however, CM is **state-dependent** – the agent's risk posture flips depending on whether the current state is above or below the safe baseline.

State-Dependent Risk Behavior: A hallmark of CM is *risk-aversion when above the baseline* and *risk-seeking when below it*. Intuitively, an agent who is "safe" (above the survival threshold) will avoid risky gambles that could dip them below the threshold, preferring sure but smaller gains to secure the status quo. Conversely, an agent who is below the threshold (or faces inevitable disaster unless fortunes improve) will embrace risk-seeking as a last resort – a long-shot chance is preferable to assured failure. This behavior mirrors the famous *"break the glass in case of emergency"* flip in risk preference. **Prospect theory** in behavioral economics provides an analogous descriptive model: human value functions are *reference-dependent and S-shaped*, being concave (risk-averse) for gains above a reference point and convex (risk-seeking) for losses

below it. **Figure 1** illustrates this classic S-shaped value curve, steeply diminishing for losses, reflecting that falling below one's reference (baseline) triggers risk-seeking in an attempt to recover.

Figure 1: Illustrative Prospect Theory value function (as a proxy for CM's state-dependent utility). The curve is S-shaped*, concave for gains (above the reference point) indicating risk aversion, and convex for losses (below reference) indicating risk seeking. A CM agent similarly becomes risk-averse once safely above a baseline threshold, but if below a critical baseline, will take risks to try to climb out of the "danger zone."*

To formalize this flip, one can imagine a utility function with a kink or threshold at baseline. For example, let T be the baseline outcome. A simple representation is a **two-part utility**:

- If outcome $x \ge T$, utility $u(x)$ is an increasing concave function (ensuring risk aversion in that region). The marginal utility of gains beyond T may diminish, reflecting contentment with securing the baseline.
- If outcome $x < T$, utility $u(x)$ might be formulated to capture desperation – for instance, **steeply negative utility** as x falls further below T, but with a shape that can favor variability. One extreme modeling choice is to assign $u(x) = -\infty$ for any x below T – effectively forbidding outcomes below the threshold (pure lexicographic safety). A softer approach might define $u(x)$ for $x<T$ as a convex function or even assign extra "gain" for reducing the shortfall (which can incentivize high-variance gambles that *might* eliminate the shortfall). In dynamic terms, below-threshold states could trigger a "survival mode" policy aimed at getting back to safety at the expense of higher risk.

An equivalent formulation uses **constrained optimization**: maximize expected reward *subject to* keeping the probability of catastrophe below some ϵ (often ϵ very small). If the constraint cannot be satisfied (i.e. catastrophe is likely inevitable), the agent "relaxes" it and focuses on maximizing the chance of survival by taking risks (since a conservative strategy would guarantee failure). This aligns with the logic of *risk-sensitive optimal foraging models*: if the safe option leads to *certain death*, the organism should gamble on the risky option that at least offers a chance of survival.

Comparison to Other Decision Strategies: It is important to distinguish CM from superficially similar strategies:

- **Risk Aversion (Constant):** Traditional risk-averse agents have a concave utility over all outcomes – they dislike variability at any wealth level. A CM agent is *not uniformly risk-averse*; rather, their risk preference is *conditional on state*. Above the baseline they behave in a highly risk-averse manner (even more extremely than standard expected-utility might predict, since falling below T carries an outsize disutility), but **below the baseline they become risk-preferring**. This pattern is sometimes called *"risk aversion with a hope of rescue"*. Empirically, this maps to observed behavior in humans and animals: for instance, recent research on health outcomes found that individuals are **risk-seeking at low levels of health** but switch to risk-aversion once health passes a moderate threshold (around 0.5 on a 0–1 scale), becoming most risk-averse when in perfect health. In other words, when one's health (or wealth, etc.) is poor (below a comfortable baseline), there's willingness to take chances, but once health is secure, people become protective of it.

- **Loss Aversion:** Loss aversion (from prospect theory) means losses loom larger than equal-sized gains – people are biased to avoid losses. CM shares the spirit of **"avoid bad consequences first"**,

but goes beyond a static bias. CM is tied to an *objective threshold* (e.g. bankruptcy, death) rather than purely subjective loss/gain framing. A loss-averse investor might refuse a fair gamble at any wealth level due to the psychological pain of losses; a CM investor will refuse gambles when already in a good position (since a loss could push them into disaster) but might **accept gambles when in dire straits**. Also, neurologically, loss aversion appears mediated by fear mechanisms (amygdala signals), which dovetails with CM's "survival circuit" basis (discussed later). But CM is conceptually more **lexicographic** (survival trumps profit), whereas loss aversion is typically modeled as a weighting in utility (losses weighted ~2× gains, for example). If loss aversion is mild, a person may still take some risks above baseline, whereas a strong-form CM agent would virtually take *no* risk that jeopardizes survival. In summary, loss aversion can be seen as a *softer* constraint (it tilts choices against losses), while CM often entails a *hard constraint* (no tolerance for catastrophic loss).

- **Minimax and Regret Minimization:** In classic game theory, **minimax (maximin)** strategies optimize the worst-case payoff. CM is essentially a *stochastic* version of maximin – avoid the outcome that is worst (e.g. ruin) with highest priority. Some complex systems and adversarial AI approaches use minimax to ensure robustness, even at expense of average performance. CM aligns with this robust stance. **Minimax regret**, on the other hand, is a criterion where one minimizes the maximum "regret" (the difference between the payoff of a chosen action and the best payoff that could have been achieved in hindsight). Regret minimization has a different motivation: it's about avoiding the pain of knowing you *could* have done better, rather than avoiding literal catastrophe. In behavioral contexts, *regret aversion* means people anticipate how bad they'd feel if a decision leads to a poor outcome and they forgo a better alternative. This can lead to **either** cautious or risky choices depending on context. For example, if only a risky choice would prevent a potential big regret (missing out on a high payoff), regret aversion might prompt risk-seeking. Researchers have shown that *regret-minimizing choices may be either risk-avoiding or risk-seeking*, depending on which action would feel worse in hindsight. CM is more straightforward – it cares about actual disastrous outcomes, not the psychological regret. Thus, CM would predict risk-seeking only in the *specific scenario of sub-baseline desperation*, whereas regret aversion could make a person gamble even from a good state if, say, not gambling might lead to future regret of a missed jackpot. One can say **CM prioritizes objective catastrophic outcomes over subjective emotional outcomes**. This suggests experimental dissociation: if given a choice where one option avoids a small chance of ruin but entails certain moderate losses (no regret because outcome is known) vs another option that has higher expected value but a tiny ruin risk (and potential regret if not chosen), a pure CM agent picks the no-ruin option, whereas a pure regret-averse agent might pick the safer option only if they anticipate regret, etc. In practice, humans no doubt combine these factors, but conceptually CM is distinct.

- **Conventional Bounded Rationality (Satisficing):** Herbert Simon's concept of *satisficing* involves setting an aspiration level and choosing the first option that meets it, rather than optimizing fully. CM can be viewed as a specific kind of satisficing where the primary aspiration is **"avoid disaster"**. In satisficing models, an individual sets a minimum threshold of acceptability (which could correspond to our baseline) and any outcome above that is "good enough." This is similar to CM's baseline, except CM specifically implies *strong aversion to falling below* that level rather than indifference above it. Indeed, Simon noted that individuals often have **aspiration levels that serve as decision criteria**. CM's baseline could be seen as a dynamic aspiration: maintain at least this level. The difference is that satisficing typically stops searching once a threshold is met (not necessarily choosing the *optimal* above-threshold outcome), whereas CM might still optimize beyond survival

once survival is secured. In other words, CM could be formalized as a **lexicographic optimization**: first satisfy the survival constraint (like satisficing to baseline), then within the safe set, maximize utility. This ties into **bounded rationality** by acknowledging cognitive limits or evolutionary pressures: organisms might not compute an ideal trade-off between risk and reward in a single utility curve; instead, they apply a heuristic rule "Don't die first, then worry about gains," which is simpler and evolutionarily sensible in many contexts.

Baseline Thresholds and Survival Zones: Central to CM is the notion of a baseline and a "survival zone." The baseline itself could be fixed (e.g. zero wealth, or subsistence level of food, or minimal reproductive success to avoid extinction) or adaptive (e.g. a running aspirational target that can move with context). Empirically, these thresholds can often be identified by observing when behavior changes. For instance, in **animal foraging** experiments, a bird at risk of starvation (energy reserves below the daily survival requirement) dramatically shifts to risk-prone foraging, whereas the same bird with ample reserves favors safe foraging. Here the survival threshold is an energy level needed to survive the night; above it, playing safe ensures survival, below it, only a risky gamble for extra food could avoid death. In humans, thresholds might be financial (avoiding bankruptcy), physiological (staying above a health deficit), or even social (e.g. maintaining a minimal status in a group).

Importantly, CM suggests *two regimes*: a **secure regime** (above baseline) where the agent's strategy is conservative, variance-minimizing, focused on **defensive** moves; and a **critical regime** (at or below baseline) where the strategy becomes aggressive, variance-seeking, focused on **offensive** or recovery moves. This can be implemented in formal models via **state-dependent utility curvature** or **threshold-triggered policy switches** (e.g. a conditional plan that says "if wealth < W *then enter gamble mode, else stay safe"). One can analyze such models in dynamic programming or evolutionary simulations to see under what conditions they outperform static risk preferences. A key result from foraging theory, for example, is that if the environment is such that falling below the threshold = death, then threshold-based strategies maximize the* probability of survival *over time, even if they do not maximize the* rate of energy gain*. In finance, this relates to avoiding ruin: maximizing long-term growth of wealth is futile if a single bad gamble can eliminate you; thus an optimal policy might be to grow wealth while constraining the probability of ruin to near zero**. CM formalizes that intuition.

