Journal of the International Society for the Systems Sciences | 70th Meeting of the International Society for the Systems Sciences, Cyprus
Wilby, J., Makar, J. and Laouris, Y. eds.

COLLECTIVE COGNITIVE COHERENCE IN HUMAN DECISION-MAKING ABOUT ARTIFICIAL INTELLIGENCE: AN APPLIED FRAMEWORK FOR OVERSIGHT, GOVERNANCE, AND HUMAN-MACHINE TEAMING

R. Eva King
Fielding Graduate University
reking@fielding.edu

Abstract
The integration of artificial intelligence into high-stakes organizational systems is reshaping how groups of humans make collective decisions, particularly under conditions of shared narrative context, ambiguity, and AI-induced uncertainty. Classical decision theory, which models group cognition as the linear sum of independent rational agents, cannot adequately capture what happens in these environments. Distributed decisions exhibit relational structure – patterns of coherence that classical aggregation cannot detect – and tracing them requires tools drawn from complexity science, information theory, and quantum-inspired multipartite analysis. This paper introduces the Quantum Forces of Change Index™ (Q-FOCI™), an applied systems instrument designed to measure multipartite cognitive coherence in human decision environments and to inform AI oversight design, alignment review architecture, and human-machine team cognition. Drawing from the mathematical formalism of quantum many-body entanglement, Q-FOCI™ rests on an empirical foundation. A mixed-design experiment immersed 40 participants, organized into eight non-communicating five-person groups, in an ambiguous AI malfunction scenario set in 2126. Responses were analyzed using multipartite correlation functions, entanglement-witness analogs, mutual information, and Shannon entropy, then benchmarked against Monte Carlo-generated classical independence baselines. Seven of eight groups exceeded the classical null on at least one measure. Subjective items elicited stronger inter-participant coupling than objective items, suggesting that contextual ambiguity organizes group cognition more powerfully than factual reasoning. A W-state structural analysis further revealed qualitatively distinct coherence classes – GHZ-dominant, W-dominant, and mixed – providing the first behavioral evidence that the GHZ-W distinction has detectable analogs in collective human cognition. Building on this foundation, Q-FOCI™ translates many-body coherence detection into a measurable framework for AI governance, oversight architectures, and human-machine teaming, with a research agenda to validate it in operational contexts.

Keywords
Quantum cognition, collective decision-making, human-machine teaming, AI governance, Q-FOCI™

1 | Introduction

The integration of artificial intelligence into high-stakes organizational systems is reshaping how groups of humans make collective decisions, and the limits of classical decision theory are becoming visible as a practical problem rather than a theoretical one. Oversight committees evaluate whether an AI system is functioning correctly. Alignment review panels assess whether a model's outputs reflect intended values. Red teams probe for failure modes. In each case, the group's formal output is a collective decision. However, the cognitive process that generates it is distributed across non-communicating or loosely coordinating individuals who are each forming judgments under conditions of shared narrative context, technical ambiguity, and AI-induced uncertainty. Classical aggregation models, which treat social group decisions as the linear sum of independent rational agents, cannot adequately capture what happens in these environments. Distributed decisions about AI systems exhibit relational structure – patterns of coherence across participants that classical aggregation cannot detect – and the consequences of missing that structure are non-trivial.

The AI safety, alignment, and oversight literatures have developed substantial machinery for evaluating AI systems. Reinforcement learning from human feedback (Ouyang et al., 2022), constitutional methods (Bai et al., 2022), and the broader alignment program (Amodei et al., 2016; Christiano, 2018; Russell, 2019) have given technical tools for shaping model behavior and probing model internals. What has received less attention is the corresponding problem on the human side: how the humans responsible for oversight, evaluation, and governance arrive at collective judgments about AI systems, and what structural dependencies exist in those judgments that classical decision-support tools do not surface. Naturalistic decision-making research (Klein, 1998) and situation awareness theory (Endsley, 1995) have built the conceptual scaffolding for individual judgment under uncertainty. Team cognition research (Cooke et al., 2013; Salas et al., 2008) extends those frames to coordinated teams. Collective intelligence research (Hong & Page, 2004; Woolley et al., 2010) addresses aggregation properties of groups. However, the specific case of distributed human cognition about AI under shared narrative framing – the case that matters most for oversight design – sits in a gap.

