Introduction to the Online Supplement for "Integrating Fragmented Risk Knowledge: Sheaf-Theory for Risk Analysts"

This Online Supplement provides the background and case discussions that underlie the concise summary presented in Table 5 of the main article. Whereas the main text highlights only the essential contrasts between enduring challenges in risk and decision science and the corresponding insights offered by a sheaf-theoretic perspective, the present Supplement expands on those points with fuller explanations, examples, and references. (The references are included at the end of the main article.)

The material is organized to parallel the structure of Table 5. Each section develops one or more "backstories" for the challenges listed in the left-hand column—ranging from incoherent individual beliefs and preferences, to difficulties in aggregating expert judgments, to failures of coordination in teams, to obstacles in causal inference, collective choice, and risk governance. For each such challenge, the Supplement illustrates in greater depth why local rationality or consistency may still fail to compose into a globally coherent picture, and how a sheaf-theoretic perspective helps to diagnose and sometimes resolve these failures.

We encourage readers who wish to move beyond the high-level overview of Table 5 to use this Supplement as a guided tour of the diverse domains where fragmentation of knowledge impedes coherence, and where sheaf-theoretic concepts provide new tools that maybe useful in understanding and potentially overcoming these barriers.

Individual-Level Incoherence in Beliefs, Preferences, Plans, and Decisions

Failures of belief updating, preference consistency, or temporal stability of preferences can lead even a single agent to violate normative coherence principles required for rational (e.g., expected utility-maximizing) risk decisions. Sheaf theory offers a way to model local belief fragments and preference relations as sections over domains of relevance and to diagnose whether they can be extended to a globally coherent belief or utility function using techniques such as solving $D^0x = \omega$ (Section 3.1), helping to clarify whether inconsistencies stem from noise, omission, or contradiction.

1. **Individual beliefs.** Subjective probabilities often fail to satisfy the axioms of probability theory—for example, by implying a probability of greater than 1 for some events or judging the probability of a conjunction (A & B) to be higher than the probability of one of its parts (B alone). A classic example

is the *conjunctive fallacy,* in which people judge the likelihood of a detailed scenario (e.g., "a thunderstorm followed by hail and a tornado") as more probable than a simpler, more general one (e.g., "a tornado") implied by it (Tversky and Kahneman, 1981; Lu et al., 2015). Such violations imply logical contradictions. If someone believes that P(A & B) > P(B), then basic probability theory implies that the conditional probability P(B | A) = P(A & B)/P(B) must exceed 1, which is impossible. In practice, the conjunctive fallacy leads to observed anomalies such as willingness to pay more for life insurance against death from a terrorist attack on a flight than for life insurance for death from all causes on the flight (Johnson et al., 1996).

Sheaf theory provides a new and useful perspective on such inconsistent probabilities. Just as a sheaf captures how locally consistent beliefs (e.g., marginal or conditional probabilities) might glue together into a globally coherent belief system (e.g., a joint distribution for all variables, corresponding to a global section of the sheaf), so, conversely, an object called a *cosheaf* (the mathematical dual of a sheaf) describes how a global section (here, a full joint distribution) can be aggregated into coherent marginal beliefs over subsets of variables. A joint probability distribution that satisfies both the local conditional probability assignments ("upward" maps forming a sheaf) and the correct marginal distributions ("downward" maps forming a cosheaf) corresponds to a global section of both the sheaf and the cosheaf. This paired sheaf-cosheaf global section encodes exactly the information in a *Bayesian network* (BN), a widely used tool in probabilistic risk analysis that generalizes fault trees, event trees, and bow tie diagrams (Aven and Thedki, 2021). represents a joint distribution as a product of local conditionals while preserving the correct marginals under marginalization (Rosiak, 2022). The sheaf represents the directed flow of conditioning (from parents to children in the graph), while the cosheaf captures marginalization (from the full joint to subsets of variables). A coherent belief system, such as a fully specified BN (or the joint probability distribution that it represents), thus corresponds exactly to a global section that satisfies both sheaf and cosheaf constraints.

Viewing belief incoherence, such as the conjunctive fallacy, as a failure to form a valid global section suggests the possibility of using sheaf-theoretic methods to help assess whether subjective beliefs (from experts, stakeholders, or individuals) can in principle be integrated into a globally coherent whole. In practice, assessing consistency between conditional and marginal judgments—e.g., checking whether the P(y) implied by P(y | x) and P(x) matches an independently assessed P(y), or whether estimated P(y | x) and P(x) values give the same product as estimated P(x | y) and P(y) values, P(y | x)P(x) = P(x | y)P(y)—can help to reveal and correct incoherent beliefs (Ravinder et al., 1988). Rather than eliciting complex joint probabilities directly, decision analysts can decompose them into marginal and conditional parts, check for internal consistency, and adjust discrepancies, e.g., by averaging estimates of the same probabilities estimated by different decompositions. Empirical studies suggest that this decomposition approach can improve reliability and reduce judgment errors in elicited subjective probabilities (Ravinder et al., 1988).

2. **Individual preferences.** Preferences for actions or outcomes often appear stable and consistent (e.g., transitive) when considered in isolation, yet shift or reverse when alternatives are framed differently or are presented together. Classic behavioral findings show that individual preferences can reverse depending on implicit reference points and framing effects introduced by logically irrelevant comparisons or variations in wording or context—for example, describing the outcomes of medical procedures as survival rates vs. mortality rates (Kahneman and Tversky, 1979; Tversky and Kahneman, 1981). Preferences may also exhibit "arbitrary coherence," in which initial, often arbitrary, valuations anchor subsequent valuations and choices without normative justification (Ariely et al., 2003). For example, people may prefer alternative A over alternative B when evaluated in isolation, but reverse that preference when the same alternatives are framed differently or presented alongside additional alternatives. Each choice may appear reasonable on its own, but when taken together, the resulting pattern of preferences can violate basic consistency principles—such as transitivity (if A is preferred to B, and B to C, then A should be preferred to C) or dominance (an alternative that is better in all respects should not be rejected). These inconsistencies reveal that locally sensible preferences may not combine into a globally coherent ordering, as required by expected utility theory and other normative models of rational choice. In short, even when each preference appears rational in context, there may be no way to combine them into a coherent overall preference structure that respects normative principles such as transitivity, dominance, or stability under logically irrelevant variations in wording or the presence of unchosen alternatives.

Merely viewing the foibles of human decision-making from a sheaf-theoretic perspective suggests a possible approach to doing better: seek, design, or adopt global sections that integrate our diverse local judgments into a coherent whole. A global section corresponding to a stable, complete, and transitive preference ordering for decisions (viewed as probability distributions over outcomes), which can be represented by a von Neumann-Morgenstern (vNM) utility function under well-known conditions. That is, a decision-maker who always acts *as if* maximizing expected utility behaves coherently in precisely the sense that their local decisions can be interpreted as restrictions of a global preference structure. Conversely, the empirical findings of Tversky and Kahneman (1981) and others—including framing effects, dynamic inconsistency, and preference reversals—show that ordinary human decision-making often violates this property: it fails to admit a global section. Even internally "reasonable" or well-justified choices, when pieced together, form globally incoherent patterns.

A possible sheaf-theory-inspired remedy is to treat decision-making not as an accumulation of context-specific choices, but as a search for a coherent global policy: a mapping from situations to actions that would remain justifiable even if all local situations were made simultaneously visible. A useful thought experiment helps motivate this idea. Suppose that instead of making each choice ourselves, we had to delegate all future decisions to an intelligent agent—such as a robot or AI—who will act on our behalf. To instruct it, we would need to encode our preferences and beliefs in a way that is unambiguous, stable, and internally consistent. The agent would need a global section, even if we ourselves do not yet have one. This inversion—figuring out how to explain a coherent self to a decision-making assistant—can guide us toward more rational behavior. It encourages the construction of simplified, consistent decision rules, utility functions, or policies that approximate our deepest goals, rather than reacting locally to salience, emotion, or framing. Once defined, we might then choose to act *as if* we were following these rules ourselves. A sheaf theory perspective thus inspires a reconceptualization of self-governance: not as a series of isolated decisions, but as the enactment of a coherent global section of values and plans, constructed with an eye toward delegability, coherence, and resilience to fragmentation.