Behavioral Economics Perspectives

Research in behavioral economics provides concepts and evidence that underpin the CM principle:

- **Prospect Theory and Reference Dependence:** Kahneman and Tversky's prospect theory (1979) introduced the idea that people evaluate outcomes relative to a reference point (often the status quo or an aspiration level) and exhibit *risk aversion in gains* vs *risk seeking in losses*. This is essentially a baseline-dependent flip in risk preference. The *prospect theory value function* (see Figure 1 above) is steeper for losses, embodying loss aversion, but also its curvature changes sign at the reference point. Empirical studies confirm these patterns. For instance, in financial decision experiments, when outcomes are framed as losses relative to current wealth, individuals often prefer risky bets (hoping to break even) rather than accept a sure loss (this is sometimes called the "break-even effect"). While prospect theory was not explicitly about catastrophic outcomes, its notion of a reference point can be connected to CM's baseline. The difference is that in prospect theory the reference point is usually *psychological* (and loss aversion is a broad trait), whereas CM's baseline is *contextual* and tied to survival/ruin. However, the experimental evidence of **risk preference reversals** lends support to

CM's core assumption. A striking recent example comes from health economics: Mulligan et al. (2023) found that people become **risk-seeking when health quality is very low**, switching to risk-averse behavior once health passes a mid-level threshold (HQoL ≈0.5) and becoming extremely risk-averse at perfect health. This suggests an intrinsic baseline around "medium health" – below it, people gamble on treatments or behaviors that might drastically improve health (because otherwise life quality is unacceptably low), but above it, they avoid anything that might jeopardize their now-decent health. Such findings mirror the idea of CM and show that **baseline-driven risk flips** occur in economic preferences.

- **Loss Aversion and "Fear of Ruin":** Loss aversion (the idea that losses hurt about twice as much as equivalent gains feel good) is pervasive in choices. It can be seen as a crude approximation to CM in everyday decision-making: people often reject positive-expected-value gambles because the potential loss triggers a strong avoidance impulse. Neuroeconomic studies suggest this loss aversion is not merely a cognitive calculation but connected to emotional brain systems that *"put a cautionary brake on behavior"*. Notably, **amygdala** activity is associated with avoiding choices that carry potential losses, and rare patients with amygdala damage show an *elimination of loss aversion*, becoming willing to gamble in ways most would find reckless. This implies the brain has specialized circuits (involving the amygdala) that prioritize avoiding potentially **"deleterious outcomes"**. In economic terms, while a fully rational agent might treat a small probability of a large loss proportionally, humans overweight that scenario (a kind of built-in worst-case focus). CM formalizes this as a strategy – effectively treating small probabilities of catastrophe as highly significant. Loss aversion is slightly different in that it applies to any loss, even small ones, whereas CM is about truly catastrophic losses (crossing a critical threshold). However, in high-stakes contexts the two coincide: a *"loss"* that threatens one's baseline (e.g. losing one's livelihood) will be weighed extremely heavily. Behavioral econ research on insurance uptake, for example, shows people pay a premium to avoid low-probability disasters (house fire, etc.), consistent with a CM mindset of paying cost to eliminate chances of ruin. Additionally, phenomena like the **Allais paradox** (where people strongly avoid a small probability of zero payoff in one scenario) reflect an aversion to uncertainty especially when it introduces a possibility of getting nothing – again highlighting how the introduction of a worst-case outcome (even if unlikely) changes preferences disproportionately. Allais-type behavior can be modeled by assuming a steeper utility drop near zero (a baseline of "no gain" or worse). CM takes this to an extreme: the disutility of absorbing states (like zero wealth if that means ruin) might be modeled as effectively infinite, to explain such choices. In summary, behavioral patterns usually considered "biases" (loss aversion, probability weighting of rare events) may actually be *adaptive heuristics* approaching a CM rule: they err on the side of avoiding scenarios that could be very bad.

- **Minimax Regret vs CM:** Economists have also considered *regret* in decision-making. *Minimax regret* is a criterion where a decision-maker imagines the regret of each possible choice in each possible state of the world (the difference between the choice's outcome and the best possible outcome in that state) and then chooses the option that minimizes the maximum regret they could face. While this is a formal decision rule (proposed by Savage, 1951), it has parallels in how consumers sometimes choose "what I won't regret later" rather than what maximizes expected value. One could confuse this with CM if one thinks "catastrophic outcome would produce enormous regret." However, regret is inherently about *comparison to a foregone alternative*, whereas CM is about the outcome itself. Behavioral experiments (e.g. by Zeelenberg and colleagues) show that anticipating regret can lead to complicated behaviors: sometimes avoiding risk (so you won't feel sorry if it goes badly), other times taking risk (so you won't feel sorry if a missed opportunity would have paid off). By

contrast, a CM-driven person is more predictable: they avoid risk if currently okay, and take risk only if not taking it is almost certainly fatal. We can see the difference in *framing effects*: in one experiment, participants were told they would or would not learn about the outcome of an unchosen option in a gamble; those expecting feedback (hence able to feel regret) often took more risk to avoid the possibility of regret from missing out. But their situation did not involve any catastrophic loss – it was purely psychological. This underscores that **CM is not about feedback or hindsight; it is a forward-looking, survival-driven criterion.** In practical terms, combining CM with regret aversion could be an interesting mix: e.g. a manager might both avoid strategies that could bankrupt the firm (CM) and also those that, if failing, would make them look especially foolish relative to alternatives (regret). But theory-wise, CM stands distinct and in some cases will conflict with regret avoidance (for example, sticking with a sure mediocre outcome to avoid ruin might invite regret if a gamble would have succeeded, but a CM agent would accept that potential regret as long as survival is assured).

- **Adaptive Aspirations and Bounded Rationality:** In many economic models of choice under uncertainty, an *aspiration level* can evolve or be learned. A person might start with a target (say, achieve at least $X income this year) and make decisions accordingly; if they consistently overshoot or undershoot, they adjust the aspiration. CM can be framed in this adaptive aspiration context: if an agent consistently stays well above the survival threshold, they may raise their threshold (e.g. once basic safety is assured, they might start treating a higher goal as critical – perhaps reflecting *Maslow's hierarchy*, discussed later). If they repeatedly fail to reach safety, they might temporarily lower aspirations in an emergency mode (e.g. focus on short-term survival, akin to a company in crisis abandoning long-term plans to just avoid bankruptcy this quarter). This dynamic is observed in behavioral experiments where individuals exhibit **"sour grapes"** or **"goal shifting"** – they devalue goals that seem unattainable (perhaps to cope), which could be seen as rational if chasing an impossible baseline only leads to ruin (so they redefine what ruin means). Game-theoretic learning models like **reference point adaptation** show that agents adjust their risk-taking as their reference point (or status quo) changes, which can create path-dependent behavior. For instance, if someone has a streak of losses dropping them below their normal baseline, they may take unusually high risks ("double or nothing" behavior) until they recover – an observation consistent with **gamblers chasing losses**. Such behavior is often maladaptive in modern settings but can be interpreted through a CM lens: the gambler might feel they must get back to baseline (break even) at all costs, a misplaced but internally coherent threshold. Behavioral economists sometimes label this the *house money effect* (being more reckless with gains) and *break-even effect* (risk-seeking to recover losses). These can be accommodated in a CM model with a moving baseline: after a windfall, baseline may stay at the old level (so above it, one is cautious, yielding house-money conservatism maybe), but after a big loss, baseline might not immediately adjust downward (psychologically, you still consider break-even as "where I need to be"), thus you act as below baseline and gamble big to try to get there. Over time, repeated failure might force a person to reset their baseline lower (perhaps in a depression or resignation). The interplay of these dynamics is complex, but the key takeaway is **economic behavior often reveals threshold-driven risk shifts**, supporting the generality of the CM concept.

Cognitive Neuroscience Perspectives

From a neuroscience standpoint, the principle of CM resonates with the brain's evolved threat-management systems. When stakes are high and threats loom, different neural circuits dominate decision-making than in

safe, ordinary conditions. Key regions include the **amygdala**, **insula**, and **anterior cingulate cortex (ACC)** – often working in concert as part of a "salience" or "survival" network that detects danger and prioritizes protective actions.

- **Amygdala and Threat Avoidance:** The amygdala is famously involved in fear conditioning and detecting threat cues. It rapidly evaluates sensory input for signs of danger (e.g. a predator-like shape, a stimulus associated with past harm) and triggers defensive responses (fight, flight, freezing). When it comes to decision-making under risk, the amygdala appears to inject a strong aversive bias, especially regarding potential losses or negative outcomes. As noted, individuals with bilateral amygdala lesions exhibit an uncanny willingness to take monetary risks that normal participants avoid – specifically, they lack the usual **loss aversion bias**. Normally, the amygdala's "cautionary brake" helps **inhibit actions with potentially deleterious outcomes**; in other words, it acts as a neural instantiation of CM by suppressing choices that carry a risk of something very bad (even if on average the choice might be good). Neuroimaging shows that amygdala activation correlates with fear responses and risk-avoidant decisions. For example, presenting people with the possibility of losing money activates the amygdala, and this activation correlates with the degree of loss aversion each person exhibits. Under acute fear or stress (like seeing frightening images or mild threats), amygdala activity can increase, potentially translating to more conservative decisions. Indeed, one study found that *fear induction* (e.g. threat of shock) led to increased loss aversion – subjects required an even higher potential gain to compensate for a given loss when they were afraid, compared to safe conditions. This suggests fear (amygdala-driven) amplifies the priority of avoiding "bad outcomes," which is very much in line with CM tendencies. We can conceptualize the amygdala as part of an evolutionarily ancient **"survival circuit"** in the brain that monitors for cues of danger and shifts the organism's behavior towards consequence minimization. Neuroscientist Joseph LeDoux has argued that what we often call "fear" is essentially the conscious registration of underlying *survival circuit* activity that has the functional purpose of avoiding threats to life and limb. These circuits (centered on the amygdala and extending to brainstem areas that control stress responses) operate rapidly and non-consciously to bias decisions towards safety. In a CM context, when a potential outcome crosses a certain threat threshold, the amygdala circuit will tend to dominate neural processing, effectively vetoing high-risk/high-reward plans in favor of sure-safe ones. Conversely, if one is already in a state of extreme threat (perhaps an amygdala-saturated state where baseline is far from achieved), this system might paradoxically allow or encourage frantic, bold actions as a "last chance" (though this is less about amygdala and more about a lack of better options; some argue the **periaqueductal gray** and midbrain regions coordinate last-ditch efforts when cornered).

