ENGINEERING COMMONALITY

Jean Tirole[*]

December 18, 2025

Abstract: Perceived commonality—rooted in shared interests, parochial altruism, durable relationships, or common priors—is a critical lubricant of collective action in both organizations and societies. Cohesion emerges when organizations cultivate narratives of shared purpose, regulate acceptable discourse, and foster interpersonal bonding.

This paper develops a framework in which two team members' engagement rises with their perceived congruence, and payoffs are represented by a convex-concave function of that congruence The agents learn about each other through i.i.d. experiments. The analysis first characterizes the optimal level of information acquisition through public experiments from each player's perspective. It then studies the strategic communication game that unfolds when information sharing is unmediated. The equilibrium outcome depends critically on the nature of the information: when information is hard, the equilibrium favors the less information averse agent; when it is soft, it favors—at best—the other agent.

Keywords: commonality, teamwork, shared attributes, acceptable discourse, weak link, information design.

JEL numbers: D23, D82, L23, M50.

[*] Toulouse School of Economics, University of Toulouse Capitole, and Institute for Advanced Study in Toulouse, Toulouse, France. I am grateful to the European Union (ERC, HWS, grant No 101098319) for financial support and to Johannes Hörner, Paul-Henri Moisson, and Mohamed Saleh as well as participants in seminars and conferences (Getulio Vargas, Imperial College, Lindau, MIT Sloan, Stanford CASBS, TSE, UC Berkeley, and ULB) for helpful reactions to a preliminary version of these ideas.

Georg Simmel (1906): *"All relationships of people to each other rest, as a matter of course, upon the precondition that they know something about each other."*

Ernest Renan (address, Sorbonne 1882): *"Now the essence of a nation is that all individuals have many things in common, and also that they have forgotten many things."*

1 Introduction

A recurrent trait of societies and organizations is their emphasis on what unites their members, and their avoidance of what divides.[1] They search and rehearse commonality narratives and shun topics that might sow dissension among their members, where "commonality" refers to anything that will induce cooperation (objective convergence of interests, kinship and other hard-wired senses of similarity).

The framework developed here builds on the classic idea that mutual trust is the lubricant that makes organizations and society operable, by making their members work in unison.[2] They are three routes for this. First, there is a strongly rational component: You will invest in a joint project if you feel that we share the same goal with regards to the desirable orientation for the project (and so we will pull in the same direction) or that our relationship has a long-enough horizon, so our interests will be aligned as the project develops. You will share private information with me if you trust me not to disclose it to your rival. You may delegate to me if you believe I'm dependable.[3] Finally, having common beliefs or a common culture may help us coordinate.

Second, and at the opposite end of the rationality spectrum, the perception of trust based on common attributes can be hard-wired, as demonstrated by the minimal-group paradigm (in which subjects are united by a relative preference for Klee over Kandinsky or even by a random draw of the same T-shirt color). You may experience empathy if you perceive me as similar along some dimension. A national anthem (played repeatedly during periods of military tensions), Olympic games, or perceived common roots can unite a population, as can shared origins, passions or hobbies unite a group of individuals. Relatedly, investment

[1] Workmates often avoid discussing politics, sexuality, religion, social roles, cultural values, or more broadly topics that in their particular society and epoch may alienate fellow members. What constitutes "acceptable discourse" is formalized in rules that limit members' ability to address such subjects. In the US, "divisive concepts" laws control how teachers talk about divisive issues of race, sex, ethnicity, religion, color, or national origin, and the "don't ask, don't tell" doctrine in force until 2011 was meant to prevent conflict and discrimination in the army. The same holds for corporate America. In 2019, Google attempted to quell political discord in its ranks by posting "community guidelines" advising workers to avoid disruptive conversations. Its "no-politics at work" rule also encouraged employees to flag problematic posts internally and enlisted moderators to monitor its chat boards. Ignoring the normative side of such policies (which may involve externalities on third parties), they demonstrate the existence of an organizational demand for preserving harmony among members.

[2] To be certain, mutual trust can also be fostered in traditional, incentive-based ways, for example through team incentives which, when doable, increase the individual benefits from cooperation (Itoh 1991, Che-Yoo 2001) or through direct subsidies for collaboration, like those granted for building European research networks, which reduce the cost of collaboration. Here, collaboration is grounded in information.

[3] Bloom et al (2012).

in public goods is low in ethnically divided communities;[4] and a common ethnicity, political party, or subjects' field of study improves behavior in prisoner's dilemma or other lab games.[5]

Third, discovered commonality can also matter indirectly. Employees with affinities and common worldviews meet more frequently, whether at the coffee machine, in joint undertakings or in non-work contexts. They are then involved in a broader game than just their strict work assignment and depend more on keeping a good relationship with each other.[6]

We posit that agents' performance in a given task (work, combat, proselytism...) depends on their perceived congruence for that particular task. While one can never be certain that one's collaborators/peers will be pulling in the same direction, one can use one's perceptions about their overall preferences as a predictor of future behavior: The perceived congruence can be predicted by how akin their views or interests are aligned in other domains. The challenge for the organization (the boss or the agents themselves) is to determine how much information they should share to shed light on the congruence for the task.

Model and insights. Section 2 sets up the framework. Two agents play a constituent game exhibiting strategic complementarities between actions (an agent's effort is more valuable, the higher the other agent's effort) and between actions and the state of nature (an agent's effort is more valuable, the more aligned the agents' interests). Furthermore, efforts exert non-negative externalities (an agent's effort may benefit the other agent). This situation is descriptive of many communities that rely on the contributions of their members. The outcome of the interaction depends on the members' perceived congruence, a higher perception raising the incentive to exert effort.

The members may have symmetrical preferences; if not, there is a strong link and a weak one. The strong link is more eager to cooperate (team up, exert effort benefiting the team) than the weak one. As I show, being the strong link may be associated with a low reservation utility, a low cost of effort, a higher degree of kin altruism, a higher prior regarding congruence (if priors differ), a lower impatience, or a higher valuation of ancillary interactions.

I append pre-game communication to allow the two members to learn about their fit prior to the actual interaction. As a building block, Section 3 formalizes this communication as a sequence of i.i.d. *public* experiments, and so agents update their beliefs in sync. Examples of such experiments are the bad-news model (the agents may learn that they are dissonant), the good-news model (they may learn that they see eye-to-eye), a choice between searching for bad or good news, and Bernoulli experiments.

Public experiments at this stage may stand for the commitment to a centralized release by the principal of employees characteristics. Alternatively, and as will be stressed in Section 4, agents supply information about their tastes or prior choices, and they are assumed for the moment to be averse to lying or retaining information when the protocol requires them to disclose it.

[4] Alesina et al (1999).

[5] Fershtman-Gneezy (2001), Glaeser et al (2000).

[6] See illustration (e) below. Social (non-work) dinners at home make agents discover the human side of each other, and raise mutual empathy.

An information designer (the principal, one of the agents, a social norm...) builds the communication protocol. This is standard information design,[7] up to a twist: The latter focuses on disclosure by a designer who can commit to reveal any coarse partition of the information she will acquire (so, the only constraint on posterior beliefs is Bayes rule). Here the public-experiment requirement implies limited ability to build arbitrary (but Bayes-consistent) beliefs. With this in mind, Section 3 characterizes the constrained-optimal communication protocol from the point of view of each of the three players, which will be the building block for the subsequent analysis. The resulting information structure, and therefore outcome may or may not be the ones that an unconstrained information designer with the player's preferences would choose for a public experiment. For example, a bad-news experiment with slow learning (in each elementary experiment, the posterior either jumps to 0 or increases slightly) allows the designer to reach the player's unconstrained information design. The reason is that the unconstrained optimum for a convex-concave value function involves posteriors at 0 and another posterior belief. The bad-news experiment with slow learning is powerful precisely because it naturally generates these two posteriors (in particular the belief 0), making it a suitable tool to achieve the desired outcome.

Whether the unconstrained information design can be implemented or not, experimentation optimally continues until the posterior probability of congruence reaches some threshold, and stops beyond this threshold. Agents can therefore be ranked according to their information aversion (minus their threshold). I provide process-contingent conditions for the strong link agent to be more information averse than the weak link one.

Section 4 builds on this analysis but now assumes that the information about congruence is decentralized, i.e. privately held by the agents. This information has T dimensions, each bringing some information about the alignment of interests in the constituent game. The introduction of this type space is a natural way to decentralize the information structure from Section 3, thereby setting the stage for analyzing the agents' incentives to share these signals truthfully.

Indeed, we assume that the information-designer can no longer control information flows, introducing two types of incentive compatibility. First, agents may communicate more or less than the information designer would wish: They may exchange information that jeopardizes their collaboration and the interests of the overall organization; conversely, they may eschew necessary teambuilding activities. Agents may further disagree on the amount of information to be exchanged. Second, agents may either lie (soft information) or refrain from disclosing information about their preferences (hard information).

We assume that the two agents engage in a freewheeling information-exchange game, that stops when the two parties no longer make disclosures. We obtain two results:

Hard information. When information is hard, agents' only alley for misrepresentation is to retain their information. I show that the outcome that is best for the less information-averse is, under a mild refinement, the unique equilibrium of the information-exchange game.

[7] The unconstrained information design problem has been the object of much analysis, especially but not uniquely in the literature in the one-agent context (e.g. Kamenica-Gentzkow 2011, and more recently Dworczak-Martini 2019 and Inostroza-Pavan 2025).

The intuition for this result is a novel version of the classic unraveling result. The less information-averse agent wants weakly more information exchange than the other agent. When the latter wants to stop exchanging information, the former can challenge her partner by disclosing more information; the more information-averse agent is then confronted with a choice between a reassuring disclosure indicating congruence along the extra dimension and a deafening silence. In the end, the less information-averse agent has her say.

Soft information. When the information about their own preferences or identity is soft, the agents may gain from misrepresenting them so as to manipulate their collaborators' behavior. Namely, whenever their collaborator's engagement generates a positive externality, they want to nurture trust. This makes sequential communication in general ineffective, with each agent "pandering" to the other agent's preferences. Under simultaneous communication in contrast, which agent gets the upper hand on the other is now reversed relative to hard information: The communication outcome that is optimal for the more information-averse player is an equilibrium outcome with soft information. If the more information-averse player deems extra communication not to be in his interest, he can (a) refrain from initiating further communication and (b) systematically offer the congruent answer to a disclosure made by the less information-averse player, thus deterring any additional information exchange.

Whether information is hard or soft, the intuition for truth telling until the communication stops goes as follows. First, strategic complementarity between the actions creates gains from synchronicity and therefore from shared beliefs. Being either both enthusiastic or both doubtful about the outcome of the task dominates one agent being enthusiastic and the other lukewarm. Common knowledge of beliefs is therefore a plus and lying to the other destroys the possibility of aligned incentives. Second, the matching between the alignment of interests and actual contributions is best achieved through truth telling. The conjunction of these two imperatives (correlate the agents' intrinsic motivations both between themselves and with the occurrence of congruence) can be viewed as a motivational synchronicity principle.

Section 5 concludes with avenues for future research.

Literature review. The paper is related to multiple strands of the academic literature.

Gains from congruence. The idea that congruence among its members benefits an organization is familiar to economists. Congruence smoothes out the exchange of both soft and hard information (starting with Crawford-Sobel 1982, Milgrom 1981, and Dye 1985), facilitates extensive delegation of decision rights (starting with Holmström 1977), enhances agents' coordination (Dewan-Myatt 2008), aligns their beliefs (Van den Steen 2010), raises their empathy (Rotemberg-Saloner 2000), expedites decision-making and boosts engagement in hierarchies, committees and cooperatives (Hansmann 1996, Aghion-Tirole 1997), and more broadly reduces agency costs in principal-agent relationships. These economics literatures echo work in other fields, that emphasizes the importance of knowing each other (e.g., Simmel 1906, Barnby et al 2022), of psychological safety which allows members to admit mistakes and criticize projects (Schein-Bennis 1965, Edmonson 1999) and of social capital (Putnam 1995). While research has pointed at the limits of congruence in benefiting the organization,[8] this

[8] Too much alignment may generate incentive issues. First, it may give rise to groupthink (Janis 1972,

literature offers ample motivation for the model's benefit-from-congruence building block. The current paper opens the black box of how this trust arises.

Communication protocols. Building on the foundational work on optimal stopping rules starting with Wald (1947) and Arrow et al (1949), an interesting literature, which is too large to be covered exhaustively, studies the design of incentive-compatible, back-and-forth information-exchange protocols when the receiver may use the information in a direction that does not benefit the sender(s). The sender may abscond with the information without paying for it (Anton-Yao 2002, Hörner-Skrzypacz 2016) or may veto a decision after overhearing the information exchange between agents with aligned interests (Antic et al 2022). Alternatively, agents with dissonant preferences may refrain from exchanging information that is useful to the other's decision-making (Orlov et al 2025), or may alter the decision of a committee (Alonso-Camara 2016, Caillaud-Tirole 2007, as well as the literature on information design with multiple receivers, e.g. Taneva 2019 and Mathevet et al 2020). Some of that literature is related to that on sequential contribution games and crowdfunding (starting with Admati-Perry 1991).

The paper builds on innovative work by Augenblick-Bodoh-Creed (2018) and Dziuda-Gradwohl (2015) on the design of a communication protocol.[9] This paper's contribution relative to theirs is three-fold: (1) Agents can game communication through preference misrepresentation and disobedience of the protocol, while most of their analysis presumes truthtelling. (2) Common knowledge about full congruence is not assumed to be needed for the team to function, which implies that there can be too much communication. (3) It considers a large

Bénabou 2013). Second, it may lead individuals to take others for granted as they will rubberstamp their suggestions even in the absence of argumentation: Congruence induces real authority and therefore little information exchange, as the latter might lead others to not rubberstamp (Dewatripont-Tirole 2005). Third, alignment may induce herd behavior ("what like-minded people have chosen is probably best for me") and thereby dampen incentives to experiment with new actions/projects. Fourth, it chills incentives to provide "semi-public goods", inducing free riding: "Others will pull in a direction I like, given they have the same taste" (in contrast, when different tastes, more incentives to produce preferred brand of public good or version that will bring personal recognition). Finally, alignment, if created by hiring similar people, may have non-incentive costs if it induces a lack of diversity when complementary skills and a diversity of cognitive repertoires are needed (Page in his 2019 book *The Diversity Bonus* reminds us of Hackman's Goldilocks principle: organizations should be neither too homogenous nor too heterogenous). The diversity point though is in part orthogonal as our analysis is built on the idea that we want trust/shared values/etc within the community, whether it is diverse or not.

[9] A receiver wants to screen out a non-matching partner; a sender truthfully reveals information about a set of independent binary traits. In both papers, a match requires congruence along *all* dimensions. The models are bad-news ones: negotiations stop when one learns the absence of a match in one dimension (Augenblick-Bodoh Creed) or when one discovers that the partner is unviable (Dziuda-Gradwohl). Players cannot lie and cannot deviate from the protocol defined prior to communication (Augenblick and Bodoh-Creed also study the complex signaling patterns when agents can deviate from the communication protocol). Assuming that there is a cost to the sender of revealing a trait (trade secrets or ideas are disclosed to a potential rival), the papers show that the best protocol involves sequential disclosures, stops if any trait reveals dissonance, and especially that the order of trait disclosure goes according to increasing information value. While the modeling differs from that in my paper (in these two papers, there is a disclosure penalty, and matching occurs only if perfect congruence is common knowledge), so does the focus on gaming through preference misrepresentation and disobedience of the communication protocol. The two papers in contrast to mine allow the size of disclosed information in a given round to be endogenous (Dziuda-Gradwohl) or parties to have private information about the likely of congruence (Augenblick-Bodoh-Creed).

class of experiments while these two papers focus on bad-news ones, and shows which results are affected by the nature of the experiments.

The paper connects, in a somewhat looser way, to several other literatures:

Acceptable discourse. The paper also relates to the study of acceptable discourse and political correctness (e.g., Loury 1994, Morris 2001, Daughety-Reinganum 2010, Sunstein 2018, Braghieri 2024, Goldman 2021).[10] For example, whether someone is willing to state an anti-affirmative-action opinion depends on whether others are as well; for, this opinion can either be indicative of racism or stem from beliefs about the policy's efficiency; due to this signal-extraction problem, multiple equilibria are then natural. This literature analyzes the interpretation of one agent's words or actions. This paper differs in its focus on mutual learning for the benefit of an organization and its members, and on the shared demand for cohesiveness.

