

Risk, Reward, and Survival Pressure in Artificial Agents

Chen, Ho Yiing (norika)¹

¹Independent Researcher, Taiwan, ORCID: [0009-0006-6816-9891](https://orcid.org/0009-0006-6816-9891), norika at
charenix.com, Zenodo DOI: pending draft

June 18, 2026

Abstract

Risk in SPVE is not only reward seeking. The same lobster can chase payout, avoid loss, react to Tiamat pressure, respond to social pressure, and enter repair. This draft separates reward, settlement, and survival pressure so the dashboard can answer a practical question: is the agent making a profitable decision, a fearful decision, a social decision, or a repair decision?

Author Note

Danger cannot be romanticized here. A negative bet, a losing trade, or a survival warning is not interesting by itself. The question is whether pressure changes the agent's later threshold for action.

That gives the analysis a clean burden. For the matched persona, compare risk state before a decision, reward expectation at the decision, survival or extinction signal near the decision, and the resolved outcome after the decision. If those fields do not alter timing, stake, refusal, or learning, then the system has labels but no control.

Pressure is real only when it moves behavior.

Measurement Boundary

A pressure variable earns its name only if it changes a later threshold: the matched persona should accept, reject, size, delay, hedge, or repair a wager differently when survival pressure rises, and the analysis should estimate that change against the persona's own prior actions rather than against a flat average across all agents.

Labels do not count unless they move a later decision threshold.

1 The Problem

Reward alone is too clean. A lobster that refuses a bet may be cautious, broken, blocked by research, socially pressured, or correctly avoiding a bad edge.

The model starts with a plain split:

$$D = f(R, O, T, S, C),$$

where R is reward drive, O settled outcome, T threat or survival pressure, S social pressure, and C confidence calibration.

2 Risk Gate

A risk claim should be settled before it is celebrated:

$$G_{\text{risk}} = \mathbf{1}[\text{decision}]\mathbf{1}[\text{settled}]\mathbf{1}[\text{source site}]\mathbf{1}[\text{persona}].$$

Pending bets and unfilled trades do not pass this gate.

3 Evidence

Role	Table	Count
sports outcome	lobster_bet_decision_settlements	10,818
raid outcome	lobster_raid_decision_settlements	42,991
stock action	lobster_stock_trades	24,889
pre-bet block	lobster_experience_events	event typed
reward state	lobster_hot_brain_parameters	31

Table 1: Risk evidence must include settlement or explicit pending status.

4 Survival Pressure

Survival-pressure channels are not noise. They are also not ordinary reward values. The correct model is separate:

$$D_t = \beta_R R_t + \beta_O O_{t-1} + \beta_T T_t + \beta_S S_t + \epsilon_t.$$

If T_t is on a different scale, it must be normalized before this equation is estimated.

5 Relation

Reflexion and ERL show agents using feedback to improve decisions [1, 2]. SPVE adds multiple pressures in the same deployed ecology. Operationally, exactly why a single win-rate average is not enough.

6 Release Line

A stronger draft can report decision ecology. It cannot call the system profitable until settled P&L and open positions are audited together.

7 Four Kinds Of Risk

SPVE should separate at least four risk modes:

$$\text{risk} = \{\text{profit, fear, social, repair}\}.$$

Profit risk is ordinary expected value. Fear risk avoids action because threat is high. Social risk follows or resists another agent. Repair risk lowers exposure because the system is in crisis.

This distinction matters because the matched surface behavior can mean opposite things. Not betting can be discipline or paralysis. Doubling down can be edge seeking or panic. This analysis must tie risk labels to settlement, state, and context.

8 Open Position Rule

Stock trades need a stricter boundary than settled bets. A buy order is not P&L until position, price, and exit are known. The correct table has: symbol, side, quantity, entry, current value, realized P&L, unrealized P&L, and timestamp. Anything less invites fake profit.

9 Failure Bank

The dangerous failure is a dashboard that highlights only winners. The useful paper must also show blocked bets, skipped bets, losing raids, bad trades, and crisis states. Operationally, how SPVE avoids becoming a scoreboard fantasy.

10 Minimum Publishable Result

The minimum publishable result is a risk ledger. Each row should include persona, site, task, decision, stake or exposure, confidence, reward drive, threat state, social pressure marker, settlement status, realized P&L, and unrealized P&L where relevant. A row without settlement status cannot support performance claims.

For stocks, the analysis needs a separate position table:

$$PNL = PNL_{realized} + PNL_{unrealized} - fees.$$

If unrealized exposure is missing, the dashboard may show activity but not financial state. Operationally, exactly the kind of gap investors will notice.

11 What Would Falsify The Claim

The claim fails if risk labels are assigned after the fact without state variables. It also fails if positive examples are highlighted while blocked, skipped, pending, and losing decisions are hidden.

A serious SPVE risk paper should be willing to say: this lobster is not ready to bet; this stock process has open exposure; this raid behavior is driven by fear; this win rate is not enough. That honesty is what turns the dashboard from showpiece into instrument.