- **Insula and Risk Signals:** The insular cortex (particularly the anterior insula) is another key region in risk processing, often associated with the subjective feeling of uncertainty, anticipation of negative outcomes, and the integration of bodily states (like the "gut feeling" of dread). Neuroimaging studies frequently report **insula activation during risky decision-making**, especially when potential losses or painful outcomes are looming. The insula seems to encode something like the "feeling of risk" or the aversive affect that accompanies the possibility of a bad outcome. In one experiment, activity in the right anterior insula was greater when individuals made **safe choices after experiencing a loss or punishment**, suggesting the insula contributes to learning to avoid negative outcomes [1]. The insula is also implicated in **harm avoidance and anxiety** traits – people with high harm-avoidance (a personality measure akin to chronic consequence-minimizing tendency) show greater insula responses during risky decisions. One study found that if the insula is disrupted or if its activation is

weakened, individuals may fail to appropriately avoid bad risks (some pathological gamblers have insular dysfunction and do not experience "gut feelings" that normally deter further risky bets). In the context of CM, the insula can be thought of as generating the *aversive conscious experience* that underlies the motivation to avoid catastrophes. It might also be involved in representing the baseline threshold in an interoceptive way – for example, sensing when current resources or states are below comfortable levels (hunger, pain, etc., which then motivates urgent action). When stress is induced, the insula often shows increased activity, correlating with heightened risk aversion or altered decision weights. Interestingly, some studies on **stress and decision-making** show that under acute stress, people's decision parameters shift: a 2025 study found that stress caused a decrease in loss aversion in many participants (leading to riskier choices), but that effect was nuanced by gender (stress impacted men's decisions more, making them significantly risk-seeking, whereas women under stress remained more cautious or improved in outcome prediction). The authors speculated an evolutionary rationale: "If you're an organism that's being hunted or chased, doing something risky might be better than doing nothing". This sounds counter-intuitive relative to CM (one might expect stress -> more caution), but it actually aligns with the **baseline-dependent** nature of CM. Acute stress can signal that one's current state is threatened (thus effectively "you are below the safe baseline"), flipping the strategy to risk-seeking because *inaction could mean certain doom*. The insula in stressed men might be signaling a switch from deliberative caution to urgent action. In contrast, chronic anxiety (which might be more akin to always feeling near a threshold) tends to increase baseline risk aversion – for example, patients with generalized anxiety show enhanced risk aversion even when potential losses are small, essentially acting as if minor outcomes are catastrophic. This again highlights that the brain has multiple pathways: acute "fight-or-flight" arousal might lower loss aversion to encourage bold escape maneuvers, whereas long-term anxious predisposition heightens avoidance. Both can be interpreted through CM: the former is *CM below baseline* (take risks to escape dire situations), the latter is *CM above baseline* (when safe, remain hyper-vigilant to avoid any slip).

- **Anterior Cingulate Cortex (ACC) and Risk/Conflict Monitoring:** The ACC is a region in the medial frontal lobe involved in monitoring performance, detecting errors or conflict, and signaling when adjustments in behavior are needed. It is part of the brain's "alarm" system in a cognitive sense. During decision tasks, the dorsal ACC often activates when outcomes are uncertain or when there is a potential for making a mistake that could lead to a bad outcome. One influential theory (Brown & Braver 2005) posits that the ACC learns to predict the *likelihood of errors* or adverse outcomes and increases its activity as that likelihood rises, presumably to engage control mechanisms that avoid the error. This fits nicely with CM: ACC would ramp up control (perhaps via connections to lateral prefrontal cortex for more careful deliberation) if an action could lead to a critical mistake. In the context of a "termination task" (where one mistake ends the game), we'd expect ACC activity to be high throughout, reflecting the high stakes on each trial. The ACC is also tightly connected to the amygdala and insula as part of the **salience network**. Studies have found that **amygdala-ACC connectivity** can modulate risk-taking; for instance, one study on social risk found that individuals with higher aversion to social risk showed stronger amygdala connectivity to ACC, suggesting the ACC was integrating the amygdala's threat signals into a decision to avoid risky social moves. Another line of work shows the ACC is involved in processing regret and learning from it – lesions to ACC in some animal studies reduce avoidance of actions that led to negative outcomes, implying ACC normally helps adjust away from risk after a bad outcome ("don't do that again" signals). Overall, ACC can be seen as the cortical executor of CM principles: it monitors when the **"survival constraint"** is at risk of being violated (e.g. when an action has a high chance of a terrible outcome or when

current state is precarious), and it then engages cognitive control or avoidance responses to minimize that risk. In fMRI, tasks that involve high potential losses or require vigilant avoidance (like not touching a risky option) often show ACC activation along with insula. These regions together ensure that the presence of a possible catastrophe heavily influences decision processing. One could say the **amygdala sounds the initial alarm**, the **insula generates the gut feeling of dread**, and the **ACC implements the behavioral adjustment** – a trio that embodies consequence minimization in neural terms.

- **"Survival Circuits" Concept:** As mentioned, neuroscientists like LeDoux have reframed discussion of fear and survival behaviors in terms of **survival circuits**. These are neural circuits that have evolved to handle challenges critical to survival: finding food, escaping predators, maintaining homeostasis, etc.. When a survival circuit (say, the defense circuit centered on amygdala) is active, it coordinates physiological and behavioral responses: release of stress hormones, orienting attention to threats, inhibiting non-urgent goals, and driving reflexive actions. In modern terms, these circuits could be seen as implementing a primal form of CM – they *prioritize avoiding immediate danger over other activities*. For example, if you are hungry (food-seeking circuit active) but suddenly a snake appears (threat circuit active), the brain very quickly suppresses the hunger motivation to deal with the threat first. That is CM in a multi-goal scenario. The **amygdala-insula-ACC network** corresponds to detection and assessment of threat within these survival circuits. The output might be something like: *"Danger high – abort normal plan, execute emergency measures."* This maps to flipping from gain-maximization to consequence-minimization mode. Researchers have also noted that these circuits operate at different levels of consciousness – the amygdala can trigger avoidance without you even being aware of the threat consciously (e.g. jump-scare reflex). Thus CM behaviors can be somewhat automatic or habit-like when it comes to obvious dangers (we don't deliberate about stepping away from a speeding car – we just do it). In decision experiments, this means that when a certain option carries a large negative consequence (even probabilistically), people might have a visceral aversion to it that doesn't require explicit calculation – an embodiment of CM's priority in neural heuristics. It's interesting to consider that **"survival intelligence"** is deeply encoded in our brains: circuits that were honed to keep our ancestors alive guide much of our behavior today, sometimes in ways that seem irrational in abstract terms but are actually rational for worst-case avoidance. For instance, an evolutionary argument explains phobias: many people have intense fear of heights or snakes (amygdala-driven), far out of proportion to the actual risk in a controlled environment, because over millennia those who erred on the side of **"better safe than sorry"** had survival advantage. This is essentially **Error Management Theory (EMT)** – the idea that when the cost of a false negative (not seeing a threat) is much higher than a false positive (thinking there's a threat when there isn't), evolution biases us to make the safer error. Our brains, by erring on the side of assuming the worst, implement a default form of consequence minimization at the perceptual and cognitive level. Thus, many cognitive biases (loss aversion, probability weighting, one-trial fear learning) can be seen as reflections of an underlying neural mandate: *minimize the chance of catastrophic error (even if it means some false alarms or lost opportunities)*.

In summary, cognitive neuroscience reveals that our decision-making is not a unitary rational process but an interplay between *"opportunity-seeking"* circuits (often dopamine/reward-related, e.g. ventral striatum for gains) and *"threat-avoiding"* circuits (amygdala/insula/ACC for losses). CM corresponds to situations where the threat-avoidance system takes the driver's seat. Understanding how these systems interact (for example, how stress or context can tip the balance) is an active area of research, and it directly informs how we model an agent that toggles between **maximize-gains** and **minimize-losses** modes.

Psychological and Sociocultural Perspectives

From a psychological perspective, CM connects to concepts of stress, motivation under pressure, and basic needs. Human behavior changes drastically when we move from comfortable to crisis conditions – reflecting the same threshold-dependent principles in more subjective terms. Additionally, individual differences (personality, culture, development) affect where our thresholds lie and how strongly we adhere to CM-like strategies.