This paper introduces the Quantum Forces of Change Index™ (Q-FOCI™), an applied systems instrument I am developing to measure multipartite cognitive coherence in human decision environments and to inform AI interpretability and alignment review, AI oversight and governance, and human-machine team cognition. The framework draws from the mathematical formalism of quantum many-body entanglement (Gühne & Tóth, 2009; Horodecki et al., 2009), and from the quantum cognition tradition (Busemeyer & Bruza, 2014; Pothos & Busemeyer, 2013), which has demonstrated that classical probability theory cannot adequately describe certain features of human judgment – order effects, contextuality, interference patterns – that the formalism of quantum probability handles natively. What the quantum cognition literature has not previously addressed is the multipartite case: what happens when three or more individuals, responding asynchronously and without communication, generate judgments under shared contextual conditions. Q-FOCI™ extends quantum-inspired cognitive analysis into that domain and translates the extension into an operational instrument.

The paper proceeds in four moves. It develops the theoretical foundation for applying many-body quantum-inspired formalism to collective decision-making, with explicit attention to the boundary between quantum-inspired analogy and quantum-mechanical claim. It presents the empirical foundation: a mixed-design experiment examining whether forty participants distributed across eight non-communicating five-person groups exhibited multipartite coherence under conditions of shared narrative context and AI-induced uncertainty. It specifies Q-FOCI™ as an applied systems instrument – what it measures, how it operates, and how its outputs inform organizational decision design. It develops the framework's applications to AI interpretability and alignment review, AI oversight and governance, and human-machine teaming. It concludes with a research-and-development program structuring the framework's near-term validation.

2 | Theoretical Foundation

The case for applying quantum-inspired formalism to collective cognition rests on a precise diagnosis of where classical probability theory becomes inadequate, not on any ontological commitment to quantum processes in the brain or social systems. The distinction is held throughout this paper. Q-FOCI™ is a quantum-inspired analytical instrument, not a quantum mechanical claim about human cognition.

Classical probability theory assumes the Kolmogorov axioms – that events have well-defined joint probabilities, that the probabilities of their conjuncts bound the probability of a conjunction, and that observations do not modify the probabilistic structure of subsequent observations. Human judgment under uncertainty routinely violates these axioms in systematic and replicable ways. The conjunction fallacy (Tversky & Kahneman, 1983), order effects in survey responses (Moore, 2002; Wang et al., 2014), and judgment patterns under ambiguity all fail classical axiomatic constraints while satisfying the more permissive structure of quantum probability theory. This finding has produced two decades of theoretical and empirical work in quantum cognition (Aerts, 2009; Busemeyer & Bruza, 2014; Khrennikov, 2010;

Pothos & Busemeyer, 2013), demonstrating that quantum-probabilistic models account for a range of cognitive phenomena that classical models cannot.

2.1 | From Bipartite to Multipartite Coherence

The quantum cognition tradition has focused almost exclusively on individual and dyadic settings. Individual decisions exhibit interference patterns; dyadic interactions produce contextuality effects. However, organizational decision environments – oversight committees, alignment review panels, interpretability rater pools – are characteristically multipartite. Three or more individuals generate judgments under shared contextual conditions, often asynchronously, often without direct communication. The relevant analytical tools are not bipartite correlation measures, but the many-body correlation functions and entanglement-witness analogs that physicists developed for multipartite quantum systems (Gühne & Tóth, 2009; Horodecki et al., 2009).

Drawing from the mathematical formalism of quantum many-body entanglement, this framework adapts three classes of analytical instruments for the cognitive case: multipartite correlation functions that quantify higher-order response coupling beyond pairwise relationships; entanglement-witness analogs that detect when group response patterns exceed what classical independence allows; and information-theoretic measures – Shannon entropy and mutual information (Cover & Thomas, 2006) – that capture distributional structure independent of any particular ontological commitment. The combined analytical layer provides convergent evidence about whether group cognition exhibits the kind of structured non-independence that classical aggregation cannot reproduce.