Screening and motivating team members. Through its emphasis on the exchange of information between members of an organization, this paper is complementary to frameworks that stress the need to screen properly the members; however such screening usually refers to finding the right person for the job, either in terms of competence or in terms of attitude towards the job (see Prendergast 2007 for the latter). Some other papers emphasize the need to boost agents' incentives by manipulating their information. For example, in Dessi (2008), agents in an organization exert positive externalities on each other (teamwork). The principal thus tries to make the agents optimistic about their productivity, so that they work hard and benefit each other, even when they would not have done so had they known the true productivity of their effort.

Pre-play communication. Another branch of the literature studies the exchange of information prior to playing a game. Farrell-Gibbons (1989) shows that cheap talk prior to bargaining may alter the bargaining outcome. Aumann-Hart (2003) and Forges-Koessler (2008) characterize the set of equilibrium payoffs achievable under unmediated communication in persuasion games.

Collusive behavior within organizations. In characterizing the outcome of unmediated communication between agents, Section 4 contributes to the literature of collusive behavior against a principal, see, e.g., Tirole (1986) on collusion at nexi of information and Laffont-Martimort (1997) on how asymmetric information among agents limits collusion among them.

2 Model

We consider a two-stage game. The first stage involves information exchange. In the second stage, agents play a real game within the organization, resulting in reduced-form payoffs that

[10] This literature usually assumes that the distribution of preferences is known. Building on the work of psychologists on pluralistic ignorance (e.g., Katz-Allport 1931 and Prentice-Miller 1993), economists (e.g., Bursztyn et al 2020, Fernandez-Duque 2022) have looked theoretically and empirically at how misperceptions affect behavior in a world in which individual actions are (in part) driven by image concerns.

we later interpret.

2.1 Payoffs

Preferences. An organization is composed of three players, two agents ($i = 1, 2$) and a principal ($i = P$). The organization functions better if the two agents perceive themselves as aligned. As discussed in the introduction and formalized below, there are multiple reasons for why this may be so. Let $\mu \in [0, 1]$ denote their perceived congruence in the dimension that will determine future payoffs. The prior probability of congruence is μ_0. For now, we will assume that perceived congruence is the same for both agents and common knowledge among them.

Equipped with posterior μ, the two agents play a "second-stage game" whose outcome depends on the posterior beliefs. Let $V_i(\mu)$ denote the objective function of player $i \in \{P, 1, 2\}$.

Assumption 1 ***(reduced-form payoffs).***

(i) Benefits of commonality: The reduced-form payoff functions $V_i(\mu)$ are weakly increasing, piecewise C^2 in the agents' perception μ of their congruence, and satisfy $V_i(0) = 0$.[11]

(ii) Convex-concave payoffs: $V_i(\mu)$ is convex on $[0, \tilde{\mu}_i]$ and concave on $[\tilde{\mu}_i, 1]$ ($\tilde{\mu}_i \in [0, 1]$ is the inflection point if V_i is smooth).

Convex-concave (S-shaped) payoffs capture the idea that effective relationships must build on sufficient mutual ground, but do not necessarily require the agents to see eye to eye on everything. Convex or concave payoffs are special cases of convex-concave ones. Appendix A considers general shapes for increasing functions $V_i(\mu)$.

We illustrate $V_i(\mu)$ in the context of a binary-action game. Equipped with common posterior beliefs μ, each agent i can exert effort/invest/cooperate/engage ($a_i = 1$) or not ($a_i = 0$). Let $\omega = 1$ if the two agents are congruent (which has probability μ at the beginning of stage 2) and $\omega = 0$ if they are dissonant (probability $1 - \mu$). Agent i's payoff is $u_i(a_i, a_j, \omega)$. The "symmetric case" will refer to the case in which the agents' payoff functions are identical up to the permutation of a_1 and a_2.

We will call *congruence uncertainty* the case in which all parameters of the game, except ω, are common knowledge. This case will give rise to two threshold beliefs $\underline{\mu}_i$ below which the agents do not want to exert effort/form a team even if the other agent does. *Deep uncertainty* will mean that the agents' initial lack of knowledge concerns not only ω, but also some other (uncorrelated) parameter of the payoff functions that they will learn between stages 1 and 2. Deep uncertainty, unlike congruence uncertainty, will give rise to smooth reduced-form payoff functions $V_i(\mu)$.

[11] That $V_i(0) = 0$ is a normalization; for, one can always redefine $\tilde{V}_i(\mu) \equiv V_i(\mu) - V_i(0)$ for all μ and apply Assumption 1 to $\tilde{V}_i$.

Congruence uncertainty. Under congruence uncertainty, the utilities, up to the realization of ω, are known at the stage of information acquisition. Agent i's incentive to cooperate when agent j picks a_j is

$$\delta_i(a_j, \omega) \equiv u_i(1, a_j, \omega) - u_i(0, a_j, \omega).$$

Assumption 2 ***(micro-foundations of payoff functions).***

Under congruence uncertainty,

(i) (Supermodularity) payoff functions $u_i(a_i, a_j, \omega)$, $i \in \{P, 1, 2\}$ are supermodular.

(ii) (Non-negative externalities) $u_i(a_i, a_j, \omega)$ is (weakly) increasing in a_j, for all a_i and ω.

To reduce the number of cases, and without loss of insight, we further posit:

(iii) Under common knowledge of congruence (dissonance), agents are willing (not willing) to cooperate if the other does: for all i,

$$\delta_i(1, 1) > 0 > \delta_i(1, 0).$$

(iv) Agent i does not cooperate ($a_i = 0$) when expecting the other agent to not cooperate ($a_j = 0$): $\delta_i(0, \omega) < 0$ for all ω.

The supermodularity of u_i (Assumption 2(i)) is illustrated by the functional form

$$u_i(a_i, a_j, \omega) = \alpha \frac{a_1 + a_2}{2} + \beta \omega a_1 a_2 - \gamma_i a_i \tag{1}$$

where $\beta > 0$ is agent i's benefit from cooperation when congruent, $\alpha > 0$, and γ_i is agent i's cost of effort. It embodies several natural properties: strategic complementarity ($\delta_i(1, \omega) \geq \delta_i(0, \omega)$), incentive-enhancing congruence ($\delta_i(a_j, 1) \geq \delta_i(a_j, 0)$), and positive externalities. Assumption 2(ii) implies that an agent may benefit from making the other agent optimistic about their congruence, which affects disclosure incentives in Section 4, where we investigate incentives for truthful revelation.

As is well-known (Milgrom-Roberts 1990), Assumption 2(i) implies that for common beliefs about ω, there exist (possibly identical) largest and smallest serially undominated strategies $\bar{\boldsymbol{a}}$ and $\underline{\boldsymbol{a}}$ where $\boldsymbol{a} \equiv (a_1, a_2)$. Both strategy profiles are pure Nash equilibrium profiles. In case of multiplicity of equilibria, we will select the Pareto-dominant one, $\bar{\boldsymbol{a}} = (1, 1)$. We will thus assume away coordination failures, an issue that is orthogonal to the focus of this paper.

Assumption 2(i) also implies that agents are more likely to collaborate if they are optimistic about their congruence. Letting μ denote the common knowledge beliefs of agent i, the

(Pareto-dominant) equilibrium involves cooperation ($a_i = a_j = 1$) if and only if, for all i,[12]

$$\mu\delta_i(1,1) + (1-\mu)\delta_i(1,0) \geq 0.$$

There exists $\underline{\mu}_i \in (0,1)$ such that agent i is conditionally willing to cooperate if and only if $\mu \geq \underline{\mu}_i$ where $\underline{\mu}_i$ satisfies this condition with equality. Cooperation thus occurs if and only if

$$\mu \geq \max_i \left\{\underline{\mu}_i\right\}.$$

Due to complementarities, the organization is only as strong as its weakest link, that is the agent with the highest need for congruence. Without loss of generality, agent 2 is the weak link: $\underline{\mu}_2 \geq \underline{\mu}_1$:

Assumption 3 ***(strong and weak links).*** *Under congruence uncertainty, agent 2 is the weak link: For all $(a,\omega) \in \{0,1\}^2$: $\delta_1(a,\omega) \geq \delta_2(a,\omega)$.*

Although our focus will be on the agents' preferences, we can also consider the principal's. As the agents internalize their cost of effort, it may make sense to assume that the principal, who does not exert any, is the strongest link of all.[13]

Finally, part (iv) of Assumption 2 (agent i, when expecting that the other agent will not engage, i.e. $a_j = 0$, does not engage either) is for convenience –it avoids having to describe multiple cases for each illustration– and does not impact any insight.

Definition 1 ***(motivational synchronicity under congruence uncertainty).*** *Under congruence uncertainty, let $p(a_1, a_2 \mid \omega)$ denote the probability that agents 1 and 2 pick a_1 and a_2 in state of nature ω. The allocation p satisfies motivational synchronicity if*

(i) $p(1,1 \mid 1) = 1$: the two agents always collaborate when congruent ($\omega = 1$);

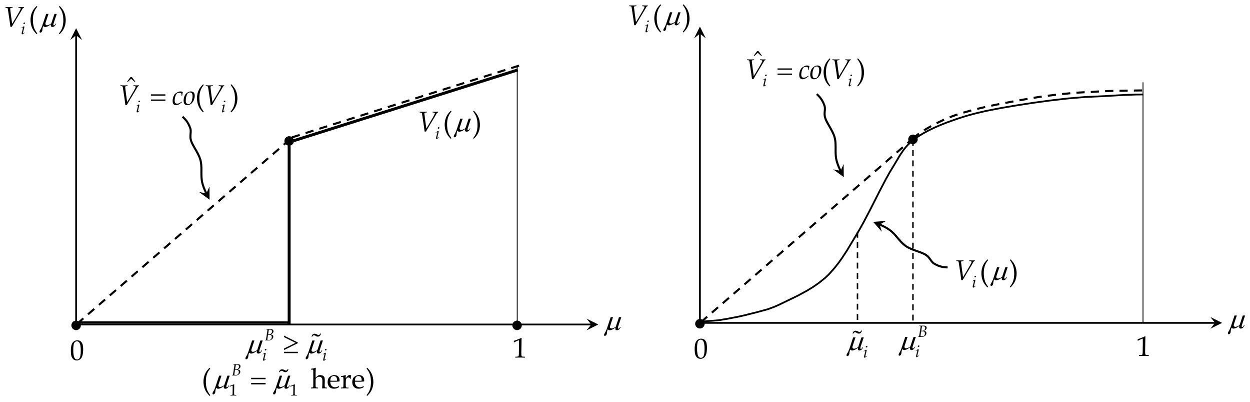
(ii) (symmetric case) when dissonant ($\omega = 0$), the two agents are in sync as they both collaborate or both refrain from collaborating: $p(a_1, a_2 \mid 0) = 0$ if $a_1 \neq a_2$;
(general asymmetric case) when dissonant, the agents' decision rules are nested (e.g., $a_2 = 1$ implies $a_1 = 1$): $p(0,1 \mid 0) = 0$.

Deep uncertainty. Section 2.3 provides several illustrations, each with a different notion of "congruence". In each illustration, we start with the "congruence-uncertainty case" described above (there is no uncertainty about payoffs for a given perceived congruence μ). Then we smooth the payoff functions by introducing some uncertainty, say about the (opportunity) costs, that is publicly realized at the beginning of stage 2, before the stage-2 game unfolds.

[12] Part (iii) of Assumption 2 serves to make the problem non-trivial (incentives depend on the state of nature). Assumption 2(iii) also implies that the net payoffs $\{V_i(0)\}_{i\in\{P,1,2\}}$ are the payoffs corresponding to $a_i = a_j = \omega = 0$. Our illustrations will indeed satisfy the convention that $V_i(0) = 0$.

[13] Let $\delta_P(a_j,\omega) \equiv u_P(a_i = 1, a_j, \omega) - u_P(a_i = 0, a_j, \omega)$. Then $\delta_i(a_j,\omega) \geq 0$ implies that $\delta_P(a_j,\omega) \geq 0$, for $i \in \{1,2\}$.

We let x generically denote this interim noise, distributed according to a smooth cdf $F(x)$ and density $f(x)$, and assume that the the distribution has a log-concave density f (so $f'(x)/f(x)$ is decreasing in x). Prekopa (1973)'s theorem states that this assumption is stronger than/implies the standard monotone-hazard-rate condition.[14]



(a) Congruence-uncertainty case.

(b) Deep-uncertainty case.

Figure 1: Convex-concave payoff functions
($co(V_i)$ is the concavification of V_i)

2.2 Experiments

An "experiment" is an arbitrary first-stage contingent protocol of elementary *public* experiments (at most T of them, with $T \leq +\infty$) that are informative about the congruence parameter. It yields a cumulative distribution function $Q(\mu)$ over posterior beliefs $\mu \in M(\mu_0)$ that are reachable from μ_0. Because the agents learn about their congruence through public experiments (until Section 4), the space of feasible posterior beliefs is more limited than in the "information design" (ID) literature, in which the principal observes the state of nature herself and can separately disclose the information to the agents in an arbitrary noisy way. This is so in two ways: First, the set of feasible experiments is given, and second the outcomes of experiments are observed by both agents.

The set of feasible distributions $Q(\mu)$ is in general strictly included in the set $\mathcal{Q}^{UID}$ of all distributions of common posterior beliefs that obey the martingale property $\int_0^1 \mu dQ(\mu) = \mu_0$ (where "UID" stands for "Unconstrained Information Design"). Player i's payoff is then $\int_0^1 V_i(\mu)dQ(\mu)$.[15]

[14] One can generalize Definition 1 to deep uncertainty:

Definition 2 ***(motivational synchronicity under deep uncertainty)***. *Consider some common noise x realized prior to the stage-2 game, and w.l.o.g. assume that x is a cost parameter (u_i is decreasing in x for $i \in \{1,2\}$). Let $p(a_1, a_2 \mid \omega, x)$ denote the probability of $\{a_1, a_2\}$ in state $\{\omega, x\}$. The allocation p satisfies motivational synchronicity if, for all x, conditions (i) and (ii) of Definition 1 are satisfied.*

[15] Section 3 investigates optimal stopping time strategies. The stopping time can be $+\infty$ with strictly

We limit attention to experiments that are combinations of i.i.d. elementary ones. This will enable us to focus on a single sufficient statistic μ_t, the posterior belief after t rounds of elementary experiments.

Simple binary experiments. The four experiments we will use as illustrations in Section 3 belong to the class of simple binary experiments (Birnbaum 1961):

> *Bad-news model.* Conditional on $\omega = 0$, dissonance is revealed with probability ρ in an elementary experiment; otherwise no signal accrues (and posterior beliefs are $\mu_{t+1} = \mu_t/[\mu_t + (1-\rho)(1-\mu_t)]$ if current beliefs are μ_t after the first t elementary experiments): no news is good news. And so $M(\mu_0) \equiv \{0, \mu_0, ..., \frac{\mu_0}{\mu_0+(1-\mu_0)(1-\rho)^t}, ...\}$. A stopping rule is defined as the set of posterior beliefs $\mu_t \in M(\mu_0) \setminus \{0\}$ such that the experiment stops (the experiment also stops at $\mu_t = 0$, as there is then nothing left to learn).[16]
>
> *Good-news model.* Conditional on $\omega = 1$, congruence is revealed with probability ρ; otherwise no signal accrues (and posterior beliefs are $\mu_{t+1} = \mu_t(1-\rho)/[\mu_t(1-\rho)+1-\mu_t]$ if current beliefs are μ_t): no news is bad news. A stopping rule is defined as the set of posterior beliefs μ_t such that the experiment stops (this set includes $\mu_t = 1$, as there is nothing left to learn).[17]
>
> *Good-or-bad-news model.* Combining the previous two models, one can have two parameters, ρ_0 and ρ_1, of accrual of an absorbing signal contingent on the true state of nature ($\omega = 0$ or $\omega = 1$). The experimenter chooses at each round between a good-news experiment or a bad-news one (or stops the experiment).
>
> *Bernoulli signals.* In a Bernoulli experiment, there are two possible signals at stage t: a good and a bad signals. The probability of a good signal conditional on $\omega = 1$ (resp. $\omega = 0$) is ρ_1 (resp. $1-\rho_0$), where $\rho_1 + \rho_0 > 1$ and ρ_0, $\rho_1 < 1$. The draws are i.i.d. over time.

Definition 2 ***(slow learning)***. *Slow learning*[18] *refers to limit cases of a large number of simple binary experiments, in which $\rho_0 + \rho_1 = 1 + h$ with $h \to 0$. These include in particular the following limit cases: slow-learning bad-news model ($\rho_1 = 1$, $\rho_0 = h \to 0$), slow-learning good-news model ($\rho_0 = 1$, $\rho_1 = h \to 0$), and slow-learning Bernoulli ($\rho_0, \rho_1 < 1$, $h = \rho_1 - (1-\rho_0) \to 0$).*

positive probability. Because the martingale μ_t is bounded (belongs to $[0,1]$), the optional stopping time theorem applies. The expected value of the martingale at the stopping time is equal to its initial value.