- **Stress and Decision Making Under Pressure:** High stress – whether due to time pressure, threat, or physiological stressors – often shifts decision patterns. As noted earlier, acute stress can sometimes produce risk-seeking (especially in men) by lowering loss aversion. On the other hand, chronic stress or anxiety typically heightens avoidance and hyper-vigilance. Psychologically, stress engages what is often called *System 1* (fast, reactive) rather than *System 2* (slow, deliberative). A stressed mind shortcuts to heuristics and often those heuristics are safety-oriented (except in fight-or-flight cases where bold action is itself a kind of safety maneuver). Experiments with stress induction (like public speaking stress, or cold pressor tasks) have found mixed but telling results: some report that stress increases **risk aversion in gains but risk seeking in losses**, essentially amplifying the typical prospect theory pattern (which is consistent with CM: under stress, people more strongly avoid risking a sure gain, but if already facing a loss they will gamble more fiercely to avoid the loss becoming realized). Stress also impairs working memory and analytical thinking, which means people rely on simpler rules – *"just don't do anything that could go horribly wrong"* might be one such rule. There's also evidence that stress hormones (cortisol, noradrenaline) affect neural decision circuits: for example, stress can potentiate amygdala activity and reduce prefrontal regulatory control, effectively biasing towards habitual or survival responses. In terms of CM, one could hypothesize that **stress pushes people into a "survival mode" sooner or more frequently**. If we design a study where participants must make a sequence of choices with accumulating rewards but a possibility of termination, inducing stress (say via noise, shocks, or evaluated performance) might cause them to set a higher baseline for "enough" and thus become overly cautious to avoid termination. Conversely, if they do fall below a target, stress might make them panic-gamble. This interplay is an open question – one we suggest explicitly testing (see Experimental Paradigms below). Clinically, understanding stress-related shifts in risk preference is important (e.g. why someone under pressure might make an uncharacteristically risky gamble, or why chronic stress leads to missed opportunities due to fear). These can be reframed as CM gone awry or overactive.

- **Maslow's Hierarchy and Motivation:** Abraham Maslow's hierarchy of needs is a classic theory in psychology which posits that basic needs (physiological and safety needs) must be satisfied before higher-level goals (esteem, self-actualization) can fully motivate behavior. In Maslow's pyramid, **safety needs** (security, stability, freedom from fear) come right after basic physical needs, and *"take precedence and dominate behavior"* once the physiological needs are met. This aligns perfectly with CM: if safety is not secured, the person is primarily concerned with avoiding harm and establishing stability. Only when one feels safe (i.e., above the baseline of security) do other motives (achievement, exploration) gain prominence. Maslow's theory, while often depicted as a strict pyramid, actually acknowledged that multiple needs can be pursued in parallel – but in times of threat, safety concerns will regress to the forefront. Empirical support for Maslow's idea comes, for example, from studies on poverty and decision-making: people in harsh economic conditions (worrying about shelter, food, personal safety) tend to discount the future more (take immediate rewards), consistent with focusing on short-term survival over long-term growth. They also often

exhibit heightened risk aversion in some domains (not risking what little they have on uncertain prospects) but sometimes risk-seeking in others (lottery play as a chance to escape poverty). This seemingly contradictory behavior can be understood through CM thresholds: when resources are just enough to survive, any gamble that could drop one below subsistence is avoided; but if one perceives themselves as stuck below an aspirational threshold (e.g. no chance of achieving comfortable life via normal means), then high-risk-high-reward options (like lotteries) become attractive as the only way out. **Figure 2** could conceptually illustrate this: imagine plotting "propensity to take risks" as a function of current resource level – it might be U-shaped, high at very low resources (desperation gambles), low at moderate resources (cautious stability), and perhaps rising again at high resources (where taking some risks is tolerable because baseline is secure – though some evidence suggests even the very wealthy can show risk aversion if they frame things as losses). Maslow's framework reminds us that **motivation is layered**. Consequence minimization corresponds to the **deficiency needs** (D-needs in Maslow's terms) – when not met, the person is anxious and laser-focused on them. Indeed, Maslow noted that if safety needs are unmet, a person will feel *"anxious and tense"* until they are resolved. Anxiety in this sense is a signal of the brain's CM circuits saying "we are below safe baseline, do something!" Conversely, when one is in the "growth needs" zone (above baseline), they might take *growth risks* (starting a new business, exploring new ideas) because even if those fail, one's basic security isn't threatened. But should a failure start to threaten their security, they will likely retreat. This can be observed in entrepreneurship: successful entrepreneurs often take calculated risks, but if initial failures start eating away their savings (their safety net), many become more conservative or exit – unless they mentally adjust their baseline (some might accept living on ramen to keep taking risks, essentially lowering their safety need temporarily in service of a long-term goal). In summary, Maslow's theory provides a qualitative backdrop that humans have an inherent prioritization of safety (avoid negative consequences) before self-actualization (maximize potential), which is precisely the trade-off CM formalizes.

- **Regret Minimization and Emotional Stress:** We discussed regret in the economic sense, but psychologically, **regret aversion** is often tied to emotions like guilt and self-blame. People not only fear bad outcomes, they also fear feeling responsible for bad outcomes. This can lead to *status quo bias* – sticking with a default or doing nothing, because taking an action that leads to a bad outcome would induce regret (whereas if the bad outcome happens by doing nothing, they feel less personal responsibility). This is related to **omission bias** in psychology. How does this tie to CM? Arguably, regret aversion could reinforce CM above baseline (don't do anything risky, you'd regret losing what you have) but conflict with CM below baseline (if you're already in trouble, regret aversion might still paralyze you from acting, whereas CM logic would say take the gamble). Some studies indeed find that **anticipated regret can cause inaction and caution** beyond what risk calculations would suggest. For example, a person might buy a very expensive insurance for a minor gadget not because of ruin risk, but because they'd *feel stupid* if it broke uninsured. That's regret avoidance, not rational CM, because the financial hit wouldn't be catastrophic. Experiments where participants expect feedback about unchosen options show that regret considerations become strong: one experiment by Zeelenberg et al. demonstrated that when people knew they'd see what would have happened had they made the other choice, they tended to choose options that minimize the chance of feeling regret – which could mean either safe or risky choices depending on which regret loomed larger. This can be seen as a different "baseline": a baseline of **self-esteem or outcome satisfaction** rather than survival. Interestingly, some individuals might treat *feeling terrible about a decision* itself as a kind of catastrophe to avoid (if they are very regret-sensitive). This could be considered a psychological implementation of CM where the "catastrophe" is an emotional state. However, for

clarity, in this review we focus on *objective negative outcomes* (financial ruin, physical harm, etc.). Still, it's worth noting that psychologically, instructing people to "imagine how you'd feel if things go wrong" often pushes them into safer choices – a strategy sometimes used in debiasing or decision counseling. This is leveraging regret aversion to achieve consequence minimization behavior. Conversely, one might leverage *hope for aspiration* (imagine how you'd feel if you missed out on a big success) to push risk-taking when appropriate. The existence of these emotional levers means that CM isn't just a cold calculation; it's supported (and sometimes overridden) by **affective forces** like fear, regret, and shame. Effective decision-making might require managing these emotions, and psychological training (e.g. for firefighters or soldiers) often involves learning to dampen panic (over-active CM that prevents action) while still respecting real dangers.

- **Developmental and Cultural Factors:** Are CM tendencies innate or learned? Evolutionary arguments suggest a strong innate component (as we see analogues in animals and very young children). Infants display wariness of heights (visual cliff experiments) even without learning – a safety bias. Toddlers and children also show some loss aversion and a sense of security-seeking (sticking with known safe options unless a situation forces them). However, *experience* and *culture* clearly shape how baselines are set and how risk is perceived. **Early-life environment** is crucial: Life History Theory in psychology posits that individuals who grow up in unpredictable, harsh conditions (e.g. poverty, high mortality neighborhoods) tend to adopt a *"fast life strategy"*: they discount the future, take more immediate risks, and have a lower threshold for acceptable security. Essentially, if early life teaches you that tomorrow is not guaranteed, you might not prioritize avoiding long-term catastrophe as much as seizing short-term gains (because the concept of a stable baseline itself is alien). On the other hand, those from stable, resource-rich upbringings often pursue a *"slow strategy"*: they plan long-term, avoid high risks, and invest in the future. Empirical studies support this: for instance, adolescents from high-stress backgrounds have been found to engage in riskier behaviors and have shorter future time horizons than their peers from low-stress backgrounds. This suggests that **baselines can be learned or inferred**. A child who never had a stable home might set their "survival baseline" at a lower level (perhaps day-to-day survival rather than year-to-year goals), and thus may not exhibit CM in the same way – they might consistently take risks that more privileged individuals see as unacceptable. Conversely, a child raised with strong security (financial, emotional) might have a *higher* baseline of what is considered necessary (e.g. they might consider a slight drop in comfort as "intolerable" and thus avoid any risk of that). Cultural norms also play a role: some cultures emphasize **collectivist safety nets** and *caution* (e.g. "better safe than sorry" proverbs, discouraging gambling), effectively raising people to favor CM. Other cultures might emphasize **entrepreneurial risk-taking** and glorify those who bounce back from failure, which could encourage people to sometimes violate CM (or at least redefine catastrophe as more recoverable). For example, in Silicon Valley culture, a failed startup is not seen as life-ruining (baseline is perhaps one's skill and network, which remain intact), so entrepreneurs might not practice CM financially (they often "risk it all" on a venture). Meanwhile, in more conservative cultures or those with less safety nets, people are taught to secure a stable job, save money – classic safety-first. These cultural differences could be explored via cross-cultural experiments on risk attitudes in scenarios with or without catastrophic downside. We predict that in cultures with strong communal support (e.g. extended families, welfare states), individuals might take more risks because the personal baseline isn't absolute (if they fail, family or society cushions them, so effective threshold of *personal* ruin is lower). By contrast, in individualistic "sink or swim" contexts with no safety net, people may be more inherently consequence-minimizing. **Developmentally**, as people age, risk tolerance often decreases – partially because wealth and responsibilities increase (so stakes of ruin

are higher), and possibly due to changes in brain chemistry and experience. Older adults might adhere to CM more (they have less time to recover from losses, so any big loss is more catastrophic). There's evidence older individuals are more loss averse and avoid high-stakes gambles more than younger folks, consistent with this idea. On the flip side, very young adults or adolescents sometimes show *inexplicable recklessness* (especially males, as per evolutionary theory focusing on mate competition, etc.). One could argue adolescents temporarily have a different baseline calculation (feeling "invincible" or simply heavily discounting worst-case outcomes due to incomplete development of the prefrontal cortex). So lifespan changes in decision-making can also be analyzed through the CM lens.