2.2 | Systems-Science Framing

For the systems science audience, the framework can be located alongside other traditions that have engaged the non-separability of complex systems. Distributed cognition (Hutchins, 1995) and sensemaking (Weick, 1995) have long argued that cognition does not reside solely within individual minds. Second-order cybernetics has insisted on the observer-dependence of measurement. Complexity science has developed tools for tracing emergent organization. Quantum-inspired multipartite analysis adds a specific analytical capability to this lineage: the ability to detect and classify forms of relational coherence that are real, measurable, and consequential, but invisible to classical aggregation. The contribution is methodological, not metaphysical.

3 | Empirical Foundation

The framework rests on controlled empirical demonstration. A mixed-design experiment examined whether small, non-communicating human groups exhibit multipartite response coherence under conditions of shared narrative context and AI-induced uncertainty.

3.1 | Method

Forty participants, organized in eight independent five-person groups, were immersed in an ambiguous artificial intelligence malfunction scenario set in 2126. The scenario placed each participant in the role of a high-stakes organizational decision-maker, confronting conflicting information about whether an autonomous AI system (Unit-34) was malfunctioning or operating within parameters. The science fiction framing was a deliberate choice of cognitive realism: it reduced the influence of prior commitments to current AI systems while preserving the structural features that matter for organizational decision-making about AI under uncertainty.

Each participant completed two counterbalanced, randomized binary-question blocks asynchronously, without communicating with other group members. Set A contained 10 objective questions (factual items about scenario details, treated as a classical control set). Set B contained 10 subjective questions (intuitive judgments under ambiguity) and was treated as the contextuality probe set. A narrative with the same factual details preceded each initial block. Participants receiving Set A (factual/control) first received a narrative that was a factual, unemotional press release. In contrast, participants who received Set B (subjective) first received a narrative that included the same facts but was designed to prime participants' emotional cognitive processes. The block order was counterbalanced across participants (A → B versus B → A), yielding the experimental conditions for order-based contextuality analysis. Responses were encoded as ±1 binary values to permit application of multipartite correlation analysis.

3.2 | Analytical Framework

Responses were analyzed using a layered analytical framework. Pairwise correlation matrices captured first-order response coupling within each group. Multipartite correlation functions extended that analysis to higher-order structure across all five group members simultaneously. Entanglement-witness analogs – adapted from the witness inequalities used to detect entanglement in physical systems (Gühne & Tóth, 2009) – tested whether observed group response patterns exceeded what classical independence allows. Shannon entropy and pairwise mutual information (Cover & Thomas, 2006) provided information-theoretic measures of response distribution and inter-participant coupling. All inferential analyses were benchmarked against Monte Carlo classical independence baselines using reproducible seeded simulations, providing a null distribution against which empirical statistics were tested.

3.3 | Findings

Seven of eight multipartite groups exceeded the classical independence null on at least one analytical measure. Five groups reached $p < .001$ for both pairwise correlations and mutual information. A more stringent witness-style threshold of 0.30, derived from entanglement detection literature (Dür et al., 2000; Gühne & Tóth, 2009), was met by two groups – Groups 1 and 8 – indicating that the elevated higher-order alignment captured by the witness analogs was selective rather than general across the sample. The pattern across analytical instruments forms a sensitivity hierarchy: the broader measures detected coherence across most groups. At the same time, the witness threshold identified the subset with the strongest higher-order structural alignment. This graded convergence supports the conclusion that the observed coherence is not an artifact of any single instrument, while characterizing the range of each instrument's operation within the distribution.

A complementary W-state structural analysis revealed qualitatively distinct classes of multipartite coherence. Groups 1 and 8 – the same two groups that exceeded the witness-style threshold – exhibited GHZ-dominant structure, with the highest GHZ coherence values in the sample (0.500 for both) and heterogeneous pairwise coupling (CV = 1.03 and 1.20, respectively), consistent with collective alignment driven by full-group consensus rather than uniform pairwise informational dependence. Group 4 exhibited a W-dominant structure, characterized by distributed coupling that is robust under perturbations of individual members. The remaining groups were mixed. To my knowledge, this is the first behavioral evidence that the GHZ-W distinction, well-established in physical many-body systems, has detectable analogs in collective human cognition. The convergence between the witness analysis and the structural classification – Groups 1 and 8 passing both – provides cross-instrument validation of the GHZ-dominant identification. The distinction matters for applied purposes: GHZ-dominant organizational groups are more fragile to membership loss than W-dominant groups.