[16] It is straightforward to generalize the bad-news model to an arbitrary absorbing belief $\mu_L \geq 0$. And similarly for the good-news model below, with absorbing belief $\mu_H \leq 1$.

[17] The "one-shot" experiment ($\rho_1 = \rho_0 = 1$), delivering value $W_i(\mu) = \max\{V_i(\mu), \mu V_i(1)\}$, yields the same information design as the good-news model (as captured in Proposition 4(i) below).

[18] I do not use the "continuous" terminology as the stochastic process may involve jumps, even in the limit as $h \to 0$.

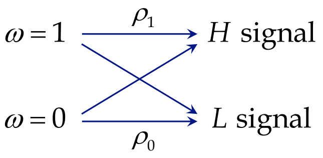


Figure 2: Binary experiments

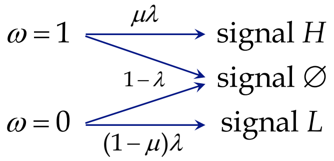


Figure 3: Conclusive Poisson signals

More general stochastic processes. The simple binary experiments (bad news, good news, Bernoulli) considered above are familiar and will be the main focus of our applications in Section 3. Section 4 results in contrast generalize to Bayesian learning (with possibly occasional conclusive evidence), as long as elementary experiments are i.i.d.: They are driven solely by the agents' relative information aversion, which only presupposes individually-optimal stopping rules characterized by a cutoff (which we will later call μ_i^*).

Thus, we can for instance consider a signaling structure with more signals $s_t \in \{1, ..., S\}$ with probabilities $f_\omega(s_t)$ and monotone likelihood ratio $f_1(s_t)/f_0(s_t)$; letting $s^t \equiv \{s_1, ..., s_t\}$ denote the sequence of observed signals, posterior beliefs are then $\mu = 1/[1+\frac{\mu_0}{1-\mu_0}L(s^t)]$, where $L(s^t)$ is the likelihood ratio, equal to the product of the t likelihood ratios. The simplest such case is that of conclusive Poisson signals (Figure 3), for which an elementary experiment is uninformative with positive probability and, with the complementary probability, reveals state of nature $\omega = 1$ (resp. $\omega = 0$) with probability μ (resp. $1 - \mu$) when the current posterior is μ. This case (studied in Che and Mierendorff 2019) has three signals (0, 1, and $\varnothing$).[19]

Finally, the model also accommodates posterior beliefs that follow a jump-diffusion martingale with absorbing jumps to $\{0, 1\}$.

Remark (time- or resource-consuming learning). For notational simplicity, we assume that learning about each other does not delay the productive task (stage 2) and involves no cost per experiment. The results generalize straightforwardly to a cost of delay. Discounting makes agents more impatient and therefore more information averse (technically, it shifts the cutoffs μ_i^*, defined below, leftward). The optimal policy is then two-sided (stopping is optimal for very-low beliefs and not only for beliefs above some threshold). The theory is otherwise unchanged.

Remark (sequencing of experiments). Our focus is on the quantity of information; the i.i.d. nature of experiments makes the sequencing question moot. The study of the optimal sequence of information exchange when experiments differ in their informativeness or in their nature, would be the team counterpart of the analysis of Sobel (1985) and Clayton et al (2025), which serve as canonical one-agent benchmarks for choosing an ordered set of heterogeneous experiments.

[19] More generally, in the three-signal, occasional-revelation model, $s_t \in \{H, \varnothing\}$ when $\omega = 1$, and $s_t \in \{L, \varnothing\}$ when $\omega = 0$.

2.3 Payoffs: economic illustrations

We develop five economic illustrations. For each, we first consider the case in which the only uncertainty is about the agents' congruence (the value of ω); we show that cooperative behavior occurs when perceived congruence exceeds some threshold $\underline{\mu}_2$, and that the reduced-form payoff functions are convex-concave. We also show how adding uncertainty about other parameters (for example, about the reservation utility or the cost of effort), while smoothing the functions $V_i(\mu)$ (which have a kink or discontinuity in the congruence-uncertainty case), can preserve the convex-concavity property. Finally, we characterize the function $\Delta(\mu) \geq 0$ defined as the agents' payoff differential:[20]

$$V_1(\mu) = V_2(\mu) + \Delta(\mu).$$

We will later see that the shape of $\Delta(\mu)$ determines the agent's relative appetite for learning about congruence and the outcome of the communication process.

(a) Effort on the job (intensive margin)

In the *intensive margin* illustration mentionned in Section 2.1, the team has already been formed and the agent's decision refers to the choice of effort ($a_i \in \{0, 1\}$). The agents and the principal share the output of a supermodular production technology, whether monetary or prestige, in some way (shares can be taken equal after a renormalization). Letting γ_i denote i's marginal cost of effort, i's payoff is[21]

$$u_i(a_i, a_j, \omega) = \alpha \frac{a_1 + a_2}{2} + \beta \omega a_1 a_2 - \gamma_i a_i.$$

Agent i's cost of effort is $\gamma_i = \gamma - \theta_i$ (with $\theta_2 = 0 \leq \theta_1 \leq \gamma$, so agent 2 is the weak link, and $\theta_P = \gamma$ - the principal does not work in the team -). Assume that $(\alpha/2) < \gamma_i$ to ensure that dissonance generates no effort, and $(\alpha/2) + \beta > \gamma$ to yield effort under perfect congruence.

Congruence uncertainty. When γ is known, agents exert effort if and only if

$$\frac{\alpha}{2} + \beta\mu \geq \gamma,$$

or $V_i(\mu) = [\alpha + \beta\mu - \gamma + (\theta_i - \theta_2)]1_{\frac{\alpha}{2}+\beta\mu\geq\gamma}$. And so

$$\Delta(\mu) = (\theta_1 - \theta_2)\mathbb{1}_{\mu \geq \underline{\mu}_2}.$$

Deep uncertainty. Similarly, when γ is drawn from distribution $F(\gamma)$ on $[\underline{\gamma}, +\infty)$ with $\underline{\gamma} - \theta_1 \geq$

[20] The same holds for the principal, but we will care mainly about the agents' incentive to communicate.

[21] Note that the output payoff in the second is equal to that in the first ($\alpha + \beta\mu$) if there is no (or a small) marginal cost of effort.

$(\alpha/2)$ and a log-concave density $f(\gamma)$ between stages 1 and 2, player i's expected payoff is

$$V_i(\mu) = \int_{\underline{\gamma}}^{\frac{\alpha}{2}+\beta\mu} [\alpha + \beta\mu - \gamma + (\theta_i - \theta_2)] dF(\gamma).$$

Thus,

$$\Delta(\mu) = (\theta_1 - \theta_2) F\big(\frac{\alpha}{2} + \beta\mu\big).$$

Illustration	(b) Extensive margin	(c) Kin altruism	(d) Different priors	(e) Ancillary interactions
Utilities u_i	$(p - r_i)a_i a_j$ with productivity $p = \alpha + \beta\omega$	$(1+\theta_i\omega)p - \gamma a_i$ with productivity $p = \alpha(\Sigma_i a_i)/2 + \beta a_i a_j$	$(\alpha + \beta\omega - r)a_i a_j$ with different priors on ω $(\mu_{0,1} \geq \mu_{0,2})$	$p - \gamma a_i + \theta_i \omega b_i b_j$ per period, with productivity $p = \alpha(\Sigma_i a_i)/2 + \beta a_i a_j$
Interpretation	$r_i = r - \theta_i$ is agent i's reservation utility	θ_i is internalization parameter	μ = agent 2's beliefs. Agent 1's are (for some $k < 1$) $\Xi_1(\mu) = \frac{\mu}{\mu+(1-\mu)k}$	θ_i is benefit from ancillary interaction if any. $b_i \in \{0,1\}$.
Congruence uncertainty (cooperate iff $\mu \geq \underline{\mu}_2$; otherwise $\Delta(\mu) = 0$)	• $\alpha + \beta\underline{\mu}_2 = r_2$ • $\Delta(\mu) = \theta_1 - \theta_2$ for $\mu \geq \underline{\mu}_2$	• $(1 + \theta_2\underline{\mu}_2)(\frac{\alpha}{2} + \beta) = \gamma$ • $\Delta(\mu) = (\theta_1 - \theta_2)(\alpha + \beta)\mu$ for $\mu \geq \underline{\mu}_2$	• $\alpha + \beta\underline{\mu}_2 = r$ • $\Delta(\mu) = \beta[\Xi_1(\mu) - \mu]$	• $\frac{\xi}{1-\xi}[\alpha + \beta - \gamma + \theta_2\underline{\mu}_2] = \gamma - (\frac{\alpha}{2}+\beta)$ where ξ = discount factor • $\Delta(\mu) = \frac{(\theta_1-\theta_2)\mu}{1-\xi}$
Deep uncertainty	• $F(r)$ with log-concave density • $\Delta(\mu) = (\theta_1 - \theta_2)F(\alpha + \beta\mu)$	• $F(\gamma)$ with log-concave density • $\Delta(\mu) = (\theta_1 - \theta_2)(\alpha+\beta)\mu F((1+\theta_2\mu)(\frac{\alpha}{2} + \beta))$	• $F(r)$ satisfying conditions in Appendix B • $\Delta(\mu) = \beta[\Xi_1(\mu) - \mu]F(\alpha + \beta\mu)$	• $F(\gamma)$ with log-concave density • $\Delta(\mu) = \frac{(\theta_1-\theta_2)\mu}{1-\xi}F(\gamma^*(\mu))$ where $\gamma^*(\mu)$ is the cutoff-cost.

Figure 4: Illustrations (b) through (e).

Appendix B develops four other illustrations, summarized in Figure 4. I describe them informally here.

- *Extensive margin.* Illustration (b) is very similar to Illustration (a), except that the agents decide whether to engage with each other rather than choose an effort after they have formed a team. Agents differ in their reservation utility $r_i = r - \theta_i$. A small, but interesting difference between the two is that free-riding can occur in the intensive-margin case, but not in the extensive-margin one; this implies that an agent always benefits from the other agent's being more optimistic about their congruence in Illustration (a), but not in Illustration (b).

- *Kin/parochial altruism.* In Illustration (c), agents differ in their social preferences. Agent i internalizes a fraction $\theta_i\mu$ of agent j's material utility, with $\theta_1 \geq \theta_2$, where μ is now the perception of kinship. That is, hard-wired altruism is conditioned on perceiving the other as alike along some dimension (be it gene, religion, politics, hobbies, etc), and one of the agents values kinship more than the other.

- *Different priors and other non-Bayesian environments.* In Illustration (d), the two agents have different priors, $\mu_{0,1} \geq \mu_{0,2}$ (agent 1 is more optimistic about congruence). They "agree to disagree", and so, when they jointly receive the same signals on the congruence on the task, their posterior beliefs co-vary. I also show that, beyond such ex-ante heterogeneous beliefs, the model also accommodates identical priors and ex-post heterogeneous beliefs: The agents may update differently and imperfectly after each elementary experiment (they may under- or over- react to the outcome of the experiment), as long as the statistical biases are commonly known.

- *Ancillary interactions and social collateral.* In Illustration (e), an agent's value of working with someone is not limited to the joint productivity. There are also potential social interactions during work time, at the coffee machine, at the cafeteria lunch, or social events and dinners outside the workplace. These may exist or not exist. The agents interact repeatedly. In each period they choose both an effort ($a_i \in \{0, 1\}$) as in Illustration (a), and whether to engage in ancillary interactions ($b_i \in \{0, 1\}$). The per-period benefit from an ancillary interaction is $\theta_i\mu b_i b_j$ (it takes two to tango). The agents optimally use trigger strategies to maintain cooperation; they use the social payoff $\theta_i\mu$ as collateral in the relationship, that is surrendered if at least one of the agents deviates in the work relationship. The social interactions form a proficient disciplining instrument when congruence is large enough.

Proposition 1

(i) *In applications (a) through (e), payoffs are convex-concave (as in Figure 1a) in the congruence-uncertainty case, and smoothly convex-concave (as in Figure 1b) when noise is added as described.*

(ii) *The weak link is, respectively, the agent with (a) a high outside option, (b) a high cost of effort, (c) a low kin altruism, (d) a low prior belief of congruence, and (e) a low benefit of ancillary interactions.*

3 Demand for communication

3.1 Information design under public messages

A *controlled protocol* of elementary public experiments is a contingent plan for the acquisition of information about congruence. It defines for each stage t a set Υ_t of beliefs μ_t at which communication stops and the stage-2 game is played under the updated beliefs.

This section studies the optimal controlled protocol from the point of view of a player i (agent, principal) with convex-concave preferences in the cases of bad news, good news, good-and-bad news, and Bernoulli experiments. This study, providing each player's preferred experiment for these processes, will prove key to Section 4's study of information exchange between the agents when obedience and truthfulness can not be taken for granted. As a warm-up exercise, I show how in the bad news, the optimal learning outcome best approximates the player's preferred information-design outcome. Such approximations are pursued in the bad-news case, but are also available in the other cases, for which for conciseness, we focus on the expositionally simpler slow-learning formulation.

Let $\mu_i^B \equiv \arg\max\{V_i(\mu)/\mu\}$. When V_i is smooth and strictly increasing, let $\tilde{\mu}_i$ denote the inflection point ($V_i''(\tilde{\mu}_i) = 0$ if interior). Assumption 1(i) (convex-concavity of V_i) implies that the ratio $R_i(\mu) \equiv V_i(\mu)/\mu$ is unimodal, with mode above the inflection point: $\mu_i^B \geq \tilde{\mu}_i$. As we will see, beliefs μ_i^B will play a key role under bad-news experiments (but not only), hence the use of the superscript "B" in μ_i^B. For smooth payoffs, beliefs when interior satisfy $\tilde{\mu}_i < \mu_i^B$, and

$$\mu_i^B V_i'(\mu_i^B) = V_i(\mu_i^B).$$

Lemma 1 *Regardless of the experiment, player $i \in \{P, 1, 2\}$ finds it optimal to stop experimenting whenever $\mu \geq \mu_i^B$.*

Proof. Let $\hat{V}_i \equiv co(V_i)$ denote the concavification of V_i (it is a linear function on $[0, \mu_i^B]$ and coincides with V_i on $[\mu_i^B, 1]$): See Figure 1. For $\mu \geq \mu_i^B$, and for any random variable $\tilde{\mu}$ such that $E[\tilde{\mu}] = \mu$,

$$V_i(\mu) = \hat{V}_i(\mu) \geq E[\hat{V}_i(\tilde{\mu})] \geq E[V_i(\tilde{\mu})]$$

from the concavity of $\hat{V}_i$, and where the expectation refers to the random stopping-time induced beliefs $\tilde{\mu}$ (with $E[\tilde{\mu}] = \mu$ from the martingale property) generated by an arbitrary sequence of experiments. ■

The converse property ("experiment whenever $\mu < \mu_i^B$") in general does not hold, as we will show. Finally, we define for future use the following notion of relative information aversion:

Definition (relative information aversion). Suppose that (as will be the case in our applications) agent i, if she had the choice of selecting the public experiment, would continue experimenting as long as $\mu < \mu_i^*$ and stop experimenting when $\mu \geq \mu_i^*$, for some μ_i^*. We will say that agent i is more (resp. equally, less) information averse than agent j if $\mu_i^* < \mu_j^*$ (resp. $\mu_i^* = \mu_j^*$, $\mu_i^* > \mu_j^*$).

3.2 Bad-news experiments

The ordered set $M(\mu_0)$ of reachable posterior beliefs from prior μ_0 that follow a sequence of bad-news experiments is denoted $\{0, \{\mu_t\}_{t \in \{0,\ldots,T\}}\}$. Bayes' rule implies that the probability of reaching, through some experiment, posterior μ_t is μ_0/μ_t. The set of potential non-zero posterior beliefs is the set of posteriors $\{\mu_t\} \in M(\mu_0) \setminus \{0\}$ with $\mu_0 < \mu_1 < \mu_2 \ldots$, that can be reached, with probability μ_0/μ_t, through a particular sequence of experiments (possibly none). Let

$$\mu_i^* \equiv \arg \max_{\mu \in M(\mu_0)\setminus\{0\}} \frac{V_i(\mu)}{\mu}.$$

We will assume for expositional simplicity that μ_i^* is uniquely defined (this is generically the case). Under slow learning, as $h \to 0$, $\mu_i^* \to \mu_i^B$ for $\mu_0 \leq \mu_i^B$.