In summary, psychology and culture shape *what we consider catastrophic, how we feel about it, and when we are willing to risk it*. The CM principle is likely universal at a basic survival level (everyone, if actually facing life-or-death stakes, becomes risk-averse about that). But in the realm of abstract or social outcomes, there is variation. Understanding these variations is important for designing interventions: e.g., teaching someone from a chaotic background to adopt a more safety-first approach in financial decisions (perhaps by providing an artificial sense of security or actual safety nets), or conversely encouraging innovation by reducing the perceived catastrophic cost of failure (like bankruptcy laws that allow second chances).

Game Theory and Complexity Science Perspectives

In game theory, complexity science, and related fields, the idea of consequence minimization appears in strategies for adversarial or uncertain environments, and in how adaptive systems ensure persistence. Several concepts echo CM:

- **Minimax and Maximin Strategies:** In zero-sum games, the minimax strategy is the one that **maximizes a player's minimum guaranteed payoff**, effectively treating the opponent as an adversarial force and focusing on the worst-case outcome. This is a direct analogue to CM in a competitive context. For example, chess programs use minimax search to avoid moves that could lead to a certain loss (checkmate) even if those moves might offer chances for quick victory – unless the potential loss can be definitively averted by good play (which is complicated, but conceptually they're looking ahead for moves that minimize the opponent's best response). In decision theory without probabilities (the domain of Knightian uncertainty or complete ignorance), **maximin** is often recommended: choose the option with the best worst-case payoff. This is a very conservative criterion, often too conservative if worst-cases are extremely unlikely. Yet, if the worst case is something like death, one might argue maximin is rational (because no amount of good outcome can compensate a fatal outcome in a one-shot deal). For instance, if you face a choice of two medical treatments, one that has a 99% chance to cure you but 1% chance of death, and another that has 100% chance to partially heal (but not fully cure), a strict maximin might pick the second (no death risk) – that's CM in action. Many real-life decisions are not zero-sum and involve probabilities, so expected utility usually dominates. But in areas like **engineering safety, military strategy, or disaster planning**, minimax thinking is common. Planners will "harden" systems against worst-case scenarios (earthquakes, attacks) even if those are rare, which might mean not maximizing efficiency in average conditions. As an example, consider infrastructure design: applying CM might mean building a bridge that can withstand a 1000-year flood (so it never catastrophically fails), even if a cheaper design could handle 100-year floods and save money (maximizing profit). The latter might statistically be fine, but CM logic – especially if human lives are at stake – often prevails in safety regulations (reflecting an implicit high weight on catastrophic failure).

- **Bounded Rationality and Heuristics:** Real-world decision makers (humans, animals, even AI in complex environments) cannot compute all probabilities and utilities. They rely on heuristics, and one powerful heuristic is **"If an action can lead to an unacceptable outcome, don't do it"**. This is simpler than optimizing expected value with a million contingencies. Nobel laureate Herbert Simon and later Gerd Gigerenzer emphasized that many heuristics are "ecologically rational" – they work well in typical environments. CM could be viewed as a heuristic: it ignores a lot of detailed trade-offs and just eliminates any option that has a non-negligible chance of disaster. This can be extremely effective in environments where disasters are irrecoverable (which is most of evolution: e.g., "Don't eat a berry that smells like this because a small chance it's poison will kill you – better to miss out on some nutritious berries than to risk one lethal one"). The **"fast and frugal"** heuristics literature might categorize CM as a one-reason decision rule: if any outcome of option A is below threshold, discard A; otherwise choose among the rest by some simpler criterion (maybe then maximize gain). This sequential rule (first eliminate dangerous options, then pick best of safe ones) is akin to lexicographic preferences. It's computationally lighter than calculating expected values for all options because it uses a *filter*. Bounded rationality often leads to **satisficing**, where an acceptable threshold is set. As discussed, that threshold can be the safety baseline. So one might say: "I'll choose the first plan that guarantees at least survival probability 99%; beyond that I don't care if it's optimal." This might not maximize expected success, but it robustly avoids failure. In complex adaptive systems (like ecosystems, markets, or AI agents), those that systematically avoid ruin tend to survive longer and thus have more representation. This is related to the concept of **selection**: in evolutionary terms, if there's a distribution of strategies, the ones that occasionally go for broke but risk extinction might yield high short-term gains but will eventually be removed from the population when that rare catastrophe hits. The more conservative strategies might grow slower but persist, and over a long time horizon can dominate (because exponential growth favors those who never crash – connections to the Kelly criterion and geometric growth, which essentially says maximizing long-run growth requires avoiding ruin, even at the expense of short-run expected value). Complexity scientists talk about **resilience** – the ability of a system to absorb shocks without collapsing. CM in an agent or policy can increase system resilience: e.g., a bank that never takes on risk that could make it insolvent is less likely to fail in a crisis, contributing to overall financial stability (though perhaps at cost of lower profits in boom times). Concepts like **Value-at-Risk (VaR)** and **stress tests** in finance are formal ways to incorporate CM: regulators require banks to hold enough capital such that even in a worst X% scenario, they survive (i.e., probability of collapse < some threshold). This is basically mandating consequence minimization up to a certain statistical level.

- **Minimax Regret and Robust Decision Making:** Another approach in decision theory and AI is *robust optimization*, which often involves minimax or minimax-regret criteria to ensure reasonably good performance across a range of models or scenarios. For example, if you are unsure of probabilities, you might use a **maximin utility** or **minimax regret** to pick a decision that won't be terrible no matter what the probabilities turn out to be. This is sometimes called **robust satisficing** – aiming for an outcome that is "good enough" across all plausible situations rather than optimal in any one predicted situation. Philosopher Nick Bostrom has applied similar reasoning to existential risks: given the huge negative utility of human extinction, even very small probabilities of it should dominate our actions (leading to a CM approach globally: focus on preventing any one catastrophe that could wipe us out, like asteroid defense, nuclear war prevention, etc., before focusing on smaller issues). In machine learning, algorithms like adversarial training also use minimax (the model optimizes performance assuming an adversary tries to maximize error). These analogies

show that in many domains, **planning against worst-cases** is a widely adopted principle when stakes are high or uncertainty is great.

- **Adaptive Systems and Game-Theoretic Equilibria:** In evolutionary game theory, we can sometimes see CM emerging from adaptation. For instance, in predator-prey dynamics, prey animals evolve strategies that minimize the chance of predation, even at cost of lost feeding opportunities (this is an **evolutionarily stable strategy** if those who take extra feeding risks get eaten and thus do not pass on genes). The **"life-dinner principle"** in ecology states that a prey (running for its life) has more to lose than a predator (running for its dinner) – hence prey are often faster or more risk-averse than predators, because the consequence for the prey (death) is higher than the consequence for the predator (missed meal). This asymmetry drives evolutionary pressure for consequence minimization on the prey's side. In complex systems, one often finds **power-law distributions of damage** – a lot of small events and a few huge catastrophes. Agents that ignore the possibility of those huge events may perform well most of the time but then get wiped out. Systems theorists like Nassim Taleb (in "Black Swan" and "Antifragile") argue for strategies that are robust to black swans (rare, extreme events). Heuristics like *"barbell strategy"* (put most resources in ultra-safe assets, and a small portion in very risky bets – thus avoiding medium risk that could wipe you out but still allowing some upside) are aligned with CM thinking: the bulk of one's effort ensures survival, a small part is for gain. That is different from the naive mean-variance optimal that might put moderate risk everywhere. This highlights a complexity insight: sometimes **combining strategies** yields a CM outcome – e.g. a population might have some members take big risks (for potential big rewards) while others play safe, so the group as a whole is unlikely to go extinct (because at least the safe ones survive if things go badly, but if things go well, the risk-takers benefit the group). Social systems often implicitly do this (some individuals are entrepreneurs, others stick in stable jobs). So consequence minimization might operate at a group level too – ensuring the group's survival by diversification of strategies and by having norms or institutions (insurance, emergency reserves, etc.) that cap the downside for individuals.

In summary, game theory and complexity science reinforce the value of CM in uncertain, high-stakes environments. The formal models (minimax, safety-first, robust optimization) give us mathematical tools to design agents that embody CM. They also highlight the potential *cost* of CM: it can be overly conservative if misapplied (the classic criticism is that minimax is too pessimistic when worst-cases are extremely unlikely or when we can recover from them). Therefore, a refined CM framework might consider *degrees* of catastrophe (not all "bad outcomes" are equal; some might be absorbed with some cost but not fatal). We might then define multiple thresholds: e.g., a **critical threshold** (ruin) to never cross, and a **serious loss threshold** that we strongly try to avoid but might accept rarely. This leads to a layered risk management approach (like Tier 1 capital vs Tier 2 in banks, or "stop-loss" rules in investment where you'll tolerate up to X% loss then cut exposure). Complexity science also tells us to watch out for **tipping points** – thresholds in systems beyond which dynamics change qualitatively (often irreversibly). A CM agent should be aware of those in its environment (e.g., if global warming passes a tipping point, catastrophe accelerates – thus policies should aim to stay below that threshold). All these considerations broaden CM from an individual decision heuristic to a principle of systemic design: design systems and strategies that avoid tipping into absorbing bad states.