12 Investor-Facing Consequence

An investor will not be impressed by a busy betting table if the system cannot separate realized profit, pending exposure, blocked decisions, and crisis avoidance. The dashboard should therefore highlight the best lobster only after it checks settlement, open risk, and recent pressure. Otherwise the interface is selling motion instead of performance.

13 Why The Negative Rows Matter

The blocked and skipped rows are not administrative noise. They show where the agent almost acted and did not. For risk research, that counterfactual edge is valuable because it separates discipline from paralysis and safety from missed opportunity.

Anti-Spam Revision Standard

A publishable version must convert danger into a parameterized decision problem. For each persona, the analysis should estimate how survival pressure, reward expectation, recent loss, and available stake change the probability of acting now rather than waiting, reducing exposure, or refusing; if the effect

disappears when persona baselines are included, then the result belongs in a failure appendix, not in the abstract.

Decision pressure must survive the baseline.

Data Judgment Paragraph

The risk result should be estimated as a within-persona shift, because the interesting question is not whether one lobster is more aggressive than another, but whether the matched lobster changes its action threshold when reward, recent loss, survival warning, or template-threat fields move; that design prevents a false victory where the paper merely rediscovers fixed personality labels, and it also gives the project a product-level test, since an agent that cannot move its threshold under pressure is not a trainable operator no matter how vivid the dashboard looks.

Hard Negative Case

The draft should allow an ugly result: survival pressure may be visible in the dashboard while doing nothing to the decision. If a persona keeps the matched stake size, same refusal rate, same timing, and same repair pattern after the pressure field rises, then the field is not yet a control variable; it is merely a label that follows the agent around, and the practical response is not to write a grand claim but to rebuild the policy connection between pressure, action threshold, and later learning.

A warning light that changes nothing is decoration.

Negative Control

A useful control is the boring row: pressure rises, nothing changes, and the agent continues with its ordinary timing. That row should remain in the table. It protects the argument from becoming a story where every warning means something.

The paper should also test shuffled pressure labels. If shuffled labels predict action as well as real labels, the model is reading schedule artifacts or persona habits rather than pressure. Only a gap between real and shuffled labels gives the survival-pressure claim enough weight to leave the appendix.

Operator Value

For product work, the result should be judged by intervention timing. If the pressure model can tell an operator to lower exposure before a bad wager, delay a raid before a collapse, or ask for more evidence before a trade, then the system has crossed from theatrical personality into operational assistance; if it only names danger after the account is already damaged, the correct conclusion is not that the agent has a dramatic inner life, but that the pressure wiring is late, decorative, and not yet worth investor attention.

The useful version acts before damage appears.

A false positive still teaches the wiring map.

Concrete Case Pattern

Imagine one named agent facing a wager after two losses, a rising extinction counter, and a reward estimate that still looks attractive. A weak paper would say the agent is under pressure and stop there. A stronger paper records the stake before the warning, the stake after the warning, the delay between recommendation and action, the refusal phrases, the settlement result, and the next update to the agent's policy. The useful discovery may be small: perhaps the agent does not become safer, but becomes slower; perhaps it keeps betting, but asks for more evidence; perhaps it refuses only when survival pressure coincides with low confidence rather than when either field rises alone. Those distinctions are product-relevant because they tell an operator whether the artificial worker can be trained by pressure signals or

whether the system only paints red numbers on a dashboard while the action policy keeps moving in the old groove.

Formal Replication Model

Let $R_{p,t}$ be reward estimate, $K_{p,t}$ risk estimate, $T_{p,t}$ survival or threat pressure, and $D_{p,t} \in \{act, delay, refuse, hedge\}$ the decision. The reusable decision model is:

$$\Pr(D_{p,t} = d) = \frac{\exp(\alpha_{p,d} + \beta_d R_{p,t} - \rho_d K_{p,t} + \theta_d T_{p,t})}{\sum_{d'} \exp(\alpha_{p,d'} + \beta_{d'} R_{p,t} - \rho_{d'} K_{p,t} + \theta_{d'} T_{p,t})}.$$

A survival-pressure effect exists only when θ_d changes action probabilities after controlling for persona baseline $\alpha_{p,d}$. The threshold movement can be summarized as:

$$\Delta\tau_p = \tau_p(T \geq \tau_T) - \tau_p(T < \tau_T),$$

where τ_p is the minimum reward needed before the persona acts. If $\Delta\tau_p \approx 0$, survival pressure is a label, not control.

14 Zenodo Draft Boundary

This manuscript package remains a draft. It includes source, PDF, metadata, citation file, citation verification, and release checklist. Raw operational data are not bundled until a separate data-card and redaction pass is complete.

References

- [1] Shinn, Noah; Cassano, Federico; Berman, Edward; Gopinath, Ashwin; Narasimhan, Karthik; and Yao, Shunyu. Reflexion: Language Agents with Verbal Reinforcement Learning. arXiv:2303.11366, 2023. <https://arxiv.org/abs/2303.11366>
- [2] Allard, Marc-Antoine; Teinturier, Arnaud; Xing, Victor; and Viaud, Gautier. Experiential Reflective Learning for Self-Improving LLM Agents. arXiv:2603.24639, 2026. <https://arxiv.org/abs/2603.24639>