Cognitive Games and Cognitive Traps[*]

Jean Tirole[†]

October 24, 2014

Abstract

The paper defines "cognitive games" as games in which players first privately choose their information structures, and then play a normal- or extensive-form game under the resulting information structures. It introduces two concepts of "expectation conformity" (which coincide under one-sided cognition): a strong version in which fixing the other players' information structures, a player has more incentive to select one information structure over another if he expected to do so; a weak version in which it is the entire vector of information structures that is the object of self-fulfilling prophecies.

The paper first shows that games of pure conflict (zero-sum games) never give rise to self-fulfilling cognition while games of pure alignment (coordination games) always do. Second, it considers "environments with a game setter" in which a player picks an information structure and one of two games to be played. A characterization of the expectation conformity property in terms of rotation points can be obtained for this class of games, which comprises many games of interest to economists, starting with the cognition-augmented lemons model.

The paper then turns to cognition-intensive contracting and shows that a single variable, the "relative exposure to the unexpected", underlies a variety of concepts such as expectation conformity, over-cognition and the desirability of mandatory disclosure laws. Finally, the paper extends the notion of expectation conformity to (signal-jamming) cognitive games in which players choose their rivals' information structure.

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

1 Introduction

Cognition is costly; thinking and memorizing, enlisting financial or engineering experts and lawyers, or brainstorming with others consumes resources (in amounts that depend on the urgency to act, the cognitive load or the context) and is strategic. This paper considers the

implications of costly cognition in a multi-player context. It defines cognitive games as two-stage games in which players at stage 1 privately select information structures for themselves or other players, and then play at stage 2 an arbitrary, normal or extensive form game under the resulting information structures. A recurrent theme is that individually optimal and anticipated cognitions can, under some circumstances, be expected to be "strategic complements": A higher level of anticipated cognitive activity raises a suspicion and incentivizes the player to indeed engage in more cognition.

Section 2 introduces expectation-conformity. (Unilateral) expectation conformity (EC) holds if there exist a player $i \in I$, two information structures $\mathcal{F}_i$ and $\widehat{\mathcal{F}}_i$ for player i, and an information structure $\mathcal{F}_{-i} = \underset{j \neq i}{X} \mathcal{F}_j$ for the other players such that player i has more incentive to acquire information $\widehat{\mathcal{F}}_i$ than to acquire $\mathcal{F}_i$ when all other players have information $\mathcal{F}_{-i}$ and expect $\widehat{\mathcal{F}}_i$ than when they all expect $\mathcal{F}_i$. For example, when information structures are ordered, the player has more incentive to acquire a fine information structure if he is expected to have a fine information structure. A "cognitive trap" is then a situation in which the player is hurt by, but cannot refrain from acquiring more costly information.

Multilateral expectation conformity (MEC) 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 $\mathcal{F}$. Collective cognitive traps then refer to all players acquiring more information in one equilibrium, but preferring another, less information intensive equilibrium. A sufficient condition for (MEC) when information structures are ordered is that, in the reduced-form game in which payoffs are written as a function of information structures, the information-acquisition strategies be either strategic complements or strategic substitutes.

These expectation-conformity conditions, which straightforwardly lead to a multiplicity of equilibrium information structures for appropriate information acquisition costs and to equilibrium uniqueness if (MEC) is never satisfied, are shown to be violated (satisfied) when the stage-2 game is zero-sum (a coordination game). The rest of the paper is then devoted to studying prominent classes of games in which the stage-2 game is neither one of pure conflict nor one of perfectly aligned preferences.

Section 3 studies the class of models with a game setter, in which one of the players acquires information and then picks one of two games; for example, he may decide whether to play a game with player 2 or to opt out. 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 condition for such games to satisfy expectation conformity.

Section 4 considers cognition-intensive contracting: parties to a contract expend resources to try to understand the likely implications of contracting. Unobserved cognition creates an adverse selection problem, as each party is concerned that the other party might enter the agreement with a better mastery of its implications. To study this, Section 4 develops a canonical model of pre-contractual cognition and obtains results for this model, which is shown to admit Gabaix and Laibson (2006)'s model of shrouded attributes and Tirole (2009)'s model of incomplete

contracting as special cases. It unveils the key role played by a single parameter, the relative exposure to an incomplete understanding of the environment (or equivalently, to contract incompleteness). This parameter determines the incentive to engage in under/over cognition, the incentive to disclose and the impact of good-faith-bargaining regulations, and, most to the point, the presence of expectation conformity. The analysis uncovers a number of new results.

Section 5 considers "signal-jamming cognition games", in which players 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 examples satisfying expectation conformity in various economic environments. Finally, Section 6 concludes with alleys for future research.

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 2013, Yang 2012) have used costly-cognition models. Our particular interest here is on strategic interactions with information acquisition. This also has been the focus of the literatures on information acquisition and aggregation in competitive markets building on Morris and Shin (2002)'s beauty-contest model (e.g., Colombo et al 2014, Hellwig and Veldkamp 2009, Llosa and Venkateswaran 2012, Myatt and Wallace 2012, Pavan 2014),[1] on information acquisition prior to an auction (e.g., Persico 2000), or on contracting (e.g. Gabaix and Laibson 2006, Dang 2008, Tirole 2009, Bolton and Faure-Grimaud 2010).

Disclosure games, in which a sender holds hard (verifiable) information and decides whether to reveal it to a receiver who then takes an action affecting both, have received particular attention.[2] The corresponding literature has investigated factors that prevent unraveling (uncertainty as to whether the sender has received information or costly disclosure) or lead to the disclosure of bad news (various forms of reputation, see Bourjade and Jullien 2011, Dziuda 2011, Grubb

---

[1]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.

[2]See Milgrom (2008) for a recent survey. Okuno-Fujiwara et al (1990) assumes that cognition is exogenous and contains a general analysis of when voluntary disclosure leads to full disclosure. Shavell (1994), who in turn credits an unpublished 1983 paper by Farrell and Sobel, provides the basic analysis of costly information acquisition prior to disclosure. He however does not emphasize strategic complementarities between anticipated and optimal cognition. Rather, he provides a careful analysis of an informal idea of Kronman, according to which mandated information disclosure is likely to reduce incentives for information acquisition; he further shows that the impact of required disclosure is rather asymmetric: it discourages buyers but not sellers from acquiring socially valuable information. Kartik et al (2014) consider a multi-sender model of disclosure of hard information; among other results, they show that disclosure strategies are strategic substitutes under a disclosure cost and strategic complements under a concealment cost. Another recent contribution to the literature is Hoffmann et al (2014) in which the sender secretly collects information about receiver preferences and selectively discloses, perhaps subject to absorption capacity constraints limiting the number of dimensions over which information can be disclosed (which is another impediment to unraveling).

2011 and Ispano 2012). Most papers on disclosure games take the information structure as given; exceptions include Matthews and Postlewaite (1985), Shavell (1994) and Dang (2008).

In parallel, the information transmission literature pioneered by Crawford and Sobel (1982) posits that the sender holds soft (unverifiable) information and transmits a message to the receiver, who then takes an action affecting both.[3] Reviewing this equally rich literature lies outside the scope of this brief overview. Most papers again take the information structure as given.[4]

2   Cognitive games

2.1   Model and expectation-conformity

There are two players, $i, j \in \{1, 2\}$. The generalization to n players poses no difficulty.[5] In the "stage-2 game", the players play an arbitrary, normal or extensive form game. Player i has action space A_i and receives gross payoff $u_i(\sigma_i, \sigma_j, \omega)$ in state of nature $\omega \in \Omega$, where (σ_i, σ_j) are mixed strategy profiles. The players have a common prior distribution Q on the state space Ω.[6] Gross expected payoffs are

$$U_i(\sigma_i, \sigma_j) \equiv E_\omega \Big[u_i(\sigma_i, \sigma_j, \omega) \Big].$$

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 $\mathcal{F}_i$ for player i, where $i \in \{1, 2\}$, and $C_i(\mathcal{F}_i)$ player i's cost of acquiring information $\mathcal{F}_{i,}$. 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.[7] 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 player's 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 $\mathcal{F}_{i,\rho_1}$ $\left(\mathcal{F}_{i,\rho_1} \subset \mathcal{F}_{i,\rho_2} \right)$. We will then assume that C_i is monotonically increasing:

---

[3]Much of the attention has focused on the polar cases of hard and soft information. In general how hard (per se informative) communicated information is depends on the – endogenous – communication efforts exerted by the sender and the receiver (Dewatripont and Tirole 2005).

[4]A recent exception is Pei (2013), who allows the sender to choose his information structure from a "rich set" (if a partition belongs to the sender's choice set, then any coarser partition also does belong to this set), shows that in contrast with Crawford and Sobel, all equilibria are such that the sender communicates all he knows to the receiver, and characterizes the equilibrium outcomes. Gentzkow and Kamenica (2012) obtain a related result in a game with hard information. They show that if information acquisition is costly and observable by the decision maker (who observes the information structure, but not the realized signal), then disclosure requirements have no effect on the set of pure-strategy equilibria.

[5]See the comment following Proposition 1, though.

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

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

a finer partition is more costly.[8]

Consider a common knowledge information structure for the two players $\mathcal{F} = (\mathcal{F}_1, \mathcal{F}_2)$. We let $\sigma^*(\mathcal{F}) = (\sigma_1^*(\mathcal{F}), \sigma_2^*(\mathcal{F}))$ 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.[9] From now on and unless otherwise stated, "equilibria" will therefore refer to pure-strategy[10] equilibria of the (stage-1) information 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

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

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 player has information $\mathcal{F}_j$.

Definition 1 (**expectation conformity**). *(Unilateral) expectation conformity for information structures* $(\mathcal{F}_i, \mathcal{F}_j)$ *and* $(\widehat{\mathcal{F}}_i, \mathcal{F}_j)$ *is satisfied if*

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

(it is strictly satisfied if this inequality holds strictly). Expectation conformity holds if $EC_{\{\mathcal{F}, \widehat{\mathcal{F}}_i\}}$ *holds for all* $\mathcal{F}$ *and* $\widehat{\mathcal{F}}_i$.

For example, when only player i acquires information and information structures are totally ordered, expectation conformity implies that player i has a higher incentive to acquire more information if player j believes he is acquiring more information.

Definition 2 (**multilateral expectation conformity**). *Multilateral expectation conformity for information structures* $\mathcal{F}$ *and* $\widehat{\mathcal{F}}$ *is satisfied if for all* i

$$V_i\big(\widehat{\mathcal{F}}_i; \mathcal{F}_i, \mathcal{F}_j\big) - V_i\big(\mathcal{F}_i; \mathcal{F}_i, \mathcal{F}_j\big) \le V_i\big(\widehat{\mathcal{F}}_i; \widehat{\mathcal{F}}_i, \widehat{\mathcal{F}}_j\big) - V_i\big(\mathcal{F}_i; \widehat{\mathcal{F}}_i, \widehat{\mathcal{F}}_j\big) \qquad \Big(MEC_{\{\mathcal{F}, \widehat{\mathcal{F}}\}}\Big)$$

Note that $EC_{\{\mathcal{F}, \widehat{\mathcal{F}}_i\}}$ is a stronger condition than $MEC_{\{\mathcal{F}, \widehat{\mathcal{F}}\}}$ in that it takes player j's information structure as fixed; that is, if $EC_{\{\mathcal{F}, \widehat{\mathcal{F}}_i\}}$ is satisfied, so is $MEC_{\{\mathcal{F}, \widehat{\mathcal{F}}\}}$ where $\widehat{\mathcal{F}} \equiv (\widehat{\mathcal{F}}_i, \widehat{\mathcal{F}}_j)$ and $\widehat{\mathcal{F}}_j \equiv \mathcal{F}_j$. This reflects the fact that multilateral expectation conformity captures strategic

---

[8]This need not be the case for all applications; consider memory management: 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).

