Cognitive Games and Cognitive Traps*

Jean Tirole†

September 21, 2016

Abstract

The paper defines "cognitive games" as games in which players choose their information structures and play a normal- or extensive-form game under the resulting information structures. It introduces the concept of "expectation conformity", according to which the vector of information structures is potentially the object of self-fulfilling prophecies. Unilateral expectation conformity holds that an information-acquiring player has more incentive to select a given information structure if he expected to do so; when multiple players acquire information, expectation conformity may also result from a second force, labeled increasing differences, reflecting strategic complementarities across players' informational choices.

For example, zero-sum games never give rise to self-fulfilling cognition while linear-quadratic games with binary information structures do, under either strategic complementarity or strategic substitutability. The paper defines a class of games for which a direct characterization of the expectation conformity property in terms of rotation points can be obtained. This class comprises many games of interest to economists, starting with the cognition-augmented lemons model.

Keywords: cognition, expectation conformity, adverse selection.
JEL numbers: C72; C78; D82; D83; D86.

1 Introduction

Cognition is costly; thinking and memorizing, obtaining financial, engineering and legal expertise and brainstorming with others consumes resources (in amounts that depend on the urgency to

*The author is very grateful to participants at conferences and seminars at MIT, EARIE, Toulouse (Games Toulouse and Toulouse conference in honor of Jacques Crémer), Game Theory Society World Congress, Econometric Society meetings, Northwestern University, the University of Chicago, and the University of Tokyo, and to Gabriel Carroll, Drew Fudenberg, Christian Gollier, Navin Kartik, Alessandro Pavan, Di Pei, Jean-Charles Rochet, Paul Seabright, Joel Sobel, Pierre-Luc Vautrey and Olivier Wang for helpful discussions and comments, and to a European Research Council advanced grant (European Community's Seventh Framework Programme (FP7/2007-2013) Grant Agreement no. 249429 and European Union's Horizon 2020 research and innovation programme, Grant Agreement no. 669217) for financial support.

†Toulouse School of Economics (TSE) and Institute for Advanced Study in Toulouse (IAST), University of Toulouse Capitole, jean.tirole@tse-fr.eu

act, the cognitive load or the context) and is most often covert. This paper considers the implications of costly cognition in a multi-player context. It defines cognitive games as games in which players privately select information structures and then play an arbitrary, normal- or extensive-form game under the resulting information structures. For convenience, we will label the latter game the “stage-2 game” and the information-acquisition stage “stage 1”. If the stage-1 choice of cognition when unobservable and the choice of stage-2 strategy can be viewed as simultaneous from a game-theoretic perspective, our characterizations will purport to classes of second-stage games, and so we keep the two stages separate for the exposition.

Individually optimal and anticipated cognitions can, under some circumstances, be expected to be “strategic complements”. Section 2 accordingly introduces expectation-conformity (EC). EC holds if there exist information structures $\mathcal{F} = \{\mathcal{F}_i\}_{i \in I}$ and $\widehat{\mathcal{F}} = \{\widehat{\mathcal{F}}_i\}_{i \in I}$ such that each player $i \in I$ has more incentive to acquire information $\widehat{\mathcal{F}}_i$ rather than $\mathcal{F}_i$ when all players expect information structure $\widehat{\mathcal{F}}$ rather than $\mathcal{F}$. A special case of interest arises when information structures are ordered and a finer information structure is costlier than a coarser one. A cognitive trap obtains when two ordered equilibria co-exist and information-acquiring players are better-off in the low-cognition equilibrium.

EC can result, for each player i, from either of two forces. The first, called *unilateral expectation conformity* (UEC), looks at how other players' expectations about player i's cognitive choice shape her actual choice of information acquisition, keeping these other players' information structures fixed. For example, if information structures are ordered, UEC requires that player i have higher incentives to acquire information if he is expected to acquire more. UEC is equivalent to EC in games with one-sided information acquisition. When multiple players acquire information, EC can result (possibly exclusively) from another force, that I label *increasing differences* (ID), since it is so closely related to the standard notion of increasing differences in supermodular games: taking the other players' expectation of player i's information structure as given, ID prevails if player i's payoff satisfies the increasing differences property relative to her and others' *actual* choices of information structures. ID is equivalent to EC for overt information structures, in which the players' choices of information structures become common knowledge at the start of the second-stage game, and in some specific games with covert information acquisition.

EC straightforwardly leads to a multiplicity of equilibrium information structures for appropriate information acquisition cost functions. By contrast, equilibrium uniqueness prevails if EC is never satisfied.

To gain some intuition about EC in games where multiple players acquire information, I first show that EC is violated when the stage-2 game is a two-person, zero-sum game. I then consider the class of games with linear-quadratic payoffs and binary information structures (players can learn the state of nature or remain ignorant). Under symmetric information, the stage-2 game exhibits either strategic substitutability or strategic complementarity and admits a unique and stable equilibrium. Under covert information acquisition, EC obtains regardless of the nature of stage-2 strategic interaction. Besides helping us forge some intuition about what drives EC, this class of preferences also shows that EC is unrelated to a possible multiplicity of equilibria in the stage-2 game or to strategic complementarity or substitutability in the corresponding interactions.

The paper then turns to a prominent class of games with one-sided information acquisition. Section 3 studies "generalized lemons environments," in which one of the players acquires information and then picks between an interaction-free option and another option whose return depends on the other player's beliefs; for example, she may decide between playing a game with the other player or opting out. The key assumption is that news that makes the first player want to interact with the second player also makes the second player react in an unfriendly manner. This class of games includes the lemons model of Akerlof (augmented with one-sided information acquisition of soft or hard information) and a number of other games of interest in economics. Section 3 derives a sufficient (and, under further assumptions, necessary) condition for such games to satisfy expectation conformity, assuming that the class of information structures is composed of rotations, and shows that EC holds for sufficient "gains from interaction," although perhaps not for low gains from interaction. The condition for EC to hold is much easier to check than verifying directly that expectation conformity prevails. To further our intuition about what drives EC in generalized lemons games, it then applies the general result to the cognition-augmented lemons game under directed and non-directed search.

The paper then derives the economic implications of this characterization, starting with three closely-related topics: cognitive traps, disclosure and cognitive styles. In generalized lemons games, the information-acquiring player is worse off in a high-information-intensity equilibrium; this cognitive trap is due to the unfriendly reaction of the other player in response to the exacerbated adverse selection. We then modify the game by assuming that the cost of acquiring information is hard instead of soft and so the information-acquiring player can choose to disclose how much attention or external resources he has devoted to the issue. Such disclosure turns out to be mostly irrelevant; the intuition is related to the cognitive-trap phenomenon: Disclosure serves to demonstrate that one is knowledgeable, which in the end is not a good idea. Along a similar vein, it is then shown that it is optimal to pose as an "informational puppy dog," i.e., to convince the other player that one is dumb or busy.

Finally Section 3 derives results for a general "anti-lemons environment" in which what makes the first player want to interact makes the second player want to behave in a friendly manner. Sufficient conditions for the presence or impossibility of EC are then supplied, as well as examples of economic games of interest that are anti-lemons games.

This paper, to the best of my knowledge, is the first to define expectation conformity (EC) and to investigate synergies in information acquisition (ID) and between expectation and actual choice (UEC). While a small literature[1] has studied specific environments in which one or several agents covertly acquire information prior to playing a game, these synergies have not been emphasized. Yet, besides being of theoretical interest, EC has important economic consequences. In particular, interactions or markets may switch behavior abruptly; for instance, the analysis of Section 3 implies that asset markets can tip from a pattern in which the assets receive little scrutiny to one in which they are heavily scrutinized by participants;[2] it also has implications for disclosure and transparency patterns.

Section 4 concludes with alleys for future research. Omitted proofs can be found in the Appendix.

Broader motivation: Covert investments

The paper's emphasis will be on information acquisition, a choice motivated both by the applications and by the fact that cognitive investments are the ultimate covert investments. But there is interest in other forms of covert investments: in capacity, learning by doing, arms buildup, and so on. To prepare the ground for the subsequent material, it is useful to consider one such environment. Suppose that there are two players, playing a "second-stage" normal-form game with actions a_i, $a_j \in \mathbb{R}$, and picking an "first-stage" investment $\rho_i \in \mathbb{R}$ at increasing investment cost $C_i(\rho_i)$. Payoffs are

$$[\phi_i(a_i, a_j) - \psi_i(a_i, \rho_i)] - C_i(\rho_i),$$

where all functions are C^2 and satisfy for all i, j:

$$\frac{\partial^2 \psi_i}{\partial a_i \partial \rho_i} < 0 \quad \text{and} \quad \frac{\partial^2 \phi_i}{\partial a_i \partial a_j} \begin{cases} > 0 & (SC) \\ \text{or} & \\ < 0 & (SS). \end{cases}$$

That is, we assume that the investment ρ_i lowers the marginal cost of action a_i, and that the strategic interaction involves either strategic complementarity (SC) or strategic substitutability (SS). One may have in mind that a_i is a quantity, ρ_i an investment that lowers the marginal cost of production, and production a_j is a strategic complement or substitute to production a_i.

Assume for simplicity that (only) player i invests and that, were her investment common knowledge, the normal-form game in (a_i, a_j) would have a unique and stable equilibrium. In particular, player j's equilibrium action $\mathsf{a}_j(\rho_i^\dagger)$ under common knowledge of cognition $\rho_i^\dagger$ is increasing in $\rho_i^\dagger$ under (SC) and decreasing in $\rho_i^\dagger$ under (SS).

[1] Persico (2000)'s work on information acquisition prior to an auction is one of the pioneering pieces in this literature. Recent entries include Pavan (2016), Zhao (2015) and Yang (2015).

[2] This mechanism differs from the ones advanced recently. For example, Dang et al (2016) emphasize the possibility that increasing the cost for investors of acquiring information about the value of an asset preserves the asymmetry of information and creates stable prices and thereby safe assets. A debt-like security may carry a low information intensity as long as no bad news accrues, and become information intensive when the environment deteriorates. This work emphasizes security design, and does not study expectation conformity.

Suppose that player i's actual choice ρ_i is not observed by j (so de facto the game is a simultaneous-move game in actions (ρ_i, a_i) and a_j, respectively). One can also define player i's optimal action when she deviates from her equilibrium investment. Our assumptions imply that player i's optimal action $\mathsf{a}_i(\rho_i, \rho_i^\dagger)$ given expected cognition $\rho_i^\dagger$ and actual cognition ρ_i is non-decreasing in $\rho_i^\dagger$ under either (SC) or (SS).

This environment is similar to that considered in the industrial organization literature on the taxonomy of business strategies,[3] except for one twist: The investment choice ρ_i is not observed by firm j and so has no commitment effect; rather, what matters for the outcome of the normal-form game is the anticipation $\rho_i^\dagger$ by j of firm i's choice as well as the actual choice ρ_i (of course, in a pure strategy equilibrium $\rho_i^\dagger = \rho_i$).

Letting

$$T_i(\rho_i, \rho_i^\dagger) \equiv \max_{\{a_i\}} \left\{ \phi_i(a_i, \mathsf{a}_j(\rho_i^\dagger)) - \psi_i(a_i, \rho_i) - C_i(\rho_i) \right\}$$

denote player i's payoff when actual cognition is ρ_i and player j anticipates cognition $\rho_i^\dagger$. The assumptions imply that, whether (SC) or (SS) prevails, for all $(\rho_i, \rho_i^\dagger)$ and $(\widehat{\rho}_i, \widehat{\rho}_i^\dagger)$ with $\widehat{\rho}_i \geq \rho_i$ and $\widehat{\rho}_i^\dagger \geq \rho_i^\dagger$, the following increasing-differences condition is satisfied:

$$T_i(\widehat{\rho}_i, \widehat{\rho}_i^\dagger) - T_i(\rho_i, \widehat{\rho}_i^\dagger) \geq T_i(\widehat{\rho}_i, \rho_i^\dagger) - T_i(\rho_i, \rho_i^\dagger).$$

We call this property "expectation conformity".

Consequently, let ρ_i (resp. $\widehat{\rho}_i$) denote i's optimal investment when player j expects investment $\rho_i^\dagger$ (resp. $\widehat{\rho}_i^\dagger$).[4] One can show that there is a complementarity between investment and anticipation of investment: $(\widehat{\rho}_i - \rho_i)(\widehat{\rho}_i^\dagger - \rho_i^\dagger) \geq 0$ whether the stage-2 game involves strategic substitutes or strategic complements.

The intuition goes as follows: Suppose that firm j expects i to invest more in capacity and therefore to produce more output. It will accordingly raise its output (SC) or decrease it (SS). Firm i is then induced in both cases to raise its output, vindicating a higher investment in the first place. It can also easily be checked that when there are two equilibria ($\rho_i = \rho_i^\dagger$ and $\widehat{\rho}_i = \widehat{\rho}_i^\dagger$), player i prefers the high investment one, again regardless of the strategic interaction (SC or SS).

Games such as this one are simple. "Cognitive games," in which the investment is in a filtration of the state space, are a priori much more complex. The action "a_i" is then an information-contingent one, i.e. a function. To obtain results on expectation conformity or its absence, we will need to put structure on the second-stage game. While no further assumption

[3]Eg. Bulow et al (1985) and Fudenberg and Tirole (1984).