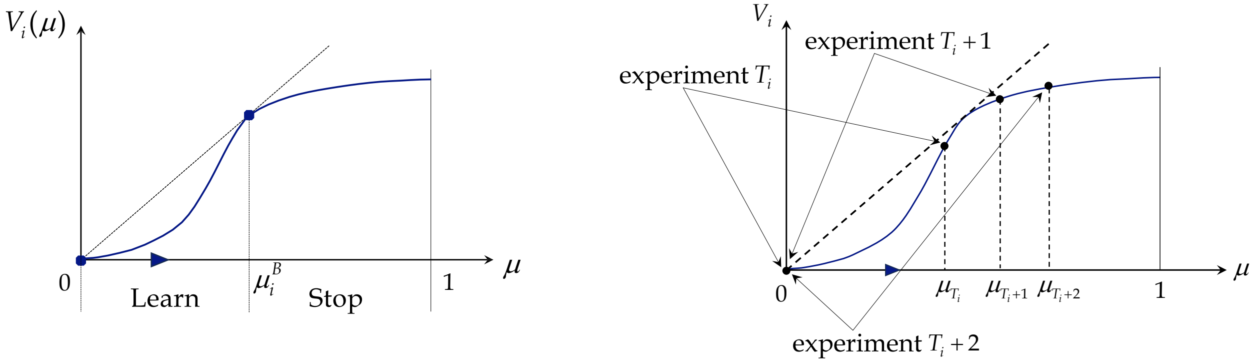
Proposition 2 ***(optimal experiment under bad-news experiments)****. Consider a convex-concave $V_i(\mu)$.*

(i) The optimal experiment for player i maximizes $\frac{\mu_0}{\mu}V_i(\mu)$ over the set of positive posterior beliefs that can be reached though some sequence of elementary experiments ($M(\mu_0) \setminus \{0\}$). [Experimenting stops after T_i^* elementary experiments, when reaching posterior beliefs (when not 0) $\mu_i^* = \mu_{T_i^*}$.]

(ii) In particular, in the limit of the slow-learning/bad-news model, player i's optimal protocol is to reach belief μ_i^B (assuming no bad news) when $\mu_0 \leq \mu_i^B$ (delivering the information design optimum), and not to experiment otherwise.

Proposition 2, illustrated in Figure 5, confirms the intuition that, because under bad news one of the posteriors, 0, of the unconstrained information design optimal policy is available, the constrained optimum for player i "best approximates" the UID outcome, and reaches it in the limit of show learning.

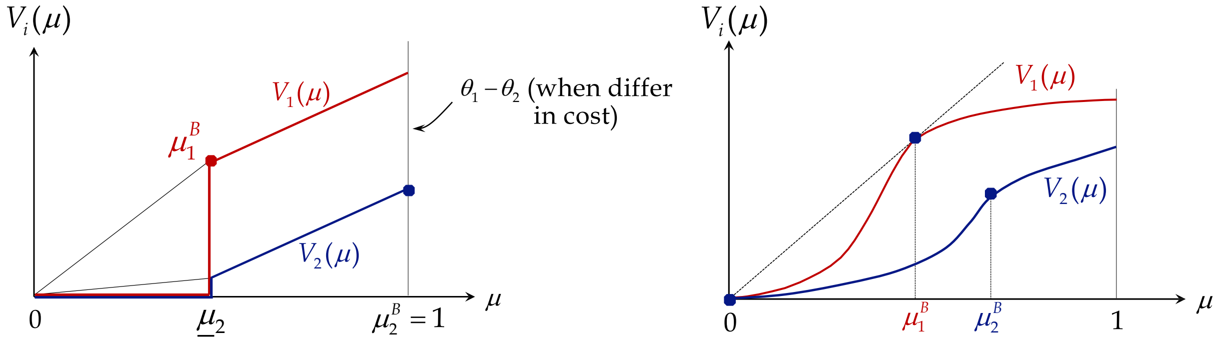
Proof of Proposition 2. The proof is straightforward, as the function $V_i(\mu)/\mu$ is unimodal (increasing from 0 to μ_i^B and then decreasing beyond μ_i^B). Under bad news, the probability of reaching $\mu_t = \mu_0/\left[\mu_0+(1-\mu_0)(1-\rho)^t\right]$ is equal to the probability that no bad news accrue up to t, or μ_0/μ_t. So the optimal stopping rule is to stop at the value of μ that maximizes $\mu_0 V_i(\mu)/\mu$ over $M(\mu_0) \setminus \{0\}$. ■



(a) Smooth payoffs, slow-learning (stop at μ_i^B)

(b) Smooth payoffs, discrete learning (stop at μ_{T_i} or μ_{T_i+1})

Figure 5: Symmetric payoffs, bad-news experiments



(a) Congruence-uncertainty case.

(b) Deep-uncertainty case.

Figure 6: Agents' preferences for information under bad news

Relative information aversion

To illustrate the agents' preferences relative to information acquisition, assume slow-learning and let $V_1(\mu) = V_2(\mu) + \Delta(\mu)$, so

$$\mu_1^B = \arg \max_{M(\mu_0)\setminus\{0\}} \left\{\frac{V_2(\mu)}{\mu} + \frac{\Delta(\mu)}{\mu}\right\},$$

where $M(\mu_0) \setminus \{0\} \equiv [\mu_0, 1]$. Let us begin with the smooth case (deep uncertainty). Unless $\mu_2^B = 1$ (in which case the comparison between μ_1^B and μ_2^B is trivial), $\frac{d(\frac{V_2(\mu)}{\mu}+\frac{\Delta(\mu)}{\mu})}{d\mu}\mid_{\mu=\mu_2^B}= \frac{d(\frac{\Delta(\mu)}{\mu})}{d\mu}\mid_{\mu=\mu_2^B}$ in the smooth (deep uncertainty) case. This implies that

$$\frac{d(\frac{\Delta(\mu)}{\mu})}{d\mu} < 0 \Rightarrow \mu_1^B \leq \mu_2^B$$

and

$$\frac{d(\frac{\Delta(\mu)}{\mu})}{d\mu} = 0 \Rightarrow \mu_1^B = \mu_2^B.$$

The proof for the congruence uncertainty case is almost identical. Convex-concavity then requires that $V_i(\mu)$ be (weakly) concave on $[\underline{\mu}_2, 1]$. When $\mu_2^B \in (\underline{\mu}_2, 1)$, the proof that $\mu_1^B \leq \mu_2^B$ is identical to that of the smooth case. Similarly, $\mu_1^B \leq \mu_2^B$ trivially holds if $\mu_2^B = 1$. So, suppose that $\mu_2^B = \underline{\mu}_2$, implying that $V_2'(\underline{\mu}_2) \leq V_2(\underline{\mu}_2)/\underline{\mu}_2$. Given that $\Delta(\mu)/\mu$ decreases, this in turn implies that $V_1'(\underline{\mu}_2) \leq V_1(\underline{\mu}_2)/\underline{\mu}_2$. And so $\mu_1^B = \underline{\mu}_2$ as well.

Finally, the same insight holds for the non-slow-learning case. Suppose, say, deep uncertainty and that $V_2(\mu_2^*)/\mu_2^* > V_2(\mu_2)/\mu_2$, where $\mu_2 \in M(\mu_0)$ and $\mu_2 > \mu_2^*$. Then $V_1(\mu_2^*)/\mu_2^* = [V_2(\mu_2^*)/\mu_2^*] + [\Delta(\mu_2^*)/\mu_2^*] > [V_2(\mu_2)/\mu_2] + [\Delta(\mu_2)/\mu_2] = V_1(\mu_2)/\mu_2$. And so, if $\Delta(\mu)/\mu$ is decreasing, $\mu_1^* \leq \mu_2^*$.

Using the expressions derived in Section 2.3, we thus obtain:[22]

[22] The only case that requires some computations is Illustration (d) (different priors). Then, $\frac{d(\frac{\Delta(\mu)}{\mu})}{d\mu} =$

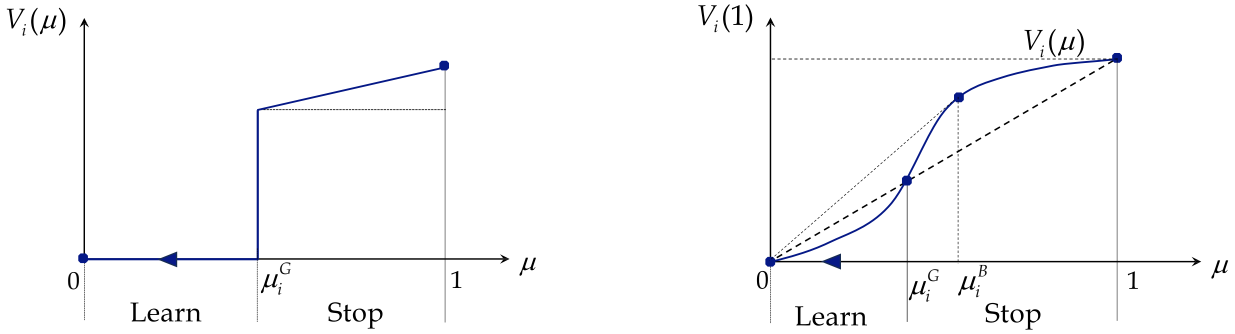
Proposition 3 ***(relative information aversion in the bad-news case)****. Under bad-news, slow-learning:*

(i) The strong link is more information averse than the weak link ($\mu_1^* \leq \mu_2^*$) if $\frac{\Delta(\mu)}{\mu}$ is a decreasing function, which is the case under congruence uncertainty in Illustrations (a) (extensive margin), (b) (intensive margin), and (d) (different priors).

(ii) The strong link and the weak link are equally information averse ($\mu_1^* = \mu_2^*$) if $\frac{\Delta(\mu)}{\mu}$ is a constant function, which is the case under congruence uncertainty in Illustrations (c) (kin altruism) and (e) (ancillary interactions and social capital).

3.3 Other simple experiments

Good-news experiments. Let us turn to good-news experiments. In contrast with the bad-news model, the origin in general does not belong to the feasible beliefs set (and when it does because an infinite sequence of elementary experiments is available, then beliefs $\mu = 1$ is also reached with strictly positive probability, which is not the case for unconstrained information design (UID)). So the UID payoff cannot be reached, even in the slow-learning case.



(a) Congruence-uncertainty case.

(b) Deep-uncertainty case.

Figure 7: Good news model

Let $M(\mu_0) \equiv \left\{\mu_t = \mu_0(1-\rho)^t/[\mu_0(1-\rho)^t + 1 - \mu_0] \text{ for some } t \geq 0\right\} \cup \{1\}$ denote the set of reachable beliefs, and let μ_i^G (where G stands for "good news") be defined by

$$\frac{V_i(\mu_i^G)}{\mu_i^G} = V_i(1) \left(= \frac{V_i(1)}{1}\right)$$

if interior ($\mu_i^G = 1$ if $V_i(\mu)/\mu \leq 1$ for all μ, and $\mu_i^G = 0$ if $V_i(\mu)/\mu \geq 1$ for all μ). There cannot be two interior solutions to this equation.[23] Let μ_i^* be defined by (in the good-news

$\beta\left(\frac{d\left(\frac{\Xi_1(\mu)-\mu}{\mu}\right)}{d\mu}\right)$, knowing that $\frac{\Xi_1(\mu)-\mu}{\mu} = \frac{(1-\mu)(1-k)}{\mu+(1-\mu)k}$ where $0 < k = \frac{1-\theta_1\mu_0}{\theta_1(1-\mu_0)} < 1$ (see Section 2.3). Thus $\text{sgn}\left(\frac{d\left(\frac{\Xi_1(\mu)-\mu}{\mu}\right)}{d\mu}\right) < 0$.

[23] Suppose there are two interior solutions, μ and μ'. Then $\frac{V_i(\mu)}{\mu} = \frac{V_i(\mu')}{\mu'} = \frac{V_i(1)}{1}$, which contradicts the fact that the function $\frac{V_i(\mu)}{\mu}$ is unimodal.

case)

$$\mu_i^* \equiv \min\{\mu \in M(\mu_0) | \mu \geq \mu_i^G\}.$$

We show that, in the slow-learning limit ($\mu_i^* = \mu_i^G$ and $M(\mu_0) \equiv [0, \mu_0] \cup \{1\}$),[24]

- *Player i would like to keep learning as long as $\mu < \mu_i^G$*
  To see this, note that for $\mu < \mu_i^G$, $V_i(\mu) < \mu V_i(1)$. Thus, by learning as long as there is no good news, beliefs converge to 1 (with probability μ) or 0 (with probability $1 - \mu$). Hence stopping is dominated by the strategy of learning as long as there is no good news.

- *Player i prefers to stop learning as long as $\mu > \mu_i^G$*
  See Appendix C.

In contrast with the bad-news case, a local analysis may deliver the wrong experiment. Indeed, at μ_0 such that $V_i'(\mu_0) = \frac{V_i(1)-V_i(\mu_0)}{1-\mu_0}$, then $\frac{d(\frac{1-\mu_0}{1-\mu}V_i(\mu)+\frac{\mu_0-\mu}{1-\mu}V_i(1))}{d\mu} |_{\mu=\mu_0}= 0$; so a local experiment has a zero return. However, μ_0 is a local minimum (indeed, it is in the convex region, in which experimenting until μ is 0 or 1 is optimal for an unbounded number of potential elementary experiments).

Good- or bad-news experiments. Suppose now that player i can pick her optimal experiment using either elementary good-news experiments or elementary bad-news ones, or else a sequential combination of both. We assume that the conditional probabilities of jumps to either $\mu = 0$ (bad-news experiment) or $\mu = 1$ (good-news experiment) are small (slow learning). We claim that the ability to run good-news experiments is valueless to player i, in the slow-learning environment. Consequently, if $\mu_0 \leq \mu_i^B$, the experiment stops when $\mu \in \{0, \mu_i^B\}$. It stops immediatly when $\mu \in [\mu_i^B, 1]$.

The proof is straightforward: Performing only bad-news experiments delivers the UID optimum. So player i cannot do better, and indeed does worse when attempting at least one good news experiment, as this generates beliefs $\mu = 1$, inconsistent with the UID optimum, with strictly positive probability.[25,26]

[24] Consider the decision of stopping experimentation with beliefs μ_t or continuing (only) one more step to beliefs either 1 or $\mu_{t+1} = (1-\rho)\mu_t/[1-\rho\mu_t]$. The choice, viewed from the perspective of player i, hinges on: $V_i(\mu_t) \gtrless \rho\mu_t V_i(1) + (1-\rho\mu_t)V_i(\frac{(1-\rho)\mu_t}{1-\rho\mu_t})$. However this exercise, which delivers in the limit of the slow-learning case an optimum at $V_i'(\mu)(1-\mu) = V_i(1) - V_i(\mu)$, yields only a local optimum, not a global one and therefore does not define an optimal stopping rule. To see this, consider Figure 7b. At μ_i^G, player i is indifferent between stopping and experimenting a bit more; however, the problem is convex, as at $\mu - \varepsilon$ the player strictly prefers to keep experimenting until learning the true ω.

[25] Section 3 has in common with Che-Mierendorff (2019) the incorporation of an explicit dynamic process of information acquisition. In their paper, a decision maker tries to match a binary decision with a binary state; discounting implies that this decision-maker stops when "confident enough" about the state. The focus of their analysis is the choice between a bad-news experiment and a good- news one as a function of the posterior belief: own-bias learning vs. opposite-bias learning. Unlike here, bad-news experiments do not necessarily dominate good-news ones. Che and Mierendorff obtain interesting insights into how decision accuracy depends on the prior beliefs.

[26] The suboptimality of good-news experiments hinges on a sufficient number of available bad-news ones. With a limited number of bad-news experiments, good-news ones, however suboptimal, may be needed to

Bernoulli experiments. Again, we will look at the case in which learning is slow and there is no bound on the number of experimentations. The learning process is in the limit a diffusion without jumps. For $\mu_0 < \mu_i^B$, agent i can reach the UID outcome by experimenting as long as $\mu < \mu_i^B$, yielding payoff $\mu_0 \left[\frac{V_i(\mu_i^B)}{\mu_i^B}\right]$. Similarly, when $\mu_0 \geq \mu_i^B$, agent i can reach the UID outcome by stopping immediately. The outcome is thus the same as under bad news.

Conclusive Poisson signals. Suppose that an elementary experiment is uninformative with positive probability and, with the complementary probability, reveals state of nature $\omega = 1$ (resp. $\omega = 0$) with probability μ (resp. $1 - \mu$) when the current posterior is μ (see Figure 3). The optimal stopping rule then reflects the trade-off between stopping learning and receiving $V_i(\mu)$, and keep learning until the state of nature is revealed, yielding $\mu V_i(1)$. Thus, agent i would like to stop (resp. continue) whenever $\mu > \mu_i^G$ (resp. $\mu < \mu_i^G$).