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

[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 his actual choice of cognition.

interactions in the choice of information structures, while expectation conformity captures only the impact of expectations; more on this below.

2.2 Expectation conformity, increasing differences and equilibrium multiplicity

The expectation-conformity condition should not be mistaken for the standard "increasing differences (ID)" condition that plays a fundamental role in monotone comparative statics. The essential difference between the two concepts is that player i's payoff in $EC_{\{\mathcal{F},\widehat{\mathcal{F}}_i\}}$ and in $MEC_{\{\mathcal{F},\widehat{\mathcal{F}}\}}$ depends not only on his information, but also on player j's anticipation of his information. More formally, the increasing differences condition, applied to ordered information structures, takes the following form: If $\widehat{\mathcal{F}}_k$ is finer than $\mathcal{F}_k$ for all k, then

$$V_i\big(\widehat{\mathcal{F}}_i;\widehat{\mathcal{F}}_i,\mathcal{F}_j\big) - V_i\big(\mathcal{F}_i;\mathcal{F}_i,\mathcal{F}_j\big) \le V_i\big(\widehat{\mathcal{F}}_i;\widehat{\mathcal{F}}_i,\widehat{\mathcal{F}}_j\big) - V_i\big(\mathcal{F}_i;\mathcal{F}_i,\widehat{\mathcal{F}}_j\big). \qquad (ID)$$

That is, expectation conformity reflects the fact that the players do not observe each other's choice of information structure while the increasing differences condition posits that information structures are common knowledge at stage 2 *on and off* the equilibrium path. Put differently, expectation conformity and increasing differences capture strategic complementarities in information acquisition, under *covert* acquisition for the former and *overt* acquisition for the latter.

Let us investigate the difference between unilateral and multilateral expectation conformity and their relationship with increasing differences a bit further. Let us decompose the difference between the RHS and the LHS of $MEC_{\{\mathcal{F},\widehat{\mathcal{F}}\}}$,

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

into three terms:

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

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

and

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

and so
$$\Gamma_i^{MEC}(\mathcal{F}, \widehat{\mathcal{F}}) = \Gamma_i^{EC}(\mathcal{F}, \widehat{\mathcal{F}}) + \Gamma_i^{ID}(\mathcal{F}, \widehat{\mathcal{F}}) + \Gamma_i^{P}(\mathcal{F}, \widehat{\mathcal{F}}).$$

The term Γ_i^{EC} takes player j's information structure $\mathcal{F}_j$ as given and measures the increase in i's incentive to acquire information $\widehat{\mathcal{F}}_i$ rather than $\mathcal{F}_i$ when player j anticipates this change in i's information acquisition strategy. The difference Γ_i^{P} measures a pure perception effect: fixing player i's actual cognition $\mathcal{F}_i$, it represents how i's gain (or loss) of being perceived as acquiring information $\widehat{\mathcal{F}}_i$ rather that $\mathcal{F}_i$ is affected by player j's actual cognition ($\mathcal{F}_j$ vs. $\widehat{\mathcal{F}}_j$). To compare multilateral and unilateral expectation conformity, one can aggregate Γ_i^{ID} and Γ_i^{P} into a "strategic interaction" term

$$\Gamma_i^{SI}(\mathcal{F}, \widehat{\mathcal{F}}) \equiv \left[V_i(\widehat{\mathcal{F}}_i; \widehat{\mathcal{F}}_i, \widehat{\mathcal{F}}_j) - V_i(\mathcal{F}_i; \widehat{\mathcal{F}}_i, \widehat{\mathcal{F}}_j) \right] - \left[V_i(\widehat{\mathcal{F}}_i; \widehat{\mathcal{F}}_i, \mathcal{F}_j) - V_i(\mathcal{F}_i; \widehat{\mathcal{F}}_i, \mathcal{F}_j) \right].$$

That is, keeping j's expectation about i's information constant, $\Gamma_i^{SI}(\mathcal{F}, \widehat{\mathcal{F}})$ depicts the increase in i's incentive to acquire $\widehat{\mathcal{F}}_i$ rather than $\mathcal{F}_i$ when j's information changes from $\mathcal{F}_j$ to $\widehat{\mathcal{F}}_j$. There is *positive strategic interaction* for information structures $(\mathcal{F}, \widehat{\mathcal{F}})$ if $\Gamma_i^{SI}(\mathcal{F}, \widehat{\mathcal{F}}) \geq 0$.

Note that for the prominent class of *one-sided cognitive games* (only player i, say, acquires information, i.e., Ψ_j is a singleton), $\Gamma_i^{ID} = \Gamma_i^{P} = 0$ and so

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

We next provide an example in which, in contrast, multilateral expectation conformity is driven solely by increasing differences. Consider a *matching model* (reminiscent of Lester et al's 2012 model of asset liquidity based on recognizability) 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, say 1 if the partner is adequate and a highly negative number otherwise, and so a match occurs only if both can ascertain it is a good one for them. 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 his expectation about the other player's information: A player who knows he receives 1 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, $\Gamma_i^{EC}(\mathcal{F}, \widehat{\mathcal{F}}) = [\widehat{\rho}_i \rho_j - \rho_i \rho_j] - [\widehat{\rho}_i \rho_j - \rho_i \rho_j] = 0$, and so expectation conformity is not strictly satisfied. Similarly, $\Gamma_i^{P}(\mathcal{F}, \widehat{\mathcal{F}}) = 0$, and so

$$\Gamma_i^{MEC}(\mathcal{F}, \widehat{\mathcal{F}}) = \Gamma_i^{ID}(\mathcal{F}, \widehat{\mathcal{F}}).$$

By contrast, $\Gamma_i^{ID}(\mathcal{F}, \widehat{\mathcal{F}}) = (\widehat{\rho}_i - \rho_i)(\widehat{\rho}_j - \rho_j)$, capturing the standard strategic complementarities that are conducive to equilibrium multiplicity.

An alternative way of rewriting the MEC condition goes as follows. Suppose that player j's *best response* to $\mathcal{F}_i$ is (for expositional simplicity) unique and denoted $\mathcal{F}_j = R_j(\mathcal{F}_i)$. Player i's

payoff depends only on his actual information structure $\mathcal{F}'_i$ and player j's information structure $\mathcal{F}_j = R_j(\mathcal{F}_i)$ when player j anticipates information structure $\mathcal{F}_i$. Let

$$T_i\left(\mathcal{F}'_i, R_j(\mathcal{F}_i)\right) \equiv V_i\left(\mathcal{F}'_i; \mathcal{F}_i, R_j(\mathcal{F}_i)\right)$$

denote player i's gross payoff when endowed with information $\mathcal{F}'_i$ and expected by played j to be endowed with $\mathcal{F}_i$.

An equilibrium of the cognitive game must satisfy (again assuming a unique best response):

$$\mathcal{F}_i = \arg\max_{\{\mathcal{F}'_i \in \Psi_i\}} T_i\left(\mathcal{F}'_i, R_j(\mathcal{F}_i)\right) \quad \text{for all } i.$$

The multilateral expectation conformity condition then becomes

$$T_i\left(\widehat{\mathcal{F}}_i, R_j(\mathcal{F}_i)\right) - T_i\left(\mathcal{F}_i, R_j(\mathcal{F}_i)\right) \leq T_i\left(\widehat{\mathcal{F}}_i, R_j(\widehat{\mathcal{F}}_i)\right) - T_i\left(\mathcal{F}_i, R_j(\widehat{\mathcal{F}}_i)\right) \quad \text{for all } i.$$

We will make use of this alternative way of writing (MEC) in the next subsection.

2.3 A differentiable version

In many applications, information structures are identified by a parameter $\rho_i \in \mathbb{R}$, with a higher ρ_i indicating a finer information structure in the sense of Blackwell. Player i's stage-2 payoff, as a function of the anticipated cognition (ρ_i, ρ_j) and an actual cognition ρ'_i by player i, can then be denoted, abusing notation slightly, by $V_i(\rho'_i; \rho_i, \rho_j)$.

Assuming differentiability, one can state (*strict*) *local versions* of the definitions above (numbered indices denoting partial derivatives):

$$EC \text{ holds if } V_{i12} > 0,$$

$$\text{Positive strategic interaction holds if } V_{i13} > 0,$$

$$MEC \text{ holds if } V_{i12} + V_{i13} > 0.$$

In the matching game just described, player j's stage-2 behavior is independent of his beliefs about i's information structure and so $V_{i12} = 0$; by contrast $V_{i13} > 0$.

Using reaction functions, we now obtain a result that sheds much light on the frequent occurrence of expectation conformity. Abusing again notation, let $T_i(\rho'_i, R_j(\rho_i))$ denote player i's gross payoff when endowed with information ρ'_i and expected by player j to be endowed with information ρ_i. We assume that T_i is strictly concave in ρ'_i.[11]

Assumption 1 (monotone best responses). *Let information structures be ordered and indexed by a parameter $\rho_i \in \mathbb{R}$, with a higher ρ_i corresponding to a finer information structure in the sense of Blackwell. Suppose that for all (i,j), player j's best response to a choice ρ_i by player i, $\rho_j = R_j(\rho_i)$, is strictly monotonic and differentiable:*

---

[11]I am grateful to Navin Kartik for posing a perspicacious question that led to the following proposition.

Either $R'_j(\rho_i) > 0$ for all (i, ρ_i) or $R'_j(\rho_i) < 0$ for all (i, ρ_i).

That is, the information structures are either strategic complements or strategic substitutes.

Proposition 1 (**monotone best responses imply** MEC). *Suppose Assumption 1 holds. Then (the strict version of) $MEC_{\{\rho, \widehat{\rho}\}}$ obtains for all $\{\rho, \widehat{\rho}\}$.*

Proof: $MEC_{\{\rho, \widehat{\rho}\}}$ strictly obtains if for all i,

$$T_i\left(\widehat{\rho}_i, R_j(\rho_i)\right) - T_i\left(\rho_i, R_j(\rho_i)\right) < T_i\left(\widehat{\rho}_i, R_j(\widehat{\rho}_i)\right) - T_i\left(\rho_i, R_j(\widehat{\rho}_i)\right),$$

or

$$\int_{\rho_i}^{\widehat{\rho}_i} \int_{R_j(\rho_i)}^{R_j(\widehat{\rho}_i)} \frac{\partial^2 T_i(x, y)}{\partial x \partial y} dx\, dy > 0,$$

which is satisfied since

$$R'_j(\rho_i) = \left(-\frac{\frac{\partial^2 T_j}{\partial \rho_i \partial \rho_j}}{\frac{\partial^2 T_j}{\partial \rho_i^2}}\right)(\rho_i, R_j(\rho_i)).$$

∎

A couple of comments on this proposition are in order. First, writing gross payoffs as reduced-form functions of information structures and reactions thereto makes transparent the relationship between models of *covert* information acquisition and the familiar application of supermodularity conditions in industrial organization and other fields. In the examples below, obtaining the reduced form will be trivial; for many games of interest for economics, though, reduced-forms are not immediately available and part of this paper's contribution will consist in looking at whether strategic complementarity/substitutatiblity obtains in some important environments. Second, and related to this first observation, we have assumed a parametrized information structure and differentiability; more generally one could endow Ψ_i with a lattice structure and make supermodularity assumptions. Most applications, though, use the simple form posited in Assumption 1. Third, this particular result extends to more than two players only in the case of strategic complements, an observation that is familiar from standard comparative statics.[12]

The intuition for Proposition 1 goes as follows: Suppose that player i is expected by player j to acquire more information. Player j then acquires more (less) information under strategic complementarity (substitutability). Player i's then optimally acquires more information, to complement j's increased knowledge in the case of strategic complements and to supplement j's shortage of knowledge in case of strategic substitutes.