[4]That is, ρ_i results from

$$\max_{\{\rho_i, a_i\}} \left\{ \left[\phi_i(a_i, \mathsf{a}_j(\rho_i^\dagger)) - \psi_i(a_i, \rho_i) \right] - C_i(\rho_i) \right\}$$

and $\{\mathsf{a}_i(\rho_i^\dagger), \mathsf{a}_j(\rho_i^\dagger)\}$ is the Nash equilibrium of the normal-form game under common knowledge that i has invested $\rho_i^\dagger$ (i.e. under symmetric information).

of the nature of information acquisition may be needed (zero-sum games), in general some regularity on the family of potential information structures needs to be imposed as well. This is for example what we will do when studying the generalized lemons environment, for which we will draw a formal analogy with the covert investment model just analyzed.

Relationship to the literature: Several literatures, including those on search (starting with Stigler 1961), on rational inattention (e.g., Maćkowiak and Wiederholt 2009, Matejka and McKay 2012, Sims 2003), or on security design with information acquisition (e.g., Dang et al 2011, Farhi and Tirole 2015) have used costly-cognition models. Our particular interest here is on strategic interactions with covert information acquisition.

This interaction has been the focus of the literatures on information acquisition in competitive markets and coordination games building on Morris and Shin (2002)'s beauty-contest model (e.g., Angeletos-Pavan 2007, Colombo et al 2014, Hellwig and Veldkamp 2009, Llosa and Venkateswaran 2012, Myatt and Wallace 2012, Pavan 2016),[5] and on information acquisition prior to an auction (e.g., Persico 2000) or to contracting (e.g. Dang 2008, Tirole 2009, Bolton and Faure-Grimaud 2010). In particular, it has been known since Hellwig-Veldkamp (2009) that strategic complementarities in actions may lead to strategic complementarities in information acquisition.

The paper's contribution is three-fold: First, it introduces the notion of expectation conformity and its impact on self-fulfilling cognition. Second, it analyses whether the condition obtains in some familiar games whose covert-information-acquisition properties had not yet been analyzed (zero-sum, linear-quadratic games with binary information structures). Finally, for an interesting class of games the paper provides a sufficient (and possibly necessary) condition for EC that is much simpler to check than verifying directly that EC obtains.

2 Cognitive games

2.1 Model and expectation-conformity

There are n players, $i \in I = \{1, \ldots, n\}$. Some of the results, and all applications will refer to two-player environments, though. In the "stage-2 game," the players play an arbitrary, normal or extensive form game. In this stage-2 game, player i has action space A_i and receives gross

[5]Hellwig and Veldkamp 2009 assume that players pay to receive signals of varying precisions and show that public signals, unlike private ones, create scope for equilibrium multiplicity; a signal serves both to better adjust one's action to the state of nature, and also, if it is public, to coordinate with the other players' actions. Myatt and Wallace (2012) demonstrate that for different information acquisition technologies, equilibrium uniqueness need not rely on private signals. In their model, players exert effort to achieve a better understanding of existing public signals (select "receiver noise"); this may naturally give rise to decreasing returns in the understanding effort. They derive a unique linear equilibrium, with interesting comparative statics. Amir and Lazzati (2010) consider general games with strategic complementarities and, for given information structures, select the equilibrium with the maximal actions. They emphasize increasing returns in information acquisition and derive existence of pure-strategy Bayesian equilibria.

payoff $u_i(\sigma_i, \sigma_{-i}, \omega)$ in state of nature $\omega \in \Omega$, where (σ_i, σ_{-i}) are mixed strategy profiles. The players have a common prior distribution Q on the state space Ω.[6] Gross expected payoffs are

$$U_i(\sigma_i, \sigma_{-i}) \equiv E_\omega \Big[u_i(\sigma_i, \sigma_{-i}, \omega) \Big].$$

Prior to playing the stage-2 game, the players *privately* choose at stage 1 their information structure. Let Ψ_i denote the set of available information structures or sigma-fields $\mathcal{F}_i$ for player i, and $C_i\big(\mathcal{F}_i\big)$ player i's cost of acquiring information $\mathcal{F}_{i,\cdot}$.[7] Player i's stage-2 strategy must be measurable with respect to the information structure $\mathcal{F}_i$ chosen by player i at stage 1.[8] The expected *net* payoffs are equal to the gross expected stage-2 payoffs minus the stage-1 information acquisition costs. In a number of applications only one of the players acquires information; this amounts to the other players' having infinite cost except for some partition; we will call this case "one-sided cognition".

A special case that is prominent in applications arises when the sets Ψ_i of information structures are totally ordered. Player i's choice of information structure is then represented by a filtration $\{\mathcal{F}_{i,\rho}\}$ where $\rho \in \mathbb{R}$ and $\mathcal{F}_{i,\rho}$ is an increasing sequence of sigma-algebras: For $\rho_1 < \rho_2$, $\mathcal{F}_{i,\rho_2}$ is finer than $\mathcal{F}_{i,\rho_1}$ $\big(\mathcal{F}_{i,\rho_1} \subset \mathcal{F}_{i,\rho_2}\big)$. We will then assume that C_i is monotonically increasing: A finer partition is more costly.[9]

For simplicity, we will be focusing on equilibria in which players use a pure strategy at the information acquisition stage.[10] All players share expectations as to how much cognition other players engage in even though they do not observe the actual realization of these cognitive choices. Thus, consider a common knowledge information structure for the players $\mathcal{F} = \big(\mathcal{F}_1, \ldots, \mathcal{F}_n\big)$. We let $\sigma^*(\mathcal{F}) = \big\{\sigma_i^*(\mathcal{F})\big\}_{i \in I}$ denote the stage-2 equilibrium strategy profile for $\mathcal{F}$; that is, we assume that either the stage-2 equilibrium is unique given the commonly-known information structure or some equilibrium selection has been performed; otherwise, the date-1 choices of information acquisition are not well defined.[11] From now on and unless otherwise stated, "equilibria" will therefore refer to pure-strategy equilibria of the (stage-1) information

[6]Che and Kartik (2009) by contrast look at incentives to acquire information in an environment with heterogeneous priors.

[7]This modeling of information structures applies not only to standard models of search and information acquisition, but also to endogenously imperfect recall, as well as to categorical thinking (e.g. Mullainathan 2002 and Mullainathan et al 2008).

[8]Messages and disclosure decisions, if any, are part of the stage-2 strategies in this formulation.

[9]This need not be the case for all applications. Consider memory management, an instance of a "signal-jamming" cognitive game as discussed in Appendix B: Increasing the probability of forgetting some information that one has received (repression) is likely to be costly. By contrast, the case in which a player receives two pieces of information simultaneously when searching and would have to pay an extra cost to receive only one (unbundling) is not problematic in our interpersonal *covert*-information-acquisition context: the unbundled information structure is simply irrelevant and can be assumed not to belong to Ψ_i (this would not be the case with overt information acquisition since we know that a player may suffer when other players know that he has more information).

[10]If stage 2 corresponds to an extensive-form game and player i's cognition is in mixed strategy, then player i's early actions in stage 2 might reveal something about her actual choice of cognition.

[11]Existence of a stage-2 equilibrium follows standard assumptions.

acquisition game.

Let $\mathcal{F}$ and $\widehat{\mathcal{F}}$ denote two arbitrary information structures, and σ and $\widehat{\sigma}$ denote the stage-2 equilibrium strategy profiles for information structures $\mathcal{F}$ and $\widehat{\mathcal{F}}$, respectively. Let

$$V_i\big(\mathcal{F}_i';\mathcal{F}\big) = \max_{\{\sigma_i \mathcal{F}_i'\text{-measurable}\}} \Big\{ U_i\big(\sigma_i\,, \sigma^*_{-i}(\mathcal{F})\big) \Big\}$$

denote player i's gross payoff from deviating to information structure $\mathcal{F}_i'$ when he is expected to choose $\mathcal{F}_i$ and the other players have information $\mathcal{F}_{-i}$.

Definition 1 (expectation conformity). *Expectation conformity for information structures $\mathcal{F}$ and $\widehat{\mathcal{F}}$ is satisfied if for all i*[12]

$$V_i\big(\widehat{\mathcal{F}}_i;\widehat{\mathcal{F}}\big) - V_i\big(\mathcal{F}_i;\widehat{\mathcal{F}}\big) \geq V_i\big(\widehat{\mathcal{F}}_i;\mathcal{F}\big) - V_i\big(\mathcal{F}_i;\mathcal{F}\big) \qquad \Big(EC_{\{\mathcal{F},\widehat{\mathcal{F}}\}}\Big)$$

A special case of expectation conformity arises when only one of the players can acquire information, i.e. the other players are endowed with a fixed information structure. We will then label the condition *unilateral expectation conformity* if there is a need to distinguish it from expectation conformity.

When information structures are indexed by a continuous parameter ρ (say, a higher ρ corresponds to a finer partition), and payoffs are a twice differentiable function $V_i(\rho_i'; \rho_i, \rho_{-i})$ (abusing notation), then we will also use the notion of *local expectation conformity* in the specific case of one-sided information acquisition, say by player i. This condition then writes $\partial^2 V_i/\partial\rho_i'\partial\rho_i > 0$.

Within and across complementarities

Let us decompose the difference between the RHS and the LHS of $EC_{\{\mathcal{F},\widehat{\mathcal{F}}\}}$,

$$\Gamma_i^{EC}\big(\mathcal{F},\widehat{\mathcal{F}}\big) \equiv \Big[V_i\big(\widehat{\mathcal{F}}_i;\widehat{\mathcal{F}}_i\,,\widehat{\mathcal{F}}_{-i}\big) - V_i\big(\mathcal{F}_i;\widehat{\mathcal{F}}_i\,,\widehat{\mathcal{F}}_{-i}\big)\Big] - \Big[V_i\big(\widehat{\mathcal{F}}_i;\mathcal{F}_i\,,\mathcal{F}_{-i}\big) - V_i\big(\mathcal{F}_i;\mathcal{F}_i\,,\mathcal{F}_{-i}\big)\Big]$$

into a within-player *unilateral expectation conformity* (UEC) term, which captures the impact of the other players' anticipation of one's information structure, fixing the other players' information structures:

$$\Gamma_i^{UEC}\big(\mathcal{F},\widehat{\mathcal{F}}\big) \equiv \Big[V_i\big(\widehat{\mathcal{F}}_i;\widehat{\mathcal{F}}_i\,,\mathcal{F}_{-i}\big) - V_i\big(\mathcal{F}_i;\widehat{\mathcal{F}}_i\,,\mathcal{F}_{-i}\big)\Big] - \Big[V_i\big(\widehat{\mathcal{F}}_i;\mathcal{F}_i\,,\mathcal{F}_{-i}\big) - V_i\big(\mathcal{F}_i;\mathcal{F}_i\,,\mathcal{F}_{-i}\big)\Big];$$

[12] Note the importance of covert investments for this condition. Were investments in information overt, the condition would become

$$V_i\big(\widehat{\mathcal{F}}_i;\widehat{\mathcal{F}}_i\,,\mathcal{F}_{-i}\big) - V_i\big(\mathcal{F}_i;\mathcal{F}_i\,,\mathcal{F}_{-i}\big) \leq V_i\big(\widehat{\mathcal{F}}_i;\widehat{\mathcal{F}}_i\,,\widehat{\mathcal{F}}_{-i}\big) - V_i\big(\mathcal{F}_i;\mathcal{F}_i\,,\widehat{\mathcal{F}}_{-i}\big).$$

That is, expectation conformity reflects the fact that the players do not observe each other's choice of information structure while the latter condition posits that information structures are common knowledge at stage 2 *on and off* the equilibrium path.

and an across-players *increasing differences* (ID) term, which by contrast captures the impact of the other players' information structures:

$$\Gamma_i^{ID}\big(\mathcal{F},\widehat{\mathcal{F}}\big) \equiv \Big[V_i\big(\widehat{\mathcal{F}}_i;\widehat{\mathcal{F}}_i\,,\widehat{\mathcal{F}}_{-i}\big) - V_i\big(\mathcal{F}_i;\widehat{\mathcal{F}}_i\,,\widehat{\mathcal{F}}_{-i}\big)\Big] - \Big[V_i\big(\widehat{\mathcal{F}}_i;\widehat{\mathcal{F}}_i\,,\mathcal{F}_{-i}\big) - V_i\big(\mathcal{F}_i;\widehat{\mathcal{F}}_i\,,\mathcal{F}_{-i}\big)\Big].$$

When expectation conformity holds, it is interesting to investigate whether this comes from unilateral expectation conformity or whether it is driven by variations in the other players' information structures. Note that for the prominent class of *one-sided cognitive games* (only player i, say, acquires information, i.e., Ψ_{-i} is a singleton), for all $\big(\mathcal{F},\widehat{\mathcal{F}}\big)$, $\Gamma_i^{ID}\big(\mathcal{F},\widehat{\mathcal{F}}\big)=0$ and so

$$\Gamma_i^{EC}\big(\mathcal{F},\widehat{\mathcal{F}}\big) = \Gamma_i^{UEC}\big(\mathcal{F},\widehat{\mathcal{F}}\big).$$

To illustrate the possibility that EC arise from ID rather than from UEC, consider a simple *matching model* in which players may invest in recognizing what's in it for them in a given partnership; that is, each potential match is characterized by a surplus for player i if the partner is adequate and a highly negative payoff otherwise, and so a match occurs only if both can ascertain it is a good one for them. Player i's payoff depends only on how much information is actually acquired by both parties, but not on the level of information that j expects i to acquire. Hence $\Gamma_i^{UEC}\big(\mathcal{F},\widehat{\mathcal{F}}\big)=0$ and so[13]

$$\Gamma_i^{EC}\big(\mathcal{F},\widehat{\mathcal{F}}\big) = \Gamma_i^{ID}\big(\mathcal{F},\widehat{\mathcal{F}}\big) \quad \text{for all} \quad \big(\mathcal{F},\widehat{\mathcal{F}}\big).$$

Another class of games in which $\Gamma_i^{EC}\big(\mathcal{F},\widehat{\mathcal{F}}\big) = \Gamma_i^{ID}\big(\mathcal{F},\widehat{\mathcal{F}}\big)$, for all $\big(\mathcal{F},\widehat{\mathcal{F}}\big)$ corresponds to *overt cognitive investments*. Suppose that all information structures are commonly observed at the beginning of the second stage, before actions are selected. Then $V_i\big(\mathcal{F}_i';\mathcal{F}_i,\mathcal{F}_{-i}\big)$ is independent of $\mathcal{F}_i$, at least if we make the Markov assumption that the equilibrium of the second-stage game depends only on the payoff-relevant state at the beginning of that stage, and therefore only on *actual* information structures, not on whether these deviate from expectations. In the overt cognitive investment context, expectation conformity can only arise from increasing differences.