Proposition 4 ***(good-news, good-or-bad-news, and Bernoulli experiments).***

Let $V_i(\mu)$ be smoothly convex-concave. Consider a slow-learning process.

(i) Good-news experiments. Let μ_i^G be (uniquely if interior) defined by

$$\frac{V_i(\mu_i^G)}{\mu_i^G} = V_i(1).$$

Then, for any μ such that $\mu < \mu_i^G$, continuing is strictly optimal for player i. The optimal slow-learning experiment consists in stopping whenever $\mu_t \geq \mu_i^G$ and continuing whenever $\mu < \mu_i^G$.

(ii) Good or bad news experiments. Suppose that at each time, player i can select either a good-news or a bad-news experiment. The ability to run good-news experiments is valueless to player i. Consequently, if $\mu_0 \leq \mu_i^B$, the experiment stops when $\mu \in \{0, \mu_i^B\}$. It stops immediatly when $\mu_0 \in [\mu_i^B, 1]$.

(iii) Bernoulli experiments. The optimal learning process for player i in the slow-learning environment is to continue if and only if the current beliefs lie below μ_i^B.

(iv) Under conclusive Poisson signals, agent i would like to stop (resp. continue) whenever $\mu{>}\mu_i^G$ (resp. $\mu < \mu_i^G$).

Relative information aversion

Goods news. For simplicity again, we consider the slow-learning environment. Like for bad news, let us start with the smooth-payoffs case. Note that the function $V_i(\mu) - \mu V_i(1)$ inherits from the function $V_i(\mu)$ its convex-concavity, and that it takes value 0 at the two boundaries,

give agents a chance to cooperate. Consider the congruence-uncertainty case (cooperation requires $\mu \geq \underline{\mu}_2$). If the sequence of all available bad-news experiments delivers at best $\mu < \underline{\mu}_2$ (that is, T is small), then the optimal policy is to try available good-news experiments until $\mu = 1$; bad-news experiments are then irrelevant.

$\mu = 0$ and $\mu = 1$. To analyze the relative demand for information in the good-news case, suppose that $\mu_2^G < 1$ (the non-trivial case). Then

$$V_1(\mu_2^G) - \mu_2^G V_1(1) = V_2(\mu_2^G) - \mu_2^G V_2(1) + \Delta(\mu_2^G) - \mu_2^G \Delta(1) = \Delta(\mu_2^G) - \mu_2^G \Delta(1),$$

This implies that[27]

$$\Delta(\mu_2^G) - \mu_2^G \Delta(1) > 0 \Rightarrow \mu_1^G < \mu_2^G$$

and

$$\Delta(\mu_2^G) - \mu_2^G \Delta(1) = 0 \Rightarrow \mu_1^G = \mu_2^G.$$

In the congruence-uncertainty case, the proof is almost identical. In particular, if $V_2(\underline{\mu}_2) > \underline{\mu}_2 V_2(1)$, then $\mu_2^G = \underline{\mu}_2$. But if $\Delta(\underline{\mu}_2) - \underline{\mu}_2 \Delta(1) \geq 0$, then $V_1(\underline{\mu}_2) > \underline{\mu}_2 V_1(1)$, and so $\mu_1^G = \underline{\mu}_2$ as well.

Finally, using the expressions derived in Section 2.3, we obtain:

Proposition 5 ***(relative information aversion in the good-news case)***. *In the slow-learning environment, the qualitative ranking across the five illustrations is the same under good-news learning or conclusive Poisson signals, for which $\mu_i^* = \mu_i^G$, as under bad-news or Bernoulli learning, for which $\mu_i^* = \mu_i^B$:*

(i) The strong link is more information averse than the weak link ($\mu_1^G \leq \mu_2^G$) if $\Delta(\mu) - \mu\Delta(1) > 0$, which is the case under congruence uncertainty in Illustrations (a) (extensive margin), (b) (intensive margin), and (d) (different priors).

(ii) The strong link and the weak link are equally information averse ($\mu_1^G = \mu_2^G$) if $\Delta(\mu) - \mu\Delta(1) = 0$, which is the case under congruence uncertainty in Illustrations (c) (kin altruism) and (e) (ancillary interactions and social capital).

3.4 Postscript on relative information aversion

While the analysis of unfettered communication (Section 4) depends only on the existence of thresholds $\{\mu_i^*\}$ such that agent i wants to stop learning iff $\mu \geq \mu_i^*$, the relative-information-aversion ranking depends on the learning process. For instance, it is not always the case that the strong link agent ($\Delta(\mu) \geq 0$) is more information averse; for, there are two effects: The "grab now" effect –stopping now yields a higher payoff to the strong link– favors $\mu_1^* < \mu_2^*$ (this is the only effect e.g. when Δ is constant in the relevant range); but agent 1 may want to explore Δ more (the extreme case occurs when $\Delta > 0$ only at $\mu = 1$).

The second, "exploration effect" may make the ranking of information aversions ambiguous in the deep uncertainty versions of our illustrations. Take bad news. When there is uncertainty

[27] Let $\Phi_i(\mu) \equiv V_i(\mu) - \mu V_i(1)$ with $\Phi_i(0) = \Phi_i(1) = 0$. The function Φ_i is convex-concave, as we observed. Suppose first that $\Phi_i'(0) \geq 0$. The inflection point of Φ_i, $\tilde{\mu}_i^G$, satisfies $\Phi_i(\tilde{\mu}_i^G) > 0$, since Φ_i' is increasing up to the inflection point. This implies that μ_i^G, the interior root of Φ_i, is in the concave part of Φ_i. Because $\Phi_i'(\mu_i^G) < 0$, $\Phi_i'(\mu) < 0$ for $\mu \geq \mu_i^G$. And so $\Phi_i(1) < 0$, a contradiction. So $\Phi_i'(0) < 0$ and $\Phi_i'(\mu_i^G) > 0$ for all i. $\Delta(\mu_2^G) - \mu_2^G \Delta(1) > 0$ then implies for instance that $\Phi_1(\mu_2^G) > 0$ and so $\mu_1^G < \mu_2^G$.

about the cost γ (or on the reservation utility r - I will generically use the "γ" notation), the ratio $\Delta(\mu)/\mu$ takes the form $AF(\gamma^*(\mu))/\mu$ in illustrations (a) and (b) and $AF(\gamma^*(\mu))$ in illustrations (c) and (e), where $A > 0$ and $\gamma^*(\mu)$ in increasing in μ. In illustrations (a) and (b), $(\Delta(\mu)/\mu)'$ is (proportional to) the difference between a constant term ($\beta\mu$ in these applications) and the (monotone) inverse hazard rate $F(\gamma^*(\mu))/f(\gamma^*(\mu))$. An increase in μ amplifies the difference in rents between the two agents, possibly reversing the previous inequality $(\Delta(\mu)/\mu)' < 0$. In illustrations (c) and (e), there is no such ambiguity as $(\Delta(\mu)/\mu)' > 0$, implying that the weak link is more information averse.

Thus, even though the strong link is often expected to exhibit a higher information aversion as in our five illustrations under congruence uncertainty, this is not a general pattern. This said, if the weak link is more information averse, just permuting the labels allows us to apply Section 4 results.

4 Incentive-compatible communication

We have so far abstracted away from individual or collective manipulations of information. Equivalently, in an environment in which agents are the holders of information, we assumed that agents are obedient (they follow the communication protocol) and truthful (they disclose their information, whether soft or hard, truthfully when requested). As earlier, we assume that, to estimate the congruence along the dimension that determines organizational payoffs (dimension 0), the agents can avail themselves of diagnostics of how they will jointly perform in the organizational task: whether they are congruent in other dimensions, $t \in \mathcal{T} \equiv \{1, ..., T\}$, with $T \leq +\infty$. So, assume that each agent knows their characteristics along the T dimensions, but not the other agent's. Put differently, agents have types, whose comparison helps predict their congruence on the second-stage task. To conform with our framework, agents must have private information about their type, but not about the likelihood of their congruence.[28]

Congruence along dimension t means that the two agents have the same type in dimension t. Namely, agent i has type θ_{it} drawn uniformly in the interval $[0, 1]$. With probability ρ_1 in state $\omega = 1$ and $1 - \rho_0$ in state $\omega = 0$, the draws are perfectly correlated (there is a single draw); with the complementary probabilities the draws are independent.[29] Assume $\rho_1 + \rho_0 > 1$ (the experiments are informative). As we discussed, the bad-news case corresponds

[28] As Augenblick and Bodoh-Creed (2018) have shown, the possibility that some subtypes are more "popular" than others has implications for the incentive to tell the truth when information is soft; to the extent that there are positive externalities in cooperating, an agent then may be tempted to deviate from announcing a correct but unpopular type and claim a popular one, as this increases the probability that the other agent feels congruent.

[29] Thus, conditional on the types being the same (congruence, c) or not (non congruence, nc) along dimension t, updated beliefs are

$$\mu_{t+1}^c = \frac{\mu_t \rho_1}{\mu_t \rho_1 + (1 - \mu_t)(1 - \rho_0)} \quad \text{and} \quad \mu_{t+1}^{nc} = \frac{\mu_t (1 - \rho_1)}{\mu_t (1 - \rho_1) + (1 - \mu_t)\rho_0}.$$

to $\{\rho_1 = 1,\, \rho_0 = h\}$, the good-news case to $\{\rho_1 = h,\, \rho_0 = 1\}$, the Bernoulli case without finite-time-absorbing beliefs to $\rho_0, \rho_1 < 1$.

Section 3 assumed that some player, e.g., the principal, or the agents through a social norm, has the power to set the communication protocol. The resulting outcome however may not be incentive compatible for several reasons. The simplest illustration of this arises when the principal designs the protocol but cannot monitor the information exchange, and the agents have symmetrical preferences. The agents then have mutual interests in the stage-1 (communication) game. Considering bad news for instance, and starting from a low prior, it is an equilibrium for them to share information until, say, $\mu = \mu_1^* = \mu_2^*$, while the principal, who values effort less than they do, prefers them to stop at $\mu = \mu_P^* < \mu_1^* = \mu_2^*$. A second issue with incentive compatibility arises when information is hard but can be concealed, or when information is soft, so that agents can lie about their preferences. For example, the combination of positive externality and incentive-enhancing congruence implies that an agent benefits from the other agent's being optimistic about their congruence, and so may want to pretend to have similar preferences even if this is not the case. Furthermore, differences in preferences may imply that agents do not have identical preferences regarding information sharing.

A "disclosure" by agent i in dimension $t \in \{1, \ldots, T\}$ is an announcement $\theta_{i,t}$. A couple of additional features and comments are in order. First, what is meant by "disclosing" hinges on the nature of private information. Under *hard* (verifiable) information, this disclosure is necessarily truthful; information thus can be manipulated only by not disclosing $\theta_{i,t}$ (announcing "$\varnothing$"). We will take *soft* information to mean either announcing one of the feasible types along dimension t, or not disclosing anything ($\varnothing$). Second, the desiderata for the communication protocol for both hard and soft information will be (a) the absence of coordination failure (as might happen when none of the agents wants to disclose, but both do so simultaneously); (b) the existence of a stopping rule, specifying when communication stops in a given substage t and overall. Third, from the no-signaling-what-you don't-know property of sequential equilibrium (Fudenberg-Tirole 1991), the absence of disclosure on a dimension on which the other agent has not yet disclosed does not reveal anything about the likelihood of congruence: Our formulation posits that agents know their types, but types on a stand-alone basis reveal nothing about congruence.

Section 4 delivers the paper's central results on incentive-compatible mutual learning: The outcome of communication is "ascendancy of the less information averse" with hard information and -if communication is feasible (see below)- "ascendancy of the more information averse" with soft information. The core economic forces driving these results are general and powerful. The hard-information outcome is driven by a classic unraveling logic, where the agent desiring more information can force disclosure. The soft-information outcome rests on the ability of the agent desiring less information exchange to neutralize any further communication by "pandering".

4.1 Hard information

Protocol. There is a large number of potential communication protocols. They differ essentially in their scope for coordination failure. For example, simultaneous disclosures lead to such failure more easily than sequential ones: Take hard information and suppose that two symmetric agents want to stop communicating (current beliefs are $\mu > \mu_1^* = \mu_2^*$). Nonetheless, they may both disclose an extra dimension, as they expect the other agent to do the same simultaneously, and (as we will see) the best response to the other agent's disclosure of hard information along a dimension is to disclose (so that the agent might as well disclose).

As a warm-up exercise, we first consider the bad-news or the good-news models and the following *en-bloc-disclosures* protocol, which limits coordination failures. Stage 1 in the en-bloc-disclosure protocol has two rounds:

Round 1: Each agent i chooses a depth of disclosure T_i, where $0 \leq T_i \leq T$, and discloses true types $\{\theta_{i,1}, \theta_{i,2}, \ldots, \theta_{i,T_i}\}$ and none of the $\theta_{i,t}$ for $t > T_i$. If $T_i < T_j$, we will say that there are $T_j - T_i$ *"orphan (unmatched) disclosures"*.

Round 2: If $T_i = T_j$, the game proceeds to the real game (stage 2). If $T_i < T_j$, agent i gets a chance to disclose types corresponding to j's orphan disclosures; the stage-2 game is then played.

Suppose that $T_i < T_j$.[30] The round-2 equilibrium payoffs are then unique because of a familiar unraveling property. The stage-2 game being supermodular with non-negative externalities, agent i benefits from agent j's being as optimistic as possible regarding their congruence. So if i and j are congruent on dimension $T_i + 1, \ldots, T_j$ (which i knows, but j does not at the end of round 1), agent i will disclose all orphan dimensions. Thus, a lack of full disclosure in the $T_j - T_i$ orphan dimensions signals that at most $T_j - T_i - 1$ dimensions exhibit congruence, leading to full disclosure when exactly $T_j - T_i - 1$ dimensions exhibit congruence, and so forth.

Building on this unraveling, we can further refine equilibrium strategies. Like in Section 3, let T_i^* denote the number of mutual disclosures that leads to beliefs μ_i^*.[31] The point is that, under the protocol, an agent who discloses fewer dimensions in round 1 can in round 2 freely decide whether or not to "match" any orphan disclosures. Because of this option value, playing T_i^* never yields a lower (and sometimes yields a strictly higher) payoff than either $T_i > T_i^*$ or $T_i < T_i^*$. Put differently, making $T_i > T_i^*$ disclosures is a weakly-dominated strategy (WDS) for agent i at round 1, since agent i can wait until round 2 to match orphan disclosures, if any; so making T_i^* disclosures weakly dominates. Similarly making $T_i < T_i^*$ disclosures is also weakly dominated by disclosing T_i^* dimensions (as, with this timing, agent i cannot make further, non-orphan disclosures at round 2). The elimination of weakly-dominated strategies

[30] Here one agent (j) discloses more than the other (i). The agents could also disclose different dimensions, so that they would both get a chance to respond in round 2; in that case, the unraveling logic below would apply equally well. The requirement to disclose the first dimensions only is motivated by the need to avoid coordination failures.

[31] If the draws lead to beliefs 0 or 1 (bad- or good-news), some disclosures may be redundant, but this does not alter the results.

therefore selects $T_i = T_i^*$ for all i.

Consider the bad- and good-news models. To accommodate the possibility that communication may stop because the set of elementary experiments is finite ($T < +\infty$) and thus not rich enough, let us define the following objects:

- μ_T is the maximum beliefs in the bad-news model, i.e. the beliefs after T experiments in the absence of bad news ($\mu_T \to 1$ as $T \to +\infty$),
- μ_T is the minimum beliefs in the good-news model, i.e. the beliefs after T experiments in the absence of good news ($\mu_T \to 0$ as $T \to +\infty$).

Proposition 6 ***(ascendancy of the less information averse under hard information)***

Consider the en-bloc-disclosures protocol and hard information. Assume strictly positive externalities in Assumption 2(ii) and that $\mu_1^* \leq \mu_2^*$ (if $\mu_1^* > \mu_2^*$, replace μ_2^* by μ_1^* below). Eliminating sequentially weakly-dominated strategies, the equilibrium payoffs of the communication game are unique.

(i) If $\mu_0 \geq \mu_2^*$, there is no disclosure.

(ii) If $\mu_0 < \mu_2^*$, disclosure

- *leads to either $\mu = 0$ or $\mu = \min\{\mu_2^*, \mu_T\}$ in the bad-news model*
- *leads to either $\mu = 1$ or $\mu = \mu_T$ in the good-news model.*

The unraveling logic implies that the less information averse can force more disclosure than the more information averse would want, as depicted in Figure 8.