The following straightforward examples satisfy Assumption 1:

Advisor/adviser game. A PhD student has written a proof for a theorem. This proof contains a flaw with probability x. If undetected, the flaw will create reputation damages $d_i > 0$ for both

---

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

the student and the advisor. Let ρ_i denote player i's probability of finding the flaw if there is one, and $C_i(\rho_i)$, an increasing function, the cost of the corresponding effort. Then player i's payoff is $-x(1 - \rho_i)(1 - \rho_j)d_i - C_i(\rho_i)$. We are in the case of strategic substitutes with $\partial^2 T_i / \partial \rho_i \partial \rho_j = -xd_i$.

Team work. Two coworkers work on distinct components of a project. Both components must be successful in order for the overall project to succeed. Let ρ_i denote the probability that player i's information leads him to find out the solution to his problem. Then payoffs are proportional to $\rho_i \rho_j$, a trivial case of strategic complementarity.

2.4 Equilibrium multiplicity/uniqueness

Revealed preference implies that a necessary condition for $(\mathcal{F}, \widehat{\mathcal{F}})$ to form two equilibria is that $\Gamma_i^{MEC}(\mathcal{F}, \widehat{\mathcal{F}}) \geq 0$ for all i. The following proposition says that the condition is also sufficient for both $\mathcal{F}$ and $\widehat{\mathcal{F}}$ to be equilibria for appropriately chosen cost functions:

Proposition 2 (multiplicity and uniqueness).
(i) If $MEC_{\{\mathcal{F}, \widehat{\mathcal{F}}\}}$ is satisfied for two distinct information structures $\mathcal{F}$ and $\widehat{\mathcal{F}}$, then there exist cost functions $\{C_i(\cdot)\}_{i=1,2}$ 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 $MEC_{\{\mathcal{F}, \widehat{\mathcal{F}}\}}$ is satisfied for no two distinct information structures $(\mathcal{F}, \widehat{\mathcal{F}})$, then there cannot exist multiple equilibria.

Proof: (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

$$V_i(\widehat{\mathcal{F}}_i; \mathcal{F}_i, \mathcal{F}_j) - V_i(\mathcal{F}_i; \mathcal{F}_i, \mathcal{F}_j) \leq C_i(\widehat{\mathcal{F}}_i) - C_i(\mathcal{F}_i) \leq V_i(\widehat{\mathcal{F}}_i; \widehat{\mathcal{F}}_i, \widehat{\mathcal{F}}_j) - V_i(\mathcal{F}_i; \widehat{\mathcal{F}}_i, \widehat{\mathcal{F}}_j). \quad (1)$$

In the absence of further requirement, pick cost functions satisfying (1) 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, and noting that $V_i(\widehat{\mathcal{F}}_i; \mathcal{F}_i, \mathcal{F}_j) - V_i(\mathcal{F}_i; \mathcal{F}_i, \mathcal{F}_j) \geq 0$, pick a cost function satisfying (1) as well as:

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

(ii) Conversely, if $\mathcal{F}$ and $\widehat{\mathcal{F}}$ were two distinct equilibria, condition (1) would be satisfied, and so $MEC_{\{\mathcal{F}, \widehat{\mathcal{F}}\}}$ would hold, a contradiction. ∎

Because $MEC_{\{\mathcal{F},\widehat{\mathcal{F}}\}}$ holds if $\widehat{\mathcal{F}}_j = \mathcal{F}_j$ and $EC_{\{\mathcal{F},\widehat{\mathcal{F}}_i\}}$ holds, Proposition 2(i) also holds if one replaces $MEC_{\{\mathcal{F},\widehat{\mathcal{F}}\}}$ by $EC_{\{\mathcal{F},\widehat{\mathcal{F}}_i\}}$.

Definition 3 (**collective cognitive trap**). *Players are exposed to a collective cognitive trap if there exist two information structures $\mathcal{F}$ and $\widehat{\mathcal{F}}$ such that*

(i) $\widehat{\mathcal{F}}_i$ is finer than $\mathcal{F}_i$ for all i,

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

(iii) $V_i\big(\mathcal{F}_i\,;\,\mathcal{F}_i\,,\mathcal{F}_j\big) - C_i\big(\mathcal{F}_i\big) > V_i\big(\widehat{\mathcal{F}}_i\,;\widehat{\mathcal{F}}_i\,,\widehat{\mathcal{F}}_j\big) - C_i\big(\widehat{\mathcal{F}}_i\big)$ for all i.

If the conditions in Definition 3 are fulfilled, cost functions are such that players conform to expectations, choosing either $\widehat{\mathcal{F}}_i$ or $\mathcal{F}_i$ when expected to (condition (ii)), and prefer the low-cognition outcome to the high-cognition one (condition (iii)).

Definition 4 (**individual cognitive trap**). *Player i is exposed to an individual cognitive trap if there exists $\big(\mathcal{F}_i\,,\widehat{\mathcal{F}}_i\,,\mathcal{F}_j\big)$ such that*

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

(ii) $\big(\mathcal{F}_i\,,\mathcal{F}_j\big)$ and $\big(\widehat{\mathcal{F}}_i\,,\mathcal{F}_j\big)$ are both equilibria,

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

Individual cognitive traps will be prominent in one-sided cognitive games (Ψ_j is a singleton).

2.5 Pure conflict and pure alignment: Zero-sum and coordination games

(a) *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); that is, the gross payoffs satisfy the zero-sum condition: for all $(\sigma_i, \sigma_j, \omega)$

$$u_i(\sigma_i, \sigma_j, \omega) + u_j(\sigma_j, \sigma_i, \omega) = 0.$$

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). Another interesting property is given by the following result.[13]

---

[13]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.

Proposition 3 (zero-sum games). *Zero-sum games satisfy for all $(\mathcal{F}, \widehat{\mathcal{F}})$*

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

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

Proof: The zero-sum property implies that for all strategies (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 $R_i(\hat{\sigma}_j)$ denote player i's best $\mathcal{F}_i$-measurable response to $\hat{\sigma}_j$ and $\widehat{R}_i(\sigma_j)$ denote player i's best $\widehat{\mathcal{F}}_i$-measurable response to σ_j. Obviously, $U_i\big(R_i(\hat{\sigma}_j), \hat{\sigma}_j\big) \geq U_i(\sigma_i, \hat{\sigma}_j)$ and $U_i\big(\widehat{R}_i(\sigma_j), \sigma_j\big) \geq U_i(\hat{\sigma}_i, \sigma_j)$. This implies that

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

for all $(\mathcal{F}, \widehat{\mathcal{F}})$. Because equilibrium multiplicity requires $\Gamma_i^{MEC}(\mathcal{F}, \widehat{\mathcal{F}}) \geq 0$ for all i, the inequality implies indifference for both players, i.e., $\Gamma_i^{MEC}(\mathcal{F}, \widehat{\mathcal{F}}) = 0$ for all i.[14] Suppose, say, that in the $(\widehat{\mathcal{F}}_i, \widehat{\mathcal{F}}_j)$ equilibrium, player i has strictly optimal strategy $\widehat{\mathcal{F}}_i$. Then

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

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

and so $\Gamma_i^{MEC}(\mathcal{F}, \widehat{\mathcal{F}}) > 0$, a contradiction. ∎

(b) *Coordination games*

In stark contrast with zero-sum games, which exhibit completely antinomic interests, coordination games (which have been the focus of much recent interest in macroeconomics) involve perfectly aligned interests. A typical coordination game has payoffs[15]

$$u_i(a_i, a_j, \omega) = -(a_i - \omega)^2 - (a_i - a_j)^2;$$

that is, each player wants to match his action both with the state of nature and with the other player's choice. The density of ω is continuous on some interval $[\omega^{\text{inf}}, \omega^{\text{sup}}]$, say.

As for the sets Ψ_i of information structures, we assume that they are totally ordered. We take the filtration to be a sequence of finer and finer (and more and more costly) information

---

[14]To illustrate this double indifference, 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.

[15]See Angeletos and Pavan (2007) for a more general version than the quadratic coordination game. Much of the macroeconomic literature analyzes the relative use of public and private signals about the state of nature. This literature often assumes Gaussian distributions; we will not need this assumption for our purposes.

structures. Furthermore, we assume that feasible information structures are the same for both players: $\Psi_i = \Psi_j$.

Suppose that player i has chosen a (weakly) finer information structure than player j: $\mathcal{F}_j \subseteq \mathcal{F}_i$. An element of $\mathcal{F}_i$, say, is then characterized by a mean ω_i and a variance $\sigma_i^2 \equiv E\big[(\omega - \omega_i)^2\big]$. Player i knows ω_j and is able to predict j's choice, but the converse may not hold. Like in the coordination games literature, the parties do not communicate prior to choosing their actions. Optimal actions are then

$$a_j = \frac{E\big(a_i(\omega_i)|\mathcal{F}_j\big) + \omega_j}{2} \quad \text{and} \quad a_i(\omega_i) = \frac{a_j + \omega_i}{2},$$

and so, using the law of iterated expectations,

$$a_j = \omega_j \quad \text{and} \quad a_i = \frac{\omega_j + \omega_i}{2}.$$

Furthermore

$$U_i = -\frac{(\omega_j - \omega_i)^2}{2} - \sigma_i^2 \;\; \text{and} \;\; U_j = -\frac{E(\omega_j - \omega_i)^2}{4} - \sigma_j^2$$

Proposition 4 (coordination games). *Coordination games with totally ordered information structures satisfy multilateral expectation conformity for any two distinct information structures $(\mathcal{F}_1, \mathcal{F}_2)$ and $(\widehat{\mathcal{F}}_1, \widehat{\mathcal{F}}_2)$ such that $\mathcal{F}_1 \subseteq \widehat{\mathcal{F}}_1$ and $\mathcal{F}_2 \subseteq \widehat{\mathcal{F}}_2$. For $\mathcal{F}_j \equiv \widehat{\mathcal{F}}_j$, they also satisfy $\Gamma_i^{EC}\big(\mathcal{F}, \widehat{\mathcal{F}}\big) \geq 0$, with strict inequality unless $\mathcal{F}_j \subseteq \mathcal{F}_i \subset \widehat{\mathcal{F}}_i$.*

The proof of Proposition 4 can be found in the Appendix. The one case in which expectation conformity is only weakly satisfied is when player i is always better informed than player j $\big(\mathcal{F}_j \subseteq \mathcal{F}_i \subset \widehat{\mathcal{F}}_i\big)$. Then player j does not adjust his strategy to the information held by player i and so the value of information for player i is independent of player j's expectation. The welfare comparison among equilibria of coordination games is in general ambiguous. On the one hand, a player may not increase his cognitive intensity by fear that the other would not, while more cognition would be beneficial for both. On the other hand, the two players may be trapped by the same coordination motive into a wastefully high-cognition state.[16]

3 Environments with a game setter

Consider now the following class of games. The stage-2 game has two sub-stages. One player, say player i, first selects between two games to play with player 2; for example, in our first three illustrations, he chooses between "out" ("null game", "exercise an outside option") and "in" ("play with player j"). Accordingly, we will label the two games the "in" and "out" games.