2.2 Equilibrium multiplicity/uniqueness

Revealed preference implies that a necessary condition for $\big(\mathcal{F},\widehat{\mathcal{F}}\big)$ to form two equilibria is that $\Gamma_i^{EC}\big(\mathcal{F},\widehat{\mathcal{F}}\big) \geq 0$ for all i. The following straightforward proposition (proved in Appendix A) says

[13]More formally, an information structure for player i here is the probability $\rho_i \in [0,1]$ that player i gets informed (at some effort cost $C_i(\rho_i)$) about what he will derive from the match: $\mathcal{F} \equiv (\rho_i\,,\rho_j)$ and $\widehat{\mathcal{F}} \equiv (\widehat{\rho}_i\,,\widehat{\rho}_j)$. At stage 2, players 1 and 2 each have a veto right on the two players' matching. Each player's stage-2 behavior is independent of her expectation about the other player's information: A player who knows he receives a surplus from the match accepts to match; one who either is uninformed or knows he receives a negative surplus does not accept the match. In this matching game, $\Gamma_i^{UEC}\big(\mathcal{F},\widehat{\mathcal{F}}\big) = \big[\widehat{\rho}_i\rho_j - \rho_i\rho_j\big] - \big[\widehat{\rho}_i\rho_j - \rho_i\rho_j\big] = 0$. By contrast, $\Gamma_i^{ID}\big(\mathcal{F},\widehat{\mathcal{F}}\big) = \big(\widehat{\rho}_i - \rho_i\big)\big(\widehat{\rho}_j - \rho_j\big)$, capturing the standard strategic complementarities that are conducive to equilibrium multiplicity.

that this condition is also sufficient for both $\mathcal{F}$ and $\widehat{\mathcal{F}}$ to be equilibria for appropriately chosen cost functions:

Proposition 1 (multiplicity and uniqueness).
(i) If $EC_{\{\mathcal{F},\widehat{\mathcal{F}}\}}$ is satisfied for two distinct information structures $\mathcal{F}$ and $\widehat{\mathcal{F}}$, then there exist cost functions $\{C_i(\cdot)\}_{i=1,\ldots,n}$ such that $\mathcal{F}$ and $\widehat{\mathcal{F}}$ are both equilibrium information structures of the stage-1 game. If furthermore Ψ_i is totally ordered and $\widehat{\mathcal{F}}$ is finer than $\mathcal{F}$, the cost functions can be chosen to be monotonic. (ii) If $EC_{\{\mathcal{F},\widehat{\mathcal{F}}\}}$ is satisfied for no two distinct information structures $(\mathcal{F},\widehat{\mathcal{F}})$, then there cannot exist multiple equilibria.

Definition 2 (cognitive trap). *Players are exposed to a cognitive trap if there exist two information structures $\mathcal{F}$ and $\widehat{\mathcal{F}}$ such that*

(i) $\mathcal{F}$ and $\widehat{\mathcal{F}}$ are both equilibria,

and for all i such that $\widehat{\mathcal{F}}_i \neq \mathcal{F}_i$:

(ii) $\widehat{\mathcal{F}}_i$ is finer than $\mathcal{F}_i$,

(iii) $V_i(\mathcal{F}_i\,;\mathcal{F}) - C_i(\mathcal{F}_i) > V_i(\widehat{\mathcal{F}}_i\,;\widehat{\mathcal{F}}) - C_i(\widehat{\mathcal{F}}_i)$.

If the conditions in Definition 2 are fulfilled, cost functions are such that players who alter their information structure conform to expectations, choosing either $\widehat{\mathcal{F}}_i$ or $\mathcal{F}_i$ when expected to (condition (i)), and prefer the low-cognition outcome to the high-cognition one (conditions (ii) and (iii)).

2.3 Zero-sum games

Before we move on to analyze classes of games that satisfy expectation conformity, it is interesting to consider an important class that does not satisfy it. Suppose that the stage-2 game is a zero-sum game (or more generally a constant-sum game) between two players; that is, the gross payoffs satisfy the zero-sum condition: for all $(\sigma_i, \sigma_j, \omega)$

$$u_i(\sigma_i, \sigma_j, \omega) + u_j(\sigma_j, \sigma_i, \omega) = k(\omega),$$

where k is an arbitrary function of the state of nature. The overall game obviously is not a zero-sum game. Any information acquisition, if costly, necessarily reduces total surplus and just amounts to pure rent-seeking.

Zero-sum games have several remarkable properties; for example, a player can only benefit from having (and being known to have) more information (Lehrer and Rosenberg 2006), a property that is well-known to be violated for general games. Another interesting property is given by the following result.[14]

Proposition 2 (two-person zero-sum games). *Two-person zero-sum games satisfy for all* $(\mathcal{F}, \widehat{\mathcal{F}})$

$$\Sigma_i \, \Gamma_i^{EC}(\mathcal{F}, \widehat{\mathcal{F}}) \leq 0.$$

As a consequence, if there are multiple equilibria, in none can a player have a strict preference for her equilibrium strategy (a fortiori, there cannot exist a strict equilibrium).

Proof: The zero-sum property implies that for any two strategy pairs (σ_i, σ_j) and $(\hat{\sigma}_i, \hat{\sigma}_j)$,

$$\Sigma_i \Big[\big[U_i(\hat{\sigma}_i, \hat{\sigma}_j)\big] - \big[U_i(\sigma_i, \hat{\sigma}_j)\big] - \big[U_i(\hat{\sigma}_i, \sigma_j) - U_i(\sigma_i, \sigma_j)\big] \Big] = 0.$$

Now consider two information structures $(\mathcal{F}, \widehat{\mathcal{F}})$ and strategies $\sigma = (\sigma_i, \sigma_j)$ $\mathcal{F}$-measurable and $\hat{\sigma} = (\hat{\sigma}_i, \hat{\sigma}_j)$ $\widehat{\mathcal{F}}$-measurable. Let $R_k(\hat{\sigma}_\ell)$ denote player k's best $\mathcal{F}_k$-measurable response to $\hat{\sigma}_\ell$ and $\widehat{R}_k(\sigma_\ell)$ denote player k's best $\widehat{\mathcal{F}}_k$-measurable response to σ_ℓ. Obviously, $V_k(\mathcal{F}_k; \widehat{\mathcal{F}}) = U_k(R_k(\hat{\sigma}_\ell), \hat{\sigma}_\ell) \geq U_k(\sigma_k, \hat{\sigma}_\ell)$ and $V_k(\widehat{\mathcal{F}}_k; \mathcal{F}) = U_k(\widehat{R}_k(\sigma_\ell), \sigma_\ell) \geq U_k(\hat{\sigma}_k, \sigma_\ell)$. This implies that

$$\Sigma_i \, \Gamma_i^{EC}(\mathcal{F}, \widehat{\mathcal{F}}) \leq 0$$

for all $(\mathcal{F}, \widehat{\mathcal{F}})$. Because equilibrium multiplicity requires $\Gamma_i^{EC}(\mathcal{F}, \widehat{\mathcal{F}}) \geq 0$ for all i, the inequality implies indifference for both players, i.e., $\Gamma_i^{EC}(\mathcal{F}, \widehat{\mathcal{F}}) = 0$ for all i. Suppose, say, that in the $(\widehat{\mathcal{F}}_i, \widehat{\mathcal{F}}_j)$ equilibrium, player k has strictly optimal strategy $\widehat{\mathcal{F}}_k$. Then

$$V_k(\widehat{\mathcal{F}}_k; \widehat{\mathcal{F}}) - V_k(\mathcal{F}_k; \widehat{\mathcal{F}}) > C_k(\widehat{\mathcal{F}}_k) - C_k(\mathcal{F}_k) \geq V_k(\widehat{\mathcal{F}}_k; \mathcal{F}) - V_k(\mathcal{F}_k; \mathcal{F}),$$

and so $\Gamma_k^{EC}(\mathcal{F}, \widehat{\mathcal{F}}) > 0$, a contradiction.[15]

■

The simplest illustration of the possibility of double indifference is the case of action-independent payoffs. More interestingly, consider the zero-sum game in which i's payoff is $(a_i - a_j)\omega$ where $a_k \in \{1, -1\}$ for all k and ω takes value 1 and -1 with equal probabilities. Each player can learn ω at cost 1. Regardless of j's behavior, i is indifferent between acquiring the information or not. There are multiple equilibria with different levels of ex-ante payoffs.

[14] I am grateful to Gabriel Carroll for prompting me to have a look at zero-sum games and for conjecturing that they do not satisfy expectation conformity.

[15] In contrast, an n-player zero-sum game ($n > 2$) may admit multiple strict equilibria. To see this, take a non-zero-sum two-player game admitting multiple strict equilibria in information acquisition (such as the coordination game studied below), Have this game played twice, by players 1 and 2 and by players 3 and 4 respectively. Use players 3 and 4 (resp. players 1 and 2) as passive "budget balancers" in the game played by 1 and 2 (resp., 3 and 4).The transformed game is a zero-sum game that admits multiple strict equilibria.

2.4 Linear-quadratic games with binary information structure

Consider a two-player game with gross payoffs

$$\phi(a_i, a_j, \omega) = a_i(1 + \omega + \beta a_j) - \frac{a_i^2}{2} + \psi(a_j, \omega) \tag{1}$$

where ψ is an arbitrary function. Stage-2 actions are strategic complements if $\beta > 0$ and strategic substitutes if $\beta < 0$. This class of games includes (up to some renormalization):

- *coordination and anticoordination games*, with payoffs

  $$\phi(a_i, a_j, \omega) = -(a_i - \omega)^2 - (a_i - a_j)^2 \text{ and } \phi(a_i, a_j, \omega) = -(a_i - \omega)^2 - (a_i + a_j)^2,$$

  respectively,

- *Bertrand and Cournot games*, with payoffs $\phi(a_i, a_j, \omega) = a_i(1+\omega-a_i+\gamma a_j)$ for the Bertrand game (with demand function $q_i = 1 + \omega - a_i + \gamma a_j$, where a_i is i's price and $\gamma > 0$: prices are strategic complements), and $\phi(a_i, a_j, \omega) = a_i(1 + \omega - a_i - \gamma a_j)$ for the Cournot game, where a_i is now i's quantity and $\gamma > 0$: quantities are strategic substitutes. In both case, one could subtract a linear-quadratic cost in quantity, while still leaving property (1) intact.

We assume that the choice set of information structures is binary. Without loss of generality, we assume that $E(\omega) = 0$, $\text{var}(\omega) = \sigma^2$ and that an agent chooses between learning ω (at some cost $C > 0$) and not learning it. We will let "1" denote the act of learning ω, and "0" that of remaining ignorant. There can be four pure strategy equilibria: $\{1, 1\}$ (both learn), $\{0, 0\}$ (both remain ignorant), $\{1, 0\}$ and $\{0, 1\}$ (asymmetric equilibria). By abuse of notation, let $V(\rho_i'; \rho_i, \rho_j)$ denote player i's second-stage payoff when the equilibrium is (ρ_i, ρ_j) and i deviates to ρ_i', where $\rho_i', \rho_i, \rho_j \in \{0, 1\}$.

We assume that $|\beta| < 1$, so that the equilibrium of the stage-2 game under symmetric information ($\{1, 1\}$ or $\{0, 0\}$) is unique. So any multiplicity can come only from the endogeneity of the information structures.