3. **Time-inconsistent plans and preferences.** An important special case of inconsistent individual preferences arises in sequential decision-making, such as planning infrastructure investments, allocating health resources, or managing long-term environmental risks. Preferences held at one point in time may contradict those held at another by the same decision-maker even if no new information has become available. For example, someone who always prefers one apple immediately to two apples after a month's delay might nonetheless prefer two apples in 101 months to one apple in 100 months – even though the latter situation will eventually (after 100 months) become the former. This phenomenon, known as *time inconsistency* or dynamic inconsistency, arises when preferences change over time in ways that are not anticipated or planned for, such as under hyperbolic discounting, perhaps reflecting short-cuts that our brains take to avoid the full complexity of evaluating delayed payoffs (Enke et al., 2025; Laibson, 1997). Neuroscientific research confirms

that this inconsistency is not merely behavioral but structural: distinct brain systems encode the value of immediate versus delayed rewards, often competing for control over behavior depending on task framing and timing and the load from other tasks (McClure et al., 2004). For example, an individual may make an explicit plan to save for a future vacation, only to spend the money on an impulsive purchase. In public life, a decision-maker for a public agency might initially prioritize supporting long-term climate resilience projects, but later reallocate funds to short-term political or economic demands, despite having previously judged the long-term investment as preferable. In such examples, each decision may appear locally justified based on the preferences of the moment, yet the resulting trajectory violates the consistency conditions required for intertemporal rationality and optimization (Strotz, 1955). In such cases, there may be no global plan that the individual endorses at all times, and no stable mapping from local preferences to a consistent long-term policy. These failures illustrate how preferences that are consistent at any given moment, but not over time, can obstruct coherent decision-making and planning over time. Dynamic inconsistency can be viewed as arising from distorted perceptions due to the use of simplifying heuristics (Enke et al., 2025), discussed next.

4. **Individual risk perceptions, heuristics and biases.** Many widely studied cognitive heuristics and biases (Berthet, 2022; Tversky and Kahneman, 1974) in individual risk perception—including scope insensitivity, overconfidence, confirmation bias, loss aversion, ambiguity aversion, and preference reversals—may arise not from isolated and independent errors, but from a deeper structural phenomenon: the affective formation and protection of *simplified but globally misleading mental models of risky prospects* (Cox, 2015). From a sheaf-theoretic perspective, the study of heuristic and biases in individual risk perception and judgments under uncertainty suggests that the *affect heuristic* (Slovic et al., 2007) plays a central role by creating a fast, emotionally coherent judgment (Loewenstein et al., 2015) that effectively serves as a *pseudo-global section*: an internally consistent but factually inaccurate assignment of beliefs and values across multiple risk attributes (e.g., costs, benefits, probabilities) based on the simplification that decision alternatives tend to be all "good" or all "bad" across their attributes and consequences. These simplified global impressions are then pushed down (in a cosheaf-like manner) to generate distorted (affect-driven) local judgments that are reflected in well-known perceptual and behavioral anomalies such as overestimated willingness to pay (WTP) for emotionally salient benefits; anchoring and irrelevant cue effects driven by initial affect-laden impressions; scope insensitivity, where the perceived benefit does not scale with the

actual magnitude of a good, since the affect does not; framing and reference point effects, due to global affective frames overriding local value assessments; endowment and status quo bias, from an over-weighted feeling of loss tied to departures from current holdings; and confirmation bias, which might be viewed as an endowment effect for already-held beliefs (Berthet, 2022; Cox, 2015; Kahneman et al., 1991). These distorted local inferences are then sustained by what might be called *learning aversion* (Cox, 2015)—a tendency to avoid potentially disconfirming feedback, reinterpret failure as exogenous, and suppress metacognitive reflection, perhaps to avoid the substantial (e.g., NP-hard) computational burden of finding a valid global section. This makes the initial pseudo-global section hard to revise, even when it causes over-spending, inconsistent preferences, or predictable decision regret.

Rather than attempting to correct each individual heuristic or bias, a sheaf-theoretic perspective suggests a more radical potential remedy: bypass all of them simultaneously by replacing the affect-driven construction of pseudo-global sections with valid global sections (if possible) or the best available, e.g., least-squares, approximations to them, deliberately satisfy a priori logical and probabilistic coherence constraints. In this view, cognitive failures are not random anomalies, but failed attempts at global synthesis—heuristic- and emotion-guided efforts to construct coherent assignments of beliefs and preferences from incomplete or affectively distorted inputs. This reframing suggests the possibility of designing an AI-based decision support system to explicitly search for global sections—coherent belief-preference-action structures—by treating subjective probabilities and utilities as partial sections subject to compatibility conditions. Unlike human cognition, such a system would not rely on affect or other heuristics, but could instead devote computational resources to evaluate whether overlapping judgments are mutually consistent and, if not, guide users toward improved structure. Such an unbiased AI assistant might offer a model of deliberative, internally consistent, unbiased risk perception and decision-making under uncertainty that has not previously been available.

5. **Causal inference.** Even excellent scientists often struggle to provide valid causal interpretations of their data and causally effective recommendations to policymakers (Perlmutter et al., 2024). Causal interpretations and conclusions are no less subject to heuristics and biases than other areas of human judgment and risk perception (ibid). In epidemiology, environmental health, and regulatory risk assessment, causal inferences are often drawn from observational data by adjusting for measured covariates and estimating measures of the "effect" (whether statistical or causal—and too

often the two are conflated) of one variable (e.g., exposure) on another (e.g., mortality or disease incidence) (Schwartz and Prins, 2024).

While each such model and effect estimate may appear locally plausible and statistically well supported, global causal conclusions can fail if unmeasured confounding, nonlinear interactions, or model misspecification are present (ibid). This reflects a broader impossibility insight that spans fields: *associational patterns observed in data collected under past conditions do not, in general, support valid predictions about the effects of interventions*. This core insight is known in economics as the *Lucas Critique* (Lucas, 1976), which warns that policy interventions change the structural context and invalidate predictions based on past correlations. In statistics, Judea Pearl (2009) emphasized a "basic distinction" between associational and causal concepts, warning that "they do not mix"—a warning that is often ignored in practice. Similar concerns arise under the heading of non-transportability in causal inference, nonstationarity in machine learning, and lack of external validity in epidemiology and social science (Schwartz and Prins 2024). All reflect the same underlying challenge: there is no purely empirical, assumption-free path from observed associations to reliable predictions of intervention outcomes when interventions change the system being observed.

An example comes from a study of the effects on all-cause mortality caused by a 1990 Dublin coal ban, and by subsequent bans on coal burning throughout Ireland. Clancy et al. (2002) initially reported dramatic declines in mortality, attributing them to the roughly 70% reduction in particulate matter pollution achieved by the ban. But subsequent follow-up efforts showed that similar mortality trends occurred in areas unaffected by the ban and that the trends started long before the bans—revealing that the original inference, though locally persuasive, was globally invalid due to lack of proper control for coincident historical trends (Zigler and Dominici, 2014). The actual reduction in all-cause (and cardiovascular) mortality rates achieved by the ban, as assessed decades later, appears to be indistinguishable from zero (Dockery et al., 2013; Lyons et al., 2023; Zigler and Dominici, 2014). This illustrates that even when local statistical associations appear to support a causal interpretation, they may not assemble into a valid global causal structure.

Formal tools such as causal graphical models (Pearl, 2009; Schwartz and Prins, 2024), invariant causal prediction (ICP) (Peters et al., 2016), and sheaf-theoretic modeling can help detect when locally plausible models fail to assemble into a valid global causal structure. Causal graphical models—often represented as directed acyclic graphs (DAGs)—map out hypothesized cause-and-effect relationships among variables, clarifying which statistical comparisons are likely to support valid causal inferences (Schwartz and Prins, 2024). However, not all DAGs encode causal structure, and not all apparent associations reflect underlying mechanisms. ICP provides a statistical method to test whether a proposed set of direct causes for a variable remains causally valid across different environments—such as across studies, policy regimes, or population strata. Specifically, it examines whether the conditional distribution of an outcome given its candidate causes remains invariant under such shifts (Cox, 2024). This invariance supports the hypothesis that the selected variables genuinely screen off the influence of background changes, as would be expected of true causal parents. If the conditional relationship fails to remain stable, the model is likely missing key confounders or incorrectly specified. In the language of sheaf theory, such a stable conditional structure functions like a “global section” that integrates locally valid relationships into a coherent, system-wide model. This sheaf-theoretic lens makes it possible to formalize and detect when causal integration across datasets or domains is possible—and when it is structurally blocked. A related constructive approach that further illustrates these principles is the cocycle framework developed by Dance and Bloem-Reddy (2024). Here, the causal effect of an intervention is modeled directly through a family of transformation maps (“cocycles”) that encode how conditional distributions change under interventions. These cocycles obey algebraic consistency conditions—analogous to sheaf-theoretic gluing constraints—and can be used to derive robust estimators for interventional and counterfactual quantities without fully specifying the observational data-generating process. This approach shares structural affinities with sheaf theory in its use of local symmetries and compatibility conditions to ensure global coherence, offering yet another illustration of how causal modeling can be grounded in constraint-based, compositional frameworks.