After choosing the "in" game, player i then choosing an action a_i and player j an action $a_j \in \mathbb{R}$ (in our examples below, only player j has an action[17]). Cognition is one-sided; prior to

---

[16]Suppose, e.g., that $\omega = -a$ with probability $1/2$ and $+a$ with probability $1/2$. The no-cognition equilibrium exists if $-a^2 \geq -(a^2/2) - c$; the high-cognition equilibrium exists if $-c \geq -2a^2$. So the two equilibria co-exist whenever $(a^2/2) \leq c \leq 2a^2$. The no-cognition equilibrium dominates for $c > a^2$ and is dominated for $c < a^2$.

[17]More generally, player i may pick an action. 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.

choosing between "in" and "out", player i selects at stage 1 an information structure. Player j, when choosing a_j, by contrast knows only that player i chose "in".

The state of nature is $\omega \in (-\infty, +\infty)$, with prior mean ω_0. An experiment, indexed by $\rho \in \mathbb{R}$, 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 $\int_{-\infty}^{+\infty} m \, dF(m; \rho) = \omega_0$ for all ρ. Player i's set of possible information structures is totally ordered: A higher ρ means a mean-preserving spread.

Assumptions on 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 $\delta_i(m, a_j)$.

Assumption 2 (game-setter environments). *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. $\delta_i(m, a_j)$ is twice differentiable, is increasing in m and a_j, and satisfies*

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

(with $\gamma > 0$ and $\tau \geq 0$).

More generally, what is needed for $F_\rho(m^*; \rho) \geq 0$ to be sufficient for expectation conformity (see Proposition 5) is, besides differentiability and monotonicity in m and a_j, that

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

Assumption 2 implies that for each a_j, there exists a cutoff strictly decreasing $m^*(a_j)$ such that player i chooses "in" if and only if $m \geq m^*(a_j)$.

As for player j, consider the stage-2 game with player j anticipating 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 (at this stage player i's decision depends only on a_j and m and no longer on the stage-1 choice of ρ).

Assumption 3 (game-setter environments).

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

where $M^+(m^*, \rho) \equiv E(m | m \geq m^*, \rho)$ denotes the truncated mean for information structure ρ.

Examples: Assumptions 2 and 3 for instance are satisfied by the following games:

(a) *Akerlof's lemons game.* Let player i be, say, the seller. The seller can sell his 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

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

as the cutoff $m^*(a_j)$ is equal to $v_i - a_j$; and so $\delta_i(m, a_j) = a_j - (v_i - m)$ (thus, $\gamma = 1$ and $\tau = 0$). The solution a_j is unique if the hazard rate of the distribution of m for parameter $\rho^\dagger$ is monotonic,[18] which we will assume, and furthermore Assumption 3 is satisfied.

(b) *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.[19] 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

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

Assuming a unique solution a_j to this equation, Assumption 3 is satisfied.

(c) *Team formation.* Player i has a project. He 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 he does not want to do if the project is a good one. Player i's payoff is $(v - \omega) - C_i$ if he does it alone and $a_j(v - \omega) - c_i$ if it is a joint project, where a_j is the value share left by (competitive) player j and $c_i < C_i$ is player i's reduced cost of project implementation. Let c_j denote player j's cost (with $c_i + c_j < C_i$). Player i chooses "in" if and only if

$$\delta_i(a_j, m) \equiv C_i - c_i - (1 - a_j)(v - m) \ge 0.$$

And so $\gamma = v$ and $\tau = 1$. Finally,

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

Provided that $c_j - \frac{\partial M^+}{\partial m^*}(C_i - c_i) > 0$ (e.g. $2c_j > C_i - c_i$ for a uniform distribution), then

---

[18]See An (1998).

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

Assumption 3 is satisfied.

(d) *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 $[0,1]$ according to distribution $F(\cdot)$, 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 $0 < \ell_k < L_k$ for all k. Player i has a v_i high enough that (s)he wants to marry regardless, may acquire information about ω and 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 will accept to marry) while $v_j - \ell_j < 0$ (player j would refuse a mariage that would end up for sure in a "smooth divorce"). Let $a_j = 1 (= 0)$ if player j accepts (refuses) to marry when the proposed contract includes the covenant. This game, which is the non-transferable-utility counterpart of the class of transferable-utility games studied in Section 4, satisfies Assumptions 2 and 3:

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

and so $\gamma = v_i$ and $\tau = \ell_i$[20]. Furthermore, $a_j = 1$ iff $v_j - M^+(m^*, \rho^\dagger)\ell_j \geq 0$, where $m^* = m^*(a_j(\rho^\dagger))$ is given by $m^* \left[L_i - a_j(\rho^\dagger)\ell_i \right] = \left[1 - a_j(\rho^\dagger) \right] v_i$. And so Assumption 3 holds.[21]

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

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

Local expectation conformity holds provided that

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

Proposition 5 (game-setter environments). *Under Assumptions 2 and 3:*
(i) Local expectation conformity holds whenever for the relevant cutoff $m^* = m^*(a_j(\rho^\dagger))$,

$$\left[(\gamma - \tau m^*) F_\rho(m^*; \rho) - \tau \int_{m^*}^{+\infty} F_\rho(m; \rho) dm \right] \left[\left[M^+(m^*, \rho) - m^* \right] F_\rho(m^*; \rho) - \int_{m^*}^{+\infty} F_\rho(m; \rho) dm \right] > 0;$$

in particular, it holds whenever $F_\rho(m^*; \rho) \geq 0$ at $m^* = m^*(a_j(\rho^\dagger))$.

(ii) When the family of distributions F is characterized by a sequence of rotations,[22] a sufficient

---

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

[21] Here $a_j \in \{0, 1\}$. But like in the other examples, one can smooth player $j's$ behavior by introducing a random shock to v_j.

[22] See e.g., Johnston and Myatt (2006). "Rotation" here refers to the existence of some m_ρ such that $F_\rho > 0$ for $-\infty < \omega < m_\rho$ and $F_\rho < 0$ for $m_\rho < \omega < +\infty$ (single crossing). Note that rotation plus mean preserving implies a mean-preserving spread but that the converse does not hold. Examples of mean-preserving spreads with

condition for local expectation conformity is therefore that the cutoff lie to the left of the rotation point, which holds when player i is sufficiently eager to play "in".

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

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

From Assumption 3 ,

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

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

Next

$$\int_{m^*(a_j(\rho^\dagger))}^{+\infty} \frac{\partial \delta_i}{\partial a_j}(m, a_j(\rho^\dagger))dF_\rho(m; \rho)$$

$$= \quad -\frac{\partial \delta_i}{\partial a_j}\big(m^*(a_j(\rho^\dagger)), a_j(\rho^\dagger)\big)F_\rho\big(m^*(a_j(\rho^\dagger)); \rho\big)$$

$$-\int_{m^*(a_j(\rho^\dagger))}^{+\infty} \frac{\partial^2 \delta_i}{\partial a_j \partial m}\big(m, a_j(\rho^\dagger)\big)F_\rho(m; \rho)dm$$

$$= \quad -\Big(\gamma - \tau m^*\big(a_j(\rho^\dagger)\big)\Big)F_\rho\big(m^*(a_j(\rho^\dagger)); \rho\big) + \tau\int_{m^*(a_j(\rho^\dagger))}^{+\infty} F_\rho(m; \rho)dm.$$

The latter term is negative as ρ is an index of mean-preserving spread. The former term is non-positive provided that $F_\rho \geq 0$ at $m^*(a_j(\rho^\dagger))$.

∎

Application to the lemons game. Under hard (i.e., verifiable) information, the seller uses acquired information to disclose to the buyer that the good for sale has a high value; under soft information (the case considered by Akerlof), the seller acquires information to withdraw from the market if the good is very valuable. We confine attention to soft information, but a similar result holds for hard information.

(a) *Non directed search.* Assume that information collection follows the standard general or non-directed search technology:

---

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 $[0,1]$).

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

That is, the seller learns the true state of nature with probability ρ and nothing with probability $1 - \rho$. It is then natural to posit an information acquisition cost $C_i^G(\rho)$. Then[23]

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

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.[24]

(b) *Directed search.* Next, suppose that search is directed. At cost $C_i^D\left(\int_{-\infty}^{+\infty} e_i^\omega dQ(\omega)\right)$, a strictly increasing and convex cost function the seller can learn ω with probability $e_i^\omega \in \{0, 1\}$; with probability $1 - e_i^\omega$, the seller receives no signal. It is easy to show that it is optimal for the seller to adopt a cutoff strategy (this result, like the others, also holds under hard information):

$$e_i^\omega = \begin{cases} 1 & \text{if } \omega \leq \rho \\ 0 & \text{if } \omega > \rho \end{cases}$$

for some ρ, which is a measure of player i's cognitive effort. In this optimal class,

$$F(m; \rho) = \begin{cases} F(m) & \text{for } m \leq \rho \\ F(\rho) & \text{for } \rho < m < M^+(\rho) \\ 1 & \text{for } m \geq M^+(\rho). \end{cases}$$

In equilibrium $v_j - a_j = M^+(\rho)$ and $m^* = v_i - a_j$. Necessarily $\rho < m^*$ (the seller acquires information about a state of nature only if he intends to withdraw). So $\rho < m^* < M^+(\rho)$, implying

$$F_\rho(m^*; \rho) > 0.$$

Thus the lemons game always satisfies expectation conformity under directed search.[25]

---

[23]To show that $H = 0$ on $[\omega_0, +\infty)$, note that $H(+\infty) = 0$ and that $dH/dm^* = 0$ on this domain.

[24]$m^* - M^+(m^*, \rho^\dagger) = -(v_j - v_i)$. 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^*.

[25]One can combine directed and non-directed search in the following way: Player i's information collection proceeds in two sub-stages in which search is first non-directed and then directed. First, the player performs some preliminary, *general-purpose search* to try to apprehend the context; this search costs $C_i^G(\rho_i^G)$ and succeeds with probability ρ_i^G. If the first stage is unsuccessful, the search process stops. If the general-purpose search is successful, the player can engage in *directed search* and pick the probability $e_i^\omega \in [0, 1]$ of learning that the state is ω; this latter search costs $C_i^D(\rho_i^D; \rho_i^G)$ where $\rho_i^D \equiv \int_{-\infty}^{+\infty} e_i^\omega dQ(\omega)$. In general, $\partial C_i^D/\partial \rho_i^D$ can be independent of ρ_i^G, decrease with ρ_i^G (benefits of acquired knowledge) or increase with ρ_i^G (fatigue, time constraints). Information structures can easily be ranked whenever directed and non-directed search efforts are "weak complements", i.e., if the cross-partial derivative of C_i^D is non-positive (an increase in general search does not discourage directed search).

Proposition 6 (**lemons game**). *The cognition-augmented lemons game always satisfies expectation conformity under directed search, and under non-directed search satisfies it if and only if the gains from trade are sufficiently large.*

The intuition why the lemons game 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 his incentives to acquire information.

A similar insight holds when it is the buyer who engages in cognition, although the treatment is more involved: Asymmetrically informed buyers are no longer competitive; furthermore the stage-2 equilibrium may be in mixed strategies. Suppose that there is a single buyer, who may acquire information. If, to parallel our treatment so far, we assume that the seller makes a take-it-or-leave-it offer, the seller may be able to extract the buyer's stage-2 surplus (for example, with directed search, charge the buyer's expected value conditional on the buyer's equilibrium cutoff). The buyer then receives a negative overall utility once the stage-1 information acquisition cost is accounted for, which is of course impossible. One can then study mixed strategies or else allow for more complex price-setting processes with divided bargaining power between the buyer and the seller. The intuition for expectation conformity is fortunately more 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. Facing a higher price, the buyer then finds it more costly to buy lemons and thus is incentivized to find out about possible bad news.