Evolutionary Biology and Ecology Perspectives

Evolution is fundamentally about survival and reproduction, so it is no surprise that the behaviors observed in the natural world often exemplify consequence minimization. Organisms that failed to avoid catastrophic outcomes (like predation, starvation, lethal injury) have been removed from the gene pool, whereas those with instincts and strategies to avert such outcomes have thrived. Here we survey how CM manifests in biology:

- **Risk-Sensitive Foraging:** As touched on earlier, animals adjust their foraging risk based on their energy budget relative to survival needs. This was first clearly demonstrated in studies with small birds (e.g., Caraco's classic experiments with juncos). At cold temperatures, when the fixed food amount might be insufficient to survive the night (negative energy budget), birds chose a risky feeding option (with variable food that could be a big payoff or zero) over a safe fixed-small portion. At warmer temperatures, when even the small portion was enough (positive budget), they chose the safe option and avoided risk. This is exactly CM: **if guaranteed survival is possible, don't risk it; if guaranteed survival is *not* possible, take the risk that at least gives a chance**. Stephens and others formalized this in the "**energy-budget rule**" and models with a **starvation threshold**. The model introduced a survival probability function – basically if the forager's intake > threshold, survival = 1, else 0 (or a sharp increase around threshold). This non-linear fitness function makes variance valuable in the deficit state and harmful in the surplus state. Many follow-up studies have explored risk-sensitive foraging across species (birds, fish, insects). While results can vary (some failed to replicate due to complexities, like how animals perceive variance), the *qualitative pattern* is well-supported and likely widespread: creatures near the edge of survival often exhibit "gambling" behaviors that those in comfort do not. Another related phenomenon is **caching or hoarding**: animals like squirrels gather and store food to reduce future starvation risk. This is a consequence-minimizing strategy – sacrificing current consumption to insure against lean times. Evolution has favored such behaviors because they mitigate the catastrophic scenario (winter with no food). Insects entering diapause (hibernation) will accumulate a threshold of fat; if they haven't, they'll take risks (extend foraging) rather than go dormant undersupplied. These are instinctual implementations of CM.

- **Predator-Prey Dynamics:** As mentioned, prey species live under constant threat of predation, and so their behaviors often prioritize avoiding that *worst-case outcome* (being killed) even at significant cost to other fitness components (like feeding or mating). Field observations and experiments show prey reducing activity in presence of predators, choosing safer (but less food-rich) habitats, and developing sentinel systems (in group-living animals) to detect predators. Lima and Dill's 1990 review famously described how animals **trade off foraging gain against predation risk**. For example, small fish might forage in deeper water where food is abundant, but if predatory fish lurk there, the small fish stay near the shallow refuge, accepting poorer feeding to not get eaten. The *"ecology of fear"* concept denotes that predators influence prey behavior far beyond actual kills – the fear (risk) is enough to change prey distribution and feeding times. Over evolutionary time, many prey adaptations are clearly CM-driven: **camouflage, keen threat detection senses, erratic escape maneuvers, protective morphologies (spines, shells)** – these all minimize the probability of the catastrophic event (predation) at various costs. "Life-dinner principle" (coined by Dawkins & Krebs) implies prey are under stronger pressure to not lose (life) than predators are to win (meal). So prey usually err on side of caution. Some prey behaviors are extreme in this sense: e.g., certain gazelles stot (jump high) when they see a predator – interpreted as a signal "I see you and I'm fit, don't bother

chasing me." They are essentially trying to avoid a chase altogether (since any chase has some probability of fatal outcome). By convincing the predator it won't succeed, they preempt the risk. Predators themselves need to manage energy, but since missing a meal is not instantly catastrophic (usually), they can be less risk-averse. However, if a predator is starving (below its own survival baseline), it may take very risky hunts (even attack larger dangerous prey – akin to desperation). So the pattern repeats on that side too depending on state.

- **Survival Intelligence and Learning:** Animals also learn from near-misses. If a monkey barely escapes a falling branch or a predator, it might develop a lasting avoidance of that location or situation (one-shot learning of a potential catastrophe). This is arguably a CM learning rule: heavily update strategy to avoid scenarios that *could have* been fatal. Such one-trial learning is often mediated by the amygdala in the brain (again showing biology's emphasis on not needing multiple experiences with disaster – once is enough to adjust permanently). This is beneficial because a creature might not survive a second encounter if it didn't change behavior after the first. This can lead to seemingly irrational fears (an animal overgeneralizing a single bad experience), but from an evolutionary view it's better to be safe than sorry (error management again). The concept of **"survival processing advantage"** in memory research shows that humans remember information better if they think about it in a survival scenario context – our cognition is tuned to pay special attention to survival-relevant details (where water is, what threats are around). That implies that our brain implicitly ranks "info that helps avoid death" as higher priority to encode and recall.

- **Evolutionary Trade-offs and Early-life Effects:** We touched on life history theory. To expand: organisms facing high extrinsic mortality (death from environment independent of their behavior, like unpredictable droughts or heavy predation) often evolve a strategy of *"live fast, reproduce early"* because playing it too safe doesn't pay if you likely die young anyway. This might seem like the opposite of CM (they take risks with the future). However, it's a rational genetic strategy given the baseline is that life will be short; their threshold is effectively set by the environment (if 90% die by age 1, then avoiding all risk won't change that much because risk is external). So, they allocate energy to reproduction rather than body repair (leading to shorter lifespans, etc.). On the other hand, species in safer niches (lower extrinsic mortality) evolve slower, more cautious strategies (long juvenile development, more investment in immune function and repair – basically "avoid dying because you have time to reap rewards of caution"). Humans are interesting – historically we had moderately high extrinsic risks, but also many cultural ways to mitigate them, and we show flexibility. Early-life cues (predictability of food, violence exposure) seem to calibrate our strategy: *If as a child you experience unpredictable trauma or resource uncertainty, your psychology shifts toward present-focused, risk-tolerant behavior; if you experience stability, you become more future-oriented and risk-averse*. This is an adaptive calibration: in a harsh world, waiting or being cautious might mean you miss your chance (or someone else takes the resource), so better to take opportunities (and risks) now. In a benign world, reckless moves could squander a long, compounding future, so better to be patient and careful. This demonstrates that **the baseline for what is considered a "catastrophe" can be environmentally determined**. In a dangerous world, *not eating today* might be catastrophic because tomorrow isn't guaranteed, so behavior becomes more opportunistic (less CM in the short term, arguably). In a stable world, *not preparing for retirement* could be catastrophic (you assume you'll live long and need resources), so people plan and avoid short-term risk that would jeopardize long-term security. This might help explain cultural differences as well (e.g., populations with history of instability might culturally emphasize living in the moment, whereas those from stable environments emphasize planning and safety).

- **Inclusive Fitness and Group Safety:** Evolution also works at the level of genes which can be shared by groups (kin selection) or even whole populations (group selection debates aside). Sometimes behaviors that look individually suboptimal are about preventing group-level catastrophe. For instance, some social insects, like honeybees, have **suicidal stinging** – a bee will sting an intruder, killing itself, to protect the hive. At the individual level, death is catastrophic, but at the gene level, saving the colony (which contains the bee's relatives) is higher priority. So one could say the bee is following a CM strategy for the colony: avoid colony destruction at cost of individual lives. In humans, we see heroism in war or disasters – risking or sacrificing one's life to save others or the community. This might be explained by kin selection, reciprocal altruism, or just culture, but it's interesting as a foil: sometimes maximizing survival of your genes means *accepting* personal catastrophe. So context matters for what "outcome" we're minimizing – self vs kin vs offspring. Generally, though, organisms evolved to preserve their direct survival strongly because if you die young you can't reproduce. Only under special circumstances do genes favor self-sacrifice (if it significantly increases relatives' survival).

To sum up, evolutionary biology provides numerous examples that mirror CM and help validate its logic. It also cautions that *if the environment changes*, a fixed CM strategy might become too conservative or too aggressive. Evolution tends to produce conditional strategies (phenotypic plasticity) – exactly what CM entails with its state-dependent flip. The natural world essentially **"tested"** strategies over eons, and the prevalence of threshold-based behavior indicates it's a winning approach when facing non-linear fitness landscapes (life or death situations). A key falsifiable point here is: if consequence minimization is fundamental, we should observe that for any critical resource or risk, organisms will have some evolved mechanism to avoid absolute depletion of that resource. Indeed, many animals stop reproduction or growth when resources are scarce, diverting energy to maintenance just to survive the bad period (avoid death, even if it means not having offspring that year – a reproductive gamble but survival-first). In contrast, organisms that do not have such restraint can crash (some species do, like certain insects will reproduce explosively and then crash – usually in environments where predation is low but resource boom-bust happens). Those strategies can work if synchronized with environment (e.g. salmon all spawn and die – but their offspring survive because predators are overwhelmed). Again, environment dictates if CM is optimal or some alternative. The principle to glean is: in *stable or moderately unpredictable environments*, evolution leans toward CM; in extremely unpredictable or one-shot environments, evolution might roll the dice in a controlled way.