Block-specific decomposition showed that the subjective items elicited stronger inter-participant coupling than objective items, and this effect was robust to order-condition control. The asymmetry is theoretically interesting: contextual ambiguity organizes group cognition more powerfully than factual reasoning under shared framing. Items requiring intuitive judgment under uncertainty produced the most structured coherence; those requiring factual recall, the least. For governance design, the implication is direct: the cognitive material most relevant to AI oversight – judgments under ambiguity, not factual reconstruction – is precisely the material most subject to structural coherence that classical aggregation cannot detect.

Exhibit 1 summarizes the analytical components, what each measures, and example findings from the empirical foundation.

Exhibit 1. Analytical Components and Sample Findings

Analytical Component	What It Measures	Sample Finding
Pairwise correlation	First-order within-group response coupling	5 of 8 groups $p < 0.001$
Multipartite correlation	Higher-order coupling across all five group members	7 of 8 groups exceeded classical null
Entanglement-witness analogs	Threshold detection of elevated higher-order alignment	2 of 8 groups (Groups 1 and 8) exceeded the 0.30 threshold
Shannon entropy	Response distribution and item-level variability	Set B < Set A; order effect $p \sim 0.056$ (objective block)
Mutual information	Inter-participant coupling magnitude	5 of 8 groups $p < 0.001$; $r = 0.89$ convergence with correlation
Monte Carlo baseline	Classical independence null distribution	Reproducible seeded simulations across all 8 groups

4 | Q-FOCI™: An Applied Systems Instrument

The empirical demonstration establishes that multipartite cognitive coherence is real, measurable, and structurally classifiable in distributed human decisions about AI under uncertainty. Q-FOCI™ translates those findings into an applied systems instrument I am developing for high-consequence distributed cognition environments – AI interpretability and alignment review, AI oversight and governance, and human-machine teaming. The four sections below specify what the instrument measures, how it operates, its boundary conditions, and its applications.

4.1 | What Q-FOCI™ Measures

Q-FOCI™ measures four dimensions of distributed group cognition that classical aggregation models do not surface. The first is the magnitude of multipartite coherence – whether the group response patterns exceed what classical independence would produce, and by how much. The second is the structural class of that coherence – GHZ-dominant, W-dominant, or mixed – corresponding to organizationally meaningful differences in robustness, distributability, and failure mode. The third is the contextual sensitivity profile – how coherence varies across objective and subjective decision domains, and how it responds to order-of-presentation effects. The fourth is the deviation pattern from classical baselines – where, specifically, the group's collective behavior diverges from what an independent-agent model would predict.

4.2 | How Q-FOCI™ Operates

The instrument's operational pipeline takes distributed binary or ordinal response data from a small group of decision-makers under defined contextual conditions. It produces a coherence profile across the four dimensions above. Inputs require minimal infrastructure: responses can be collected through standard survey instruments, asynchronous decision platforms, or structured deliberation processes. The analytical layer applies multipartite correlations and witness-style metrics, information-theoretic measures, and the Monte Carlo baseline comparisons described in Section 3. Outputs are interpretable at the organizational level: coherence magnitude expressed as deviation from classical null; structural class expressed as GHZ/W/mixed taxonomy with associated robustness implications; contextual sensitivity expressed as an objective-versus-subjective coupling differential; deviation pattern expressed as a localized profile of where collective behavior departs from independence.

4.3 | Boundary Conditions

Q-FOCI™ has explicit boundary conditions, stated here because the framework's credibility depends on its discipline about what it does not claim. The instrument is quantum-inspired, not quantum-mechanical. It detects structural coherence in collective cognition; it does not adjudicate the underlying cognitive or social mechanisms that produce the coherence. It applies to small group settings under shared contextual conditions; it does not scale arbitrarily to large populations without further methodological development. It requires bounded response domains (binary or ordinal); it does not currently handle open-ended response data. Moreover, it measures coherence rather than correctness. Q-FOCI™ reports on the structural properties of group cognition, not on whether the group has arrived at the right answer.

These boundaries are not limitations to be apologized for. They define the conditions under which Q-FOCI™ produces interpretable, defensible results, and they distinguish the framework from approaches that overclaim explanatory scope.