Sheaf-theoretic modeling is starting to shed new light on causal modeling. For example, Krasnovsky (2025) shows how sheaf-based representations of local-to-global data constraints can be combined with flow-based simulations and information-theoretic “causal emergence metrics” to assess when macro-level descriptions—such as aggregate population models—actually have more causal explanatory power than fine-grained individual-level models. This phenomenon—known as scale-dependent causality—occurs when interventions at the macro level (e.g., on aggregated variables or “macro-nodes”) produce more informative, stable, or predictive effects than interventions at the micro level (e.g., on individual variables or nodes). In Krasnovsky’s framework, this is quantified using Effective Information (EI), an information-theoretic measure of how much knowing an intervention improves prediction of its effects. When EI is higher for a macro-node than for its constituent micro-nodes, the macro-level model is said to exhibit causal emergence. Sheaf-theoretically, this corresponds to identifying macro-assignments that preserve or enhance global coherence relative to their more fragmented subcomponents. The key insight is that causality can be scale-dependent, meaning that macro-level (coarse-grained or aggregated) models of a system can provide stronger or more informative causal explanations than models based on micro-level (fine-grained) components. Particularly in systems with internal redundancy, noise, or feedback, aggregated causal models can better support robust causal inferences and predictions—and hence decisions intended to cause desired outcomes—than more detailed models. Sheaf theory provides a principled way to detect and validate this kind of causal emergence by testing whether macro-level assignments form consistent global sections that outperform their finer-scale analogs.

This view has practical implications. In contexts such as public health, environmental monitoring, or infrastructure resilience, sheaf-theoretic methods can help analysts discover when high-resolution data are less useful than more abstracted or aggregate summaries for capturing the key causal structures that guide effective interventions. The sheaf-theoretic approach enables analysts to test whether causal models composed from different levels or domains (e.g., exposure $\rightarrow$ biomarker $\rightarrow$ disease) are mutually compatible, or whether conflicting assumptions or information gaps obstruct global integration. In this way, sheaf-theoretic tools extend beyond consistency checks to support principled causal synthesis across fragmented observational data, model components, and scientific disciplines.

6. **Individual reasoning with confirmation bias (Wason selection task).** In this classic logical reasoning experiment, people are asked to test the rule "If a card has a vowel on one side, then it has an even number on the other." Four cards are shown face-up, such as E, K, 4, 7, and subjects are asked which ones logically must be turned over to verify whether the rule holds. Most people choose to flip E (which is correct) and 4—seeking confirmation—rather than the correct answer of flipping E and 7, since a vowel on the other side of the card with a 7 could potentially falsify the rule. The widespread failure to detect which combinations actually test the rule reflects an inability to reason accurately about the compatibility of local information with a global rule.

   Unfortunately, confirmation bias is prevalent in applied science and risk analysis (Perlmutter et al., 2024). For example, in the previously discussed example of causal inference for the effects of Ireland's coal-burning bans on mortality rates, extensive data show no detectable effect on all-cause mortality rates (or on cardiovascular mortality rates) despite a large reduction in particulate air pollution, even after decades of surveillance (Lyons et al., 2023). A decade after the original study, the original researchers and colleagues noted that "Mean [black smoke] concentrations fell in all affected population centers post-ban compared with the pre-ban period, with decreases ranging from 4 to 35 microg/m^3 (corresponding to reductions of 45% to 70%, respectively), but we observed no clear pattern in SO2…. In comparisons with the pre-ban periods, no significant reduction was found in total death rates associated with the 1990 (1% reduction), 1995 (4% reduction), or 1998 (0% reduction) bans, nor for cardiovascular mortality." Nonetheless, rather than acknowledging that the original confident causal interpretation was simply an instance of the *post hoc* fallacy based on local information that was falsified by a wider view based on data from counties outside the ban area (Dominici and Zigler, 2014), advocates of tighter regulation simply ignored the disconfirming evidence, e.g., suggesting that "It is likely that the ban or a change in policy was independent of measured or unmeasured variables that confounded the association between air pollution and mortality" (Wang et al., 2016). This suggests the remarkable power of confirmation bias and fragmented views in creating and sustaining erroneous causal conclusions in practice (Perlmutter et al., 2024). From outside the confirmation bias frame, it is perhaps obvious that the passage of time (and with it, the improvement of methods for detecting, preventing, and treating cardiovascular disease) is a measured variable that did in fact confound the association between air pollution and mortality – and that explained away the entire exposure-mortality association (Dominici and Zigler, 2014).

Yet, the perception that reductions in particulate matter have saved lives prevails and drives policy to this day, with disconfirming evidence playing no role in policy-making or in PR statements to the public. For example, the Irish government's plan to phase out smoky coal nationally (Department of the Environment, Climate and Communications, 2021) states that "Research indicates that the introduction of the Smoky Coal Ban in Dublin in 1990 has resulted in approximately 350 fewer mortalities per year, reducing cardiovascular, cerebrovascular and respiratory mortality in the general population," disregarding the subsequent clarification from the original authors that, when control groups were considered, no effect of the bans on all-cause (or cardiovascular) mortality was found (Dockery et al., 2013; Dominici and Zigler, 2014). Instead, Ireland's "recent Clean Air Strategy contains a commitment to achieve the WHO guideline limits [of 5 μg/m^3] for PM2.5 by 2040… [which] will require policy measures to decarbonise home heating, promote active travel and the transition to electric vehicles, and further regulations on burning fossil fuels and enforcing environmental regulations more tightly" (Hoy et al., 2024). The continued propagation of the message that reducing particulate air pollution saved lives, notwithstanding ignored evidence to the contrary, shows that regulatory risk analysis and messaging can itself be a source of widely shared disinformation (Thekdi and Aven, 2023).

Such examples underscore how confirmation bias can undermine not only individual judgment, but collective understanding and policy coordination—especially when different actors interpret partial or time-lagged evidence through locally consistent but globally incompatible mental models (Perlmutter et al., 2024). This sets the stage for the next challenge: *how can teams or organizations with distributed knowledge and responsibilities align their beliefs and decisions to function as a coherent whole?*

Coordinated Teams: Communication and Alignment

When agents share goals but differ in what they know or when they act, coordination depends on whether their decision rules and beliefs can be assembled into a consistent joint strategy or set o behaviors. Sheaf theory provides a way to formalize local information sets and decision functions as stalks and sections, and to ask whether a global plan exists that aligns these local views. The following examples show that failures to glue local views into coherent shared views and effective plans or joint actions in team decision-making—such as those due to asynchronous delays, misinformation, or observational ambiguity—can be precisely modeled as cohomological obstructions, revealing whether and how coordination can be achieved or fails.

7. **Aggregating expert probabilities.** In engineering risk settings, expert judgments are often used to estimate system reliability or failure probabilities. It may seem reasonable to aggregate these judgments, e.g., by averaging or otherwise combining them, especially if the experts provide the same or similar probabilities (Dietrich, 2007). But this can be misleading if the aggregation ignores the details of how each expert's belief depends on what that expert privately observed. For example, consider a two-component system that works if at least one component functions. Each component has an 80% chance of working, and they are independent, so the system has a 96% chance of success (1-0.2*0.2). Expert 1 inspects one component and sees that it has failed. She reports that the system has an 80% chance of working (based on the other, unseen component). Expert 2 also sees a failed component and also reports 80%. If both saw the same component, then 80% is the correct group belief. But if they saw different components, the system has a 0% chance of working. Unanimity or averaging would misleadingly suggest 80% confidence in system success.

This example shows that *expert probabilities cannot always be directly aggregated without knowing the structure of what each expert observed and how their observations overlap.* Each expert's belief (i.e., probability) is locally valid based on what the expert has seen, but these individual beliefs cannot be "glued" into a valid globally consistent view using aggregation rules (e.g., unanimity or averaging) without understanding the dependency relationships among the observations. Valid group belief aggregation requires more than agreement on numbers—it requires shared context and coherent integration of partial perspectives (Dietrich, 2010).