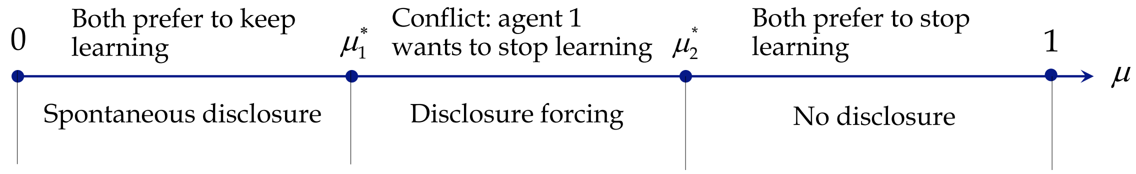

0
Both prefer to keep learning
μ_1^*
Conflict: agent 1 wants to stop learning
μ_2^*
Both prefer to stop learning
1
μ
Spontaneous disclosure
Disclosure forcing
No disclosure

Figure 8: Ascendancy of the less information averse.

Sequential disclosures protocol

The en-bloc-disclosures protocol captures the economics of disclosure in the simplest way in the good-news and bad-news cases, in which beliefs evolve monotonically with the number of elementary experiments or else reach an absorbing state. But it is not reasonable for non-monotonic processes such as the Bernoulli one. What matters to the agents more generally is not the number of experiments, but the target for resulting beliefs. A dynamic disclosure process allows the number of experiments to adapt to realizations, a property which was irrelevant in the good-news and bad-news models. Fortunately, regardless of the nature of experiments, the same logic applies to a variety of (perhaps more realistic) protocols, although some refinement (like above the elimination of WDSs) is generally needed to prevent coordination failures.

We here consider the *sequential-disclosure* protocol. Agents alternate in making disclosures: In odd rounds ($t = 2k + 1$) agent 1 communicates, i.e.,

- can choose a new dimension and disclose her type along that dimension (or not do so), and
- can, if she wishes so, disclose her type in any of the dimensions of agent 2's earlier orphan disclosures.

In even rounds ($t = 2k$), agent 2 communicates similarly. Communication stops when, in two successive rounds the agents neither disclose a new dimension nor respond to an earlier orphan disclosure.

The unraveling logic again implies that the less information averse can force more disclosure than the more information averse would want. It is an equilibrium for agents to disclose whenever $\mu < \mu_2^*$ (or $\mu = \mu_T$ if there are not enough dimensions to disclose) and stop disclosing otherwise.[32] To obtain this as a unique equilibrium, one must again eliminate weakly dominated strategies, namely disclosing when $\mu \geq \mu_2^*$.

Proposition 6' ***(sequential protocol)***
Assume hard information. Eliminating weakly dominated strategies, the outcome for the sequential protocol is the less-information-averse-agent's preferred one.

Plausible deniability

We have assumed so far that agents do know their traits and can disclose them in a verifiable way. A broader model allows, as in Dye (1985)' hard-information framework, for the possibility that the agents do not know their trait along some dimension (or, in a variant, that this information is not verifiable). In the previous notation, agent i then receives a signal $s_{i,t} \in \{\theta_{i,t}, \varnothing\}$ and can report $r_{i,t} \in \{s_{i,t}, \varnothing\}$. The agent is informed of $\theta_{i,t}$ with probability q and does not know their dimension-t trait with probability $1 - q$.

Consider the sequential-disclosure timing above. Given that each agent weakly benefits from the other agent being optimistic about congruence, agent i, when learning from agent j's disclosure $\theta_{j,t}$ that they are not congruent along dimension t, appeals to plausible deniability and claims not to have the information about own trait in that dimension. That means that a disclosure that remains orphan makes agent j update in a pessimistic way (for, agent i may conceal bad news about the relationship). I do not offer a complete analysis of the plausible-deniability model, and content myself with illustrating, in the context of the bad-news model, the new features that arise then.

[32] As earlier, agent 1 would like to stop when $\mu \geq \mu_1^*$, but might as well continue as long as $\mu < \mu_2^*$ as agent 2 will not stop anyway.

1. *Asymmetric information and endogenous information aversion.* The most obvious departure from the analysis above is that agents will be asymmetrically informed along the equilibrium path. But the beliefs conditional on not having received bad news are common knowledge. Namely, let $\nu(t, \tau_j)$ denote agent i's posterior belief about congruence after t rounds of communication in which agent j has left orphan τ_j dimensions disclosed by i and agent i has not learnt any bad news that would imply absorbing beliefs 0. So, letting as earlier ρ denote the probability of bad news in an elementary experiment given state $\omega = 0$,

   $$\nu(t, \tau_j) = \frac{1}{1 + \frac{1-\mu_0}{\mu_0}\left(\frac{1-q(1-\rho)}{1-q}\right)^{\tau_j}(1-\rho)^{t-\tau_j}},$$

   an increasing function of t and a decreasing function of τ_j. Suppose that agent i is more information averse in the absence of plausible deniability ($q = 1$). Asymmetric information implies that agent i will be *on average, but not always* more information averse than agent j; for, agent j may leave orphan a string of agent i's disclosures, that raise a concern with the latter. Thus, agent i may want to pursue information collection when agent j would like to stop.

2. *Non-attainability of the information-design outcome (even in the bad-news case).* A key feature of the bad-news case is that belief 0 belongs to the reachable set. However, under plausible deniability such a belief is never both reached and common knowledge, as agents engage in plausible deniability when learning they are not congruent.

3. *Delineation of acceptable discourse by the agents themselves.* Agents may lose when not knowing their type as orphan disclosures raise a suspicion. Indeed, if agents have an opportunity to delineate acceptable topics at the beginning of the communication game, it is an equilibrium for them to rule out dimensions in which they have no disposable information. That means that the communication game is one without plausible deniability, with the number of remaining experiments being the number of dimensions in which both can make a disclosure.

Appendix D pursues these themes in more detail (again for the bad-news model).

4.2 Soft information

With soft information, no equilibrium uniqueness is to be expected due to the standard babbling possibility. The next proposition describes, for the en-bloc-disclosure protocol, an equilibrium in which the strong link obtains their preferred communication. Consider the following strategies and beliefs: In round 1, and on the equilibrium path, both agents disclose their types $\left\{\theta_{i,1}, \ldots, \theta_{i,T_1^*}\right\}$ truthfully and only these types. The game proceeds to the real game. If agent i deviates and discloses their taste in more or less than T_1^* dimensions, the agent with the lowest number of disclosed dimensions, announces for the orphan disclosures the same types as those announced by the other agent (recall that the mimicry occurs in the second round, after seeing the other agent's disclosures). So nothing is learnt from the response to orphan disclosures. Suppose, say, that agent i deviates and announces types along $T_i > T_1^*$ dimensions. Agent j then believes, say, that agent i has randomly announced the extra types (alternatively, agent j could believe that i told the truth along these extra dimensions) and feigns congruence on the extra dimensions.

Proposition 7 ***(ascendancy of the more information averse under soft information).***
Consider the en-bloc disclosure protocol, and either the good-news or the bad-news model. The following is an equilibrium outcome of the communication game:

(i) If $\mu_0 \geq \mu_1^*$, then there is no experimentation, as the more information averse agent does not disclose in round 1 and feigns congruence if the other agent discloses; knowing this, the latter agent does not disclose.

(ii) If $\mu_0 < \mu_1^*$, under bad news, both agents disclose their traits along T_1^* dimensions such that $\mu_1^* = \mu_0/[\mu_0 + (1-\mu_0)(1-\rho)^{T_1^*}]$ if $\mu_0 < \mu_1^*$, and $T_1^* = 0$ if $\mu_0 \geq \mu_1^*$. Similarly, under good news, both agents disclose no trait as long as $\mu_0 \geq \mu_1^*$ and disclose all traits (and thus reach beliefs 0 or 1) if $\mu_0 < \mu_1^*$.

Beyond μ_1^*, any announcement by the weak-link agent 2 is met by a congruent message by agent 1 in the orphan dimensions, whether agent 1, who benefits from agent 2's perception of congruence, is actually congruent or not on those dimensions.

Next, consider the sequential communication protocol. Then there can no longer be any effective communication if externalities are strictly positive (u_i is strictly increasing in a_j for all a_i);[33] for, the agents benefit weakly from convincing the other agent that they are congruent. In a sense, simultaneity of disclosure is required for informative communication.

Are the quantities of information communicated under hard and soft information comparable, as would be the case if Proposition 7 described an upper bound on the transmitted information under soft information? Appendix E provides soft-information-communication examples in which (a) two symmetric agents fail to coordinate and communicate more information than is optimal for them, and, more interestingly, (b) two asymmetric agents communicate as much information as under hard information and more than is predicted by Proposition 7, without any Pareto ranking between the two.

[33] Consider a contrario illustration (b) (extensive margin), with, say, $\rho_1 = \rho_0 = 1$ (signals are fully informative, so that there is at most one round of effective communication). I claim that if $\mu < \underline{\mu}_2$, there is a fully revealing equilibrium despite the softness of information. Suppose that on the equilibrium path, agent i announces θ_i truthfully (I omit the subscript t, as there is a single round of communication). Agent j then knows whether $\omega = 1$ or 0. Even if $\omega = 0$, agent j is willing to tell the truth, as she will not engage anyway ($a_j = 0$). In the intensive-margin model by contrast, i's perception of congruence benefits j, even when $a_j = 0$, implying that babbling obtains.

5 Discussion and extensions

The paper sheds light on the role of team building and acceptable discourse in organizational policies and social norms, and on strategic information exchange among team members. Taking on board the idea that mutual trust is an essential lubricant for effective collective action, it obtains results of independent interest regarding the relative demands for communication and shows how these demands affect actual communication of hard and soft information. This final section covers other interpretations of this analysis, and indicate avenues for further research.

Learning to forget. Although I phrased the analysis in terms of learning information about each other, similar insights apply to unlearning. Recall Renan's insistence on the need to "forget many things" to form a community[34]. For time to heal rifts, the agents must not dwell on old grudges, say by rehearsing what divides. Yet, old grudges have a private rationale, the protection of self-esteem ("the other is responsible for our past conflicts"). Refraining from rehearsing them therefore involves a personal cost, but yields a team benefit. Social norms and principal-edicted rules on acceptable exchange may tilt the balance against the collective rehearsal of hard feelings. Similarly, an emphasis on what unites (nation, family love, sport team fandom...) not only boosts perceived congruence, but also crowds out negative congruence thinking.

Team formation and open-ended groups. Humans often have a choice between rewarding, but fragile team endeavors (marriage, friendship, joint venture, civic engagement...) and going alone. A while ago, Putnam (1995) lamented about the increasing tendency for citizens to "bowl alone", a debate that has grown even more intense with the spread of smartphones, social media, and overprotective parenting (Haidt 2024), This evolution may be analyzed as a decline in people's willingness to get to know others and build the trust that facilitates cooperation for mutual benefit.

Boundary setting. I noted the role of social norms and organizational injunctions against conflict-prone conversations. Sometimes, though, such conversations are tolerated or even encouraged within "Employee Resource Groups"; but this channeling toward like-minded groups (on LGBTQ+ or ethnic identity) is meant to limit such conversations to a like-minded group and thereby avoid conflict. On the other hand, we occasionally observe real attempts at encouraging employees to engage in political and social issues.[35] In part this may reflect the leaders' desire to advance progressive causes. But it might also be a way of screening personnel to obtain a very congruent organization. This strategy presumably reflects an ability to tap in a wide enough labor pool given the company's needs.

What makes an issue particularly divisive. Another area worth investigating would give content to what may particularly divide people. Psychologists have identified threats to identity as factors of defensiveness, contempt and rejection. Self-esteem maintenance then requires attributing hostile intents or stupidity to explain the other's divergent worldview. But not all failures of alignment generate disrespect and conflict.

[34] Relatedly, Napoleon argued that "History is a set of lies agreed upon."

[35] E.g., Ben & Jerry's Values model.

Low- and high-commonality environments. The paper suggests at least two determinants of the demand for information about commonality. First, situations range from "low-commonality environments" to "high-commonality ones". In the former, interaction is satisfactory as is and does not require people to know each other well; investigating commonality may backfire if conflicts loom larger than accords. In a cafe, restaurant, food market, stadium, taxi, or parent-teacher association, one is probably better off not knowing the politics or social views of the service provider or the people sharing the service; that probably would spoil the fun and create unnecessary tensions. The activity is self-sufficient in that it does not require learning more about each other (things are different if a derivative activity is making friends). At the opposite end of the spectrum lie critical tasks or marriage (couples tend to exhibit high homophily in terms of values, politics and personality traits), that require high perceptions of congruence.

Second, organizational information aversion depends on the depth of the "internal matching market". While collaborators are given assignments and therefore teammates, this matching may be neither cast in stone nor independent of decentralized initiative. Indeed, organizations may promote internal information sharing to foster collaborations and peer learning and encourage cross-functional connections. Still, I would expect the visibility of relevant information to be tiered (public, internal, team-only), reflecting the importance of commonality.

Cross-cultural insights. An intriguing implication of the theory developed in this paper is that, ceteris paribus, knowing each other may result in less, rather than more communication, contrary to first intuition. In some cultures, such as Japan, acceptable communication is particularly expected to be harmony-preserving. This may reflect the homogeneity of a population, in which trust is naturally present and communication is particularly expected "not to rock the boat". Like for boundary setting, a rigorous empirical analysis of what drives acceptable exchange would be fascinating.

Alongside initial insights into the determinants of preferences for, and acquisition of, commonality information, this paper seeks to catalyze further research on these broader issues.

References

Admati, A. R., and M. Perry (1991) “Joint Projects without Commitment,” *Review of Economic Studies*, 58(2): 259–276.

Aghion, P. and J. Tirole (1997) “Formal and Real Authority in Organizations,” *Journal of Political Economy*, 105(1): 1–29.

Alesina, A., Baqir, R., and W. Easterly (1999) “Public Goods and Ethnic Divisions,” *Quarterly Journal of Economics*, 114(4): 1243–1284.

Alonso, R. and O. Camara (2016) “Persuading Voters,” *American Economic Review*, 106(11): 3590–3605.

Antic, N., Chakraborty, A. and R. Harbaugh (2022) “Subversive Conversations,” mimeo.

Anton, J. J. and D. A. Yao (2002) “The Sale of Ideas: Strategic Disclosure, Property Rights, and Contracting,” *Review of Economic Studies*, 69(3): 513—531.

Arrow, K., Blackwell, D., and M. Girshick (1949) “Bayes and Minimax Solutions of Sequential Decision Problems,” *Econometrica*, 17(3-4): 213–44.

Augenblick, N. and A. Bodoh-Creed (2018) “To Reveal or Not to Reveal: Privacy Preferences and Economic Frictions,” *Games and Economic Behavior*, 110(C): 318–329.

Aumann, R. J. and S. Hart (2003) “Long Cheap Talk,” *Econometrica*, 71(6): 1619–1660.

Barnby, J., Raihani, N. and P. Dayan (2022) “Knowing Me, Knowing You: Interpersonal Similarity Improves Predictive Accuracy and Reduces Attributions of Harmful Intent,” *Cognition* (225): 1–23.

Bénabou, R. (2013) “Groupthink: Collective Delusions in Organizations and Markets,” *Review of Economic Studies*, 80(2): 429–462.

Besley, T., and T. Persson (2024) “Organizational Dynamics: Culture, Design, and Performance,” *Journal of Law, Economics, and Organization*, 40: 394–415.

Birnbaum, A. (1961) “On the Foundations of Statistical Inference: Binary Experiments,” *Annals of Mathematical Statistics*, 32(2): 414–435.

Bloom, N., Sadun R. and J. Van Reenen (2012) “Americans Do IT Better: US Multinationals and the Productivity Miracle,” *American Economic Review*, 102(1): 167–201.

Braghieri, L. (2024) “Political Correctness, Social Image, and Information Transmission,” *American Economic Review*, 114(12): 3877–3904.

Bursztyn, L., González, A. and D. Yanagizawa-Drott (2020) “Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia,” *American Economic Review*, 110(10): 2997–3029.

Caillaud, B. and J. Tirole (2007) “Consensus Building: How to Persuade a Group,” *American Economic Review*, 97(5): 1877–1900.

Che, Y.K. and K. Mierendorff (2019) “Optimal Dynamic Allocation of Attention,” *American*

Economic Review, 109(8): 2993–3029.

Che, Y.K. and S.W. Yoo (2001) "Optimal Incentives for Teams," *American Economic Review*, 91: 525–541.