5 | Applications to AI Governance and Oversight

The integration of AI into organizational decision systems has created a category of cognitive labor – distributed human evaluation of AI systems under uncertainty – for which the existing decision-support toolkit is inadequate. Q-FOCI™ addresses that gap. Three application domains are developed below, in order of technical specificity: AI interpretability and alignment review, AI oversight and governance, and human-machine teaming.

5.1 | AI Interpretability and Alignment Review

AI interpretability and alignment review processes share a structural feature that makes them the most direct application of Q-FOCI™: distributed human raters, working asynchronously from shared rubrics and example sets, produce judgments about model outputs under genuine ambiguity. The structural conditions closely match the experimental conditions of the empirical foundation. Whether the question is interpreting a model's internal representations, evaluating whether a model's outputs reflect intended values, or assessing the reliability of a model's behavior under distributional shift, the cognitive labor is multipartite, asynchronous, and shared-context, the regime in which classical aggregation fails most predictably.

Q-FOCI™ is being developed to operate in two roles within interpretability and alignment review pipelines. As a diagnostic, it indicates whether a rater pool exhibits structural coherence that the alignment metric does not reflect – substantial multipartite coupling among raters means the alignment score is not the independent aggregation it is assumed to be, but a partially coherent collective judgment whose aggregation properties differ from the classical case. As a design instrument, it informs decisions about rubric framing, example curation, and rater selection to produce review architectures with desired coherence properties: deliberately engineered W-dominant rater coherence (robust, distributable) over GHZ-dominant coherence (fragile, dependent on framing). For the interpretability research specifically, the instrument provides a measurement layer on the human side of the loop that the technical side of interpretability work has not previously had access to. This is the most operationally specific application of Q-FOCI™ currently being scoped for pilot evaluation.

5.2 | AI Oversight and Governance

AI oversight committees, ethics review boards, and institutional governance bodies face a recurring methodological problem distinct from but related to the alignment review case. The classical assumption – that committee members evaluate AI systems independently, and that the committee's decision reflects the aggregation of independent assessments – is empirically questionable. Committee members share narrative context, technical framing, and exposure to the same documentation. The empirical findings reported above suggest that under such shared conditions, distributed judgment exhibits multipartite coherence that the committee's formal aggregation process does not surface.

Applied to oversight bodies, Q-FOCI™ provides a measurement capability that complements the formal committee process. It surfaces whether members' nominally independent judgments are exhibiting structured coherence beyond what independence would predict, and it identifies the structural class of that coherence: a GHZ-dominant committee, in which all-or-nothing convergence depends on full participation, signals a different organizational risk profile than a W-dominant committee, in which distributed coherence is robust to membership change. The same instrument extends naturally to higher-stakes governance decisions about AI deployment. Whether to integrate a new model into production, authorize an autonomous system for a specific operational role, or escalate observed model behavior to incident response. Q-FOCI™ informs decision-making leadership about the structural properties of the decision process, enabling intervention before the decision is finalized. The instrument does not replace committee judgment; it informs the committee about the cognitive structure of its decision-making process, a piece of information committees routinely lack.

5.3 | Human-Machine Teaming

The framework's natural extension is to human-machine teaming environments, in which human team members work alongside AI agents under conditions of shared situational framing (Endsley, 1995) and distributed task allocation (Salas et al., 2008). The applied scenario in the empirical foundation involved decisions about an AI system; extending this to decisions involving AI systems is methodologically natural but requires further empirical work. The relevant question is whether the multipartite coherence detected in human-only groups generalizes – and how it generalizes – when AI agents are introduced as cognitive participants rather than evaluation objects. This is the most active extension of the research program, addressed in Section 6.

6 | Research Agenda and Recommendations

The empirical demonstration and applied framework presented above structure in an active research-and-development program organized around four parallel tracks. The framework is under development, with the empirical foundation, the operational scope, and the validation path already in place.

The first replication is at scale. The eight-group sample establishes the phenomenon under controlled conditions; scaled replication across operational settings – actual AI oversight committees, alignment review and interpretability rater pools, deployed decision-support contexts – characterizes the framework's operational sensitivity and refines the analytical layer for applied use.

The second is the interpretability and alignment review pilot deployment. The structural match between the empirical conditions and the alignment review workflow makes this the most operationally specific track. Pilot validation in the actual rater-pool context – quantifying coherence structure, classifying GHZ versus W-dominant configurations, and testing whether deliberate architecture choices produce predicted coherence profiles – is currently being scoped.