From a sheaf-theoretic perspective, identical beliefs (probabilities) of different experts cannot be glued into a valid global section without understanding the underlying information structure. This reframes the design of belief aggregation not as choosing an opinion-pooling function (e.g., average, weighted average, or median) that satisfies abstract axioms, but as designing restriction maps that relate local belief domains in a way that permits global consistency. Sheaf models make explicit the structure of local domains (observations or contexts), the overlaps among them, and the conditions under which locally valid beliefs can be extended to a coherent group view. This perspective encourages closer scrutiny of axioms like unanimity and independence, which may seem appealing in the abstract but fail to preserve coherence when applied to fragmented or context-sensitive expert knowledge. Sheaf theory helps analysts focus on whether and how different local beliefs can be *jointly satisfied* across shared domains—not merely whether they agree on point values.

Recent work in Bayesian aggregation theory (Dietrich, 2010) reinforces this view. Dietrich shows that even when agents share a common prior and receive conditionally independent signals, averaging posteriors may be unjustified, and instead proposes a multiplicative pooling rule based on priors and likelihoods. His framework makes explicit the epistemic structure behind belief formation and avoids misleading aggregation by incorporating how beliefs are formed—not just their numerical values. This complements the sheaf-theoretic insight: belief aggregation should reflect how knowledge is distributed and related, not just how beliefs align.

8. **Team decisions with trustworthy members.** Suppose that two or more individuals share identical goals and utility functions for outcomes. Thus, their interests and incentives are perfectly aligned: what one prefers, all prefer, and so each can fully trust the others to try to do what is best for all. Such a group is called a *team* (Marschak and Radner, 1972; Malikopoulos, 2022). Despite their shared goals and preferences, the members of a team may obtain different local information and face different opportunities to act. In such settings, if early decisions influence what later-acting

members observe, earlier actors face a dual challenge: choosing an action that advances the team's goal and conveys helpful information to later actors. This implicit communication—known as signaling—can make the structure of optimal decision rules too complex to calculate (Witsenhausen, 1968; Papadimitriou and Tsitsiklis, 1985). Even if each team member uses a decision rule that is optimal given its own local information—i.e., one that minimizes expected cost or maximizes expected utility conditional on available information—the joint implementation of these locally optimal rules by the team members may still fail to achieve the globally optimal team performance. This failure arises because the information available to each team member may not be rich enough to support coordination: the locally optimal rules, though well-adapted to each member's view, may not fit together into a globally coherent plan that achieves the best joint outcome. In such decentralized settings, local optimality with respect to each individual's information set does not imply global optimality for the team, even under full alignment of preferences and goals.

Recent work by Felber et al. (2025) offers a constructive response to this problem by modeling decision-making tasks in distributed systems as *task sheaves* defined over *system frames* that represent agents' local information via indistinguishability relations. In this framework, a team decision problem is solvable—that is, admits a protocol that guarantees task completion—if and only if the associated task sheaf admits a global section. The key insight is that coordination failure corresponds to the presence of nontrivial cohomology in the task sheaf. The first cohomology group H^1 provides a precise algebraic characterization of the obstructions to consistent coordination, even when agents act optimally from their own perspectives. When the cohomology vanishes, global coordination is possible, and the corresponding global sections describe protocols that guarantee team success. This allows analysts not only to detect where and why coordination fails but to use homological methods to redesign the system by modifying information flows, overlap conditions, or communication protocols to eliminate obstructions and restore global coherence.

9. **Team decision rules with unreliable members.** In many health, safety, and environmental (HSE) contexts, effective coordinated action depends on reaching a shared decision across a distributed team. Examples include regional emergency response units deciding whether and/or how to evacuate an area, interagency panels determining risk classifications for novel or emerging hazards (e.g., FDA panels during COVID, chemical regulation, AI safety), or independent safety monitors coordinating shutdown decisions for complex facilities during or following a cascading catastrophic accident sequence. These systems often operate under time pressure, with uncertain or delayed

communication, and with the possibility that one or more members may become temporarily unreachable or effectively non-responsive due to communication failure, overload, or cascading system disruption.

The Fischer-Lynch-Paterson (FLP) Theorem (Fisher et al., 1985) from fault-tolerant distributed computing shows that in any asynchronous system where message delays are unpredictable and at least one member can fail silently (i.e., crash) and stop responding, no protocol can guarantee that the remaining agents will always reach consensus in a finite amount of time—and hence by when decisions must be made. While human teams often mitigate this by escalating decisions or using informal communication channels (e.g., phone calls, direct overrides), such workarounds may fail or be infeasible in high-stakes, time-critical scenarios—especially when AI systems, automated controls, or strict procedural rules are involved. In such cases, each agent's decision rule may appear locally sound based on available information, but coordination can still break down due to the structural limits imposed by timing uncertainty and the possibility of dropout—especially when agents are waiting on responses from one another to reach consensus, and no participant can be sure whether others are still active or merely delayed. The FLP result serves as a cautionary insight: even fully aligned teams can fail to coordinate effectively when the system lacks guarantees about who is still participating and what others have heard or decided.

10. **Teams with disinformation.** In many organizations and group decision problems, not every member can fully trust all others to pursue the same goals. The notorious Byzantine Generals Problem (Lamport et al., 1982), originally posed in distributed computing, illustrates how coordinated action can fail even when most agents are trustworthy and share the same objectives. If a sufficiently large minority of team members are compromised—specifically, one-third or more—they can send conflicting messages to different teammates (e.g., telling one to "attack" and another to "retreat"), making it impossible for the loyal members to determine whether they will act together at the crucial moment. Even when each team member has a message history that appears internally consistent, the members may lack any assurance that their teammates received the same information and will make the same decision. Without verifiable communication or a trusted shared reference, the team cannot reliably coordinate their actions.

    While solutions to this problem are known in synchronous systems, they require that less than one-third of the participants are faulty; beyond that, reliable agreement and coordination become impossible (ibid). In contrast to the FLP result, which shows that consensus may fail due to uncertainty and silence, the Byzantine case shows how actively conflicting information can also prevent coordinated action—even when communication is timely. Both settings illustrate how local consistency among agents may still be insufficient for global alignment of decisions and actions.

The negative results in these examples may seem daunting. Yet, recent work in applied sheaf theory has shown that these same local-to-global obstructions that prevent effective team coordination in these and other examples can also help to guide the design of improved communication protocols and information structures to enable improved – and in some cases even optimal – team decision-making. For example, Schmid (2025) proposes a general "homological programming" framework in which coordination objectives—such as decentralized consensus or cooperative optimization—are modeled as optimization problems on sheafs that can be constructively solved using existing distributed optimization algorithms and software. Failures of coordination correspond to nontrivial cohomology classes, revealing exactly where and how system redesign is needed. In this way, the sheaf-theoretic framework not only detects and explains where coordination fails, but also suggests the possibility of using optimization methods to identify minimal and targeted structural changes (such as adding communication links, adjusting update rules, or redefining overlaps between agent responsibilities) to restore coherence. Thus, while perfect coordination may be undecidable or unreachable in the worst case, sheaf cohomology may help overcome such obstacles by providing a principled way to work backward from inconsistency to remediation by redesigning or upgrading systems with additional information and communication links to bridge fragmented knowledge and align multi-agent decisions under uncertainty.

The potential to re-frame coordination problems—ranging from inconsistent team decisions to infeasible system trajectories—as sheaf assignments and solve them via optimization over global sections has been demonstrated recently in the context of optimal distributed network control. For example, Kearney et al. (2020) show that the problem of controlling a distributed network to meet joint goals can be framed as minimizing the consistency radius of a sheaf encoding both local constraints and global objectives. When no global section exists, their framework identifies which constraints or structural elements (e.g., control inputs, communication links, or neighborhood definitions) need to be modified to restore coherence. In this way, sheaf theory becomes not just a diagnostic tool, but a design guide: it pinpoints where to improve observability, inject redundancy, or restructure decision flows to make coordination feasible. By quantifying how close partial solutions are to full coherence and offering bounds on approximation error, it enables practical planning in imperfect, asynchronous, or discretized systems.

4.3 Strategic Interactions and Asymmetric Information

So far, we have considered agents who are all on the same team and who seek to cooperate to manage risky or uncertain systems when each agent has only a partial view of its state. When agents have diverging interests or possess private information, compositionality can break down not due to any lapse in rationality, but because of misaligned incentives. Building effective cooperation in such settings may require not just aligning knowledge, but realigning motivations.