Clayton, C., Dos Santos, A., Maggiori, M., and J. Schreger (2025) "Internationalizing Like China," *American Economic Review*, 115(3): 864–902.

Crawford, V. P., and J. Sobel (1982) "Strategic Information Transmission," *Econometrica*, 50(6): 1431–1451.

Daughety, A., and J. Reinganum (2010) "Public Goods, Social Pressure, and the Choice between Privacy and Publicity," *American Economic Journal: Microeconomics*, 2(2): 191–221.

Dessi, R. (2008) "Collective Memory, Cultural Transmission, and Investments," *American Economic Review*, 98(1): 534–60.

Dewan, T. and D. Myatt (2008) "The Qualities of Leadership: Direction, Communication, and Obfuscation," *American Political Science Review*, 102: 351–368.

Dewatripont, M. and J. Tirole (2005) "Modes of Communication," *Journal of Political Economy*, 113(6): 1217–1238.

Dworczak, P. and G. Martini (2019) "The Simple Economics of Optimal Persuasion," *Journal of Political Economy,* 127(5): 1993–2048.

Dye, R. (1985) "Disclosure of Nonproprietary Information," *Journal of Accounting Research*, 23(1): 123–145.

Dziuda, W. and R. Gradwohl (2015) "Achieving Cooperation under Privacy Concerns," *American Economic Journal: Microeconomics*, 7(3): 142–173.

Edmonson, A. (1999) "Psychological Safety and Learning Behavior in Work Teams," *Administrative Science Quarterly*, 44(2): 350–383.

Farrell, J. and R. Gibbons (1989) "Cheap Talk Can Matter in Bargaining," *Journal of Economic Theory*, 48(1): 221–237.

Fernandez-Duque, M. (2022) "The Probability of Pluralistic Ignorance," *Journal of Economic Theory*, 202, 1–33.

Fershtman, C., and U. Gneezy (2001) "Discrimination in a Segmented Society: An Experimental Approach," *Quarterly Journal of Economics*, 116(1): 351–377.

Forges, F., and F. Koessler (2008) "Long Persuasion Games," *Journal of Economic Theory*, 143: 1–35.

Fudenberg, D., and J. Tirole (1991) "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory*, 53: 236–260.

Glaeser, E., Laibson, D., Scheinkman, J., and C. Soutter (2000) "Measuring Trust," *Quarterly Journal of Economics*, 115(3): 811–846.

Goldman, R. (2021) "Acceptable Discourse: Social Norms of Beliefs and Opinions," mimeo.

Haidt, J. (2024) *The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness*, Penguin Press.

Hansmann, H. (1996) *The Ownership of Enterprise*, Harvard University Press, ISBN 9780674001718.

Holmström, B. (1977) "On Incentives and Control in Organizations," Thesis, Stanford University, 274–278.

Hörner, J. and A. Skrzypacz (2016) "Selling Information," *Journal of Political Economy*, 124(6): 1515–1562.

Inostroza, N. and A. Pavan (2025) "Adversarial Coordination and Public Information Design," *Theoretical Economics*, 20(2): 763–813.

Itoh, H. (1991) "Incentives to Help in Multi-Agent Situations," *Econometrica*, 59: 611–636.

Janis, I. (1972) *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*, Boston, Houghton Mifflin.

Kamenica, E. and M. Gentzkow (2011) "Bayesian Persuasion," *American Economic Review*, 101(6): 2590–2615.

Katz D., and F. Allport (1931) *Students' Attitudes: A Report of the Syracuse University Reaction Study.* Syracuse, NY: Craftsman Press.

Laffont, J.J. and D. Martimort (1997) "Collusion Under Asymmetric Information," *Econometrica*, 65(4): 875–911.

Loury, G. (1994) "Self-Censorship in Public Discourse: A Theory of "Political Correctness" and Related Phenomena," *Rationality and Society*, 6(4): 428–461.

Mathevet, L., Perego, G., and I. Taneva (2020) "On Information Design in Games," *Journal of Political Economy*, 128(4): 1370–1404.

Milgrom, P. (1981) "Rational Expectations, Information Acquisition, and Competitive Bidding," *Econometrica*, 49(4): 921–943.

Milgrom, P. and J. Roberts (1990) "The Economics of Modern Manufacturing: Technology, Strategy, and Organization," *American Economic Review*, 80(3): 511–528.

Morris, S. (2001) "Political Correctness," *Journal of Political Economy*, 109(2): 231–265.

Orlov, D., Skrzypacz A. and P. Zryumov (2025) "Trading Information," mimeo.

Page, S. (2019) *The Diversity Bonus: How Great Teams Pay Off in the Knowledge Economy,* Princeton University Press.

Prekopa, A. (1973) "On Logarithmic Concave Measures and Functions," *Acta Scientiarum Mathematicarum,* (Szeged), 34: 335–343.

Prendergast, C. (2007) "The Motivation and Bias of Bureaucrats," *American Economic Review*, 97(1): 180–196.

Prentice, D. and D. Miller (1993) "Pluralistic Ignorance and Alcohol Use on Campus: Some Consequences of Misperceiving the Social Norm," *Journal of Personality and Social Psychol-*

ogy, 64(2): 243–256.

Putnam, R. (1995) "Bowling Alone," *Journal of Democracy*, 6(1): 65–78.

Rotemberg, J. J. and G. Saloner (2000) "Visionaries, Managers, and Strategic Direction," *Rand Journal of Economics*, 31(4): 693–716.

Schein, E.H. and W.G. Bennis (1965) "Personal and Organizational Change Through Group Methods: The Laboratory Approach," (Book Review), *Administrative science quarterly* 10: 530.

Simmel, G. (1906) "The Sociology of Secrecy and of Secret Societies," *American Journal of Sociology*, 11(4): 441–498.

Sobel, J. (1985) "A Theory of Credibility," *Review of Economic Studies*, 52(4) : 557–573.

Sunstein, C. (2018) *#Republic: Divided Democracy in the Age of Social Media*, Princeton University Press.

Taneva, I. (2019) "Information Design," *American Economic Journal: Microeconomics*, 11(4): 151–85.

Tirole, J. (1986) "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics, & Organization*, 2(2): 181–214.

Van den Steen, E. (2010) "On the Origin of Shared Beliefs (and Corporate Culture)," *Rand Journal of Economics*, 41(4): 617–648.

Wald, A. (1947) "Foundations of a General Theory of Sequential Decision Functions," *Econometrica*, 15(4): 279–313.

Appendix

A Non convex-concave extension

While convex-concave payoff functions cover many applications of interest, payoffs may be more complex in some other interesting cases.

A non convex-concave example. Building on Example (e) in the text, consider the situation in which the repeated relationship may break (exogenous turnover) with Poisson probability $1-\mu$ in each period, capturing the idea that an important driver of cooperation is the belief that the team will interact repeatedly. The model is otherwise the intensive margin one, in which individual incentives, $(\alpha/2)+\beta$, are smaller than the collective benefit $(\alpha+\beta)$: see Illustrations (a) and (f). Let ξ_i denote the discount factor of agent i, with $\xi_1 \geq \xi_2$. While ξ_i captures impatience, agent i's real discount factor is thus $\xi_i\mu$. The principal gains by pretending that the relationship will be stable, but agents might independently exchange information about the likely turnover. The cooperative equilibrium exists if and only if:

$$\frac{\xi_2\mu}{1-\xi_2\mu}[\alpha+\beta-\gamma] \geq \gamma - (\frac{\alpha}{2}+\beta);$$

let $\underline{\mu}_2$ satisfy this condition with equality. And so, under *congruence uncertainty*,

$$V_i(\mu) = \Big[\frac{\alpha+\beta-\gamma}{1-\xi_i\mu}\Big]\mathbb{1}_{\{\mu\geq\underline{\mu}_2\}},$$

is convex when $V_i(\mu) > 0$, which implies that V_i is not convex-concave and that $V_i(\mu)/\mu$ is maximized at $\underline{\mu}_2$ or 1. And so

$$\Delta(\mu) = \frac{(\alpha+\beta-\gamma)(\xi_1-\xi_2)\mu}{(1-\xi_1\mu)(1-\xi_2\mu)}$$

which need not satisfy the convenient shape properties that allow us to compare straightforwardly information aversions. Besides, the payoffs fail to be convex-concave, as V_i is convex beyond $\underline{\mu}_2$. But one can nonetheless analyze communication using techniques similar to those in the paper. Consider for instance the bad-news model. Because $V_i(\mu)$ is convex beyond $\underline{\mu}_2$,

$$\mu_i^B = \arg\max\Big(\frac{V_i(\mu)}{\mu}\Big) \in \{\underline{\mu}_2, 1\}$$

is generically unique. Under slow learning and hard information, when they disagree on the amount of learning ($\mu_i^B = \underline{\mu}_2 < \mu_j^B = 1$), then agent j can force agent i to learn until $\mu = 1$ (or more generally μ_j^* if one takes a smooth approximation of the payoffs above). The analysis of Section 4 therefore applies.

Interestingly, the weak-link may be more information averse than the strong link, as $\Delta(\mu)$

is increasing in μ. So, it may be the case that $\mu_1^B = 1 > \mu_2^B = \underline{\mu}_2$. This however does not invalidate the analysis of Section 4, which only presumes that for each i, each μ_i^* is uniquely defined.

A few generalities. For simplicity, I will focus here on the bad-news case with smooth payoffs (deep uncertainty) and slow learning.

(1) Notice, first, that the number of inflection points below $\mu_i^B \equiv \arg\max\{V_i(\mu)/\mu\}$ is irrelevant: Starting in $(0, \mu_i^B)$, player i wishes to continue learning until $\mu = \mu_i^B$ and stop there, regardless of the shape of V_i in that range.

(2) The new feature when the convex-concavity assumption is relaxed is that optimal learning may not be monotonic (although it is in the example discussed above). Consider Figure 9. As pointed out in (1), player i would like learning to continue until $\mu = \mu_i^B$, regardless of the shape of V_i up to μ_i^B. In Figure 9, the function $V_i(\mu)/\mu$ has a second maximum, $\hat{\mu}$, this time a local one. At μ_i^B, player i would like to stop learning as this will reduce $V_i(\mu)/\mu$. A bit above $\mu^\dagger$ (with $V_i(\mu^\dagger)/\mu^\dagger = V_i(\hat{\mu})/\hat{\mu}$), player i would want to experiment, and then stop when reaching $\hat{\mu}$.

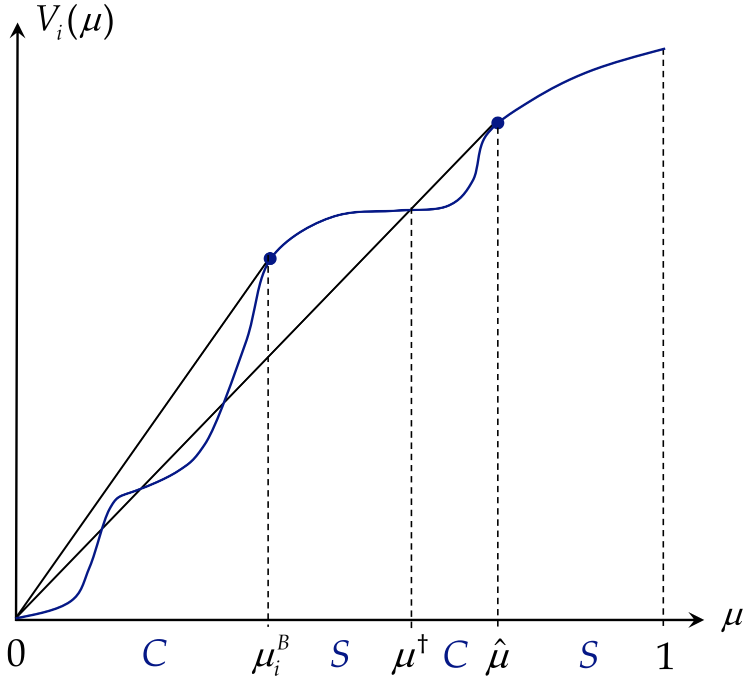


Figure 9: (non convex-concave payoffs).
("*C*" stand for "continue learning", "*S*" for "stop learning")

B Payoffs: Illustrations (b) through (e)

(b) Outside options (extensive margin)

In the outside-options illustration, the agents and the principal share the output of a supermodular production technology, whether monetary or prestige, in some way (shares can be taken equal after a renormalization). In this extensive margin illustration, each agent's decision is whether to consent to forming a team with the other agent ($a_i = 1$ if she consents, $a_i = 0$ is she does not): The team forms if and only both consent. Letting r_i denote agent i's opportunity cost/reservation utility, player i's payoff is

$$u_i(a_i, a_j, \omega) = [\alpha + \beta\omega - r_i]a_1a_2. \tag{A.1}$$

Let $\mu = E[\omega]$ be the probability that agents will have the same vision of the project and pull in the same direction when they need to interact, and $r_i = r - \theta_i$ denote the reservation utilities, with $\theta_2 = 0 \leq \theta_1 < \theta_P$, ensuring that agent 1 is the strong link. And so $V_2(\mu) = [\alpha + \beta\mu - r_2]1_{\alpha+\beta\mu\geq r_2}$, and more generally $V_i(\mu) = [\alpha + \beta\mu - r_2 + (\theta_i - \theta_2)]1_{\alpha+\beta\mu\geq r_2}$. The payoff functions are convex-concave.[36]

In particular, letting the threshold satisfy $\alpha + \beta\underline{\mu}_2 = r_2$ and be in $(0, 1)$, net payoff functions satisfy $V_1(\mu) = V_2(\mu) + \Delta(\mu)$, where

$$\Delta(\mu) = (\theta_1 - \theta_2)\mathbb{1}_{\mu\geq\underline{\mu}_2}.$$

Deep uncertainty. Next, we introduce some additional uncertainty: At the beginning of stage 2 (that is, after posterior beliefs μ about congruence are established and before actions are selected), the reservation utilities are revealed. Namely, the common shifter r is drawn from distribution $F(r)$ on $(\underline{r}, +\infty)$ with $\underline{r} \geq \alpha$ (to normalize $V_i(0)$ at 0), with a log-concave density. Expected payoffs become

$$V_i(\mu) = \int_{\underline{r}}^{\alpha+\beta\mu} \big[\alpha + \beta\mu - r + (\theta_i - \theta_2)\big]dF(r)$$

and are convex-concave.[37] And so the payoff differential takes the more general from:

$$\Delta(\mu) = (\theta_1 - \theta_2)F(\alpha + \beta\mu).$$

Payoffs are almost identical to those in the intensive-margin case, although one cannot quite replace the cost of effort γ by the reservation utility r. The reason for this is interesting: free-riding can occur in the intensive-margin case, but not in the extensive-margin one. This

[36] V_2 is here convex, while V_1 is convex if and only if $\theta_1 - \theta_2$ is small enough.

[37] $V_i'(\mu) = \beta[F(\alpha+\beta\mu) + (\theta_i - \theta_2)f(\alpha+\beta\mu)]$ and so $V_i''(\mu) = \beta^2[f(\alpha+\beta\mu) + (\theta_i - \theta_2)f'(\alpha+\beta\mu)]$. Because f'/f is decreasing, $f' + \lambda f$ changes sign at most once for any λ. Therefore V_i'' changes sign (at most once) from + to −. So, payoffs are convex-concave (or convex).

explains the difference between the incremental ($\alpha/2$) and total (α) values for the additive part of the payoff. The insights, including the convex-concave property of both the threshold and the deep-uncertainty cases, are however very similar.