4 Cognition-intensive contracting

4.1 Description and illustrations

This section studies an environment in which cognition changes the nature of the contract between two parties. Its contribution is two-fold. First, it brings together and generalizes a number of otherwise disconnected contributions within a unified framework. Second, it obtains new results, in particular by extending the analysis to the absence of good-faith bargaining requirement and to two-sided cognition.

In a number of applications, the sample space, with generic element $\widetilde{\omega}$ for the sake of this section, is binary: $\Omega = \{\omega, \widehat{\omega}\}$ (with common knowledge prior probabilities q and $\widehat{q}$ such that $q + \widehat{q} = 1$). If $\widetilde{\omega} = \omega$, the players' initial and final information is Ω, regardless of any cognitive effort incurred (there is no snag/flaw to be discovered, say). If $\widetilde{\omega} = \widehat{\omega}$, then player i learns $\widehat{\omega}$ with probability ρ_i and nothing (keeps information set Ω) with probability $1 - \rho_i$. Information $\widehat{\omega}$ is hard information and therefore can be disclosed to player $j \neq i$ if player i decides so. The cognition cost is $C_i(\rho_i)$ with $C_i' > 0$, $C_i'' > 0$, $C_i'(0) = C_i(0) = 0$, $C_i'(1) = \infty$.

A prominent interpretation of this information structure goes as follows: Knowing the state of nature allows a trade, a technological choice or a contract design to match the state. Furthermore, state $\widehat{\omega}$ is initially off the radar of players, although they are aware that "they may not

have thought about something". In state ω, a known (say, "business as usual", or "boiler plate") choice is fine. By contrast, state $\widehat{\omega}$ requires an original and yet unknown response; furthermore the very act of conceptualizing state $\widehat{\omega}$ also reveals the nature of this unknown response. The model is Bayesian: Players are uninformed, but rational, in that they know that they don't know.[26]

Contracts are most often "incomplete": They fail to describe some states of nature or actions to be undertaken in certain states of nature. Incompleteness is often motivated by the presence of "unforeseen contingencies". The question as to whether omitted contingencies are truly unforeseeable or just extremely costly to foresee may not be worthy of interrogation, though. An approach in terms of "I did not think about/I did not have this in mind when making the decision" seems more fruitful.[27]

Two risk-neutral parties can jointly contract on an action in $\{a, \hat{a}\}$ and transfer money between themselves. Action a (respectively $\hat{a}$) is jointly optimal, i.e., maximizes the joint surplus, in state ω (respectively $\widehat{\omega}$). Initially only the sample space and action a are known to the two players. Searching for information leads them to either learn nothing, or to learn both $\widehat{\omega}$ and $\hat{a}$: Becoming aware of the state of nature $\widehat{\omega}$ also reveals what's to be done in that state of nature, and conversely.[28]

Figure 1 represents the joint surplus of the two players. $\delta \geq 0$ here stands for the deadweight loss associated with choosing action a in state $\widehat{\omega}$.

state of nature

	ω	$\widehat{\omega}$
a	U	$\widehat{U} - \delta$
$\hat{a}$		$\widehat{U}$

action

Figure 1: joint surplus

We let $U_i^{\widetilde{\omega}}$ denote player i's *gross* surplus if the optimal action is chosen in state of nature $\widetilde{\omega}$: $\Sigma_i U_i^{\widetilde{\omega}} = U^{\widetilde{\omega}}$. Similarly, we can decompose the respective losses (or gains) of both players when the wrong action is selected: $\Sigma_i \delta_i = \delta$.

The following notation will play an important role in what follows:

- w_i denotes player i's bargaining power in a negotiation; that is, player i reaps a fraction w_i of gains from trade: $\Sigma_i w_i = 1$.

---

[26]This approach therefore uses the familiar state-space representation. As Dekel et al (1998) show in their eponymous paper, "standard state-space models preclude unawareness". In any state of nature in which a player does not know some event, he knows that he does not know it.

[27]The cognitive approach to incomplete contracting is taken up for example in Bolton and Faure-Grimaud (2010), Tirole (2009), Von Thadden and Zhao (2012) and Zhao (2014).

[28]This can easily be extended to multi-stage cognition: The first search may reveal that "something is fishy" with design a. The player can then continue searching, and so forth.

An even richer environment would add a trust dimension as in Dziuda (2011), who shows that in the presence of multi-dimensional adverse selection, a sender may want to engage in partial disclosure of information that she would not normally disclose, so as to inspire trust.

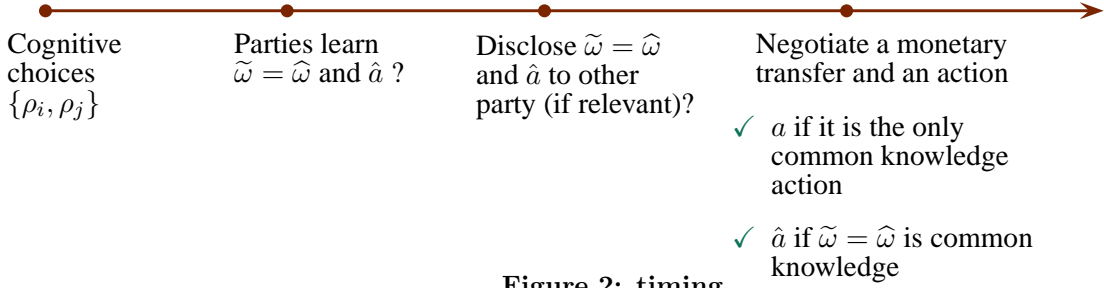- σ_i denotes player i's *relative exposure to the unexpected*:

$$\sigma_i \equiv \left[U_i - \left(\widehat{U}_i - \delta_i \right) \right] - w_i \left[U - \left(\widehat{U} - \delta \right) \right],$$

and so

$$\Sigma_i \sigma_i = 0.$$

Intuitively, player i loses (or gains) gross surplus $U_i - \left(\widehat{U}_i - \delta_i \right)$ when a is selected and the realized state is $\widehat{\omega}$ rather than ω; from an ex-ante viewpoint, his internalization of the anticipated total loss (or gain) depends on this bargaining power w_i.

The timing is summarized in Figure 2.

Cognitive choices $\{\rho_i, \rho_j\}$	Parties learn $\widetilde{\omega} = \widehat{\omega}$ and $\hat{a}$?	Disclose $\widetilde{\omega} = \widehat{\omega}$ and $\hat{a}$ to other party (if relevant)?	Negotiate a monetary transfer and an action

✓ a if it is the only common knowledge action

✓ $\hat{a}$ if $\widetilde{\omega} = \widehat{\omega}$ is common knowledge

Figure 2: timing

Example 1: the buyer-seller game.

Consider the celebrated buyer-seller paradigm. The seller's cost of supplying the buyer is known and equal to c. In the paradigm's "symmetric version", the buyer's gross surplus is B if the design matches the state of nature, but only $b < B$ if design a is chosen in state $\widehat{\omega}$. If design a is chosen and state $\widehat{\omega}$ is revealed, the buyer can enjoy full surplus B only if the seller incurs some adjustment cost $\alpha \geq 0$. Figures 3 and 4 summarize the gross payoffs of the buyer and seller, respectively.

(a) *No renegotiation (or large ex-post adjustment cost $\alpha \geq B - b$)*

	ω	$\widehat{\omega}$
a	$B, -c$	$b, -c$
$\hat{a}$		$B, -c$

**Figure 3: Buyer-seller game
in absence of renegotiation**

Then

$$U = \widehat{U} = B - c \quad ; \quad \delta_B = \delta = B - b \quad \text{and} \quad \delta_S = 0$$

and

$$\sigma_B = \delta_B - w_B \delta = w_S \delta = w_S(B - b).$$

(b) *Ex-post renegotiation* (adjustment cost $\alpha < B - b$)

	ω	$\widehat{\omega}$
a	$B, -c$	$b + w_B(B - b - \alpha),$ $-c + w_S(B - b - \alpha)$
$\hat{a}$		$B, -c$

Figure 4: Buyer-seller game with renegotiation

Then

$$\delta = \alpha \qquad \text{and} \qquad \sigma_B = w_S(B - b).$$

This model allows one to get at the notion of contract incompleteness in an otherwise familiar environment. "Contract incompleteness" is then related to the amount of pre-contractual cognition and is measured by the probability that the wrong design is adopted in the bad state of nature; this is a rather natural definition, especially when some adjustment/renegotiation occurs in that configuration.

Figures 3 and 4 describe *symmetric* versions of the buyer-seller game: The players' payoffs are state independent provided that the action matches the state of nature. More generally, one can allow these payoffs to depend on the state of nature; a case in point is the shrouded attributes model, to which we now turn.

Example 2: The generalized Gabaix-Laibson (2006) shrouded attributes model

Gabaix and Laibson (GL) analyze a seller's incentive to disclose to a buyer the possibility that the satisfactory consumption of a "basic good" requires access to an "unanticipated" add-on also controlled by the seller. Although their model is phrased in terms of a boundedly rational behavior, Gabaix and Laibson's key insights can be illustrated in our framework. Furthermore, their model can be extended to allow for pre-contracting cognition.

In the GL model, the seller sells a basic good to a buyer; this basic good's unit production cost is denoted c. An add-on may or may not be needed to be able to enjoy the basic good. The prior probability that an add-on is needed is denoted $\hat{q}$. If needed, the unit cost of the add-on is $\hat{c}$; thus the add-on is bad news (it involves an extra cost). The buyer knows neither the state of nature nor the nature of the add-on. By contrast, the seller is perfectly informed. That is, GL assumes exogenous one-sided cognition, with $\rho_S = 1$ and $\rho_B = 0$, and focus on the disclosure decision.

The timing is described in Figure 5. Note that we assume that in the bad state of nature (the add-on is needed) the basic good brings no value unless combined with the add-on. The motivation behind a random willingness to pay for the buyer is that it generates a downward sloping demand (and hence a monopoly distortion) without introducing adverse selection on the buyer side at the ex-ante contracting stage. The lack of ex-ante adverse selection implies that if
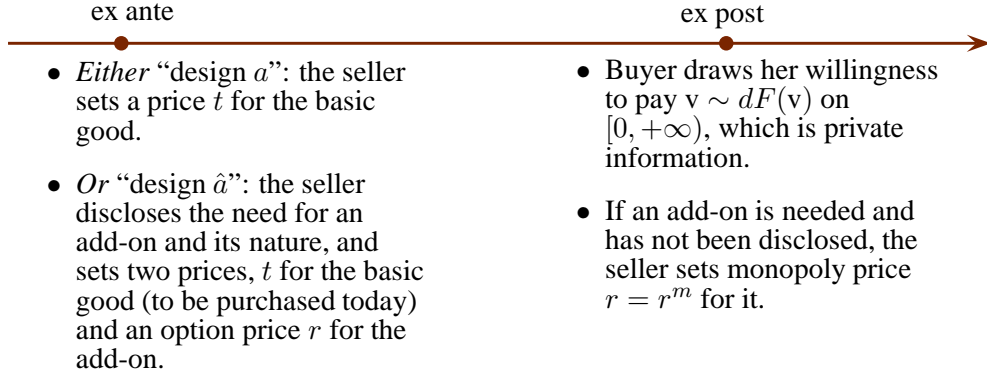
Figure 5: Timing in the shrouded-attributes model

the add-on is disclosed, then it is optimal for the seller to price it at marginal cost ($r = \hat{c}$), and so there is then no distortion. Let $\mathcal{S}(r) = \int_r^\infty (\mathrm{v} - r)dF(\mathrm{v})$ denote the buyer's net surplus.[29]

GL's model is a buyer-seller game, asymmetric as long as $\hat{c} > 0$, as then the payoffs are not the same in both states of nature for the appropriate design:

	ω	$\widehat{\omega}$
a	$\int_0^\infty \mathrm{v}dF(\mathrm{v}),\ -c$	$\int_{r^m}^\infty (\mathrm{v} - r^m)dF(\mathrm{v}),$ $-c + (r^m - \hat{c})[1 - F(r^m)]$
$\hat{a}$		$\int_{\hat{c}}^\infty (\mathrm{v} - \hat{c})dF(\mathrm{v}),\ -c$

Figure 6: Payoffs in the GL model
(these payoffs do not include the ex-ante transfer)

In the general notation:

- $w_S = 1$ (the seller is a price setter),

- $\delta = \big[\mathcal{S}(\hat{c}) - \mathcal{S}(r^m)\big] - (r^m - \hat{c})\big[1 - F(r^m)\big]$ is the deadweight loss associated with non-disclosure and ex-post monopoly pricing (so the "wrong design" de facto corresponds to a contractual failure in which the seller, by not disclosing, fails to commit to the cost-based add-on price),

- $\sigma_B = \int_0^{r^m} \mathrm{v}\,dF(\mathrm{v}) + \big[1 - F(r^m)\big]r^m$, the buyer's relative exposure to the unexpected, is equal to the buyer's loss of utility between the good state and the bad state under monopoly pricing; this loss is decomposed into foregone consumption ($\mathrm{v} \leq r^m$) and extra payment ($\mathrm{v} > r^m$).