This section explores two central sources of strategic friction in regulatory and policy settings: (1) *Principal–Agent relationships*, where hidden actions or effort levels and private information can undermine compliance and trust; and (2) *mechanism design problems*, in which institutions must be structured to elicit truthful information and cooperative behavior from self-interested actors—even though those actors hold private knowledge and may have incentives to misreport. This branch of game theory studies how to design rules or contracts so that individually motivated decisions jointly produce good outcomes for the system as a whole (where "good" is defined explicitly by properties such as Pareto-efficiency, incentive-compatibility, and individual rationality, as discussed in the following examples). Recent work by Frey et al. (2023) shows how such strategic interactions can be modeled using composable game structures, in which local games (each representing a piece of a larger institutional arrangement) are linked via mappings that must satisfy coherence constraints. Their work suggests that modular, formally structured approaches to mechanism design—akin to sheaf-theoretic constructions—can clarify how large-scale institutions succeed or fail depending on the compatibility of their components.

We believe that sheaf models have the potential to help in these settings—not by assuming cooperation, but by distinguishing structural non-integrability (i.e., no global agreement is possible due to strategic incompatibility) from mere epistemic asymmetry (i.e., local sections exist but cannot be shared or reconciled without violating incentives). These tools help map how lack of common knowledge, trust, or verifiability obstructs contractibility or enforceability in regulatory and organizational contexts.

While relatively little published work has yet applied sheaf theory directly to strategic game-theoretic settings with private information, the conceptual tools—particularly the ability to model overlapping but misaligned information and incentive structures—offer promising foundations. The compositional game framework proposed by Frey et al. can be seen as a step in this direction, showing how strategic interactions with partial observability and local rules can be assembled into a global institutional architecture. The discussion that follows includes both well-documented examples of strategic failure and emerging ideas for how sheaf-theoretic diagnostics and representations might help identify, diagnose, and possibly remediate these failures.

10. **Principal–Agent relationships, moral hazard, and regulation.** In many regulatory settings—such as environmental cleanup, workplace safety, food safety inspection, or hazardous materials handling—regulators must rely on others (contractors, firms, or local agencies) to carry out actions that protect public health and safety. These arrangements often take the form of Principal–Agent relationships, in which one party (the "principal," such as the EPA or OSHA) sets rules and incentives, while the other (the "agent," such as a firm or site manager) makes decisions and takes costly actions that may not be fully observable by the principal but that affect the outcomes that the principal cares about (Bier and Lin, 2013). If the principal cannot freely and accurately observe the agent's private information about the true level of effort, diligence, care, or compliance exerted, the agent may rationally provide less effort – and the principal may rationally be willing to pay less for it – than if this private information were common knowledge. For example, firms may follow only the letter—not the spirit—of OSHA safety rules if they believe that exceeding minimum standards won't be recognized or rewarded. Manufacturers may take less care than consumers would willingly pay for, if the resulting product safety or quality is not transparent. Employees may provide less value than employers would gladly compensate, and insured parties may take less care and be charged higher premiums than both insurer and insured would prefer, if levels of effort or care are difficult or expensive to observe.

    In each of these and countless other cases, both sides would be better off if the principal could trust the agent to exert more effort in return for greater reward (or less monitoring and enforcement burden). But each side recognizes that private information creates incentives for the agent to overstate its true level of effort or compliance. The result is often a more burdensome regime of enforcement or lower trust-based compensation, not because of bad faith, but because the agent's true effort cannot easily be credibly verified, so that the principal and agent lack a common view of the agent's true level of effort.

11. **Mechanism design and limits to agreement.** Even when all parties are rational and would prefer to reach agreement, disagreements can and do persist in economic transactions – and also in

regulation, scientific risk assessment, and policy design. Failures to reach agreement may arise due to information asymmetries and strategic incentives. For example, a classic result in information economics, the Myerson–Satterthwaite Theorem (Myerson and Satterthwaite, 1983), shows that if buyers and sellers have private information about their own values (e.g., true willingness to pay or accept compensation), then no mechanism can guarantee that a deal will be struck whenever it would be Pareto-efficient ("win-win") to do so—while also ensuring incentives for truthful reporting ("incentive-compatibility"), voluntary participation ("individual rationality"), and no outside subsidies ("balanced budget").

Similar impossibility results apply to collective choice procedures ("mechanisms") in regulatory and risk settings where stakeholders report estimates of costs, risks, or benefits that are difficult to verify directly (Maskin, 2007). In the jargon of economics and collective choice theory, a "mechanism" is a formal system, institution, or procedure that maps individuals' inputs (such as their expressed preferences, costs, risk estimates, willingness-to-pay, willingness to accept, votes, offers, or signals) into collective decisions and outcomes. A mechanism specifies who communicates what to whom, when and under what rules, and how allocations, payments, or other decisions are determined based on those inputs. Impossibility results show that, under certain reasonable conditions, no mechanism can simultaneously satisfy all desirable properties—such as efficiency, incentive-compatibility, individual rationality, and budget balance—across all private-information settings. This means that disagreement or inefficiency may persist even when all parties would benefit from cooperation, simply because the information and incentives cannot be aligned through any single, universally acceptable rule.

More recent work in dynamic mechanism design shows that these impossibility results can sometimes be avoided in repeated or multi-stage interactions with sufficiently patient players who value future payoffs (and hence long-term relationships) (Athey and Segal, 2013; Bergemann and Välimäki, 2019). In many real-world situations—including public-private partnerships, long-term contracts, and environmental agreements—people make decisions over time, and each person may know things that others don't. This creates a challenge: How can we design fair, efficient agreements when no one sees the whole picture, and each player seeks his own best interest? Athey and Segal (2013) tackle this challenge by designing a *dynamic mechanism*—a kind of long-term agreement or protocol—that evolves over time. It rewards people based on how much their information helps others; spreads payments out over time so that honest and helpful participants benefit in the long run; and adjusts decisions and payments as new information emerges, maintaining fairness and incentives throughout. If players can commit to longer-term contracts or reveal information over time, then such dynamic mechanisms can sometimes achieve efficient outcomes and truthful communications without external subsidies and with all players expecting to gain by participating.

A closely related line of work arises from mechanism design in organizational risk management. Alpcan (2012) develops a game-theoretic framework in which decentralized organizational units—modeled as noncooperative agents with private utility functions—allocate resources to mitigate security risks. The central risk manager functions as a designer who cannot dictate decisions directly but can shape outcomes via incentive mechanisms (e.g., subsidies or penalties). These mechanisms are engineered to satisfy efficiency, preference-compatibility, and strategy-proofness. Notably, Alpcan shows that iterative, information-limited protocols can converge rapidly to equilibrium, even when individual preferences are hidden. Still, dynamic mechanisms such as those proposed by Alpcan (2012) and Athey and Segal (2013) typically require complex structures—such as transfer payments, reputation tracking, or repeated commitments. They may be challenging to design, implement, and enforce. The computational complexity of optimally designing even relatively simple dynamic mechanisms is known to be high (NP-hard), making the use of advanced algorithms a practical necessity (Papadimitriou et al., 2022.

A sheaf-theoretic lens can help reframe these challenges. Rather than assuming that valid aggregation or agreement must arise from combining all local inputs (e.g., via simple averaging or consensus), sheaf models require analysts to define *restriction maps*—rules for how local information relates to overlapping and higher-level domains. From this perspective, impossibility results like Myerson–Satterthwaite may reflect overly simplistic or misaligned restriction maps: attempts to glue together inherently incompatible information structures without accounting for the strategic and epistemic constraints on each agent's inputs. This insight suggests a constructive design principle: if consistent global decisions or agreements are desired, one should design local-to-global information structures (restriction maps) to satisfy anticipated incentive and information constraints. In this view, sheaf theory offers more than a metaphor—it becomes a guide to structuring negotiations, regulatory processes, or decision protocols so that desired global properties (like efficiency, fairness, or enforceability) can emerge from locally valid contributions.

A common theme in these examples (and some of the others we discuss in this section) is that sheaf-theoretic tools suggest practical paths forward for designing institutions and contracts that are more robust to asymmetric information and strategic misalignment. One approach might be to frame the incentive compatibility constraints of strategic agents as gluing conditions across overlapping local models—each representing an agent's private information and decision space. When global sections fail to exist, the structure of the resulting cohomology can reveal not only the incompatibility of preferences or beliefs, but the precise information gaps, incentive misalignments, or communication constraints that block agreement. This shifts the analysis from “is agreement possible?” to “what would need to change to make agreement possible?”