Experimental Paradigms and Falsifiability

To robustly test the principle of consequence minimization and distinguish it from other decision strategies, we propose several experimental paradigms along with predicted outcomes (falsifiable predictions):

1. **Termination Task Paradigm:** Design a decision task where participants accumulate rewards over trials, but a single "catastrophic" mistake ends the game and zeros out earnings (or ends the session). For example, on each trial a participant can choose a safe option (small sure gain) or a risky option (higher gain but some probability of "game over"). Crucially, include states where the participant's accumulated earnings are above a target vs states where they are below a target they hope to achieve. **Predictions:** Under CM, participants will be **risk-averse when ahead** (above target) – likely sticking to safe choices to lock in their baseline – but **risk-seeking when behind** (below target, especially as trials run out) – willing to take the risky option because otherwise they can't reach the target. This would manifest as a dynamic change in risk preference as a function of current

earnings vs desired threshold. If participants were simply loss-averse or regret-averse, one might not see such a clean flip; for instance, loss aversion would predict aversion to the risky option's loss component even when behind (unless the reference point shifted). By manipulating whether the "game over" truly wipes out earnings or just stops further gains, we can see if it's the catastrophic aspect (complete termination) driving behavior or just the loss amount. **Falsifiability:** If CM is a valid model, when a single mistake ends all earnings, we should see extreme caution after a certain point of success – more than in a control condition where mistakes just cost some points but you continue. Conversely, if someone has almost no earnings and only a few trials left, CM predicts a high proportion of risk-taking (because safe choices yield a sure low total which is "failure" relative to baseline). If experiments show people *not* doing this – e.g., they remain risk-averse even when clearly below an attainable goal, or they take wild risks even when doing well – then the CM model would be challenged. Some prior studies in economics with similar setups (sometimes called "aspiration levels" or "target earning" tasks) have indeed found target-based shifts in risk preference, but we'd refine it by adding the *absorbing failure* condition.

2. **Stress Induction and CM:** Conduct an experiment where participants perform a risk-reward task under different conditions: a high-stress condition (e.g. threat of electric shock, time pressure, noise, social evaluative threat) vs a low-stress control. We measure changes in the point at which they switch from risk-taking to risk-avoidance as their state changes. For instance, use a task with a gradually changing reference point: participants start at zero, can gamble to increase a score, but if score ever drops below –100 (a catastrophic debt), something aversive happens (or they lose a bonus). See how stress modulates strategy. **Predictions:** If stress amplifies CM tendencies, stressed participants should show a **greater emphasis on avoiding the catastrophic boundary** – they may gamble less frequently if any gamble could push them into "debt." However, following the earlier reasoning, acute stress might also raise the subjective baseline (people might feel more is at stake), causing mixed effects: some might actually become *more* risk-averse under stress (to avoid mistakes), which is straightforward CM above baseline; others might become *more* risk-seeking if they perceive themselves in a dire situation due to stress (like a "now or never" mindset – though this is more likely if stress is paired with being in a losing position). This paradigm might require careful design to interpret, but one marker could be physiological measures (cortisol, heart rate) as covariates. **Falsifiability:** If CM is truly a distinct process, we'd see interactions: for example, stress might not just uniformly scale risk aversion up or down, but rather sharpen the state-dependent flip. Perhaps stressed individuals stick even more tightly to safe choices when doing well, and flip to risk-taking more abruptly when doing poorly (a kind of all-or-nothing strategy). Alternatively, if stress mainly operates through other channels (like impaired calculation or amplified loss aversion in general), we might see different patterns (like uniformly more caution or uniformly erratic behavior). By including conditions or questionnaires to measure *regret sensitivity* and *loss aversion* separately, one could attempt to dissociate those from pure CM. For example, include a scenario where one option has high outcome variance but no chance of catastrophe (e.g. large gain or moderate loss that isn't fatal) and another with a small chance of a huge catastrophe but better average – a stressed CM-focused person should avoid the latter at all costs, while a purely loss-averse person might treat them more on par in terms of variance.

3. **Regret vs Catastrophe Dissociation:** Construct choices where the worst-case outcome either is framed as a *foregone better alternative* (high regret) or an *absolute loss* (catastrophe) while keeping probabilities the same. For example, two conditions: In one, choosing Option A over B could lead to a outcome that is a bit lower than B would have given (high regret if that happens, but not

catastrophic), whereas in another, choosing A could lead to the same numerical outcome but framed as "a critical failure" (perhaps triggering a penalty or end of task). By measuring choices, we see whether people avoid A more in the catastrophic framing than in the regret framing. **Predictions:** If objective catastrophic risk outweighs regret, participants will be more averse to A in the catastrophic frame than in the regret frame, even if in the regret frame B would clearly outperform A in some cases. Conversely, if regret aversion dominates, they might avoid A similarly in both frames (as long as they know B would have been better in hindsight). Another approach: have a condition where a potential loss is either one they'll find out about (causing regret if they chose wrong) or one they'll never know (so no regret, just an objective loss). If CM is driving behavior, the knowledge of the outcome shouldn't matter – only the presence of the loss. If regret is driving, then choices will differ depending on expected feedback. Prior research (Zeelenberg et al.) indeed shows feedback expectations alter risk choices. So we can use that to factor out the emotional component. **Falsifiability:** If we find that even when no feedback about the forgone option is given (so regret is minimized), people still avoid options with a small probability of a large loss far more than options with a higher probability of a smaller loss (with equal EV), that indicates a special concern for catastrophic outcomes beyond just regret or variance. If, however, removing regret (feedback) significantly increases people's willingness to take low-probability-big-loss gambles, it suggests that emotional factors might be more at play than a lexicographic safety rule.

4. **Developmental and Cross-Cultural Tests:** To see if CM baselines are learned, one could compare populations. For example, take two groups of participants: one from a volatile background (unstable income or high childhood adversity) and one from a stable background. Have them perform a task with a clear baseline goal and risks. **Predictions:** The volatile-background group might (a) set lower baselines or give up on reaching high safety thresholds (thus appearing more risk-tolerant in loss domain because they consider failure almost inevitable or not as shocking), or (b) alternatively, they might be more sensitive to any chance of immediate ruin because they've experienced it (this could go either way depending on personal strategies). We might also see cultural differences: in a collectivist society, participants might take a bit more risk personally, perhaps subconsciously expecting social backup (thus not treating personal ruin as total ruin), whereas individualists might be more careful since they're on their own. Children could be tested in a simplified CM scenario (like a game where if they pull too hard on a candy dispenser it breaks and they get nothing vs gently get some candy) to see at what age they understand "don't risk everything for more candy." We might find that younger kids focus on immediate reward (risk-seeking even when they shouldn't), but as they age into later childhood, they learn the concept of a "bird in hand" baseline. If CM tendencies strengthen with age and experience, that suggests a learned or rational component; if even very young kids avoid catastrophic outcomes (maybe by parental instruction or innate fear), that suggests an innate predisposition.

5. **Neuroscientific Experiment:** Using fMRI or EEG, one could replicate the termination task and look for neural correlates when people shift from risk-taking to risk-avoidance or vice versa around the baseline. **Predictions:** If the amygdala/insular "survival circuit" is critical, we'd expect increased activity in these regions when participants are above baseline and considering a risky option (the idea of a catastrophic loss might trigger a strong aversive signal), whereas when below baseline, perhaps the nucleus accumbens/striatum (reward circuit) might be relatively more engaged because the focus shifts to potential gain (since not acting has negative certainty). One might also see changes in ACC signals: above baseline, risky choices might produce a spike in ACC (error likelihood warning) that drives people to choose safe, whereas below baseline that ACC alarm might actually

quiet down (since the error of *not* trying might be encoded). These are speculative, but measurable. A key falsifiable outcome would be: does the brain treat a potential loss differently if it crosses a perceived critical threshold? E.g., losing $50 might be not a big deal if you have $1000 (baseline not threatened) but huge if you have $50 (zero left = "ruin"). We can directly compare neural response to a possible $50 loss in those contexts. A pure expected-utility framework would mostly care about proportions, whereas a CM framework would show qualitatively different activation (fear circuitry firing when $50 loss means bankruptcy vs not when it just means a dent). Some studies indirectly support this: one found insula response to monetary loss predicted safe choices subsequently, *especially* in people for whom that loss carried high personal significance [1].

Concrete Test Designs: To ground some of the above in a specific example, consider this experimental design:

- Participants play 20 rounds of a gambling task. They start with a stake of 100 points. In each round, they choose either a sure +5 points or a gamble that gives +15 points with 90% chance and **-50 points with 10%** chance. Critically, if their point total ever falls to 0 or below, the game ends and they lose all accumulated points (catastrophe). Tell participants that accumulating at least 150 points is the "target" for a good bonus (baseline goal), but anything above 0 is still kept if the game finishes normally. Track choices across rounds and across different point totals. **Hypothesis:** Early on, if a participant's points drop near 50 (danger zone), we expect them to either start gambling a lot (if they're below a feasible trajectory to 150 with only safe choices) or, if they manage to get to say 130 points by mid-game, to possibly switch entirely to safe choices to guarantee surpassing 150. If many participants follow this pattern (risk when behind, caution when ahead), that supports CM. If they don't modulate and either keep gambling or keep playing safe regardless, it would be inconsistent with CM. One can introduce a control group where the game doesn't end at 0 (they just go negative and continue, perhaps with a debt, but still could climb back). CM theory would predict much more gambling in that control (since a -50 doesn't permanently end the game, some might risk going into debt to later climb out). If we see that difference – the presence of an absorbing failure state changes strategy significantly – it confirms people are specifically averse to *absorbing states*, not just losses per se.