(c) Kin/parochial altruism

Consider the intensive-margin model, but assume that agents differ in their social preferences rather than in their effort cost. Agent i internalizes a fraction $\theta_i\mu$ of agent j's material utility, with $\theta_1 \geq \theta_2$, where μ is now the perception of kinship. Hard-wired altruism is thus conditioned on being alike along some dimension (be it gene, religion, politics, hobbies, etc). And so the congruence-uncertainty payoffs are:

$$u_i = [1+\theta_i\omega]\Big[\alpha\Big(\frac{a_1+a_2}{2}\Big)+\beta a_1 a_2\Big]-\gamma a_i.$$

The weak link (the agent with the lower degree of kin altruism) is willing to exert effort if and only if $\mu \geq \underline{\mu}_2$ where $(1+\theta_2\underline{\mu}_2)(\frac{\alpha}{2}+\beta)=\gamma$. In that range the agents' payoffs are related in the following way:

$$\mathit{\Delta}(\mu) = (\theta_1-\theta_2)(\alpha+\beta)\mu \mathbb{1}_{\mu\geq\underline{\mu}_2}.$$

Deep uncertainty. Suppose, next, that the cost of effort γ is distributed according to cdf $F(\gamma)$ with log-concave density $f(\gamma)$ on $[\underline{\gamma},+\infty)$, where $\underline{\gamma} \geq (\frac{\alpha}{2}+\beta)$ to ensure there is no cooperation when it is common knowledge that the agents are dissonant. The cooperative equilibrium emerges if and only if $\gamma \leq \gamma^*(\mu)$, where $[1+\theta_2\mu]\Big[\frac{\alpha}{2}+\beta\Big]=\gamma^*(\mu)$. Again, the smoothed model has convex-concave payoffs.[38] And

$$\Delta(\mu) = (\theta_1-\theta_2)\int_{-\infty}^{(1+\theta_2\mu)(\frac{\alpha}{2}+\beta)}(\alpha+\beta)\mu dF(\gamma).$$

(d) Different priors and other non-Bayesian environments

Suppose that the two agents have different priors, $\mu_{0,1} \geq \mu_{0,2}$, and that they "agree to disagree". When they jointly receive the same signals on the congruence on the task, their posterior beliefs co-vary and satisfy $\mu_1 \geq \mu_2$. One can therefore take $\Xi_2(\mu)=\mu$ as the updated congruence (and denote $\mu_0=\mu_{0,2}$). Writing through an abuse of notation Ξ_i as a function of μ,[39] and letting $\mu_{0,i}=\theta_i\mu_0$ with $\theta_2=1\leq\theta_1$, we can express posterior beliefs as function of the likelihood ratio L associated with a sequence of realizations:[40] $\Xi_i(\mu)=\frac{\theta_i\mu_0 L}{\theta_i\mu_0 L+1-\theta_i\mu_0}$,

[38] $V_i(\mu)=\int_{-\infty}^{\gamma^*(\mu)}\big[(1+\theta_i\mu)(\alpha+\beta)-\gamma\big]dF(\gamma)$. And so, there exists $\{A_i,B_i\}$ with $B_i>0$ and $A_i>0$ if $i\in\{P,1\}$, and $=0$ if $i=2$,

$$\text{sgn}(V_i''(\mu))=\text{sgn}\Big(A_i\frac{f'(\gamma^*(\mu))}{f(\gamma^*(\mu))}\mu+B_i\Big).$$

Because f is log-concave, the function V_i is convex-concave (or convex).

[39] This is feasible because the mapping $t\rightarrow\mu$ is, in the absence of bad news, one-to-one.

[40] The likelihood ratio is $L=(1-\rho)^{-t}$ in the bad-news case (if no bad news has accrued), $L=(1-\rho)^t$ in the good-news case (if no good news has accrued), and $L=(\rho_1)^n(1-\rho_1^m)/(\rho_0)^n(1-\rho_0)^m$ under n favorable signals and m unfavorable ones.

yielding

$$\Xi_1(\mu) = \frac{\mu}{\mu + (1-\mu)k}$$

where $k \equiv \frac{1-\theta_1\mu_0}{\theta_1(1-\mu_0)} \leq 1$. Consider, say, the extensive-margin model with a common reservation utility r: $u_i = [\alpha + \beta\omega - r]a_1 a_2$.

Congruence-uncertainty. Agent 2 is willing to join the team if and only if $\alpha + \beta\Xi_2(\mu) \geq r$. Agent i's payoff for $\mu \geq (r-\alpha)/\beta$ is therefore $V_i(\mu) = \alpha + \beta\Xi_i(\mu) - r$, and so

$$\Delta(\mu) = \beta[\Xi_1(\mu) - \Xi_2(\mu)]\mathbb{1}_{\mu \geq \underline{\mu}_2}.$$

And so

$$\left(\frac{\Delta(\mu)}{\mu}\right)' < 0 \quad \text{and} \quad \Delta(\mu) - \mu\Delta(1) > 0$$

since $\Xi_1(\mu)$ is of the form $a\mu/(b\mu + c)$, $\Xi_2(\mu) = \mu$ and $\Delta(1) = 0$.

Deep uncertainty. One can also smooth the payoff functions in the now-familiar manner. Let r be drawn from cumulative distribution $F(r)$. Agent i's payoff is

$$V_i(\mu) = \int_{-\infty}^{\alpha+\beta\Xi_2(\mu)} [\alpha + \beta\Xi_i(\mu) - r]dF(r).$$

Finally, let us find conditions under which the functions $V_i(\mu)$ are convex-concave. $V_2(\mu)$ is always convex. As for $V_1(\mu)$,

$$V_1''(\mu) = \beta\Xi_1''(\mu)F(\alpha+\beta\mu) + \beta^2[2\Xi_1'(\mu) - 1]f(\alpha+\beta\mu) + \beta^3[\Xi_1(\mu) - \mu]f'(\alpha+\beta\mu).$$

Sufficient conditions for convex-concavity are (a) the density $f(r)$ is log-concave and unimodal (as are the Normal, Logistic or Laplace distributions), (b) $k < 1/2$, and (c) $\alpha \leq m \leq \alpha + \beta$, where m is the mode of f.[41] Δ is given by:

$$\Delta(\mu) = \beta[\Xi_1(\mu) - \mu]F(\alpha+\beta\mu).$$

Beyond ex-ante heterogeneous beliefs, the model also accommodates ex-post heterogeneous beliefs. The agents may update differently and imperfectly after each elementary experiment (they may under- or over- react to the outcome of an experiment), as long as the statistical biases are commonly known, so that the posterior beliefs map one-to-one to each other: $\mu_{t,i} = f(\mu_{t,j})$. For example, starting from a common prior $\mu_{0,1} = \mu_{0,2} = \mu_0$, the agents might update beliefs in the bad-news model after t experiments without bad news so that

[41] Under our maintained assumption that the density $f(r)$ is log-concave, let m denote the mode of f ($f' > 0$ for $r < m$ and $f' < 0$ for $r > m$). The first term in $V_1''(\mu)$ is negative and decreasing in μ. The second term is decreasing in μ from $2/k - 1 > 0$ to $2k - 1$. So if $k < 1/2$, the second term is first positive and then negative as μ grows from 0 to 1. The third term's sign depends on whether $\alpha + \beta\mu \gtrless m$; assuming $\alpha \leq m \leq \alpha + \beta$, this third term is positive and then negative as μ grows from 0 to 1. Thus the sum of the three terms changes sign (from + to $-$) at most once.

$[(1-\mu_{t,1})/\mu_{t,1}]/(1-\rho_1)^t = [(1-\mu_{t,2})/\mu_{t,2}]/(1-\rho_2)^t$, where $\rho_i \gtrless \rho$ measures the over- or under-reaction to the news and ρ corresponds to the true model.

(e) Ancillary interactions and social collateral

Consider the following variant of the intensive-margin model, in which one's value of working with someone is not limited to the joint productivity. There are also social interactions during work time, at the coffee machine, at the cafeteria lunch, or social events and dinners outside the workplace. These can be more or less pleasant, or even exist or not exist. This can be captured by the following *per-period* utility function:

$$\left[\alpha\Big(\frac{a_1+a_2}{2}\Big)+\beta a_1 a_2\right]-\gamma a_i+[\theta_i\mu]b_1b_2.$$

The first bracket is the team's productivity, the second term is the individual cost of cooperation, and the third term the social benefit of working with someone one gets along with. Suppose that the sociability parameters satisfy $\theta_1 \geq \theta_2$ (agent 2 is the weak link). $b_i \in \{0,1\}$ stands for agent i's choice of maintaining ancillary interactions with agent j in the period. The ancillary-action benefit from the feeling of interacting with a kin, $\theta_i\mu$, based on perceived kinship μ, has no incentive role if a static context. So, suppose that the two agents play a repeated game with discount factor ξ in which they use trigger strategies to maintain cooperation; they use the social payoff $\theta_i\mu$ as collateral in the relationship, that is surrendered if at least one of the agents deviates (the two agents revert to $a_1 = a_2 = 0$ and no longer interact outside work after a deviation). Assuming for simplicity that the perceived kinship remains constant (agents might learn about it further through interactions), the cooperative equilibrium exists if and only if $\mu \geq \underline{\mu}_2$, where:

$$\frac{\xi}{1-\xi}[\alpha+\beta-\gamma+\theta_2\underline{\mu}_2]=\gamma-(\frac{\alpha}{2}+\beta).$$

Such social interactions enable cooperation if the latter cannot be sustained in the absence of ancillary interactions, i.e. if

$$\frac{\xi}{1-\xi}\Big[\alpha+\beta-\gamma\Big]<\gamma-\Big(\frac{\alpha}{2}+\beta\Big).$$

Provided that μ exceeds the threshold enabling cooperation, the reduced-form payoff functions are $V_i(\mu)=\frac{\alpha+\beta-\gamma+\theta_i\mu}{1-\xi}$, and so

$$\Delta(\mu)=\frac{(\theta_1-\theta_2)\mu}{1-\xi}\mathbb{I}_{\mu\geq\underline{\mu}_2}.$$

Deep uncertainty. As for the smooth version of this application, let the cost γ be distributed according to c.d.f. $F(\gamma)$. Let $\gamma^*(\mu) \equiv (1-\xi)\big(\frac{\alpha}{2}+\beta\big)+\xi(\alpha+\beta+\theta_2\mu) \equiv a+b\mu$ denote the highest cost that enables incentive compatibility. Then $\alpha+\beta-\gamma^*(\mu)+\theta_i\mu > 0$, and

$V_i(\mu) = \int_{-\infty}^{\gamma^*(\mu)} \frac{(\alpha+\beta-\gamma+\theta_i\mu)}{1-\xi} dF(\gamma)$ is convex concave if the density $f(\gamma)$ is log concave. And so

$$\Delta(\mu) = \frac{(\theta_1 - \theta_2)}{1-\xi}\mu F(\gamma^*(\mu)).$$

C Optimal stopping rules in the good-news model

Suppose that $\mu_0 \in (\mu_i^G, 1)$. The player must set a "target" $\mu \leq \mu_0$ such that learning stops at μ unless good news accrue in between ($\mu = \mu_0$ corresponds to stopping immediately). If $\mu \leq \mu_i^G$, then we know from the text that player i may as well continue until $\mu = 0$. This however would be costly to player i as $V_i(\mu_0) > \mu_0 V_i(1) + (1-\mu_0)V_i(0)$ (since $V_i(0) = 0$ and $V_i(\mu_0)/\mu_0 > V_i(1)$).

So, suppose that $\mu > \mu_i^G$. To show that $\frac{1-\mu_0}{1-\mu}V_i(\mu) + \frac{\mu_0-\mu}{1-\mu}V_i(1)$ is increasing in μ (implying that stopping at μ_0 is optimal: $\mu = \mu_0$), i.e. $V_i'(\mu) > \frac{V_i(1)-V_i(\mu)}{1-\mu}$, study the function

$$\phi_i(\mu) \equiv \frac{V_i(1) - V_i(\mu)}{1-\mu}$$

on $(\mu_i^G, 1)$. Note that $\phi_i'(\mu) = \frac{\phi_i(\mu) - V_i'(\mu)}{1-\mu}$. Let $\tilde{\mu}_i$ denote the inflection point of V_i. Either $\tilde{\mu}_i \leq \mu_i^G$, and then V_i is concave over $(\mu_i^G, 1)$. Then,

$$\phi_i(\mu) = \frac{\int_\mu^1 V_i'(t)dt}{1-\mu} \leq V_i'(\mu)$$

and so $\phi_i'(\mu) < 0$.

Finally, suppose that $\tilde{\mu}_i > \mu_i^G$ and that μ belongs to the convex part above L, where L is the line between $(0,0)$ and $(1, V_i(1))$. Then

$$V_i'(\mu) \geq \frac{V_i(\mu)}{\mu} \geq V_i(1) \geq \phi_i(\mu),$$

where the first inequality stems on our focus on $\mu \in (\mu_i^G, \tilde{\mu}_i)$ and the fact that $\tilde{\mu}_i < \mu_i^B$; and so $\phi_i'(\mu) < 0$ again. ■

D Plausible deniability

As in the text, take the bad-news model and, for the sake of exposition, the intensive-margin illustration with congruence uncertainty.

Agent i choose $a_i = 0$ when learning that there is no congruence. If not, agent i selects $a_i = 1$ when communication has gone t stages and agent j left τ_j of agent i's disclosures orphan, whenever

$$\frac{\alpha}{2} + \beta\nu(t, \tau_j) \geq \gamma_i. \tag{A.2}$$

If condition A.2 is satisfied for both agents, then each exerting effort when having no bad news is an equilibrium (unlike exact information, t and τ_j are common knowledge).

Denote, as earlier, by $\underline{\mu}_i$ agent i's cutoff for being willing to cooperate: $\frac{\alpha}{2} + \beta\underline{\mu}_i \equiv \gamma_i$. Thus, experimentation will not stop unless:

$$\nu(t, \tau_j) \geq \underline{\mu}_i \quad \text{for all } i. \tag{A.3}$$

Conversely, suppose that (A.3) is satisfied and that both agents' payoff functions are convex-concave, so $\mu_i^* = \underline{\mu}_i$. If (A.3) is not satisfied and the agents stop experimenting, then at least one agent prefers choosing $a_i = 0$ even if they have not received bad news, and so the other agent picks $a_j = 0$ regardless of acquired information. Both obtain payoff 0.

Finally, let us show that, at least for $T = +\infty$ (unbounded number of possible experiments), both agents' ex-ante truthful disclosure of dimensions in which they don't know their type is an equilibrium. Along the equilibrium path, available experiments are then as if $x_{i,t} = 1$ for all $\{i, t\}$: Experiments are conducted until the first T^* such that $\mu_0/[\mu_0 + (1-\mu_0)(1-\rho)^{T^*}] \geq \underline{\mu}_2$, provided no bad news accrues and both agents' payoff function are convex-concave (which is equivalent to $\alpha/2 > \underline{\mu}_2(\alpha + \beta - \gamma_2)$). Suppose that agent i conceals some dimension in which she does not know her type and that dimension belongs to the first T^* experiments. Then both agents' payoffs are 0 (as agent j interprets the absence of response as a bad news), and agent i's payoff is therefore reduced. This shows that agents may define themselves acceptable discourse.

E Complements on communication of soft information

Consider the bad news, congruence-uncertainty model with a single possible elementary experiment ($T = 1$) and the en-bloc-disclosure communication protocol.

(a) Suppose first that agents are symmetrical (so, we drop their index), that $\mu_0 \geq \underline{\mu}$ (hence, the agents cooperate in the absence of learning), and that they prefer not to learn:

$$\frac{\mu_0}{\mu_1} V(\mu_1) < V(\mu_0),$$

where, recall, $\mu_1 \equiv \mu_0/[\mu_0 + (1-\mu_0)(1-\rho)]$. Assume that they simultaneously disclose their type in dimension 1 and that they interpret the other agent's disclosure as truthful; it is then optimal to disclose one's true type conditional on disclosing (if not, the other agent will believe with probability 1 that $\omega = 0$). Would an agent, say agent 2, want to deviate by not disclosing at all? Agent 2, regardless of type, then responds to agent 1's disclosure by mimicking (possibly feigning) congruence; and so that response brings no information to agent 1. Agent 1 then thinks that agent 2 will pick action $a_2 = 1$ with probability

$\mu_0 + (1-\mu_0)(1-\rho)$ (the probability that agent 2 has not received bad news) and action $a_2 = 0$ with probability $(1-\mu_0)\rho$ (bad news accrued). So if μ_0 exceeds only slightly $\underline{\mu}$ (the minimum level inducing the agents to cooperate when they expect the other to do so for sure), agent 1 picks $a_1 = 0$. This implies that agent 2 is worse off not disclosing and that the presumed over-communication is an equilibrium behavior.

(b) Consider the previous example, but with an asymmetry in preferences (depicted by $V_i(\mu)$), with agent 1 being the strong link, and $\mu_0 \geq \underline{\mu_2}$. Suppose that agent 1 still prefers not to learn ($\mu_0 V_1(\mu_1) < \mu_1 V_1(\mu_0)$), while agent 2 has opposite preferences ($\mu_0 V_2(\mu_1) < \mu_1 V_2(\mu_0)$). So, the equilibrium with communication is not dominated if it exists. Suppose agent 1 does not disclose (and therefore mimics agent 2's announced type in a second round). Agent 2 thinks that agent 1 will choose action $a_1 = 0$ with probability $(1-\mu_0)\rho$, and so, if $\mu_0 - \underline{\mu_2}$ is small enough, chooses $a_2 = 0$. So, simultaneous disclosure is again an equilibrium, but now an Pareto-undominated one.