Proposition 3. *The equilibrium set for linear-quadratic games with binary information structures is represented in Figure 1. Furthermore, $\Gamma_i^{UEC} \geq 0$ under both strategic complementarity and strategic substitutability; and $\Gamma_i^{ID} > 0$ for strategic complements and $\Gamma_i^{ID} < 0$ for strategic substitutes.*

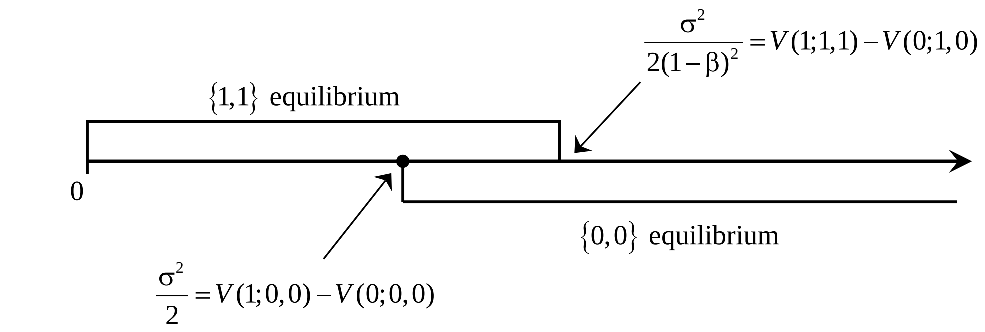

Figure 1(a): Strategic complements ($\beta > 0$)

- no asymmetric equilibria
- multiple symmetric equilibria over a range

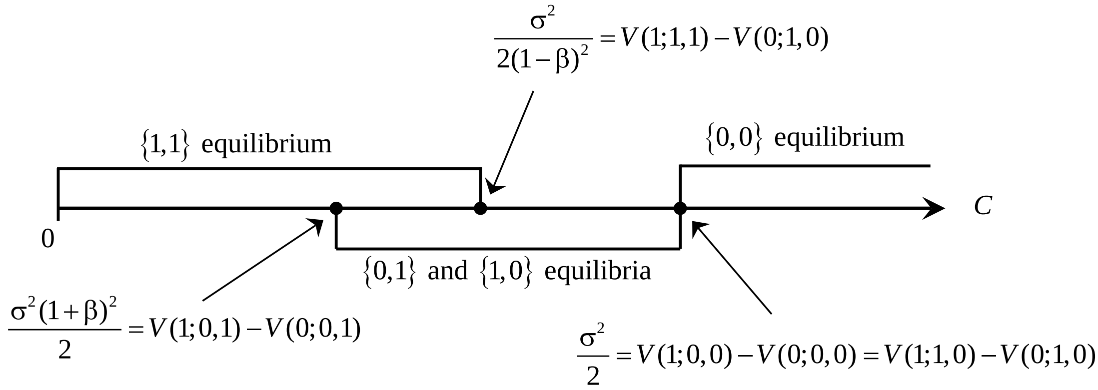

Figure 1(b): Strategic substitutes ($\beta < 0$)

- coexistence of (two) asymmetric equilibria over a range
- never a coexistence of symmetric equilibria.

Let us develop some intuition for these results (proved in the Online Appendix), beginning with the strategic complement case. Player j, when informed, scales her action up with the state of nature. If $\beta > 0$, this implies that $\omega + \beta a_j$ is more sensitive to the state of nature when player j is informed; it is as if the variance of the state of nature had increased, raising player i's incentive to get informed, and so EC is satisfied. Conversely, under strategic substitutes, player j's reaction to the state of nature dampens the overall effect of ω, lowering player i's incentive to acquire information. Again, there is co-existence of equilibria, in this case asymmetric ones.[16]

[16]More subtle is the interpretation of the range over which the asymmetric equilibria co-exist with the $\{1,1\}$ equilibrium. To understand this compare two candidate equilibria $\{1,1\}$ and $\{1,0\}$. The first player, call him i, gets informed in both; but he reacts less to ω in the former than in the latter, due to the dampening effect of player j's reaction to the state of nature. So from the point of view of player j, $\omega + \beta a_i$, has a higher variance in the $\{1,1\}$ configuration than in the $\{1,0\}$ one; and so player j has more incentive to acquire information in the $\{1,1\}$ configuration.

For the specific case of coordination games, one can show that EC is particularly pervasive; namely if information structures are ordered, then expectation conformity is satisfied for any two distinct information structures $(\mathcal{F}_1, \mathcal{F}_2)$ and $(\widehat{\mathcal{F}}_1, \widehat{\mathcal{F}}_2)$ such that for all $i \in \{1, 2\}$, $\widehat{F}_i$ is finer than $\mathcal{F}_i$: See the Online Appendix.

3 Generalized lemons environments

This section considers a class of Stackelberg games in which one player, player i ("she"), chooses between two actions. One is "adverse-selection sensitive", in that the reaction of the other player, player j ("he"), depends on his beliefs about what motivated i's choice of the action. The other action available to player i is "adverse-selection insensitive" in that player j's inferences following this action do not affect player i. Player i covertly acquires information about the state of nature before making her decision; her set of possible information structures is indexed by a rotation parameter. In the lemons (anti-lemons) environment, information that makes player i eager to engage with player j would, if it were known, trigger a hostile (friendly) action from player j. We provide sufficient and/or necessary conditions for EC to hold in this class of games, which is shown to encompass a number of economic games of interest.

3.1 Description

(a) Actions

One player, say player i (the "leader"), first chooses between the adverse-selection insensitive action, labelled "out" ("outside option") and the adverse-selection sensitive action, "in" ("interact with player j", the "follower"). When player i picks "in", player j chooses an action $a_j \in \mathbb{R}$.[17] Cognition is one-sided; prior to choosing between "in" and "out", player i selects an information structure. Player j, when choosing a_j, by contrast knows only that player i chose "in".

To take a classical example, player i might be the seller of a used car; $a_i =$ "in" if the seller decides to sell the car, and $a_i =$ "out" if he keeps the car. Then a_j is the price that a competitive buyer offers for the car (i.e. its expected value conditionally on the car being in the market). More generally, we will normalize player j's action so as to be a friendly one: player i's utility is increasing in a_j (conditional on playing "in"). More generally, one should think of "in" as an action that makes player i's payoff more sensitive to player j's perceived adverse selection.

Prior to choosing between "in" and "out", player i acquires information about the state of nature. We will assume that news that makes player i want to interact with player j also makes the latter choose an unfriendly action (a friendly action, when later on we turn to the

[17] More generally, player i may pick an action after picking "in". Because a_i will be a best reaction to a_j, the envelope theorem implies that what matters is the impact of a_j on player i's payoff.

anti-lemons environment). In that sense, the game generalizes the lemons game, in which the seller is more keen on parting with a low-quality car.

(b) Information

The state of nature is $\omega \in (-\infty, +\infty)$, with prior mean ω_0 and distribution $Q(\cdot)$. We will assume that the two players' preferences are quasi-linear in the state of nature, so they care only about the posterior mean of the state. An experiment, indexed by $\rho \in [0, +\infty)$,[18] will be taken to be the choice of a distribution $F(m; \rho)$ in a differentiable family of distributions over the posterior mean m, satisfying the martingale property $\int_{-\infty}^{+\infty} m\, dF(m; \rho) = \omega_0$ for all ρ. The cost of acquiring information is, by abuse of notation, $C(\rho)$ (with $C(0) = 0$, $C' > 0$). A higher index ρ corresponds to a finer information structure in the following sense:

Assumption 1 (rotations). *Player i's set of possible information structures is indexed by the parameter ρ in the sense of "rotations" (or "simple mean-preserving spreads" or "single-crossing property"); that is, there exists m_ρ such that $F_\rho(m; \rho) \geq 0$ for $-\infty < m \leq m_\rho$ and $F_\rho(m; \rho) \leq 0$ for $m_\rho \leq m < +\infty$ (with some strict inequalities).*[19]

Non directed search example. To illustrate assume that information collection follows the standard non-directed search technology:

$$F(m; \rho) = \begin{cases} \rho Q(m) & \text{for } m < \omega_0 \\ \rho Q(m) + 1 - \rho & \text{for } m \geq \omega_0. \end{cases}$$

That is, the seller learns the true state of nature with probability ρ and nothing with probability $1 - \rho$. Note that the rotation point is equal to the prior mean ω_0. It is then natural to posit an increasing information acquisition cost $C_i^{ND}(\rho)$.

(c) Preferences

Player i's payoff difference between "in" and "out" depends on a_j, and on player i's posterior beliefs only through the posterior mean m. This difference will be labeled $\delta_i(m, a_j)$, with $\partial\delta_i/\partial m > 0$ and $\partial\delta_i/\partial a_j > 0$ from our previous sign convention. The strict monotonicity of δ_i in m implies that player i plays "in" if and only if m exceeds some cutoff $m^*(a_j)$ where $\delta_i(m^*(a_j), a_j) = 0$. The cutoff is decreasing in a_j.

[18] We will not need to index by i here, since only player i acquires information.

[19] See e.g. Diamond and Stiglitz (1974) and Johnston and Myatt (2006). A simple mean-preserving spread (MPS) is a MPS, but the converse does not hold as a MPS may have multiple points of intersections. A combination of two rotations need not be a rotation, unless of course they have the same rotation point. But as is well known, any MPS can be obtained through a sequence of simple MPS. Examples of mean-preserving spreads with a rotation include: the case of a normally distributed state of nature ω together with a signal that is normally distributed around the true state (ρ is then the precision of this signal); the class of triangular distributions on $[0, 1]$ with uniformly distributed underlying state (so $\rho = +\infty$ corresponds to $F(m; \rho) = m$ on $[0, 1]$). We will provide other examples of rotations later on. In these examples (Pareto, exponential, directed and non-directed search) as well as the two just discussed, information structures are ordered.

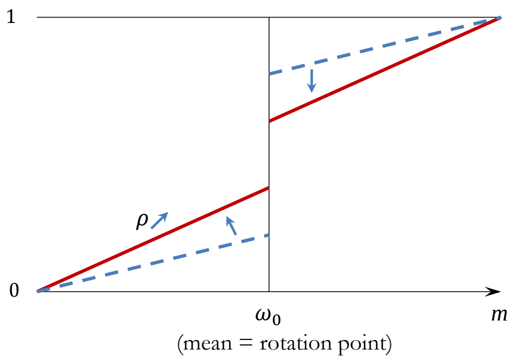

Figure 1: (cumulative distribution $F(m;\rho)$ for non-directed search)

Assumption 2 (leader's preferences). *Player i's net payoff from playing "in", $\delta_i(m, a_j)$, depends on i's posterior mean m about ω and on j's action a_j. It is twice differentiable, is increasing in m and a_j, and is such that the marginal impact of a friendly action ($\partial\delta_i/\partial a_j$) is weakly decreasing and weakly concave in the expected state of nature:*

$$\frac{\partial^2\delta_i}{\partial a_j \partial m} \leq 0 \quad \text{and} \quad \frac{\partial^3\delta_\mathrm{i}}{\partial \mathrm{a_j} \partial \mathrm{m}^2} \leq 0. \tag{2}$$

Condition (2) in Assumption 2 will be used to determine whether cognition becomes more or less attractive to player i when player j behaves in a more friendly way (see the proof of Part (iii) of Proposition 4). In the five illustrations described in Section 3.2[20], δ_i will take the more specific functional form over the relevant range:

$$\frac{\partial\delta_i}{\partial a_j} \equiv \gamma - \tau m \tag{3}$$

with $\gamma > 0$ and $\tau \geq 0$, and so Assumption 2 is satisfied. For example, in the classic lemons game, the benefit of selling depends only on a_j, and the utility from keeping the car depends only on m. So $\partial^2\delta_i/\partial a_j \partial m = \partial^3\delta_i/\partial a_j \partial m^2 = 0$.

Let player j anticipate cognition $\rho^\dagger$ by player i ($\rho^\dagger$ out-of-equilibrium can differ from actual cognition ρ). Let $a_j(\rho^\dagger)$ denote the resulting equilibrium choice. As stated above, we also assume that player j cares only about the posterior mean of m. And so, due to player i's use of a cutoff rule, a_j depends only on $M^+(m^*(a_j), \rho^\dagger)$, where $M^+(m^*(a_j), \rho) \equiv E(m|m \geq m^*(a_j), \rho)$ denotes the truncated mean for information structure ρ. An increase in M^+ can be viewed as an exacerbation of the adverse selection problem.

[20]This will also hold for the warfare game of Section 3.4. By contrast, this functional form holds, but with $\gamma = 0$, $\tau = -1$ for the leadership game considered in that section.