Another example: **"Cliff-edge" experiment** – gradually increase the potential loss of a gamble while adjusting the potential gain to keep expected value the same, and see at what point each participant opts out. If CM-driven, there may be a sharp threshold where as soon as the loss could push them below a critical level (like wiping their current bankroll), they stop, regardless of EV. Loss aversion would also cause stopping, but more gradually based on loss amount; regret might depend on if they know what could have happened; risk aversion (curvature) would also be gradual. A sharp, almost binary change when a loss becomes "ruinous" would support threshold-based modeling.

Mapping to Falsifiability: Each of these paradigms provides potential outcomes that could contradict CM. For instance, if in the termination task people don't become particularly cautious even when they have a lot to lose (they gamble away a secure win more often than predicted), then CM alone isn't explaining behavior (maybe thrill-seeking or probability neglect is at play). Or if under stress people do not alter their safety-first behavior at all, then stress might not interact with CM as hypothesized (or our stress manipulation might not simulate the right kind of perceived threat). The key is that CM is a specific *conditional* strategy – so data should show *interaction effects* (state x choice pattern interactions). If only main effects of, say, loss amount or stress are seen, simpler theories might suffice.

Finally, it's important to note that **real-world validation** is also possible. We could examine archival data: e.g., do companies on the brink of bankruptcy make more volatile strategic moves than those in solid financial shape? (Case studies suggest yes – e.g. struggling firms often take on risky projects as a Hail Mary, whereas cash-rich firms are often more conservative.) Do people with terminal illness (nothing to lose health-wise) take more financial risks or try unproven treatments? (Anecdotally yes, many will try experimental cures – risk-seeking – because the alternative is certain death.) These observations, while not controlled experiments, strengthen the case that consequence minimization (or its mirror, nothing-to-lose risk-seeking) is a real phenomenon across domains.

Conclusion

In this report, we presented a formal modeling framework for **Consequence Minimization (CM)** – an agent strategy that lexicographically prioritizes avoiding catastrophic outcomes over maximizing gains – and reviewed evidence and theory across disciplines that underpin this principle. The formal model captures *state-dependent risk attitude*, with a baseline threshold demarcating a switch from **risk-averse behavior in safe zones** to **risk-seeking behavior in danger zones**. We compared CM to related decision strategies: while sharing features with loss aversion, regret minimization, and maximin approaches, CM is distinguishable by its focus on objective survival thresholds and lexicographic ordering of priorities (safety before profit).

The literature review found convergent support for the CM concept: Behavioral economists document reference-point flips in risk preference and emphasize "safety first" heuristics in portfolio selection. Cognitive neuroscience reveals dedicated "survival circuits" (amygdala-insula-ACC) that enforce avoidance of threats [1], essentially implementing CM at a neural level by inhibiting risky actions that carry serious danger. Psychology shows that under stress or when basic needs are threatened, human decision-making shifts toward securing safety (Maslow's hierarchy), and that our emotions (fear, regret) often serve as internal guides to prevent disastrous choices. Game theory and complexity science formalize the value of worst-case optimization and robust strategies in uncertain environments, paralleling CM's logic in algorithms and systemic design. Evolutionary biology and ecology demonstrate through countless adaptations – from risk-sensitive foraging to predator avoidance – that avoiding "game over" events is a dominant evolutionary driver.

We also outlined how future research can test CM rigorously: through behavioral experiments like termination tasks (one mistake = ruin) to observe threshold-triggered risk aversion/risk-seeking flips, stress-modulated decision trials to see if survival concerns trump other factors, and developmental or cross-cultural studies to examine how baselines are learned or vary. Neuroimaging could further validate if distinct neural signatures accompany CM-type decisions (e.g. heightened insula/ACC activation at the prospect of breaching a critical threshold). These studies would solidify whether CM is a fundamental decision principle or an epiphenomenon of other processes.

Practically, understanding CM has important implications. It can inform better design of incentives and information: for instance, communicating risks in terms of potential catastrophic outcomes may engage people's CM mindset and promote caution where needed (e.g., climate change communication focusing on worst-case scenarios to overcome complacency). In finance and public policy, recognizing when agents will flip to risk-seeking (e.g. a "gamble for resurrection" by a failing bank or a regime in decline) could lead to preemptive measures to stabilize them before they reach that desperation point. In personal decision-making, being aware of one's baseline and how it shifts under stress could improve choices – for example,

not panicking into a big gamble when one's chips are down, or conversely, not becoming too timid when one can afford to take opportunities.

In conclusion, the principle of consequence minimization offers a unifying framework to understand behaviors across scales: from neurons firing in fear of a snake, to a parent sacrificing for their child's safety, to a company hedging against rare disasters. It reminds us that **not all outcomes are created equal** – some carry a weight that defies simple expected-value logic. By formally modeling and empirically probing CM, we gain insight into the profound influence of potential catastrophes on decision-making. The evidence strongly suggests that both humans and other organisms have evolved (and learned) to follow a guarded maxim: *"First, survive – then, seek reward."* This maxim, encoded in our brains, behaviors, and perhaps even societal norms, captures the essence of consequence minimization.

References (Annotated):

- Mulligan, K. et al. (2023). *Risk Preferences Over Health: Empirical Estimates.* – Found that individuals' risk tolerance depends on health state: **risk-seeking at low health, switching to risk-averse at a mid-health threshold**, and very risk-averse at perfect health. Supports state-dependent CM behavior in a health context.

- Bateson, M. (2002). *Recent advances in our understanding of risk-sensitive foraging.* – Reviews risk-sensitive foraging in animals. Notably, **birds on a negative energy budget choose high-variance food options (risk-prone) because the safe option means certain death, whereas birds on a positive budget choose low-variance (risk-averse)**. Introduces the energy-budget rule with a survival threshold.

- Roy, A.D. (1952). *Safety First and the Holding of Assets.* – Pioneering work proposing the **safety-first criterion**: select investment that minimizes the probability of disaster (returns falling below a threshold). This formal lexicographic approach in finance mirrors CM's priority structure.

- Wikipedia: *Minimax* – Definition of minimax decision rule: **minimize possible loss for worst-case scenario; maximize minimum gain**. Relates to CM's worst-case focus (maximin). Also mentions extension to general decision-making under uncertainty.

- Fiveable Library (Class Notes): *Minimax decision rules* – Explains minimax regret: **minimize the worst-case regret**. Useful for contrasting regret vs absolute outcomes; notes that regret-based choices can differ from utility-based ones.

- Lopes, L. (1987). *SP/A Theory (Security-Potential/Aspiration).* – Introduced a dual-criterion model where decisions consider both a security level (safety) and an aspiration level. Noted that **aspiration acts as a second criterion beyond expected utility**. This is conceptually aligned with CM: a security threshold (don't go below) combined with potential (achieve above threshold).

- De Martino, B. et al. (2010). *Amygdala damage eliminates loss aversion.* (PNAS) – Showed that patients with amygdala lesions **no longer exhibit loss aversion and take risks that controls wouldn't**, suggesting the amygdala generates the avoidance of potentially harmful outcomes. Implies a neural basis for CM in loss contexts.

- Li, Y. et al. (2024). *Role of loss aversion in social conformity.* – (As cited in [26]) Noted that **anterior insula is more activated under risky choices and is associated with loss aversion**. Also that higher insula activation links to choosing non-punished (safe) outcomes [1], indicating insula's role in avoidance learning (consistent with CM).

- University of Arkansas News (2025). *People make riskier choices when stressed.* – Reports a study where acute stress led to **riskier decisions via decreased loss aversion**. Suggests evolutionary rationale: in dire situations, risk-taking might increase (an adaptive shift – which we interpret through CM lens: if stressed = environment signals threat, maybe trigger "nothing to lose" behavior).

- Wikipedia: *Maslow's hierarchy of needs* – States that once basic physiological needs are met, **safety needs take precedence and dominate behavior**. Aligns with CM: safety (avoid catastrophe) is prioritized before higher goals.

- Lima, S. & Dill, L. (1990). *Behavioral decisions made under the risk of predation.* – Highlights how animals often **forage less optimally in return for safety**. The need to avoid predators (catastrophic outcome) constrains their behavior, exemplifying CM in ecology.

- Haselton, M. & Buss, D. (2000). *Error Management Theory.* – (Referenced via Wikipedia) Proposes that cognitive biases exist because **making the less costly error was favored by evolution**. E.g., false alarms over misses. This supports the prevalence of "better safe than sorry" heuristics – essentially a rationale for CM biases.

- Wu, J. & Wu, Q. (2024). *Early-life stress and risk-taking (Frontiers).* – Found that adolescents from harsh, unpredictable early-life environments adopt **faster life-history strategies with more risk-taking and shorter future orientation**. Shows developmental calibration of risk preferences, relevant to how baseline and CM tendencies can be shaped by environment.

- Zeelenberg, M. (1996). *Consequences of Regret Aversion on Decision Making.* – Demonstrated that **anticipated regret can lead to both risk-averse and risk-seeking choices**; specifically, people make **regret-minimizing choices which might be either the safest or the riskiest option** depending on which would reduce potential regret. This underscores that regret aversion is not identical to always avoiding variance, highlighting the need to distinguish regret from pure catastrophic outcome avoidance.

---

[1] Increased Activation in the Right Insula During Risk-Taking Decision Making Is Related to Harm Avoidance and Neuroticism | Request PDF
https://www.researchgate.net/publication/10588342_Increased_Activation_in_the_Right_Insula_During_Risk-Taking_Decision_Making_Is_Related_to_Harm_Avoidance_and_Neuroticism