Moreover, sheaf-theoretic representations support modular institutional design. By isolating subcomponents of a mechanism—such as reporting structures, enforcement rules, or monitoring protocols—as stalks over different information domains, analysts might test whether local design choices support global consistency. For instance, if truthful reporting fails due to misaligned payoffs in a specific subregion of the state space, one might use cohomological diagnostics to identify where incentive compatibility breaks down and how local rules might be adapted to enable gluing. This could guide the redesign of scoring rules, audit policies, or communication links that make decentralized mechanisms work as intended.

In regulatory and contracting environments, this perspective also suggests a new design goal: *build systems whose information structures and incentive rules are gluable.* Rather than forcing conformity across all participants, institutions can be structured to respect heterogeneity while still enabling compatible behavior. Sheaf theory offers a language, concepts, and a computational toolbox for exploring this delicate balance—bridging local autonomy with global functionality.

4.4 Collective Choice, Aggregation, and Emergent Opinion Dynamics

Public policy and risk regulation often require fusing divergent individual preferences, beliefs, or values. The challenge is not only to make (and perhaps to define) optimal individual decisions under uncertainty, but to combine diverse judgments, preferences, or actions into a coherent collective outcome. While the previous section focused on *strategic interactions* among agents with private information—highlighting how incentive misalignments can obstruct cooperation—this section will describe how, even in the absence of explicit strategic manipulation, failures of coordination and of effective collective decision-making may still arise from structural, informational, or normative conflicts among well-intentioned actors. Locally rational contributions—from citizens, scientists, or stakeholders—can fail to compose into globally coherent group decisions. These failures include well-known paradoxes from social choice theory (e.g., Arrow’s impossibility theorem), aggregation pathologies in group belief formation, value conflicts in moral and policy reasoning, and dynamic failures such as informational cascades or herding.

Some problems discussed here, such as collective action dilemmas, occupy a conceptual middle ground between strategic interaction and collective choice. In these cases, participants may share goals or public values but still act in ways that are globally suboptimal—often because institutional structures or communication and commitment constraints fail to enable the coordination, trust, or mutual knowledge needed for beneficial collective outcomes to emerge. These settings illustrate how **locally** justified or individually optimal behavior can produce globally unstable, inconsistent, or Pareto-inefficient results—not because of irrationality, but due to failures of integration across local decision contexts.

Sheaf-theoretic methods provide a framework for analyzing such failures as obstructions to gluing: situations where local consistency across individuals or groups does not extend to a consistent global judgment, decision, or policy. They also offer tools—such as cohomological diagnostics and structured restriction maps—to identify the specific structural misalignments that block coherent aggregation and suggest design interventions to improve collective decision-making.

12. **Collective choices: limits to combining individual preferences into fair and consistent group decisions.** Public decisions—such as whether to fund a flood protection system, approve a new medical intervention, or regulate a potentially hazardous chemical—often require aggregating the preferences and judgments of multiple individuals or stakeholder groups. While each person's input may be consistent and well-informed, combining these into a single group decision can lead to outcomes that violate basic standards of fairness or logic (Arrow, 1951). For example, under majority rule, a group of 3 people choosing among 3 alternatives, A, B, and C, may collectively prefer alternative A to B, and B to C, but then irrationally prefer C to A. (This result, Condorcet's Paradox, is produced if individual 1 prefers A to B and B to C, abbreviated A > B > C, while individuals 2 and 3 have preference orderings B > C > A and C > A > B, respectively. In this case, each of the local preferences A > B, B > C, C > A is supported by a majority of 2 out of 3.) Or a voting procedure might deliver different results depending on whether an irrelevant alternative is added or removed. In other cases, people may vote strategically—misrepresenting their true preferences to influence (or "manipulate") the outcome (Gibbard, 1973). Even more troubling, granting individuals basic rights or veto power over specific decisions can sometimes prevent the group from making any choice that benefits everyone, even when such a choice clearly exists (Sen, 1970). These problems have been

formalized in social choice theory (Arrow, 1951), which shows that no decision rule can satisfy normative principles of fair, efficient, consistent, and manipulation-proof outcome when combining multiple people's unrestricted rankings or judgments. In health, safety, and environmental (HS&E) contexts, these negative results may help to explain why advisory panels, expert committees, or participatory processes sometimes fail to reach stable or broadly acceptable recommendations—not because people are irrational, but because the structure of group decision-making itself imposes hard limits on what is possible.

A sheaf-theoretic perspective suggests a fruitful way to re-think, and perhaps bypass (without contradicting) such impossibility results. From this perspective, it is natural to wonder whether the impossibility arises not because normatively desirable aggregation is inherently unworkable, but because Arrow's framework pays insufficient attention to the details of information flows and restriction maps – what exactly we really want an aggregation mechanism to do, and using what kinds of individual-level information. Perhaps it assumes the wrong types of inputs, outputs, and consistency requirements for creating a global section – and hence for informing many real-world decisions. Specifically, we might postulate that the following aspects can be re-designed to create possibility instead of impossibility:

- Inputs: Full ordinal rankings of all alternatives are often unnecessary and unrealistic; people rarely hold or express complete and stable preference orderings.
- Outputs: Real-world decisions usually aim to select a single "good" (normatively desirable) action or policy—not to construct an entire group-wide ranking over all alternatives.
- Goals: The normative aims of collective choice often concern coherence with ethical principles (e.g., fairness, benefit to all), incentives, or risk tolerances—rather than satisfying formal mathematical axioms such as Arrow's "independence of irrelevant alternatives" (Arrow, 1951).

Asking for too much to go into a collective choice procedure (complete rankings) and too much to come out of it (full social orderings of all alternatives), and insist on strict structural consistency (e.g., across all subsets of alternatives and all possible preference orderings), we may be setting ourselves up for impossibility. But if we design aggregation procedures that are tailored to the actual decision context—e.g., based on cardinal utilities, relevant constraints, and well-defined normative goals—then aggregation often works well.

Indeed, a subsequent result, Harsanyi's utilitarian theorem (Harsanyi, 1955; Hammond, 1992) shows that under plausible assumptions, consistent and ethically meaningful aggregation is not only possible but yields a simple and intuitive rule: if individuals and society both maximize expected utility, and society respects Pareto indifference (i.e., it is indifferent between lotteries whenever all individuals are), then the social welfare function must be a weighted sum of individual utilities:

$$W(x) = w_1 \cdot u_1(x) + w_2 \cdot u_2(x) + \ldots + w_n \cdot u_n(x)$$

Here, *x* is a *social state*—that is, a possible complete specification of relevant outcomes affecting individuals in the society (e.g., who receives what benefits or harms, or what policy consequences occur). In Harsanyi's framework, society faces uncertainty over which social state will occur, and collective choices are modeled as *lotteries* (i.e., probability distributions) over these social states. Each $u_i(x)$ is individual *i*'s von Neumann-Morgenstern utility function representing i's preferences under uncertainty, and the weights w_i reflect ethical or policy priorities (such as equity, anonymity, or priority to the worse-off). This result shows that aggregation can succeed when we choose inputs and outputs that make sense for utility-based decision-making under uncertainty.

From a sheaf-theoretic perspective, Arrow's impossibility reflects the use of overly rigid restriction maps—rules about how group preferences must relate to individual preferences across all possible subsets of alternatives. These restriction maps encode consistency requirements (like IIA) that are often unnecessary or counterproductive in practical decision-making. By contrast, the success of Harsanyi-style aggregation suggests that better-designed restriction maps—focused on preserving what truly matters, such as expected utility or ethical comparability—can support the existence of coherent global sections. That is, we can sometimes glue individual-level information into a meaningful collective decision, if we are careful about what information we ask for and what constraints we impose.

These lessons are especially relevant in real-world risk decisions involving uncertainty or ambiguity where different experts may have diverging beliefs or attitudes toward unknown risks. Recent work by Nascimento (2012) shows that aggregation can be designed to balance utilitarian averaging with precautionary or Rawlsian caution when experts disagree or when their opinions are incomplete. The resulting aggregation rules support compromise decisions that hedge against uncertainty that is assessed differently by different participants. It does so without requiring full agreement on probabilities or risk attitudes. Moreover, aggregation of individual preferences among risky collective choices can still succeed even under 'deep' uncertainty when different individuals hold different beliefs and ambiguity attitudes (Pivato and Tchouante, 2024). Rather than assuming shared priors or full von Neumann-Morgenstern utility functions, their framework allows individuals to express qualitative preferences over uncertain acts, consistent with some Bayesian interpretation. They show that individual utility functions can then still be meaningfully aggregated via weighted summation even when collective beliefs remain undefined or diverge across agents. This decoupling of ethics (what outcomes people value) from epistemics (what they believe about the world) reflects many real-world decision contexts in risk analysis, where experts may disagree about probabilities but still agree on normative goals, such as reducing harm or protecting vulnerable populations.