Assumption 3 (**lemons**). *The friendliness of player j's reaction to an increase in player i's cognition depends negatively on the impact of this cognition on adverse selection, as measured by the conditional mean of the state of nature:*

$$\mathrm{sign}\left(\frac{\mathrm{da_j}}{\mathrm{d}\rho^\dagger}\right) = -\mathrm{sign}\left(\frac{\partial}{\partial\rho^\dagger}\left(\mathrm{M}^+\left(\mathrm{m}^*(\mathrm{a_j}), \rho^\dagger\right)\right)\right). \tag{4}$$

3.2 Examples

Assumptions 2 and 3 may look somewhat abstract. Let us now illustrate them through a number of games of interest. *The reader may want to skip this section and content him/herself with the following quick description.*

The first three examples are variants of Akerlof's model, augmented by the seller's covertly acquiring information. In the classic variant, the seller holds soft information about the quality of the good and must decide whether to sell the good or keep it for own usage or future resale. In the second, the good is divisible (is a share in a project); the owner benefits from synergies associated with taking an associate in the project, but is hesitant about sharing the proceeds if he knows the project to be highly profitable. In the third variant, the seller may have hard information about the quality of the good and chooses whether to disclose it to the buyer. The fourth illustration describes herding with interdependent payoffs; for example, a firm may by entering a market encourage a rival to follow suit. The fifth illustration is the marriage game, in which covenants may smooth the hardship of a subsequent divorce, but also may signal bad prospects about the marriage's prospects.

a) *Akerlof's lemons game.* Let player i be, say, the seller. The seller can sell her good in the market ("in") or not sell it ("out"). Player j is then a set of competitive buyers who choose a price a_j equal to the value for a buyer conditional on the good being put in the market. Suppose that the players' utilities from the good are $v_i - m$ for the seller and $v_j - m$ for the buyer, with $v_i < v_j$ (gains from trade). Then, a_j is the price offered by competitive buyers

$$a_j = E\Big(v_j - m | v_i - m \leq a_j \, , \rho^\dagger\Big) = v_j - M^+\big(v_i - a_j \, , \rho^\dagger\big)$$

as the cutoff $m^*(a_j)$ is equal to $v_i - a_j$; and $\delta_i(m, a_j) = a_j - (v_i - m)$ (thus, $\gamma = 1$ and $\tau = 0$). We assume that the solution a_j is unique, which is indeed the case if the hazard rate of the distribution of m for parameter $\rho^\dagger$ is monotonic.[21] Furthermore Assumption 3 is satisfied.

b) *Team formation.* Player i has a project. She can associate player j to the project or do it alone. Bringing player j on board creates synergies (lowers the cost of implementation), but forces i to share the gains, which she does not want to do if the project is a good one. Player

[21] Then $\partial M^+/\partial m^* \in (0,1)$. See An (1998).

i's payoff is $(v-\omega)-d_i$ if she does it alone and $a_j(v-\omega)-c_i$ if it is a joint project, where a_j is the value share left to i by (competitive) player j and $c_i < d_i < v$ is player i's reduced cost of project implementation. Let c_j denote player j's cost (with $c_i + c_j < d_i$). Player i chooses "in" if and only if

$$\delta_i(m, a_j) \equiv d_i - c_i - (1-a_j)(v-m) \geq 0.$$

And so $\gamma = v$ and $\tau = 1$. Finally, $a_j = a_j(\rho^\dagger)$ solves:[22]

$$(1-a_j)\left(v - M^+\left(v - \frac{d_i - c_i}{1-a_j}, \rho^\dagger\right)\right) = c_j.$$

Provided that $c_j - \frac{\partial M^+\left(v - \frac{d_i - c_i}{1-a_j}, \rho^\dagger\right)}{\partial m^*}(d_i - c_i) > 0$ (e.g. $2c_j > d_i - c_i$ for a uniform distribution), which guarantees a unique solution if it exists (otherwise the team does not form), then Assumption 3 is satisfied.

c) *Disclosure game*. In another important variant of Akerlof's game the seller wants to sell for sure (she has no value for the good, say), and either has no information about ω (with probability $1-\rho$) or knows ω (with probability ρ). So $F(m;\rho) = \rho Q(m)$ for $m < \omega_0$ and $F(m;\rho) = \rho Q(m) + 1 - \rho$ for $m \geq \omega_0$, where ω_0 is the prior mean of ω. In contrast with Akerlof's soft-information lemons game, the seller's decision is not whether to put the good in the market (a foregone conclusion), but whether to reveal the state of nature when knowing it. A well-established literature surveyed by Milgrom (2008), has studied such an incentive to disclose. A natural extension of the disclosure model consists in thinking of ρ (the precision of information) as endogenous. The family $F(m;\rho)$ is a family of rotations, with constant rotation point the prior mean ω_0.

In order to apply the general results, we must define the "in" and "out" actions and the corresponding δ_i function. Let "in" stand for the act of non-disclosure and "out" the act of disclosing the state of nature; the rationale for this choice is that player j's beliefs about player i's cognition level matter only if there is no disclosure. Letting (as in (a)) $v_j - m$ denote the buyer's utility and a_j the price in the absence of disclosure, the seller obtains $v_j - m$ if he discloses m and a_j in the absence of disclosure, and so

$$\delta_i = a_j - (v_j - m),$$

and so $\gamma = 1$ and $\tau = 0$, a special case of Assumption 2.

To compute a_j, note that the seller discloses if and only if $m \leq m^*$ for some cut-off m^*

[22] Assuming that the solution to this equation satisfies $a_j(v - \omega^{\max}) \geq c_i$ (here we must assume a finite support).

uniquely defined

$$m^* = \frac{(1-\rho)\omega_0 + \rho \int_{m^*}^{+\infty} m dQ(m)}{1-\rho+\rho[1-Q(m^*)]}$$

leading to $dm^*/d\rho > 0$ (a standard "partial unraveling" result stating that a higher probability of knowing leads to more disclosure).

Finally, define

$$a_j = v_j - (M^+)^{-1}(M^+(m^*)),$$

and so Assumption 3 is satisfied.

d) *Interdependent herding game.* Player i decides on whether to enter a market. Player j, a rival, then decides whether to follow suit. Player j uses the information revealed by player i's decision, but in contrast with most herding models, payoffs are interdependent and so externalities are not purely informational. Suppose for instance that i and j are rivals, with per-customer profit π^m under monopoly and $\pi^d < \pi^m$ under duopoly.[23] The state of nature ω here indexes (minus) the fixed cost of entry or opportunity cost of firms i and j. Let a_j denote the probability of non-entry by firm j. Then

$$\delta_i(m, a_j) \equiv \Big[a_j \pi^m + (1-a_j)\pi^d\Big] - (k_i - m),$$

where $k_i - m$ is firm i's entry cost. So $m^*(a_j) = k_i - a_j\big(\pi^m - \pi^d\big) - \pi^d$, $\gamma = \pi^m - \pi^d > 0$ and $\tau = 0$; and so Assumption 2 is satisfied.

Firm j has entry cost $k_j - m$, where, say, $k_j \in (0, +\infty)$ with distribution $G(k_j)$. The realization of k_j is unknown to player i. Then $a_j(\rho^\dagger)$ is the solution to

$$a_j = 1 - G\Big(\pi^d + M^+\big(m^*(a_j), \rho^\dagger\big)\Big).$$

Assumption 3 is satisfied whenever the solution to this equation is unique (which is the case if the density g satisfies, at least at the cutoff, $g(\pi^m - \pi^d) < 1$ and the distribution of m satisfies the monotone-hazard-rate condition, which implies that $\partial M^+/\partial m^* < 1$).

e) *Marriage game.* Consider the following variant of Spier (1992)'s model, augmented with cognition. Players i and j decide whether to get married. Getting married has value v_i and v_j if all goes well; with probability ω distributed on $[0,1]$, things will go wrong (divorce), generating utility $v_k - L_k$ for $k = i, j$. The divorce can, however, be made less painful (utility $v_k - \ell_k$) through a covenant spelling out the outcome in case of divorce, where the losses satisfy $0 < \ell_k < L_k < v_k$ for all k. Adding the covenant costs a fixed $c_k < L_k - \ell_k$ to player k (so it is efficient to add the covenant if the parties want to marry but are certain to divorce). Player i has a v_i high enough that she wants to marry regardless ($v_i \geq L_i$), while player j's value v_j is distributed on

[23] One can also perform the analysis for complementors, with $\pi^d > \pi^m$.

$[\underline{v}_j, +\infty)$ according to c.d.f. G and is private information. Player i may acquire information about ω and then chooses between a contract with ("in") and without ("out") covenant. Player j then decides whether to accept to marry. Assume that $v_j - \omega_0 L_j \geq 0$ (so in the absence of any information, player j always accept to marry). Let a_j denote the probability that player j accepts to marry when the proposed contract includes the covenant. This game also satisfies Assumptions 2 and 3. Note first that

$$\delta_i(m, a_j) = a_j[v_i - m\ell_i - c_i] + (1 - a_j) \cdot 0 - (v_i - mL_i)$$

is increasing in m and satisfies condition (3) with $\gamma = v_i - c_i$ and $\tau = \ell_i$[24]. Furthermore, $m^* = m^*(a_j(\rho^\dagger))$ is given by $m^* \left[L_i - a_j(\rho^\dagger)\ell_i\right] = \left[1 - a_j(\rho^\dagger)\right] v_i + c_i$.[25] To see if Assumption 3 holds, note that

$$a_j = 1 - G(M^+(m^*(a_j), \rho^\dagger)\ell_j + c_j);$$

and so Assumption 3 holds under a condition analogous to that found in Example (d).

3.3 Expectation conformity

Adverse selection is here captured by the truncated mean $M^+(m^*, \rho^\dagger)$:

Definition 3 (impact of cognition on adverse selection). *An increase in cognition*

(i) aggrevates adverse selection if sign$\left(\frac{\partial}{\partial \rho^\dagger}(M^+(m^*, \rho^\dagger))\right) > 0$

(ii) alleviates adverse selection if sign$\left(\frac{\partial}{\partial \rho^\dagger}(M^+(m^*, \rho^\dagger))\right) < 0$.

Simple computations show that

$$\mathrm{sign}(\frac{\partial}{\partial \rho^\dagger}(\mathrm{M}^+(\mathrm{m}^*, \rho^\dagger)) = \mathrm{sign}(\mathrm{A}),$$

where

$$A \equiv \left[M^+(m^*, \rho^\dagger) - m^*\right] F_\rho(m^*; \rho^\dagger) - \int_{m^*}^{+\infty} F_\rho(m; \rho^\dagger) dm.$$

From Assumption 3, $A > 0$ ($A < 0$) means that a higher level of anticipated cognition triggers an unfriendly (a friendly) action.

[24] The absence of covenant is "good news" about m. And so, $v_j - \omega_0 L_j \geq 0$ for all v_j implies that the contract without covenant is always accepted.

[25] Such a cut-off is guaranteed to exist if $[1 - G(\ell_j + c_j)][v_i - \ell_i - c_i] > v_i - L_i$, since the contract with covenant is always accepted if $v_j \geq \ell_j + c_j$.

Let B denote the sensitivity of i's marginal benefit of cognition to an unfriendly action:

$$\begin{aligned} B &\equiv -\int_{m^*(a_j(\rho^\dagger))}^{+\infty} \frac{\partial \delta_i}{\partial a_j}(m, a_j(\rho^\dagger)) dF_\rho(m;\rho) \\ &\equiv \frac{\partial \delta_i}{\partial a_j}\Big(m^*(a_j(\rho^\dagger)), a_j(\rho^\dagger)\Big) F_\rho\Big(m^*(a_j(\rho^\dagger));\rho\Big) + \int_{m^*(a_j(\rho^\dagger))}^{+\infty} \frac{\partial^2 \delta_i}{\partial a_j \partial m}(m, a_j(\rho^\dagger)) F_\rho(m;\rho) dm. \end{aligned}$$

Definition 4 (cognition incentive effect of unfriendly actions). *An unfriendly action raises (lowers) the incentive for cognition if $B > 0$ (respectively $B < 0$).*

The monotonicity of δ_i in m implies that player i's relative payoff of choosing "in" rather than "out" when her information is indexed by ρ and player j expects a choice of $\rho^\dagger$ is

$$V_i(\rho, \rho^\dagger) \equiv \int_{m^*(a_j(\rho^\dagger))}^{+\infty} \delta_i\big(m, a_j(\rho^\dagger)\big) dF(m;\rho).$$

(Strict) expectation conformity holds locally provided that

$$\frac{\partial^2 V_i}{\partial \rho \partial \rho^\dagger} > 0 \text{ at } \rho = \rho^\dagger.$$

Proposition 4 (generalized lemons environments). *Under Assumptions 1 through 3:*
(i) Local EC holds if (and only if) $AB > 0$.
(ii) Cognition always aggrevates adverse selection for the uniform, Pareto and exponential distributions. For other distributions, a sufficient condition for cognition to aggrevate adverse selection is that the cutoff be to the left of the rotation point:

$$F_\rho(m^*;\rho^\dagger) > 0 \ \text{ at } \ \mathrm{m}^* = \mathrm{m}^*(\mathrm{a_j}(\rho^\dagger)).$$

(iii) A sufficient condition for an unfriendly action to raise the incentive for cognition is that $F_\rho(m^*;\rho) > 0$ at $m^* = m^*(a_j(\rho^\dagger))$.
(iv) Therefore a sufficient condition for local expectation conformity is therefore that the cutoff lie to the left of the rotation point: $m^*(a_j(\rho^\dagger)) < m_\rho$.
(v) Suppose that $M^+(m^*;\rho)$ is always increasing in ρ (as is the case for the uniform, Pareto and exponential distributions), implying that $A > 0$. If furthermore $\partial^2\delta_i/\partial a_j \partial m = 0$ (as is the case for examples a), c) and d)), then $m^*(a_j(\rho)) < m_\rho$ is a necessary and sufficient condition for local EC.

