From Safety Paradoxes to Safety-II: A Dialectical Method

Safety paradoxes are situations where efforts to increase safety can unintentionally increase risk. For instance, safer equipment may encourage riskier behavior, while increased automation can reduce system resilience. This paper shows how such effects can be systematically transformed into actionable Safety-II mechanisms using a dialectical framework. The underlying structure of these paradoxes can be expressed using the Eye Opener approach (Fig. 1).

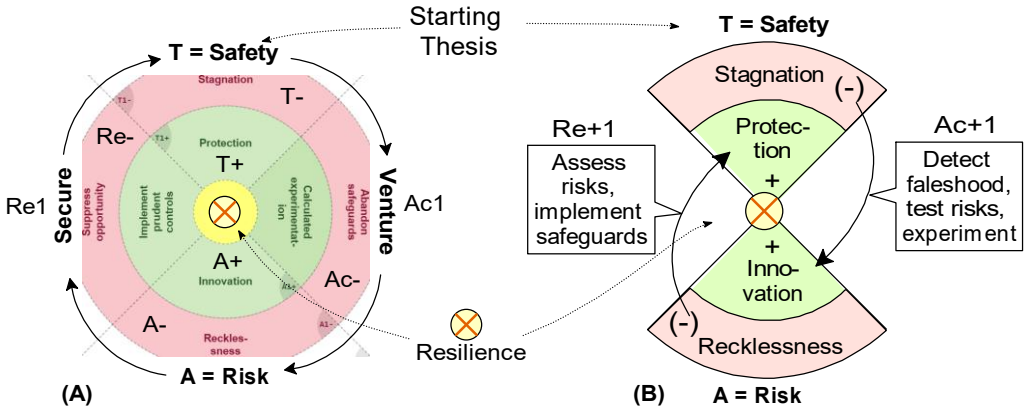

Fig. 1

Entering a simple thesis (T) = "Safety" yields an antithesis (A) = "Risk". Resilience emerges in the center as a dynamic balance between two positive aspects: T+ (Protection) and A+ (Innovation).

When this balance is lost, the system moves into negative states:
• T+ without A+ leads to T− (Stagnation)
• A+ without T+ leads to A− (Recklessness)

Maintaining resilience requires circular causation (T → Ac → A → Re → …), where action (Ac) introduces variation and reflection (Re) stabilizes it.

Scheme A describes this dynamic as:
• Ac (Venture): moving from safety toward risk
• Re (Secure): restoring safety from risk

Scheme B refines this into constructive mechanisms:
• Ac+ (Detect falsehood, test risks, experiment): transforming stagnation into innovation
• Re+ (Assess risks, implement safeguards): transforming recklessness into protection

This structure captures the core logic of Safety-II and can significantly accelerate the development of practitioner expertise

The Fifth Paradox

To illustrate the approach, we analyze Erik Hollnagel's "Fifth Paradox" (see also video).

As we strive to make safety more measurable (using KPIs, incident rates, and checklists), we often make it harder to manage. By focusing on what can be counted (usually things that go wrong), we lose sight of the processes that make things go right. As a result, we end up managing the metrics of safety rather than the capacity for safety.

This paradox can be expressed in dialectical form, as shown in Fig. 2.