From a sheaf-theoretic perspective, this reinforces the theme that aggregation succeeds when restriction maps are well-designed—that is, when we allow for structured flexibility in what must be preserved across individuals and subdomains. Arrow's theorem imposes structural constraints that are too rigid. Pivato and Tchouante offer a constructive alternative: allow individuals to disagree on fine-grained beliefs while agreeing on enough to justify collective action. Impossibility results like Arrow's are not the end of the story. Rather, they inspire a more refined and constructive understanding of what kinds of aggregation are both possible and desirable in uncertain, high-stakes settings.

In these collective choice settings, a sheaf-theoretic perspective does more than diagnose where consistency fails. It encourages us to ask: what structure are we trying to preserve? What relationships really need to hold across decision contexts? And how can we design aggregation rules that respect those relationships without imposing unnecessary constraints? We believe that in this way, sheaf-theoretic reasoning supports a broader shift from impossibility theorems to constructive design principles grounded in and aligned with the real-world goals of decision support.

13. **Collective beliefs.** In probabilistic decision-making contexts, combining individual beliefs and values into a group decision can also violate desired normative principles even when everyone is rational and well-informed. Hylland and Zeckhauser (1979) showed that if a group decision rule aggregates beliefs and utilities separately while also respecting unanimity and Pareto-efficiency, then it must adopt the beliefs of a single individual, implying a "dictatorship of beliefs." Nehring (2005) further demonstrated that even if everyone agrees on rules for aggregating their individual beliefs (expressed as probabilities) and preferences (expressed as utilities) and also agrees on how to use the results to make decisions, the resulting process may select decisions that no one agrees with. Suppose a group decides whether to fund a levee based on its expected benefit using the unanimously agreed-to rule: build levee if and only if P(flood) × P(levee works | flood) > 0.2. They further agree to average their probabilities to obtain the group probabilities to be used in decision-making. Half the group believes that these probabilities are high/low (e.g., 0.8 for P(flood) and 0.2 for P(levee works | flood)); the other half believes the reverse (0.2 and 0.8). All individuals agree that

the joint probability is 0.8*0.2 = 0.2*0.8 = 0.16 which is below the acceptance threshold of 0.2, so no one supports funding the levee. But if the group averages the components separately—P(flood) = (0.8 + 0.2)/2 = 0.5; P(levee works) = (0.2 + 0.8)/2 = 0.5—then the group concludes that the expected benefit is 0.25 and decides to fund the project. The outcome is one that no individual believes is justified. Nehring shows that such failures of probabilistic compositionality are not rare anomalies but structurally inevitable under plausible aggregation rules. Even when individuals are unanimous in their overall decisions, aggregating probabilistic beliefs component-by-component can produce group decisions that everyone individually rejects. This reflects a broader theme: locally rational or consistent beliefs, when aggregated mechanically, may not yield globally acceptable collective decisions. (However, Pivato (2022) shows that if individuals learn from shared evidence and their beliefs converge over time, then it is possible to justify utilitarian group decisions based solely on preferences. This avoids the need to aggregate probabilistic beliefs and sidesteps the paradoxes identified by Hylland–Zeckhauser and Nehring.)

14. **Moral and value aggregation failures.** In public policy and risk regulation, decisions often involve balancing fundamentally different types of harm—such as deaths, economic losses, disabilities, or environmental degradation. While each stakeholder or expert may offer consistent and well-reasoned judgments within a specific domain, attempts to combine these judgments into a single decision or ranking can fail. For example, Thomas Schelling (1984) described how students judged the fairness of a tax policy involving children. When framed as a deduction for families with children, students objected to giving wealthier families larger deductions. But when reframed as a surcharge on childless couples, they supported higher “surcharges” for wealthier families who remained childless—albeit this is logically equivalent to higher “deductions” for those families if they had children. This illustrates how people’s fairness judgments can shift dramatically depending on how a policy is described. Similar effects arise in policy settings where multiple outcomes must be weighed—such as prioritizing climate adaptation, public health, or biodiversity. The result is often unstable or conflicting judgments, where small changes in framing or weighting can lead to sharp disagreements or reversals in collective priorities. These breakdowns show that even when individual judgments are consistent, combining them into a coherent, widely accepted decision may not be possible.

15. **Persistent disagreement despite shared evidence.** Even when parties agree on goals and share available data, differences in assumptions, models, or frameworks can make persistent disagreement inevitable. Aumann (1976) showed that fully rational agents with a shared understanding of prior probabilities and evidence who communicate their posterior probabilities to each other (thereby potentially precipitating further updates in the posterior distributions) cannot continue to disagree about posterior probabilities. Their posterior beliefs must end up being identical. Therefore, enduring disagreement often signals deeper differences in background beliefs or institutional perspectives. Yet persistent disagreement is common in health, safety, and environmental (HS&E) risk analysis—about, for example, the cancer risk from a chemical, the strength of a dose-response relationship, or the credibility of an epidemiological association. Aumann's result suggests that these disagreements may reflect differences in background assumptions, modeling choices, or prior beliefs, not irrationality. Even when all sides are acting in good faith and share common goals and information, incentive mismatches, private information, and divergent epistemic frameworks can block consensus and coordinated action, as when agencies hesitate to lower regulatory thresholds without strong causal evidence, while community groups rely on precautionary reasoning or experiential knowledge. These insights may help to explain real-world impasses in policy design, regulation, and risk communication. Trust, transparency, and aligned incentives may be prerequisites—not consequences—of successful risk governance.

16. **Collective action failures.** In many public health, environmental, and safety contexts, individuals or institutions must make decisions that affect not only themselves but also others who share common risks or resources. Yet when each actor chooses the alternative that is best for itself individually—such as avoiding costly pollution controls, conserving less water, or neglecting maintenance on shared infrastructure—the collective result can be worse for everyone. These dynamics, formalized in classic game-theoretic models such as the Prisoner's Dilemma (Axelrod, 1984; Ostrom, 1990), the free rider problem (ibid), and the tragedy of the commons (Hardin, 1968), show how locally rational decisions (e.g., making "dominant" choices that are best responses no matter what others do) can lead to globally irrational (Pareto-dominated) outcomes (i.e., outcomes that are less good for everyone than those that would have been obtained had each individual selected dominated choices). In these settings, the structure of incentives discourages cooperation—even when all parties would benefit from coordinated restraint or investment. These dilemmas are central to risk governance challenges ranging from climate change to chemical regulation to pandemic response.

Local communities can sometimes overcome such dilemmas by developing shared norms, monitoring, and enforcement mechanisms (Ostrom, 2000), but success requires institutional arrangements that support trust, reciprocity, and accountability. In formal terms, each actor's strategy may be locally optimal given its own information, choices, and incentives, but the resulting set of behaviors cannot be combined into a globally coherent and beneficial (Pareto-efficient) outcome—a classic failure of compositionality between individual rationality and collective welfare.

Again, however, sheaf-theoretic thinking suggests that this failure should not be, and need not be, the end of the story. Just as Arrow's impossibility result can be overcome by carefully redesigning the inputs, outputs, and structural constraints of aggregation, so too can the apparent inevitability of collective action failure in one-shot games be overcome by rethinking the temporal and informational structure of interaction. In particular, the Folk Theorems of repeated game theory (Fudenberg and Maskin, 1986; Tsodikovich et al., 2024) show that, under the right conditions, sustained cooperation is not only possible but rational—even when defection would be dominant in a single play. If interactions are repeated indefinitely (or with uncertain duration), and if players can observe and respond to each other's actions over time, then a wide range of cooperative outcomes can emerge as stable equilibria. These results apply not only to stylized settings like the repeated Prisoner's Dilemma, but to many real-world challenges—such as climate policy, arms control, and pandemic containment—where the same actors face interdependent risks repeatedly and can build trust, reciprocity, and reputational incentives over time.