Proof. (i) Using the condition that $\delta_i\big(m^*(a_j(\rho^\dagger)), a_j(\rho^\dagger)\big) = 0$,

$$\frac{\partial^2 V_i}{\partial \rho \partial \rho^\dagger} = \left[\int_{m^*(a_j(\rho^\dagger))}^{+\infty} \frac{\partial \delta_i}{\partial a_j}\big(m, a_j(\rho^\dagger)\big) dF_\rho(m;\rho)\right] \frac{da_j}{d\rho^\dagger}.$$

So local EC holds if (and only if) $AB > 0$.
(ii) From Assumption 3,

$$\text{sign}\left(\frac{da_j}{d\rho^\dagger}\right) = -\text{sign}\left[\left[M^+(m^*, \rho^\dagger) - m^*\right]F_\rho(m^*; \rho^\dagger) - \int_{m^*}^{+\infty} F_\rho(m; \rho^\dagger)dm\right]$$

Because ρ indexes a mean-preserving spread, $\displaystyle\int_{m^*}^{+\infty} F_\rho(m; \rho^\dagger)dm \leq 0$. And thus $da_j/d\rho^\dagger < 0$ whenever $F_\rho(m^*(a_j(\rho^\dagger)), \rho^\dagger) > 0$.

$F_\rho(m^*(a_j(\rho^\dagger)), \rho^\dagger) > 0$ is only a sufficient condition for local EC. The condition that the cutoff be to the left of the rotation point ($F_\rho > 0$) is by no means necessary for $A > 0$, that is for an increase in cognition to exacerbate the adverse selection problem. For a number of distributions indeed, $\frac{\partial M^+}{\partial \rho}(m^*, \rho) > 0$ regardless of the value of m^*; these include:

- the *uniform distribution*: m is uniformly drawn from $[\omega_0 - \rho, \omega_0 + \rho]$. Then for $|m - \omega_0| \leq \rho$, $F(m; \rho) = \frac{1}{2} + \frac{m - \omega_0}{2\rho}$. The rotation point is $m_\rho = \omega_0$ for all ρ and $M^+(m^*, \rho) = \frac{m^* + (\omega_0 + \rho)}{2}$ is increasing in ρ for all ρ.

- the *Pareto distribution*: m is distributed on $[1/\rho, +\infty)$ according to the survival function $1 - F(m; \rho) = (1/\rho m)^{\alpha(\rho)}$ where $\omega_0 = \frac{\alpha(\rho)}{\alpha(\rho)-1}\frac{1}{\rho}$ (changes in ρ are mean preserving). One can check that an increase in ρ induces a rotation, with rotation point $m_\rho = \frac{e^{\omega_0 \rho - 1}}{\rho} > \omega_0$. Furthermore,

$$M^+(m^*; \rho) = \rho\omega_0 m^* \qquad \text{and so} \qquad \frac{\partial \mathrm{M}^+}{\partial \rho} > 0.$$

- the *exponential distribution*: m is distributed on $[1/\rho, +\infty)$ according to the survival function $1 - F(m; \rho) = e^{-\lambda(\rho)(m - (1/\rho))}$ where $\frac{1}{\rho} + \frac{1}{\lambda(\rho)} = \omega_0$ (mean preservation). Then

$$M^+(m^*; \rho) = m^* + \frac{1}{\lambda(\rho)} \quad \Rightarrow \quad \frac{\partial M^+}{\partial \rho} = -\frac{\lambda'(\rho)}{\lambda^2(\rho)} = \frac{1}{\rho^2} > 0,$$

  and the rotation point is $m_\rho = \omega_0$.

(iii) Next

$$\begin{aligned}\int_{m^*(a_j(\rho^\dagger))}^{+\infty} \frac{\partial \delta_i}{\partial a_j}(m, a_j(\rho^\dagger))dF_\rho(m; \rho) \quad &= \quad -\frac{\partial \delta_i}{\partial a_j}\Big(m^*(a_j(\rho^\dagger)), a_j(\rho^\dagger)\Big)F_\rho\Big(m^*\big(a_j(\rho^\dagger)\big); \rho\Big) \\ &\quad - \int_{m^*(a_j(\rho^\dagger))}^{+\infty} \frac{\partial^2 \delta_i}{\partial a_j \partial m}\big(m, a_j(\rho^\dagger)\big)F_\rho(m; \rho)dm.\end{aligned}$$

The latter term is negative as ρ is an index of mean-preserving spread and $\partial^2 \delta_i / \partial a_j \partial m$ is negative and weakly decreasing in m. The former term is negative provided that $F_\rho > 0$ at $m^*(a_j(\rho^\dagger))$.

■

Example: the cognition-augmented lemons model under non-directed search

We have seen that under non-directed search the rotation point is the prior mean. Proposition 4 applied to Akerlof's game states that local EC holds whenever the cutoff m^* is to the left of the prior mean. Expectation conformity thus arises when gains from trade $v_j - v_i$ are large and so $m^* \leq \omega_0$, but not when they are small.[26] The seller puts her car up for sale when the state is above some m^*: When m^* is smaller than the rotation point, that is the mean ω_0, then the seller enters the market both when she is uninformed and when she learns that the state m is above m^*. Hence, as $\rho^\dagger$ increases, the expected quality conditional on entry goes down, so the price paid by buyers goes down as well. But then it becomes even more important for the buyer to learn the value of the car, that is, to increase ρ: So local EC is satisfied. When $\omega_0 < m^*$, the seller entering the market only if she learns that $m > m^*$ implies that the value of the car conditional on entry is the same independent of ρ: So a higher $\rho^\dagger$ does not affect the price obtained, and hence does not increase the incentives to search.

To sum up the cognition-augmented lemons game satisfies expectation conformity under non-directed search if and only if the gains from trade are sufficiently large. The general intuition as to why the lemons game often satisfies expectation conformity goes as follows: Suppose that the seller is expected to engage in a high level of cognition; then adverse selection is a serious concern for the buyers, who are therefore willing to pay only a low price. A low price in turn makes it particularly costly for the seller to part with a valuable item, raising her incentives to acquire information.

A similar insight holds when it is the *buyer* who engages in cognition (is player i). Suppose that there is a single buyer, who may acquire information. If there is a single, price-setting seller, the model is not the mirror image of the one considered so far. The treatment is sightly different as price setting involves market power. Because our objective here is merely to show that similar considerations apply to buyer information acquisition, suppose instead that sellers (player j now) are competitive. The price offered by the sellers is $-a_j$. The buyer has utility 0 when not buying and $\delta_i = v_i + m + a_j$ when buying. Think of m as (minus) the scope for ex-post holdup by the selected seller, where the holdup comes from a renegotiation following the discovery that the initial specification is not optimal for the buyer. So $\gamma = 1$, $\tau = 0$ and $m^* = -v_i - a_j$. A representative seller's utility is $v_j - m - a_j$, and so the competitive outcome

[26] Let

$$H(m^*, \rho^\dagger) \equiv F_\rho(m^*; \rho^\dagger)\left[M^+(m^*, \rho^\dagger) - m^*\right] - \int_{m^*}^{+\infty} F_\rho(m, \rho^\dagger) dm \quad \begin{cases} > 0 & \text{if } m^* \leq \omega_0 \\ = 0 & \text{if } m^* > \omega_0. \end{cases}$$

To show that $H = 0$ on $[\omega_0, +\infty)$, note that $H(+\infty) = 0$ and that $dH/dm^* = 0$ on this domain. Also, $m^* - M^+(m^*, \rho^\dagger) = -(v_j - v_i)$. Because $\partial M^+/\partial m^* \in (0, 1)$ under a monotone hazard rate, m^* decreases with $(v_j - v_i)$, i.e., with the gains from trade. For $m^* > \omega_0$, $M^+(m^*, \rho^\dagger)$ is invariant to m^*.

is $a_j = v_j - M^+(m^*; \rho^\dagger)$. The results apply to buyer information acquisition as well.

The intuition for expectation conformity is again straightforward. Suppose that the seller anticipates more cognition; he then raises price to reflect the fact that the buyer has ruled out more bad news for himself. Facing a higher price, the buyer then finds it more costly to enter disadvantageous deals and this is incentivized to find out about possible bad news.

3.4 Discussion and implications

3.4.1 Cognitive traps

Corollary 1 (cognitive trap). *Consider two equilibria $\{\rho_1, a_j(\rho_1)\}$ and $\{\rho_2, a_j(\rho_2)\}$ with $\rho_1 < \rho_2$ and assume that $A > 0$ (which is guaranteed if $F_\rho(m^*(a_j(\rho); \rho)) > 0$ for $\rho \in [\rho_1, \rho_2]$). Then player i is better off in the low-cognition equilibrium $\{\rho_1, a_j(\rho_1)\}$.*

Proof: $A > 0$ on $[\rho_1, \rho_2]$ and part (i) of Proposition 4 together with Assumption 3 ensure that $da_j/d\rho \leq 0$ for $\rho \in [\rho_1, \rho_2]$. For a given a_j, player i's welfare is $\mathcal{V}(a_j)$ given by

$$\mathcal{V}(a_j) = \max_{\{\rho\}} \left\{ \int_{m^*(a_j)}^{+\infty} \delta_i(m, a_j) dF(m; \rho) - C_i(\rho) \right\}.$$

The envelope theorem and the property that a_j is a friendly action imply that $d\mathcal{V}/da_j > 0$. Because $a_j(\rho_2) \leq a_j(\rho_1)$, player i is better off in the $\{\rho_1, a_j(\rho_1)\}$ equilibrium.

■

Cognitive traps do not result just from a higher cost of cognition (actually, at the margin player i's gain from a finer information structure is equal to the increase in the cost of information acquisition). Rather, player j's anticipating an exacerbated adverse selection reacts in a way that hurts player i.

3.4.2 Disclosure and cognitive styles

We have so far assumed that information acquisition is covert. Suppose that it is indeed covert, but that some form of disclosure prior to j's action is feasible: player i can disclose how much attention or external resources he devoted to the issue (but not what she actually learnt). Namely, she can disclose any[27] level of cognition $\widehat{\rho}$ as long as $\widehat{\rho} \leq \rho$, where ρ is the actual cognition level. The disclosed information is hard information. For an arbitrary disclosure $\widehat{\rho}$, one can consider the "$\widehat{\rho}$-constrained cognition game", namely the no-disclosure game with modified cost function $D(\rho, \widehat{\rho}) = C(\rho)$ if $\rho \geq \widehat{\rho}$ and $D(\rho, \widehat{\rho}) = +\infty$ if $\rho < \widehat{\rho}$. Let $E(\widehat{\rho})$ denote the corresponding equilibrium set. Note that for $\widehat{\rho} < \widehat{\rho}'$, then $E(\widehat{\rho}') \subseteq E(\widehat{\rho})$. We will say that the equilibrium selection $e(\widehat{\rho}) \in E(\widehat{\rho})$ (for all $\widehat{\rho}$) is *monotonic* if for all $\{\widehat{\rho}, \widehat{\rho}'\}$ such that $\widehat{\rho} < \widehat{\rho}'$, then

[27] Considering only a subset of disclosable levels in $[0, \rho]$ including $\hat{\rho} = 0$ would only reinforce Corollary 3 below.

$e(\widehat{\rho}) \leq e(\widehat{\rho}')$. In words, a higher level of disclosure cannot result in a lower equilibrium level of cognition. This property is trivially satisfied if $E(\widehat{\rho})$ is a singleton for all $\widehat{\rho}$. We will say that the equilibrium of the disclosure game is *regular* if it is monotonic and $e(0)$ is an equilibrium of the no-disclosure game.[28]

Corollary 2 (disclosure). *Assume that* $A > 0$*.*

(i) Any equilibrium without disclosure is still an equilibrium with disclosure.

(ii) Conversely, the largest and smallest cognition levels in the regular equilibrium set with disclosure are the same as without disclosure.

Proof. (i) Corollary 3 admits the same logic as Corollary 2. Consider an equilibrium ρ^* of the game without disclosure and introduce the possibility of disclosure. On the equilibrium path, let player i pick ρ^* and disclose nothing (or any level below ρ^*). Let player j form beliefs $r(\widehat{\rho}) = \rho^*$ whenever $\widehat{\rho} \leq \rho^*$, and arbitrary beliefs $r(\widehat{\rho}) \geq \widehat{\rho}$ when $\widehat{\rho} > \rho^*$. Because $A > 0$, $a_j(r(\widehat{\rho})) \leq a_j^* \equiv a_j(\rho^*)$. Thus

$$\begin{aligned}
&\max_{\{\rho, \widehat{\rho} \leq \rho\}} \left\{ \int_{m^*(a_j(r(\widehat{\rho})))}^{+\infty} \delta_i(m, a_j(r(\widehat{\rho})))dF(m;\rho) - C(\rho) \right\} \\
&\quad \leq \max_{\{\rho\}} \left\{ \int_{m^*(a_j^*)}^{+\infty} \delta_i(m, a_j^*)dF(m;\rho) - C(\rho) \right\} \\
&\quad = \int_{m^*(a_j^*)}^{+\infty} \delta_i(m, a_j^*)dF(m;\rho^*) - C(\rho^*),
\end{aligned}$$

using the fact that $\{\rho^*, a_j^*\}$ is an equilibrium of the no-disclosure game.