The third is the human-machine teaming extension. Whether multipartite coherence holds when AI agents are introduced as cognitive participants is an open empirical question that the existing framework is well-positioned to answer. Pilot studies in human-AI mixed teams, using the same analytical framework, will determine whether Q-FOCI™ extends naturally to teaming contexts or requires methodological adaptation.

The fourth is a structural class characterization across organizational and institutional types. The GHZ/W/mixed taxonomy is currently a behavioral observation; its organizational correlates – what kind of institutional structures, framing conditions, and selection processes produce which coherence classes – require systematic study across the operational contexts in which Q-FOCI™ is being deployed.

These four tracks are mutually informing, and progress is parallel. They are also explicitly oriented toward the institutions for which Q-FOCI™ is being developed: AI labs conducting alignment review and interpretability research, government and applied research organizations overseeing AI, and operational settings developing human-machine teaming infrastructure. The framework is, intentionally, an instrument under active development for operational deployment by the author and collaborating institutions, with the research agenda above structuring its near-term validation path.

7 | Conclusions

The integration of artificial intelligence into high-stakes organizational systems has exposed a methodological gap in how distributed human cognition about AI is measured and understood. Classical aggregation models, which treat collective decisions as the linear sum of independent rational agents, cannot adequately capture what happens when groups make judgments under conditions of shared narrative context, ambiguity, and AI-induced uncertainty. Distributed decisions exhibit relational structure that classical instruments cannot detect, and the consequences of missing that structure are non-trivial for oversight, governance, and human-machine teaming.

This paper has presented Q-FOCI™ - the Quantum Forces of Change Index™ - as an applied systems instrument I am developing to measure multipartite cognitive coherence in high-consequence distributed cognition environments. The framework is grounded in empirical demonstrations showing that seven of eight non-communicating five-person groups exhibited multipartite coherence exceeding the classical independence baselines; that the GHZ-W structural distinction has detectable analogs in collective human cognition; and that contextual ambiguity organizes group cognition more powerfully than factual reasoning under shared framing. Q-FOCI™ translates this foundation into an operational instrument with applications to AI interpretability and alignment review, AI oversight and governance, and human-machine teaming.

The framework is quantum-inspired, not quantum-mechanical. Its discipline is what makes it deployable. Its applications are concrete, its boundary conditions explicit, and its development program active.

References

Aerts, D. (2009). Quantum structure in cognition. *Journal of Mathematical Psychology, 53*(5), 314–348.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. *arXiv preprint* arXiv:1606.06565.

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., … Kaplan, J. (2022). Constitutional AI: Harmlessness from AI feedback. *arXiv preprint* arXiv:2212.08073.

Busemeyer, J. R., & Bruza, P. D. (2014). *Quantum models of cognition and decision*. Cambridge University Press.

Christiano, P. (2018). Clarifying "AI alignment." *AI Alignment Forum*.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive Science, 37*(2), 255–285.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Wiley-Interscience.

Dür, W., Vidal, G., & Cirac, J. I. (2000). Three qubits can be entangled in two inequivalent ways. *Physical Review A, 62*(6), 062314.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32–64.

Gühne, O., & Tóth, G. (2009). Entanglement detection. *Physics Reports, 474*(1–6), 1–75.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences, 101*(46), 16385–16389.

Horodecki, R., Horodecki, P., Horodecki, M., & Horodecki, K. (2009). Quantum entanglement. *Reviews of Modern Physics, 81*(2), 865–942.

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Khrennikov, A. (2010). *Ubiquitous quantum structure: From psychology to finance*. Springer.

Klein, G. (1998). *Sources of power: How people make decisions*. MIT Press.

Moore, D. W. (2002). Measuring new types of question-order effects. *Public Opinion Quarterly, 66*(1), 80–91.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., … Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems, 35*, 27730–27744.

Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences, 36*(3), 255–274.

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control.* Viking.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors, 50*(3), 540–547.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*(4), 293–315.

Wang, Z., Solloway, T., Shiffrin, R. M., & Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences, 111*(26), 9431–9436.

Weick, K. E. (1995). *Sensemaking in organizations*. Sage.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science, 330*(6004), 686–688.