---

[29]This is an ex-post surplus, in that it does not include the purchase price for the basic good.

4.2 One-sided cognition

Suppose, first, that only player i can search for state of nature $\widehat{\omega}$ ("j" will then stand for the other player) and so unilateral and multilateral expectation conformity coincide. For example, only the seller is able to find out whether a design flaw may prevent the buyer from fully enjoying the consumption experience. Conversely, only the buyer may invest in learning whether the design matches her own needs. Let $\rho_i = \rho^*$ in equilibrium. Let us assume for the moment that it is optimal for player i to disclose the need for an add-on when he learns it.[30]

Following a lack of disclosure, the other player, player j, forms posterior beliefs that the state of nature is $\widetilde{\omega} = \widehat{\omega}$:

$$\hat{q}' = \frac{\hat{q}(1-\rho^*)}{1-\hat{q}\rho^*} \leq \hat{q}.$$

We assume "passive beliefs" in the bargaining process: Player j sticks to equilibrium posterior beliefs $\hat{q}'$ and so demands a share w_j of the corresponding expected surplus. Such passive beliefs seem reasonable given that on the equilibrium path player i discloses $\widehat{\omega}$ when he learns it.

4.2.1 Disclosure decision

In this subsection we take the cognitive effort ρ^* and thus posterior beliefs $\hat{q}'$ as exogenous, and we investigate player i's incentive to disclose. Suppose player i learns that $\widetilde{\omega} = \widehat{\omega}$. By disclosing this information, he obtains $w_i\widehat{U}$, that is share w_i of the total surplus $\widehat{U}$. By not revealing his information, he obtains

$$\left[\widehat{U}_i - \delta_i\right] + t_i$$

where t_i is the monetary transfer to i when agreeing on design a ($t_i + t_j = 0$). Given posterior beliefs $(q', \hat{q}')$, this transfer is given by the equalization of each player's expected utility with his due share of total surplus:

$$q'U_i + \hat{q}'\big(\widehat{U}_i - \delta_i\big) + t_i = w_i\Big[q'U + \hat{q}'\big(\widehat{U} - \delta\big)\Big].$$

After some manipulations, player i discloses if and only if:

$$w_i\delta \geq q'\sigma_j \tag{2}$$

where

$$\sigma_j \equiv \left[U_j - \big(\widehat{U}_j - \delta_j\big)\right] - w_j\left[U - \big(\widehat{U} - \delta\big)\right],$$

is player j's relative exposure to the unexpected.

Condition (2) says that player i is willing to disclose state $\widehat{\omega}$ whenever his share of the deadweight loss, $w_i\delta$, exceeds the cross-subsidy embodied in the transfer, namely $q'\sigma_j$. This cross-subsidy is proportional to the other party's posterior probability of the erroneous state ω, q', that player j faces when not informed that the true state of nature is $\widehat{\omega}$.

---

[30]There are two possible motivations for player i to acquire information. The first (on which we focus here) is that it leads to communication and a different design/contract. The second is to decide whether to interact at all with player j (this alternative motivation can be ruled out if there is enough surplus).

If (2) is violated, then disclosure with probability 1 is not an equilibrium. If $q\sigma_j \geq w_i\delta$, then the (unique) equilibrium has no disclosure at all. If $q\sigma_j < w_i\delta < q'\sigma_j$, then the (unique) equilibrium has player i randomize between disclosing and not disclosing; posterior beliefs that the state is $\widetilde{\omega} = \omega$ in the absence of disclosure are then q'' such that $q''\sigma_j = w_i\delta$.

Proposition 7 (**incentive to disclose prior to contracting**). *Player i's willingness to disclose is driven by two factors: his share, w_i, of the resulting deadweight loss, δ, and player j's relative exposure to the unexpected, $\sigma_j = \left[U_j - \left(\widehat{U}_j - \delta_j\right)\right] - w_j\left[U - \left(\widehat{U} - \delta\right)\right]$, times the extent $q' \geq q$ of player's j misperception in the absence of disclosure (when disclosure occurs with probability 1).*

(i) Player i discloses if $w_i\delta \geq q'\sigma_j$

(ii) Player i does not disclose if $w_i\delta \leq q\sigma_j$

(iii) Player i plays a unique mixed disclosure strategy if $q\sigma_j < w_i\delta < q'\sigma_j$.

4.2.2   Applications

• *Price setting.*

In some applications, the informed player is a price setting seller, and so $w_i = 1$. Formula (2) then simplifies to:

$$\delta \geq q'\left[U_j - (\widehat{U}_j - \delta_j)\right].$$

The right-hand side of this inequality is equal to the posterior probability of state ω times player j's "disappointment" or loss of utility when the state turns out to be $\widehat{\omega}$ rather than ω.

• *Symmetric buyer-seller game.*

In the symmetric buyer-seller game , $\sigma_B \equiv w_S(B - b)$. So, in the absence of renegotiation (2) boils down to:

$$\begin{cases} w_B \geq -q'w_S & \text{if} \quad i = B \\ w_S \geq q'w_S & \text{if} \quad i = S \end{cases}$$

Player i, whether he is the buyer or the seller, always discloses.

In the presence of renegotiation, the buyer always discloses, but the seller discloses only if the deadweight loss (then equal to the adjustment cost α) is large enough: $\alpha \geq q'(B - b)$.

• *Gabaix-Laibson model.*

The general formula implies that the seller in the GL model opts for shrouded attributes (does not disclose the existence of the add-on when one is needed) if and only if $w_S\delta \leq q\sigma_B$, or, applying the specific expressions for these variables:

$$\int_{\hat{c}}^{r^m} (\mathrm{v} - \hat{c})dF(\mathrm{v}) \leq q\left[\int_0^{r^m} \mathrm{v}dF(\mathrm{v}) + \left[1 - F(r^m)\right]r^m\right]. \tag{3}$$

The left-hand side of this inequality is equal to the deadweight loss associated with non-disclosure and the concomitant monopoly pricing; this loss is entirely borne by the seller who has full

bargaining power as a price setter. The right-hand side is the product of the posterior probability of the good state (which is equal to the prior probability in a no-disclosure equilibrium) and the buyer's loss of utility between state ω and state $\widehat{\omega}$ under monopoly pricing; this loss is decomposed into foregone consumption ($v \leq r^m$) and extra payment ($v > r^m$).

4.2.3 Choice of cognition

Next, we can analyze player i's choice of cognition, assuming disclosure upon learning that the state is $\widehat{\omega}$.[31] Letting

$$t_i(\rho^*) \equiv w_i U - U_i + \hat{q}'(\rho^*)\sigma_i$$

(where the posterior belief $\hat{q}'$ is decreasing in ρ^*), player i's utility is:

$$
\begin{aligned}
\mathcal{U}_i(\rho^*) \equiv \max_{\{\rho_i\}}\Big\{ &\hat{q}\Big[\rho_i\big(w_i\widehat{U}\big) + (1 - \rho_i)\big(U_i - \delta_i + t_i(\rho^*)\big)\Big] \\
&+ q\big[U_i + t_i(\rho^*)\big] - C_i(\rho_i)\Big\}
\end{aligned}
\tag{4}
$$

yielding in equilibrium $\rho_i = \rho^*$, where

$$C_i'(\rho^*) = \hat{q}\Big[w_i\widehat{U} - \big(\widehat{U}_i - \delta_i + t_i\big)\Big]$$

or

$$C_i'(\rho^*) = \hat{q}\Big[w_i\delta - q'(\rho^*)\sigma_j\Big], \tag{5}$$

whenever the right-hand side is positive. Unsurprisingly, this right-hand side is positive if and only if (2) is satisfied: The stake in acquiring information is positive if and only if disclosure brings a benefit to player i. Note, furthermore, that ρ^* is unique if $\sigma_j \geq 0$.

Cognitive traps. How fine the information structure is here is captured by the search effort ρ_i. Thus consider two levels $\widehat{\rho}_i > \rho_i$. Simple computations show that

$$\Gamma^{EC}(\rho_i, \widehat{\rho}_i) = q\hat{q}(\widehat{\rho}_i - \rho_i)\left[\frac{1}{1 - \hat{q}\widehat{\rho}_i} - \frac{1}{1 - \hat{q}\rho_i}\right]\sigma_i$$

Thus expectation conformity is satisfied provided that $\sigma_i \geq 0$. The higher the anticipated cognition effort, the higher the probability of the bad state in the absence of disclosure and so the less favorable the agreement to player i whenever the latter is relatively exposed to the unexpected. This raises player i's stake in information acquisition.

Proposition 8 (necessary and sufficient condition for expectation conformity).
Consider one-sided cognition in the incomplete contract game.
(i) Cognition level: Player i acquires information and discloses for sure if and only if (2) is satisfied for equilibrium cognition effort ρ^*.
(ii) Expectation conformity. Player j's anticipation of a higher cognitive effort by player i

---

[31]That is $w_i\delta \geq q'(\rho^*)\sigma_j$. Note that the mixed-strategy region exhibited in (iii) of Proposition 7 can exist only if information is exogenous. Player i will put zero effort in acquiring information if one of his optimal ex-post strategies is not to disclose.

increases the latter's stake in cognition (raising the possibility of multiple equilibria) if and only if player i is relatively exposed to the unexpected ($\sigma_i > 0$) .

(iii) Cognitive traps: In case of multiple equilibria, player i is better off in a low-cognition equilibrium.

To prove the last point in the Proposition, note that

$$sgn\Big(\frac{d\mathcal{U}_i}{d\rho^*}\Big) = sgn\Big(\frac{dt_i}{d\rho^*}\Big) = -sgn(\sigma_i).$$

4.2.4 Strategic delay in disclosure and good faith in bargaining

Until now, a party acquiring information disclosed it before contracting if he disclosed it at all; in the absence of intent of pre-contractual disclosure, he would not even try to acquire this information. Either the parties have a reputation for not coming up with bad surprises just after contracting (before reliance), or a mandatory-disclosure/good-faith-bargaining law can be enforced.[32] Would the analysis be altered if we allowed player i to delay disclosure until after the contract is signed? That is, suppose that after agreeing on design a, player i can disclose that the state of nature is $\widehat{\omega}$ and offer to costlessly renegotiate to design $\widehat{a}$.