(ii) Conversely, consider the disclosure game and an equilibrium ρ^* with disclosure $\widehat{\rho} \leq \rho^*$ and suppose that $\rho^* < \underline{\rho}$, the lowest equilibrium level without disclosure. Then ρ^* would be an equilibrium of the no-disclosure game $D(\rho, 0)$, a contradiction. Next, suppose that player i chooses not to disclose ($\widehat{\rho} = 0$). Then monotonicity implies that the highest equilibrium cognition level is $\bar{\rho}$. So player i can guarantee at worst $a_j(\bar{\rho})$.

■

In the same vein, one can consider the possibility of *transparency*, namely a commitment to reveal the exact amount of cognitive resources. Then $\widehat{\rho} = \rho$ (overt information acquisition). It is clear that player i is always better off committing ex ante as compared to the case of voluntary ex-post disclosure just studied or the complete absence of disclosure. More interestingly, under

[28] This "Markov" requirement rules out equilibria of the following type: Player i may pick $\rho^* = \widehat{\rho} > \bar{\rho}$ (the highest equilibrium cognition level in the no-disclosure game) fearing that player j forms beliefs $\rho = +\infty$ if anything else than $\widehat{\rho}$ is disclosed.

transparency, player i will choose a level $\rho \leq \underline{\rho}$, that is lower than the lowest equilibrium cognition levels.

Another focus of comparative statics concerns the player i's *cognitive style*. We here provide only an informal account. In the generalized lemons model, suppose that the cost-of-cognition function $C(\rho, \xi)$ depends on a parameter ξ, interpreted as cognitive ability. A higher-ability player i has a lower marginal cost of cognition: $C_{\rho\xi} < 0$ (and $C(0, \xi) = 0$, $C_\rho(0, \xi) = 0$, $C_\rho > 0$, $C_{\rho\rho} > 0$). As player i's ability increases, the equilibrium cognition increases (if unique, or in case of multiple equilibria, in the sense of monotone comparative statics: the minimum and maximum of this set both increase). Put differently, player i's ability, while directly beneficial, indirectly hurts him as player j becomes more wary of adverse selection. This suggests that if player i has side opportunities to signal cognitive ability, he will want to adopt a dumbed-down profile.

Suppose indeed that player i can be bright or dumb. A bright person can demonstrate she is bright (and can always mimick a dumb one), but the reverse is impossible. The set of equilibrium cognition levels is monotonically increasing in the posterior probability that $\xi = \xi_H$. Let us assume a monotone selection in this equilibrium set: Player j's action a_j is decreasing in the probability that $\xi = \xi_H$ (a property automatically satisfied if the equilibrium is unique). Then if we add a disclosure game prior to the cognitive game in which player i can disclose she is bright if this is indeed the case, the equilibrium is a pooling one, in which the bright player i does not disclose her IQ. Conversely, player i will disclose, if she can, that she is overloaded with work (assume that she cannot prove that she has a low workload), and therefore that her marginal cost of investigation is high.

Corollary 3 (cognitive style). *Suppose that player i can acquire information cheaply (ξ_H) or expensively (ξ_L) and that prior to playing the cognitive game, player i can disclose only ξ_H (in the IQ interpretation) or only ξ_L (in the work load one). Then player i does not disclose in the IQ interpretation and discloses in the work load one. In either case, player i poses as an informational puppy dog.*

3.4.3 Role of gains from interaction

It is interesting to note that EC is more likely to be satisfied when gains from interaction are large.

Corollary 4 (gains from interaction). *Suppose that player i's gain from playing in is $\delta_i(m, a_j) + \theta$ where θ is a parameter in $\mathbb{R}$.*[29] *Then if $F_\rho(m^*(a_j(\rho^\dagger), \theta); \rho^\dagger) > 0$ a fortiori $F_\rho(m^*(a_j(\rho^\dagger), \theta'); \rho^\dagger) \geq 0$ for $\theta' > \theta$ and so local EC prevails as well. So for all ρ, there exists*

[29]For instance, an increase in θ corresponds to a decrease in v_i (example a)), $c_i - d_i$ (example b)), $\pi^m - \pi^d$ (example d)), ℓ_i or $-L_i$ (example e) and a decrease in v_j (example c)).

$\theta^*(\rho)$ such that for all $\theta \geq \theta^*(\rho)$, EC prevails locally at ρ: Expectation conformity is more likely, the larger the gains from interaction.

Corollary 4 says that higher gains from interaction reinforce expectation conformity. By contrast, it should be noted that they lower player i's incentive to acquire information under the sufficient condition ($F_\rho \geq 0$) for expectation conformity:

$$\frac{\partial^2}{\partial\theta\partial\rho}\left[\int_{m^*(a_j(\rho^\dagger),\theta)}^{+\infty}[\delta_i(m, a_j(\rho^\dagger)) + \theta]dF(m, \rho)\right] = -F_\rho(m^*(a_j(\rho^\dagger), \theta), \theta) \leq 0.$$

3.5 Anti-lemons

Finally, note that the results can be applied with slight modifications to environments that do not satisfy the assumptions above. Suppose that an increase in anticipated cognition generates a friendly reaction, i.e., Assumption 3 is reversed:

Assumption $3'$ (anti-lemons).

$$\text{sign}\left(\frac{da_j}{d\rho^\dagger}\right) = \text{sign}\left(\frac{\partial M^+}{\partial\rho^\dagger}\right).$$

Under Assumption $3'$, local EC requires that $AB < 0$. Suppose that as in the examples above (uniform, Pareto, exponential distributions), $\partial M^+/\partial\rho^\dagger$ is always strictly positive and so $A > 0$. Then under Assumption 2, $B \geq 0$ as long as $F_\rho \geq 0$ and so local EC cannot hold. An illustration satisfying Assumption 2 and Assumption $3'$ is the warfare game:

f) *Warfare.* Country i is a potential invader and must decide whether to engage in a fight ("in" action). The state of nature ω here represents the probability that country i wins in case of a fight. Let a_j denote the probability that country j surrenders without fighting back, 1 the payoff in case of victory and c_i the cost in case of defeat:

$$\delta_i(m, a_j) = a_j + (1 - a_j)[m - (1 - m)c_i]$$

and so

$$\gamma = \tau = 1 + c_i \quad \text{and} \quad m^*(a_j) = \frac{c_i - (1 + c_i)a_j}{(1 - a_j)(1 + c_i)}.$$

Letting, similarly, country j's payoff from victory be equal to 1 and its loss in case of defeat be equal to c_j, country j fights back if and only if

$$\left[1 - M^+(m^*(a_j(\rho^\dagger)); \rho^\dagger)\right] - M^+(m^*(a_j(\rho^\dagger)); \rho^\dagger)c_j \geq 0.$$

Assuming that c_j is drawn from some cumulative distribution H,

$$a_j(\rho^\dagger) = 1 - H\left(-1 + \frac{1}{M^+(m^*(a_j(\rho^\dagger)); \rho^\dagger)}\right).$$

Next, consider games in which $\partial^2\delta_i/\partial a_j \partial m \geq 0$. Then, assuming again that A is always strictly positive, local EC is satisfied if the cut-off lies to the *right* of the rotation point: $m^* = m^*(a_j(\rho^\dagger)) > m_{\rho^\dagger}$. To see this, recall that (omitting the arguments), $B \equiv \frac{\partial \delta_i}{\partial a_j} F_\rho + \int_{m^*}^{+\infty} \frac{\partial^2 \delta_i}{\partial a_j \partial m} F_\rho dm$. Suppose that $m^* = m^*(a_j(\rho^\dagger)) > m_{\rho^\dagger}$. Then $F_\rho \leq 0$ for all $m \geq m^*$, and so $B < 0$. An illustration satisfying these assumption is the leadership game:

g) *Leadership game.* Like in Hermalin (1998)'s theory of leadership, a leader shares in the team's output and has information about the return to effort. The information structure is here endogenous. Player i decides to undertake a costly project or not. Her gain from the project depends on whether player j gets on board and on the quality m of the project:

$$\delta_i(m, a_j) = a_j m - c_i.$$

Note that $\partial^2\delta_i/\partial a_j \partial m = 1$. Player j observes whether i undertakes the project and decides whether to get on board, yielding payoff $m - c_j$, where c_j is drawn from some cumulative distribution H, and so

$$a_j = H(M^+(c_i/a_j))$$

is increasing in M^+.

Proposition 4′ (generalized anti-lemons environment). *Suppose that Assumptions 1 and 3′ hold and that $\partial M^+/\partial \rho^\dagger$ is always strictly positive.*
(i) If Assumption 2 holds (like in the warfare game), local EC cannot hold if the cutoff is to the left of the rotation point: $m^*(a_j(\rho^\dagger)) \leq m_{\rho^\dagger}$.
(ii) If $\partial^2\delta_i/\partial a_j \partial m \geq 0$ (like in the leadership game), then local EC is satisfied if the cutoff is to the right of the rotation point: $m^*(a_j(\rho^\dagger)) \geq m_{\rho^\dagger}$.

3.6 Relation to the covert investment game

Let us draw a formal analogy between the generalized lemons game and the covert investment game described in the introduction. The investment ρ_i in the latter is the cognitive investment in the generalized lemons game. To interpret the generalized lemons game as a covert investment game, let us define the "stage-2 action" a_i as being just matching this investment: $a_i \equiv \rho_i$.[30]

[30]Thinking of a_i as the information used by player i in the stage-2 game, $a_i = \rho_i$ simply means that player i makes full use of the acquired information. Using the notation in the introduction, $\psi(a_i, \rho_i) = 0$ if $a_i = \rho_i$, $= -\infty$ otherwise, which is a discontinuous version of the complementarity relationship $\partial^2\psi_i/\partial a_i \partial \rho_i < 0$ of complementarity between action and cognition.

Then one can define

$$\phi_i(a_i, a_j) \equiv \int_{m^*(a_j)}^{+\infty} \delta_i(m, a_j) dF(m; a_i)$$

and so $\partial^2 \phi_i / \partial a_i \partial a_j < 0$ whenever condition $B > 0$ in Proposition 4 is satisfied.

Finally, the exact expression of $\phi_j(a_i, a_j)$ is application-specific, but the condition $A > 0$ in Proposition 4 expresses the condition that $da_j/da_i < 0$, and so strategic substitutability (SS) prevails.

Similarly, part (ii) of Proposition 4$'$ shows that some anti-lemons environments, like under some conditions the leadership game, satisfy the strategic complementarity assumption (SC) of the covert investment model of the introduction.

4 Concluding remarks

Economic agents manage their information in multiple ways: allocation of scarce cognitive resources, brainstorming, search and experimentation, hiring of engineering, financial or legal experts. Such "cognitive activities" determine information structures and are often the essence of adverse selection; they thereby condition the functioning of contracts and markets, and more broadly of social interactions. This motivates the study of "cognitive games," defined as games in which a normal- or extensive-form game is preceded by players' selecting their or their rivals' information structures.

Expectation conformity arises when players have an incentive to comply with the level of cognition they are expected to engage in. We showed that zero-sum games never give rise to self-fulfilling cognition while linear-quadratic games with binary information structures do. We then considered a generalized lemons (or anti-lemons) environment, which comprises many games of interest to economists such as the cognition-augmented lemons model. A characterization of the expectation conformity property in terms of rotation points was obtained for this class of games.

Because of their importance for economics, cognitive games need to be better understood, and there are multiple alleys for future research. For instance, cognition may occur in multiple stages as an extensive-form game unfolds. In multi-stage cognition, players may learn with an endogenous lag about their rivals' choice of cognitive strategies.

Cognitive strategies are particularly relevant to contracting environments. In such environments, parties may acquire information not only to reach more efficient agreements, but also to either design self-advantageous covenants or, as in the lemons game, eschew undesirable trades. Contracts may then be too complete or too incomplete from a social welfare viewpoint, where incompleteness refers to the need for ex-post adjustments or more generally to inefficiencies generated by imperfect contracting. The relationship between EC, cognitive traps and excess cognition in contracting environments is developed in a follow-up paper (Tirole 2016).

We have assumed that players choose their own information structure, Appendix B considers "signal-gamming cognition games," in which players covertly choose their rival's information structure rather than their own. Its main purpose is to adapt the definition of expectation conformity to this context. It also provides various economic environments satisfying expectation conformity.

References

Amir, R., and N. Lazzati (2010) “Endogenous Information Acquisition on Bayesian Games with Strategic Complementarities,” mimeo, University of Arizona.

An, M.Y. (1998) “Logconcavity versus Logconvexity: A Complete Characterization,” *Journal of Economic Theory*, 80(2): 350–369.

Angeletos, G.M., and A. Pavan (2007), “Efficient Use of Information and Social Value of Information,” *Econometrica* 75: 1103–1142.

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

Bénabou, R., and J. Tirole (2002) “Self-Confidence and Personal Motivation,” *Quarterly Journal of Economics*, 117(3): 871–915.

Bénabou, R., and J. Tirole (2006) “Belief in a Just World and Redistributive Politics,” *Quarterly Journal of Economics*, 121: 699–746.

Bolton, P., and A. Faure-Grimaud (2010) “Satisficing Contracts,” *Review of Economic Studies*, 77(3): 937–971.

Bulow, J., Geanakoplos, J. and P. Klemperer (1985) “Multimarket Oligopoly: Strategic Substitutes and Complements,” *Journal of Political Economy*, 93(3): 488–511.

Che, Y.K., and N. Kartik (2009) “Opinions as Incentives,” *Journal of Political Economy*, 117(5): 815–860.