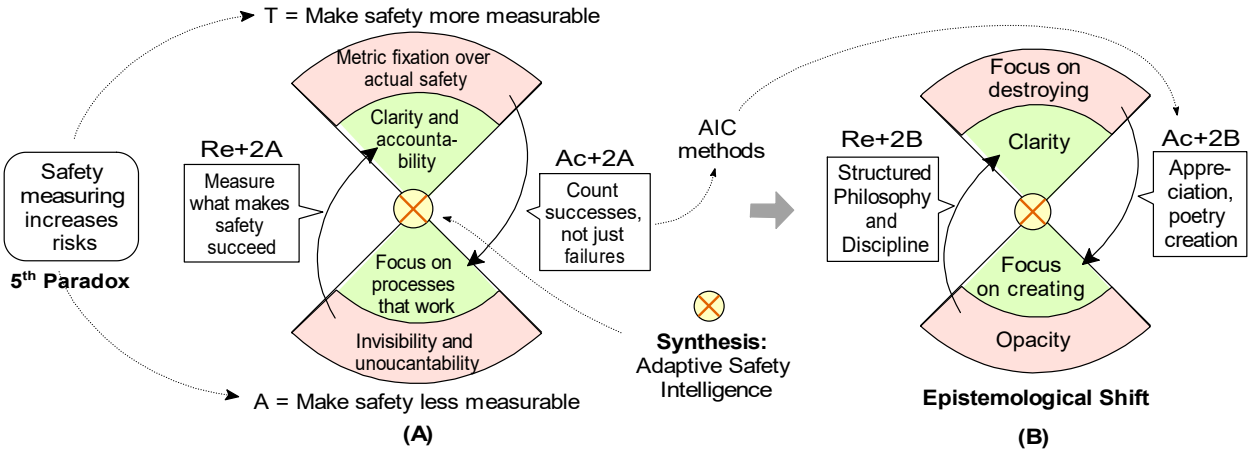

Fig. 2

The Ac+2A and Re+2A steps shift attention from failure to successful performance. Instead of asking what went wrong, the focus moves to how work succeeds under varying conditions.

This challenges the traditional Safety-I approach, where control is based on measurement and correction. In contrast, effective control requires prior understanding: one cannot influence what has not first been appreciated. This aligns with the AIC (Appreciate–Influence–Control) framework and Appreciative Inquiry, which seek to identify the system's Positive Core.

Table 1. Summary of the shift for HSE Specialists

Aspect	Safety-I (Destructive/Stagnant)	Safety-II (Appreciative/Creative)
Primary Tool	Checklists & Audits	Dialogue & Learning
Focus	How do we stop failure?	Why does it usually go right?
View of Human	A source of error to be controlled.	A source of resilience to be appreciated.
Outcome	Metric Fixation / Opacity	Systemic Clarity / Adaptive Intelligence

Beyond this operational shift, the analysis also reveals a deeper change in how safety is understood. Scheme B represents this epistemological shift. While Scheme A changes what is observed, Scheme B changes the basis on which safety is understood. Here, Ac+2B emphasizes appreciation, and Re+2B provides a structured foundation through philosophy and discipline. This shift underpins approaches such as AIC, where control is achieved through understanding

and influence rather than compliance. The practitioner's role shifts from enforcing control to supporting the conditions for successful performance.

This reframing is not limited to the Fifth Paradox, but extends to other safety paradoxes.

Operationalizing Safety-II Through Safety Paradoxes

Table 2 summarizes operational mapping across selected paradoxes, while Table 3 provides the underlying dialectical analysis.

Table 2. Strengthening Safety-II through Targeted Ac+/Re+ Integration

Paradox Type	**Ac+ / Re+**	**Figs. 1-2(A-B)**	**Safety-II Interpretation**	**Professional Practice**
Definition: safety is measured by absence of failure rather than presence of success*	Track prevention, not just incidents. Mine success for hidden lessons.	**Ac+2A (Success)**, Re+2A (Clarity)	Safety is the presence of adaptive capacity, not absence of failure	Learning teams analyzing successful operations, resilience indicators dashboards
Variability: Reducing variability undermines the adaptability*	Empower adjustments within safety limits. Codify adaptation into evolving protocols.	**Ac+1 (Experiment)**, Re+1 (Safeguard)	Humans are sources of resilience, not error	Structured debriefs on how work succeeds under pressure, adaptive procedure design
Zero Safety: absence of incidents reduces sensitivity to signals	Learn from near-misses and small failures.	**Re+2A (Clarity)**, Ac+2A (Success)	Safety emerges from sensitivity to early signals, not zero outcomes	Near-miss learning systems, weak signal reviews, "safe-to-fail" probes
Absolute Safety: pursuit of perfection discourages transparency*	Reward transparency, reporting vulnerabilities without fear	**Re+2B (Philosophy)**, Ac+2B (Appreciation)	Safety depends on openness about system limits	Just culture practices, confidential reporting, leadership vulnerability reviews
Compliance: strict adherence to rules can reduce resilience	Assess context, adapt when rules harm. Structure exceptions with explicit protocols	**Re+1 (Safeguard)**, Ac+1 (Experiment)	Rules are resources, not constraints; adaptation is disciplined	Controlled flexibility frameworks, "guided discretion" training, exception logs
Automation: increased automation reduces human readiness	Drill manual control regularly.	**Re+2B (Philosophy)**, Ac+2B (Appreciation)	Resilience requires maintained human competence in rare events	Scenario-based simulations, degraded-mode training, human-in-the-loop design
Transparency: more information can obscure actual risk	Expose gaps between alerts and reality. Signal danger clearly, not complexity	**Re+2A (Clarity)**, Ac+2A (Success)	Safety depends on interpretable signals, not more signals	Simplified dashboards, signal-to-noise audits, frontline validation of alerts