In other words, we here study the other polar case of strategic delay, which corresponds to a situation in which such good-faith bargaining laws are unenforceable (it is hard to prove that the party had the information prior to contracting) and reputation concerns are weak. This sub-section can thus be viewed as studying the impact of a mandatory disclosure law.

Upon disclosing immediately after contracting that $\widetilde{\omega} = \widehat{\omega}$, player i receives

$$\Big[\big(U_i - \delta_i\big) + t_i\Big] + w_i\delta.$$

The incentive to disclose then no longer accounts for the realized deadweight loss, as no deadweight loss materializes. It is optimal for player i to disclose prior to contracting if and only if:

$$0 \geq q'\sigma_j,$$

or

$$\sigma_i \geq 0. \tag{6}$$

Thus player i discloses prior to contracting if and only if he is relatively exposed to the unexpected. Note that (6) is just a special case of (2), for deadweight loss $\delta = 0$. If $0 < q'\sigma_j < w_i\delta$, then player i discloses after contracting while he would disclose prior to contracting if delayed disclosure were unfeasible.[33]

---

[32]The notion of mandatory disclosure is a complex one, and has been the object of tensions in contract law for a long time (see Kronman's 1978 seminal paper on the topic). For example, in Macquarie International Health Clinic Pty Ltd v Sydney South West Area Health Service (2010, NSWCA 268), the Court held that the obligation of "good faith" does not require parties to compromise their own commercial interests, but that parties must cooperate, including disclosing information, in a reasonable way to achieve the contract's objectives.

[33]Suppose the absence of disclosure and that parties contract on a. Then the transfer is given by

$$t_i + qU_i + \hat{q}[\widehat{U}_i - \delta_i + \rho^* w_i\delta] = w_i\big[(qU + \hat{q}[\widehat{U} - (1 - \rho^*)\delta]\big].$$

Proposition 9 (absence of good faith bargaining). *Under early renegotiation (the design can be costlessly altered after the contract is signed):*

(i) Player i discloses prior to contracting if and only if he is relatively exposed to the unexpected: $\sigma_i \geq 0$. The equilibrium set is then the same as under mandatory disclosure.

(ii) If $\sigma_i \leq 0$, the equilibrium cognition in the absence of mandatory disclosure is unique and given by

$$C_i'(\rho^*) = \hat{q}[w_i \delta]. \tag{7}$$

It exceeds the (also unique) equilibrium level of cognition under mandatory disclosure. Both parties are better off in the absence of a good faith bargaining requirement.

Part ii) of Proposition 9 fits well with Kronman (1978)'s and Eisenberg (2003)'s informal argument that mandatory-disclosure laws must distinguish between the cases of casually acquired information and information that results from deliberate search (which according to Kronman, must benefit from a legal non-disclosure privilege, in effect a property right). When $\sigma_i < 0$, i.e. when mandatory disclosure matters, mandatory disclosure reduces the incentive of party i to acquire information. In the end, player i receives a share w_i of total surplus, minus the information cost:

$$w_i\big[qU + \hat{q}\widehat{U} - \hat{q}(1 - \rho^*)\delta\big] - C_i(\rho^*),$$

which is maximized at the level given by (7). Player j's welfare,

$$w_j\big[qU + \hat{q}\widehat{U} - \hat{q}(1 - \rho^*)\delta\big]$$

is obviously increasing in ρ^*.

By contrast, if information were *exogenous*, mandatory disclosure would be irrelevant if the contract can be renegotiated before reliance (as in this section), but would improve welfare in the benchmark case in which a deadweight loss δ is actually incurred if no disclosure of $\widetilde{\omega} = \widehat{\omega}$ occurs before contracting.

4.3 Two-sided cognition

4.3.1 Overcognition implies one-sided cognition

The first best level of cognition ρ^{FB} under one-sided cognition is given by

$$C_i'(\rho^{FB}) = \hat{q}\delta.$$

Under good faith bargaining, one-sided cognition by player i results in excessive cognition

---

Agent i strictly prefers not to disclose if

$$t_i + \widehat{U}_i - \delta_i + w_i\delta > w_i\widehat{U} \iff \sigma_i < 0.$$

($\rho^{FB} < \rho^*$, the contract is "too complete") if and only if

$$\delta < w_i \delta - q' \sigma_j$$

or

$$w_j \delta < q' \sigma_i. \tag{8}$$

Note that condition (8) is exactly the condition under which, under player i cognition, player j would not want to acquire information, however small the cost of doing so: see equation (5). The intuition for this result is that excess cognition by player i occurs when the private benefit of information exceeds the social benefit. Put differently, information, at the margin, reduces j's welfare; and so player j has no incentive to acquire this information, however cheap.

A simple corollary is that there can be overcognition only by the player who is relatively exposed to the unexpected (i.e., $\sigma_i > 0$), for example the buyer in the symmetric buyer-seller game.[34]

Proposition 10 (overcognition and two-sided cognition).
Whether disclosure is mandatory or not:
(i) Under one-sided cognition, player i engages in excess cognition if and only if (8) is satisfied.
(ii) Under two-sided cognition, player j does not want to engage in cognition, even when $C_j'(0) = 0$, if and only if (8) is satisfied.

4.3.2 Actual cognition by both players

Next, we look for an equilibrium in which both sides invest in cognition (and disclose, as the two co-vary), which as we saw requires that $w_i \delta + q' \sigma_i > 0$ for all i. Assume that the search outcomes are independent. Such an equilibrium must satisfy:

Privately optimal cognition
Under mandatory disclosure or if $\sigma_i \geq 0$: either $\rho_i = 0$ if $w_i \delta \leq q' \sigma_j$ or

$$C_i'(\rho_i) = \hat{q}(1 - \rho_j)(w_i \delta - q' \sigma_j) \tag{9}$$

In the absence of mandatory disclosure and if $\sigma_i < 0$:

$$C_i'(\rho_i) = \hat{q}(1 - \rho_j) w_i \delta. \tag{10}$$

Bayes rule

$$q' = \frac{q}{q + \hat{q}(1 - \rho_i)(1 - \rho_j)}. \tag{11}$$

---

[34]Condition (8) applied to the buyer takes the following form:
- in the absence of ex-post renegotiation: $1 < q'$, which is impossible, so that there is never overcognition;
- under ex-post renegotiation: $\alpha < q'(B - b)$, which is satisfied if the adjustment cost α is small enough. The buyer then engages in overcognition and the seller does not exert any cognitive effort at all.

Rewrite the first-order condition (9) as:

$$C_i'(\rho_i) = \hat{q}\left[(1-\rho_j)w_i\delta + (-\sigma_j)\frac{\hat{q}}{\dfrac{q}{1-\rho_j} + \hat{q}(1-\rho_i)}\right].$$

If $\sigma_j \leq 0$, player i's reaction curve (ρ_i as a function of ρ_j) is obviously downward-sloping. But strategic substitutability holds also if $\sigma_j > 0$: the derivative of the RHS with respect to ρ_j is $\hat{q}\left[-w_i\delta + \sigma_j(q')^2\right] < 0$, from the active cognition condition.

We will assume that the functions C_i are sufficiently convex so that the reaction curves cut only once, defining a stable equilibrium.

Proposition 11 (actual two-sided cognition). *Assume independent searches and a stable equilibrium. Then, in the two-sided cognition region, cognitive efforts are locally strategic substitutes, reflecting the public good nature of information. Suppose that player i is relatively more exposed to the unexpected ($\sigma_i > 0$); then lifting the mandatory disclosure requirement increases j's cognition and reduces i's cognition.*

4.4 A graphical summary

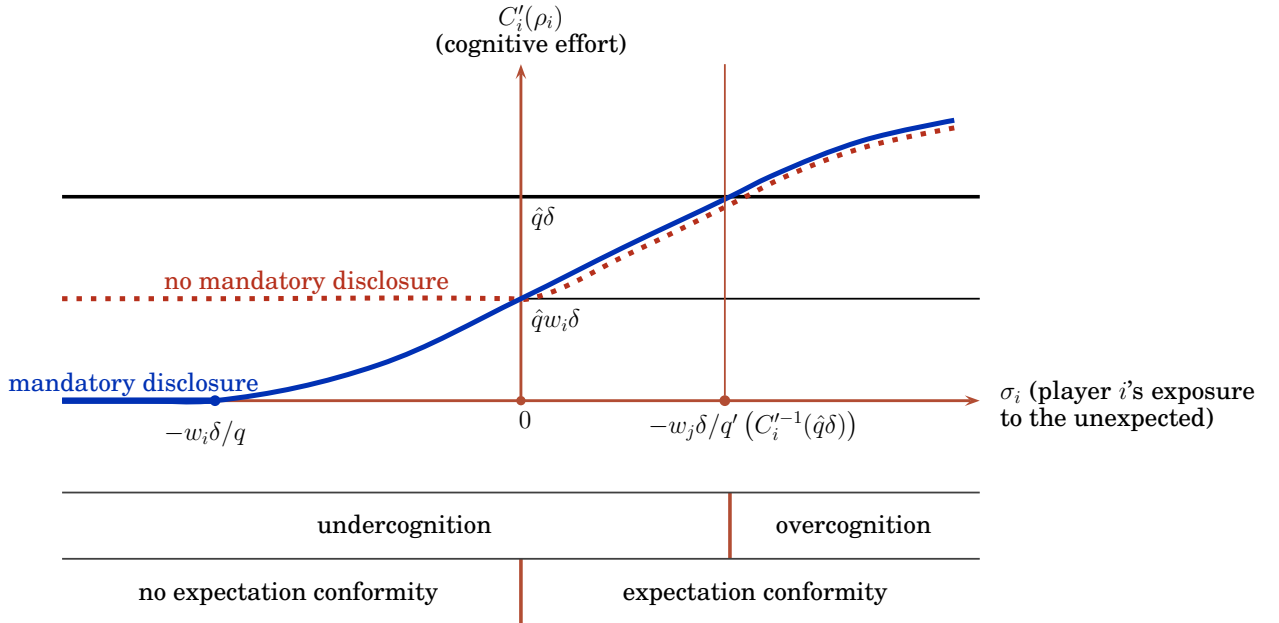Figure 7 captures some of the main insights of this section.

Figure 7: Comparative statics
[(1) Vary σ_i (e.g., by changing δ_i), keeping $w_i\delta$ constant; (2) ignore cognitive traps: C_i'' large enough.]

5 Signal-jamming cognitive games

So far we have presumed that players choose their *own* information structure. In a number of economic games, though, players choose *their opponents'* information structure. Such signal

jamming 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.

5.1 Defining expectation conformity in signal-jamming cognition games

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 $q(s_j|\omega)$ over the signal s_j received by player j in state of nature ω. Player j then plays a stage-2 (mixed) strategy $\alpha_j(s_j)$. The overall strategy under $\mathcal{F}_j$ is then an "$\mathcal{F}_j$-measurable" strategy $\sigma_j^{\mathcal{F}_j}$, defined by:

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

Let $\{\mathcal{F}_i, \mathcal{F}_j\}$ denote a common-knowledge choice of information structures and $\{\alpha_i, \alpha_j\}$ denote the corresponding equilibrium strategies.