Colombo, l., Femminis, G., and A. Pavan (2014) “Information Acquisition and Welfare,” *Review of Economic Studies*. DOI: 10.1093/restud/rdu015 First published online: May 9, 2014.

Dang, T.V. (2008) “Bargaining with Endogenous Information,” *Journal of Economic Theory*, 140: 339–354.

Dang, T.V., Gorton, G. and B. Holmström (2011) “The Information Acquisition Sensitivity of a Security,” mimeo.

Dang, T.V., Gorton, G., Holmström B. and G. Ordonez (2016) “Banks As Secret Keepers, ” mimeo.

Dessi, R. (2008) “Collective Memory, Cultural Transmission and Investments,” *American Economic Review*, 98(1): 534–560.

Dewatripont, M., Jewitt, I., and J. Tirole (1999) "The Economics of Career Concerns. I: Comparison of Information Structures," *Review of Economic Studies*, 66 (1): 183–198.

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

Diamond, P. and J. Stiglitz (1974) "Increases in Risk and Risk Aversion," *Journal of Economic Theory*, 8: 337–360.

Farhi, E., and J. Tirole (2015) "Liquid Bundles," *Journal of Economic Theory*, 158(B): 634–655.

Fudenberg, D., and J. Tirole (1984) "The Fat-Cat Effect, the Puppy-Dog Ploy, and the Lean and Hungry Look," *American Economic Review*, 74(2): 361–366.

Gentzkow M. and E. Kamenica (2016) "Bayesian Persuasion," *American Economic Review*, 101: 2590–2615.

Gentzkow M. and E. Kamenica (2016) "Bayesian Persuasion with Multiple Senders and Rich Signal Spaces," working paper.

Gottlieb, D. (2014a) "Imperfect Memory and Choice under Risk," *Games and Economic Behavior*, 85: 127–185.

Gottlieb, D. (2014b) "Will you Never Learn? Self-Deception and Biases in Information Processing," mimeo, Wharton School, University of Pennsylvania.

Hellwig, C., and L. Veldkamp (2009) "Knowing What Others Know: Coordination Motives in Information Acquisition," *Review of Economic Studies*, 76(1): 223–251.

Hermalin, B. (1998) "Toward an Economic Theory of Leadership: Leading by Example," *American Economic Review*, 88: 1188–1206.

Holmström B. (1999) "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66(1): 169–182.

Johnston, J., and D. Myatt (2006) "On the Simple Economics of Advertising, Marketing, and Product Design," *American Economic Review*, 96(3): 756–784.

Lehrer, E., and D. Rosenberg (2006) "What Restrictions do Bayesian Games Impose on the Value of Information?," *Journal of Mathematical Economics*, 42: 343–357.

Llosa, L. G., and V. Venkateswaran (2012) "Efficiency under Endogenous Information Choice," mimeo Penn State.

Maćkowiak, B., and M. Wiederholt (2009) "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 99(3): 769–803.

Matejka, F., and A. McKay (2012) "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 105(1): 272-298.

Milgrom, P. (2008) "What the Seller Won't Tell You: Persuasion and Disclosure in Markets," *Journal of Economic Perspectives*, 22: 115–131.

Milgrom, P. and C. Shannon (1994) "Monotone Comparative Statics," *Econometrica*, 62: 157–180.

Morris, S. and H. Shin (2002) "Social Value of Public Information," *American Economic Review*, 92: 1521–1534.

Mullainathan, S., Schwartzstein, J. and A. Shleifer (2008) "Coarse Thinking and Persuasion," *Quarterly Journal of Economics*, 123: 577–619.

Mullainathan, S. (2002) "Thinking through Categories", mimeo, MIT.

Myatt, D. P., and C. Wallace (2012) "Endogenous Information Acquisition in Coordination Games," *Review of Economic Studies*, 79: 340–374.

Pavan, A. (2016) "Attention, Coordination, and Bounded Recall," mimeo, Northwestern University.

Persico, N. (2000) "Information Acquisition in Auctions," *Econometrica*, 68(1): 135–148.

Sims, C. (2003) "Implications of Rational Inattention," *Journal of Monetary Economics*, 50(3): 665–690.

Spier, K (1992) "Incomplete Contracts and Signaling," *Rand Journal of Economics*, 23: 432–443.

Stigler, G. (1961) "The Economics of Information," *Journal of Political Economy*, 69: 213–225.

Tirole, J. (2009) "Cognition and Incomplete Contracts," *American Economic Review*, 99(1): 265–294.

Tirole, J. (2016) "Cognition-intensive Contracting," mimeo, Toulouse School of Economics.

Yang, M. (2015) "Coordination with Flexible Information Acquisition," *Journal of Economic Theory*, 158: 721–738.

Zhao, X. (2015) "Strategic Mis-selling and Pre-Purchase Information," mimeo.

A. Proof of Proposition 1

(i) Assume that $EC_{\{\mathcal{F},\widehat{\mathcal{F}}\}}$ is satisfied. For $\mathcal{F}$ and $\widehat{\mathcal{F}}$ to be both equilibrium information structures, it is necessary that for all i

$$V_i\big(\widehat{\mathcal{F}}_i;\mathcal{F}\big) - V_i\big(\mathcal{F}_i;\mathcal{F}\big) \leq C_i\big(\widehat{\mathcal{F}}_i\big) - C_i\big(\mathcal{F}_i\big) \leq V_i\big(\widehat{\mathcal{F}}_i;\widehat{\mathcal{F}}\big) - V_i\big(\mathcal{F}_i;\widehat{\mathcal{F}}\big). \tag{5}$$

In the absence of further requirement, pick cost functions satisfying (5) as well as $C_i(\widetilde{\mathcal{F}}_i) = +\infty$ if $\widetilde{\mathcal{F}}_i \notin \{\mathcal{F}_i, \widehat{\mathcal{F}}_i\}$.

When Ψ_i is totally ordered and $\mathcal{F}_i \subseteq \widehat{\mathcal{F}}_i$ for all i, pick a cost function satisfying (5) as well as:

$$C_i\big(\widetilde{\mathcal{F}}_i\big) = \begin{cases} C_i\big(\mathcal{F}_i\big) \quad \text{for} \quad \widetilde{\mathcal{F}}_i \subseteq \mathcal{F}_i \\ C_i\big(\widehat{\mathcal{F}}_i\big) \quad \text{for} \quad \mathcal{F}_j \subset \widetilde{\mathcal{F}}_i \subseteq \widehat{\mathcal{F}}_i \\ +\infty \quad \text{for} \quad \widehat{\mathcal{F}}_i \subset \widetilde{\mathcal{F}}_i. \end{cases}$$

From (5) and the fact that $V_i\big(\widehat{\mathcal{F}}_i;\mathcal{F}\big) - V_i\big(\mathcal{F}_i;\mathcal{F}\big) \geq 0$, $C_i(\widehat{\mathcal{F}}_i) \geq C_i(\mathcal{F}_i)$ and so $C_i(\cdot)$ is indeed monotonic. Because more information cannot hurt if covertly acquired, $\mathcal{F}$ and $\widehat{\mathcal{F}}$ are indeed both equilibria for these cost functions.

(ii) Conversely, if $\mathcal{F}$ and $\widehat{\mathcal{F}}$ were two distinct equilibria, condition (5) would be satisfied, and so $EC_{\{\mathcal{F},\widehat{\mathcal{F}}\}}$ would hold, a contradiction. ■

B. Signal-jamming cognitive games

The paper has assumed that players choose their *own* information structure. In a number of economic games, though, players choose *their opponents*' information structure. Such signal jamming[31] has been studied for example in industrial organization, as when a firm secretly cuts its price so as to convince its rivals that demand is low and induce their exit. Furthermore, cognitive traps are common in such games as well as we will shortly observe.

Defining expectation conformity in signal-jamming cognitive games Consider a two-player game. In a signal-jamming cognitive game, player i chooses player j's information structure $\mathcal{F}_j$ at cost $C_i(\mathcal{F}_j)$. We have to be a bit careful with regards to measurability, as a deviation from $\mathcal{F}_j$ to $\widehat{\mathcal{F}}_j$ is not observed by player j. Thus, think of $\mathcal{F}_j$ as a conditional distribution $q_j(s_j|\omega)$ over the signal s_j received by player j in state of nature ω. Player j then plays a stage-2 (mixed)

[31] Signal jamming here is covert, unlike in the literature on Bayesian persuasion, where a player – the principal – commits to a specific information structure for the other player – the agent (e.g. Gentzkow-Kamenica 2011).

strategy $\alpha_j(s_j)$. Player j's stage-2 strategy under $\mathcal{F}_j$ is then an "$\mathcal{F}_j$-measurable" strategy $\sigma_j^{\mathcal{F}_j}$, defined by:

$$\sigma_j^{\mathcal{F}_j}(\omega) = \Sigma_{s_j} q_j(s_j|\omega)\alpha_j(s_j).$$

Let $\{\mathcal{F}_i, \mathcal{F}_j\}$ denote a common-knowledge choice of information structures with signal distributions $q_i(s_i|\omega)$ and $q_j(s_j|\omega)$ and $\{\alpha_i, \alpha_j\}$ denote the corresponding equilibrium strategies. Suppose that player i deviates and picks information structure $\widehat{\mathcal{F}}_j$ (corresponding to contingent signal distribution $\widehat{q}_j(s_j|\omega)$). Let

$$V_i(\widehat{\mathcal{F}}_j; \mathcal{F}_i, \mathcal{F}_j) \equiv \max_{\{\alpha_i'(\cdot)\}} \left\{ \Sigma_{\omega, s_i, s_j} q(\omega) q_i(s_i|\omega) \widehat{q}_j(s_j|\omega) u_i(\alpha_i'(s_i), \alpha_j(s_j), \omega) \right\}.$$

Definition 5 (expectation conformity under signal jamming). *$EC_{\{\mathcal{F}, \widehat{\mathcal{F}}\}}$ is satisfied if for all i,*

$$V_i(\widehat{\mathcal{F}}_j; \mathcal{F}_i, \mathcal{F}_j) - V_i(\mathcal{F}_j; \mathcal{F}_i, \mathcal{F}_j) \le V_i(\widehat{\mathcal{F}}_j; \widehat{\mathcal{F}}_i, \widehat{\mathcal{F}}_j) - V_i(\mathcal{F}_j; \widehat{\mathcal{F}}_i, \widehat{\mathcal{F}}_j). \tag{6}$$

Examples of signal-jamming games satisfying expectation conformity One-sided[32] signal-jamming environments (described rather informally below) exhibiting expectation conformity include:

(a) *Imperfect persuasion.* Consider the trading game when the seller with strictly positive probability knows the buyer's willingness to pay. For simplicity, suppose that the state of nature can take one of two values $\{\omega, \widehat{\omega}\}$ and is the buyer's utility, with $\omega > \widehat{\omega}$ and that the seller does not value the good. By exerting more effort, the seller can increase the probability that the buyer understands the argument and thereby learns the true state of nature: information is "semi-hard" in that it can be disclosed, but the amount of disclosure depends on the seller's communication effort.[33] The seller's effort (understood as the effort incurred prior to actual communication with the buyer) increases the probability that the buyer identifies the true state and is unobserved by the buyer. Clearly, the seller exerts no effort if the state is $\widehat{\omega}$. By contrast, convincing the buyer that the state is ω is profitable and so in general elicits effort.

It is easy to check that cognitive traps quite similar to those for the lemons game arise naturally: If the seller is expected by the buyer to exert substantial effort to communicate that the state is ω, the price p in the absence of persuasion is low (the state of nature is unlikely to be ω), and then it is particularly profitable for the seller to convince the buyer that the state is ω. In case of multiplicity, the seller is better off in a lower-effort equilibrium.

(b) *Career concerns.* In Holmström (1999)'s celebrated career-concerns model, an agent's current performance depends on talent, effort and noise. The agent does not know her talent and tries to

[32] That is, only one player, player i, manipulates the other player's information structure: $\widehat{\mathcal{F}}_i = \mathcal{F}_i$ in condition (6). Again, multilateral and unilateral expectation conformity coincide in such environments.

[33] In this simplified model, only the seller exerts effort; in general communication involves moral hazard in team (see Dewatripont and Tirole 2005).

convince future employers that she is talented by secretly exerting more effort to boost current performance. The signal jamming cost is here the cost of effort in the current task. When talent and effort are complements, such signal jamming often generates information conformity and traps (e.g., Dewatripont et al 1999). Indeed, suppose that the labor market expects a higher effort; then employers put more weight on performance when updating their beliefs about talent, as performance is more informative about talent. The increased performance sensitivity of future compensation then boosts the agent's incentive to exert effort. Again, in case of multiplicity, the agent is better off in the low-effort equilibrium.

(c) *Memory management game*. Another class of signal-jamming games giving rise to expectation conformity is the class of memory-management games.[34] This class of games describes situations in which a player receives information that he may try to remember or repress. The individual may find himself in a self-trap, in which repression or cognitive discipline are possible self-equilibria with distinct welfare implications.

[34] Introduced in Bénabou-Tirole (2002). See also Gottlieb (2014a,b). Dessi (2008) applies similar ideas in the context of cultural transmission with multiple agents. Bénabou (2013) and Bénabou-Tirole (2006) show how memory management and collective decisions interact to produce collective delusions.