Mechanisms such as altruistic punishment and shared moral norms, reputation systems, binding commitments, enforceable contracts, and transparent monitoring can all expand the set of achievable cooperative outcomes (Fudenberg and Maskin, 1986; Ostrom, 1990; Mailath and Samuelson, 2006; Gintis, 2009). From a sheaf-theoretic perspective, these results reflect a shift in the underlying restriction maps: instead of treating each actor's decision in isolation, repeated games embed choices in a broader temporal and informational context that enables the construction of globally coherent strategies. In this way, the path from local incentives to global cooperation depends not only on payoff structures, but on the design of communication channels, observation rules, institutional scaffolding, and memory—all of which help make cooperation sustainable and rational. As with other areas of decision analysis, the key is to recognize and repair the structural misalignments that prevent local choices from composing into globally desirable outcomes.

17. **Informational cascades and herding.** In decision settings with strong social influences on risk perceptions and behaviors—such as vaccine adoption, financial investments, or responses to public warnings about environmental or natural disaster risks—individuals may rationally base their actions in part on the observed behavior of others rather than on their own private information. If early decisions are visible and later agents believe those earlier actors acted on useful information, a phenomenon known as an *informational cascade* can occur (Bikhchandani et al., 1992). Once a few early decisions point in one direction, subsequent actors may ignore their own evidence and follow the apparent trend. For example, early conflicting or dismissive statements by health authorities about the effectiveness of face masks against COVID-19 led to widespread public skepticism—even after new evidence emerged and official recommendations shifted. Many individuals, observing the choices of peers and officials, ignored their own concerns or data. Each such decision may be locally rational, given the available information and observed behavior, but the resulting cascade can entrench globally suboptimal outcomes (Banerjee, 1992). Moreover, cascades can be fragile: minor changes in early signals or perceived consensus can reverse them.

    This dynamic aligns with the Social Amplification of Risk Framework (SARF) of Kasperson and colleagues (1988), which explains how psychological, social, and institutional factors can amplify or dampen public responses to risk signals. SARF posits that messages about hazards are filtered through a network of "social stations" (e.g., media, government agencies, peer networks), where they may be exaggerated, suppressed, or reframed. These distortions can lead to disproportionate public reactions, secondary effects (like stigma or market volatility), or lasting changes in behavior. As reaffirmed in a 2022 retrospective (Kasperson et al., 2022), SARF continues to provide a powerful integrative framework for understanding how psychological, social, and institutional factors interact with technical risk information to amplify or attenuate public responses—especially in a digitally mediated world. The 2022 update to SARF recognizes social media as a dominant 'amplification station,' blurring traditional boundaries between expert, media, and public roles in shaping and spreading risk perceptions. Computational amplification (algorithm-driven ranking and personalization) adds layers of complexity to risk signal processing not envisioned in the original 1988 SARF. Amplified perceptions and behaviors can feed back into real-world impacts—what SARF terms 'risk ripples'—including economic disruption, institutional distrust, and policy shifts, via societal responses such as stigma, trust loss, and political backlash.

    Milli (2021) provides a formal agent-based model that gives operational structure to many of these SARF concepts. In the model, individuals update their opinions based on others' views, but only if those views fall within their bounded confidence interval—representing a cognitive threshold of interpretability. Social influence is modulated by trust and homophily, meaning agents are more influenced by those they already agree with. The model shows how informational cascades, persistent polarization, and belief fragmentation can arise endogenously—even when agents are rational and start from well-calibrated beliefs. Trust structures, information echo chambers, and the presence of "bridge agents" who connect otherwise disconnected groups all influence whether society converges on shared understanding or fractures into competing narratives.

These findings reinforce and refine the SARF prediction that amplification is not merely a media effect, but a function of the structure and content of interpersonal and institutional communication networks. They also suggest that belief updating in social systems is not inherently self-correcting, especially when early signals are noisy, agents overweight social consensus, or network structures prevent integration of divergent information. Indeed, ass Lyons et al. (2021) show, individuals most confident in their ability to detect misinformation are often the most vulnerable to spreading it—suggesting that even well-intentioned local agents may produce system-wide incoherence. From a sheaf-theoretic perspective, these patterns exemplify a failure of compositionality in collective inference: locally justified beliefs, when passed through distorted or incomplete restriction maps (e.g., trust-weighted social influence), do not glue into a consistent global understanding. Instead, partial observations can propagate misinformation, suppress contradictory evidence, or overfit to early noise. As such, even rational decision-makers can contribute to and sustain globally irrational or unstable information states.

Recent work by Jeong and Shin (2022) suggests constructive methods for repairing such failures by designing optimal interventions in influence networks. They apply a linear algebra technique (singular value decomposition (SVD) of the influence matrix) to data on who influences whom to identify key structural features—"hub" and "authority" vectors—that determine how opinions spread and are received. Targeting interventions (e.g., implanted messages or informational nudges) toward agents with high hub centrality aligned with desirable target beliefs, and away from those whose initial beliefs reinforce misinformation, can shift the global opinion trajectory more effectively. This mathematical formalism complements SARF's sociological framing by quantifying how budget-constrained interventions can steer network learning dynamics toward more desired equilibria. Identifying and modifying the structural properties of belief propagation—such as strengthening bridge nodes, tuning influence weights, or inserting corrective agents—may thus offer pathways to reducing informational cascades and improving collective resilience to misinformation.

These insights offer a promising integration of computational and sociological models for understanding and shaping public responses to risk.

As mentioned in Section 3, a closely related line of work by Ghrist and Hansen (2021) introduces the concept of *discourse sheaves*, a topological framework for modeling how opinions, evidence, and arguments propagate through social networks. In this framework, beliefs are treated as structured data attached to nodes, while communication constraints (e.g., norms, translation effects, censorship, strategic misrepresentation) are modeled as restriction maps. Agents may engage in strategic filtering or distortion of expressed opinions: rather than transmitting full private beliefs, agents may soften, omit, or reframe what they share depending on social context, audience receptivity, or strategic goals. This generalization allows sheaf-theoretic models to capture both unintentional miscommunication and deliberate misinformation. It further illustrates how failures of gluing may result not only from epistemic fragmentation or semantic mismatch, but from strategic filtering behaviors that prevent belief alignment, even when underlying opinions are compatible. These extensions offer a powerful conceptual bridge between topological representations of knowledge and empirical models of trust, deception, and belief manipulation in risk communication.

In slightly more detail, and using the concepts and terminology of sheaf theory from Section 3, agents (represented by nodes in the social network) hold private beliefs. These beliefs may be represented by full vectors, rather than as single numbers, as in simpler and less general models. The agents send filtered or transformed versions of their true private beliefs to neighbors with whom they communicate via restriction maps on network edges. The model defines a sheaf Laplacian to simulate belief updating as a diffusion process constrained by these communication structures. This allows the system to converge—when feasible—to a globally consistent belief profile (a global section). When convergence fails, the sheaf cohomology can reveal structural obstructions to agreement, such as “opinion singularities” where beliefs cannot be reconciled. The framework also accommodates strategic communication behaviors, in which agents adjust not only their beliefs but how they express their beliefs to others—such as softening, omitting, or rephrasing opinions to align with norms or to manipulate perceptions—to avoid conflict or achieve apparent consensus. These dynamics allow sheaf-theoretic models to capture both unintentional miscommunication and deliberate information distortion, and to diagnose how failures of gluing may arise from strategic filtering, even when underlying beliefs are locally compatible. As such, discourse sheaves offer a powerful bridge between formal representations of knowledge and real-world phenomena of trust, deception, polarization, and belief manipulation in risk communication.

The preceding examples show how failures of global coherence can arise in collective decision-making—even when local choices are individually rational and evidence-based. If information is not only fragmented but also actively distorted—as in cases of misinformation or disinformation—the structural barriers to coherence become even more severe. In such contexts, sheaf-theoretic tools may help analysts move from diagnosing information misalignment to designing risk analysis and decision support systems that are robust to adversarial manipulation, semantic incompatibility, and selective trust.

Concluding Note

This Supplement has provided the background narratives and illustrative cases that lie behind the summary in Table 5 of the main article. By tracing each challenge from its origins in fragmented knowledge or inconsistent local reasoning to its diagnosis through sheaf-theoretic analysis, we have sought to make visible the rationale for the entries in Table 5 and to show how a sheaf-theoretic perspective may help risk analysts understand why coherence so often breaks down, and how it might sometimes be restored. Together, Table 5 and this Supplement are intended to offer examples and points of departure for those interested in applying sheaf-theoretic ideas to their own domains of risk analysis and decision-making.