***** James Reason Safety Paradoxes and Safety Culture

Table 3 lists various safety paradoxes with suggested T/A pairs and Ac+/Re+ steps. Each paradox can yield several T/A pairs, whilst we used what looked for us the simplest.

Table 3. Analysis of other paradoxes

Paradox	Description	Dialectic Components
1. Definition*	Safety is defined and measured more by its absence (accidents, incidents) than by its presence. We only "see" safety when it fails	T = Safety is defined by its absence A = Safety is defined by its presence Ac+ = Track prevention, not just incidents Re+ = Mine success for hidden lessons
2. Defense*	The same defenses, barriers, and safeguards designed to protect a system can also be the source of its catastrophic breakdown (e.g., a safety valve that fails and creates a new pressure hazard)	T = Safeguards protect A = Safeguards cause breakdown Ac+ = Stress-test safeguards, reveal boundaries Re+ = Learn from failure, redesign safer
3. Variability*	Organizations try to limit human variability to minimize error, but it is this same human variability (adaptability and adjustments) that maintains safety in a dynamic, changing world	T = Limit human variability A = Increase human variability Ac+ = Empower adjustments within safety limits Re+ = Codify adaptation into error-proof protocols
4. Absolute Safety*	An unquestioning belief in the attainability of absolute safety (Target Zero) can actually impede real safety goals by encouraging people to hide data or ignore real risks.	T1 = Absolute safety pursuit impedes real safety A1 = Absolute safety pursuit enhances real safety Ac+ = Reward transparency, not perfection Re+ = Report vulnerabilities without fear.
5. Risk Compensation	When people feel safer (e.g., wearing a helmet), they tend to take greater risks, often negating the safety benefit of the equipment.	T = Safety measures reduce risk A = Safety measures increase risk Ac+ = Acknowledge hazards despite protection. Re+ = Build confidence through incremental protection
6. Zero Safety	The closer an organization gets to "Zero Incidents," the more vulnerable it may become to a "Big Bang" event, because it has lost the ability to detect small signals of failure.	T = Eliminating incidents creates safety A = Eliminating incidents creates blindness Ac+ = Learn from near-misses, not perfection. Re+ = Learn from small failures safely.
7. Automation	The more reliable an automated system is, the less prepared the human operator is to take over when that system eventually fails (the "Ironies of Automation").	T = Reliable automation A = Human preparedness Ac+ = Drill manual control regularly. Re+ = Drill overrides before failures occur.
8. Compliance	Strict adherence to rules can make a system brittle. In a crisis, following the rules exactly as written may lead to disaster, whereas "intelligent non-compliance" might save it.	T = Strict adherence to rules A = Intelligent non-compliance Ac+ = Assess context, adapt when rules harm. Re+ = Structure exceptions with explicit protocols.
9. Transparency	Much of what is done in the name of safety (like complex warning systems) can mask the true level of risk, creating a "false sense of security."	T = Warning systems reveal danger T = Warning systems mask risks Ac+ = Expose gaps between alerts and reality. Re+ = Signal danger clearly, not complexity.

* James Reason Safety Paradoxes and Safety Culture