Let

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

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

$$V_i\big(\widehat{\mathcal{F}}_j\,;\mathcal{F}_i\,,\mathcal{F}_j\big) - V_i\big(\mathcal{F}_j\,;\mathcal{F}_i\,,\mathcal{F}_j\big) \le V_i\big(\widehat{\mathcal{F}}_j\,;\widehat{\mathcal{F}}_i\,,\widehat{\mathcal{F}}_j\big) - V_i\big(\mathcal{F}_j\,;\widehat{\mathcal{F}}_i\,,\widehat{\mathcal{F}}_j\big). \tag{12}$$

Definition 6 (**expectation conformity under signal jamming**). $EC_{\{\mathcal{F},\widehat{\mathcal{F}}_j\}}$ is satisfied if (12) is satisfied for $\mathcal{F}_i = \widehat{\mathcal{F}}_i$. Expectation conformity is satisfied if for all i, $\mathcal{F}$ and $\widehat{\mathcal{F}}_j$, $EC_{\{\mathcal{F},\widehat{\mathcal{F}}_j\}}$ is satisfied.

5.2 Examples of signal-jamming games satisfying expectation conformity

One-sided[35] signal-jamming environments (described rather informally below) exhibiting expectation conformity include:

(a) *Imperfect communication.* Consider the trading game when the seller with strictly positive probability knows the buyer's willingness to pay. For simplicity, suppose that $\widetilde{\omega} \in \{\omega, \widehat{\omega}\}$, that $\widetilde{\omega}$ 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 $\widetilde{\omega}$: information is "semi-hard" in that it can be disclosed, but the amount of disclosure depends on the seller's communication effort.[36] The seller's effort (understood as

[35]That is, only one player, player i, manipulates the other player's information structure: $\widehat{\mathcal{F}}_i = \mathcal{F}_i$ in condition (12). Again, multilateral and unilateral expectation conformity coincide in such environments.

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

the effort incurred prior to actual communication with the buyer) is unobserved by the buyer. Clearly, the seller exerts no effort if $\widetilde{\omega} = \widehat{\omega}$. By contrast, convincing the buyer that $\widetilde{\omega} = \omega$ is profitable.

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 $\widetilde{\omega} = \omega$, 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 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.[37] 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.

6 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. They also manipulate other agents' information by jamming the latters' signals. 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 distinguished between unilateral and multilateral expectation conformity, where the latter may be driven, depending on the game, by unilateral expectation conformity, positive strategic interaction, or a mixture of the two. We first obtained

---

[37]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.

a sufficient condition for multilateral expectation conformity.

The paper then showed that games of pure conflict (zero-sum games) never give rise to self-fulfilling cognition while games of pure alignment (coordination games) always do. Second, it considered environments with a game setter, in which a player picks an information structure and a game to be played. A characterization of the expectation conformity property in terms of rotation points can be obtained for this class of games, which comprises many games of interest to economists such as the cognition-augmented lemons model.

The paper then turned to cognition-intensive contracting and showed that a single variable, the "relative exposure to the unexpected", underlies a variety of concepts such as expectation conformity, over-cognition and the desirability of mandatory disclosure laws. Finally, the paper extended the notion of expectation conformity to (signal-jamming) cognitive games in which players choose their rivals' information structure.

Because of their importance for economics, cognitive games need to be better understood and there are multiple alleys for future research. For instance, we have assumed that cognition is unobservable; one would like to investigate how its equilibrium level is affected by the ability, if any, to disclose to other players its intensity. Relatedly, cognition often occurs in multiple stages as an extensive form game unfolds.[38] Multi-stage cognition offers new features; players may learn progressively about their rivals' choice of cognitive strategies; furthermore, the possibility of cognitive traps suggests that players in a variety of environments will want to develop a reputation for being cognitively limited or overloaded with work.

We noted that multilateral expectation conformity results from a combination of effects: increasing differences (the standard form of strategic complementarity in information structure choices, when these are publicly observable), unilateral expectation conformity (a player's incentive to conform to his rivals' expectation) and the impact of the rivals' information on a player's preferred perception of his information by the rivals. While some of these effects but not others are at play in each environment, a better understanding of their relative importance and of when they are likely to hold would bring a deeper understanding of cognitive games and cognitive traps.

---

[38]The framework of this paper accommodates this possibility (player i's filtration can describe a history-dependent cognitive strategy), but we must then assume "passive beliefs": i's beliefs about $\mathcal{F}_j$ do not change while observing j's behavior.

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.

Athey, S., Milgrom, P., and J. Roberts (1998) "Robust Comparative Statics," book draft.

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.

Bénabou, R., and J. Tirole (2011) "Identity, Morals and Taboos: Beliefs as Assets," *Quarterly Journal of Economics*, 126(2): 805–855.

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

Bourjade, S., and B. Jullien (2011) "The Roles of Reputation and Transparency on the Behavior of Biased Experts," *Rand Journal of Economics*, 42: 575–594.

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

Carrillo, J.D., and T. Mariotti (2000) "Strategic Ignorance as a Self-Disciplining Device," *Review of Economic Studies*, 67(3): 529–544.

Chakravary, S., and B.W. MacLeod (2009) "Optimal Contracting in the Shadow of the Law," *Rand Journal of Economics*, 40(3): 533–557.

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.

Crawford, V., and J. Sobel (1982) "Strategic Information Transmission," *Econometrica*, 50: 1431–51.

Crémer, J., Khalil, F., and J.C. Rochet (1998) "Strategic Information Gathering before a Contract Is Offered," *Journal of Economic Theory*, 81(1): 163–200.

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

Dekel, E., Lipman, B. L., and A. Rustichini (1998) "Standard State-Space Models Preclude Unawareness," *Econometrica*, 66(1): 159–173.

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.

Dziuda, W. (2011) "Strategic Argumentation," *Journal of Economic Theory*, 146(4): 1362–1397.

Eisenberg, M. (2003) "Disclosure in Contract Law," *California Law Review*, 91: 1645–1691.

Farhi, E., and J. Tirole (2013) "Liquid Bundles," forthcoming, *Journal of Economic Theory*.

Farhi, E., Lerner, J. and J. Tirole (2013) "Fear of Rejection? Tiered Certification and Transparency," *Rand Journal of Economics*, 44(4): 610–631.

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

Fishman, M.J., and J.A. Parker (2010) " Valuation, Adverse Selection, and Market Collapses," NBER Working Paper 18358.

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.

Gabaix, X., and D. Laibson (2006) "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 121(2): 505–40.

Gentzkow, M., and E. Kamenica (2012) "Disclosure of Endogenous Information," mimeo, University of Chicago.

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.

Green, J. R., and J.-J. Laffont (1986) "Partially Verifiable Information and Mechanism Design," *Review of Economic Studies*, 53, 447–56.

Grossman, S.J. (1981) "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, 24: 461– 483.

Grubb, M. (2011) "Developing a Reputation for Reticence," *Journal of Economics, Management and Strategy*, 20: 225–268.

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

Hirshleifer, J. (1971) "The Private and Social Value of Information and the Reward to Inventive Activity," *American Economic Review* , 61(4): 561–574.

Hoffmann, F., Inderst, R., and M. Ottaviani (2014) "Persuasion through Selective Disclosure: Implications for Marketing, Campaigning, and Privacy Regulation," mimeo, Goethe University Frankfurt, Bocconi University and IGIER.

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

Ispano, A. (2012) "Information Acquisition and Voluntary Release of Bad News," mimeo, TSE.

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

Kamenica, E. (2008) "Contextual Inference in Markets: On the Informational Content of Product Lines," *American Economic Review*, 98(5): 2127–49.

Kartik, N. (2009) "Strategic Communication with Lying Costs," *Review of Economic Studies*, 76(4): 1359–1395.

Kartik, N., Lee, F.X. and W. Suen (2014) "Does Competition Promote Disclosure?", mimeo.

Kronman, A. (1978) "Mistake, Disclosure, Information and the Law of Contracts;" *Journal of Legal Studies*, 7: 1–34.

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

Lester, B., Postlewaite, A., and R. Wright (2012) "Information, Liquidity, Asset Prices, and Monetary Policy," *Review of Economic Studies* , 79 (3): 1209–1238.

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

Macauley, S. (1963)"Non Contractual Relations in Business: A Preliminary Study," *American Sociological Review*, 28: 55–69.

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," mimeo.

Matthews, S., and A. Postlewaite (1985) "Quality Testing and Disclosure," *Rand Journal of Economics*, 16: 328–340.

Milgrom, P. (1981) "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12: 380–391.

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

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

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

Okuno-Fujiwara, M., Postlewaite, A., and K. Suzumura (1990) "Strategic Information Revelation," *Review of Economic Studies*, 57(1): 25–47.

Pavan, A. (2014) "Attention, Coordination, and Bounded Recall," mimeo, Northwestern University.

Pei, D. (2013) "What Do You Want To Know? Communication with Information Acquisition," mimeo TSE.

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

Reis, R. (2006) "Inattentive Producers," *Review of Economic Studies*, 73(3): 793–821.

Shavell, S. (1994) "Acquisition and Disclosure of Information Prior to Sale," *Rand Journal of Economics*, 25(1): 20–36.

Shin, H.S. (2003) "Disclosures and Asset Returns," *Econometrica*, 71(1): 105–133.

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.

Von Thadden, E.-L., and X. Zhao (2012) "Incentives for Unaware Agents," *Review of Economic Studies*, 79 (3): 1151–1174.

Zhao, X. (2014) "Strategic Mis-Selling and Pre-Contractual Cognition," mimeo. Available at SSRN: http://ssrn.com/abstract=2131315 or http://dx.doi.org/10.2139/ssrn.2131315

Appendix.   Proof of Proposition 4

Define $\omega_i \equiv E(\omega|\mathcal{F}_i)$ and $\sigma_i^2 \equiv E((\omega - \omega_i)^2|\mathcal{F}_i)$.

For any $\mathcal{F} = (\mathcal{F}_i, \mathcal{F}_j)$ such that $\mathcal{F}_j \subseteq \mathcal{F}_i$, best responses of the stage-2 game are:

$$a_j \equiv a_j(\mathcal{F}) = \frac{E(a_i(\mathcal{F})|\mathcal{F}_j) + \omega_j}{2} \quad \text{and} \quad a_i \equiv a_i(\mathcal{F}) = \frac{a_j + \omega_i}{2}$$

hence

$$a_j = \omega_j \quad \text{and} \quad a_i = \frac{\omega_j + \omega_i}{2}$$

Expected gross payoffs conditional on $\mathcal{F}_i$, resp. $\mathcal{F}_j$, are:

$$\frac{-(\omega_j - \omega_i)^2}{2} - \sigma_i^2 \quad \text{and} \quad U_j(\sigma_i, \sigma_j, \mathcal{F}_j) = -\frac{E((\omega_j - \omega_i)^2)|\mathcal{F}_j)}{4} - \sigma_j^2$$

while ex-ante gross payoffs, defined as in the general model, are:

$$-E_\omega\left[\frac{(\omega_j - \omega_i)^2}{2} + \sigma_i^2\right] \quad \text{for player } i \quad \text{and} \quad - E_\omega\left[\frac{(\omega_j - \omega_i)^2)}{4} + \sigma_j^2\right] \quad \text{for player } j.$$

Consider two information structures $\mathcal{F}$ and $\widehat{\mathcal{F}}$. Let

$$\Delta_i \equiv V_i\big(\widehat{\mathcal{F}}_i; \mathcal{F}_i, \mathcal{F}_j\big) - V_i\big(\mathcal{F}_i; \mathcal{F}_i, \mathcal{F}_j\big)$$

and

$$\widehat{\Delta_i} \equiv V_i\big(\widehat{\mathcal{F}}_i; \widehat{\mathcal{F}}_i, \widehat{\mathcal{F}}_j\big) - V_i\big(\mathcal{F}_i; \widehat{\mathcal{F}}_i, \widehat{\mathcal{F}}_j\big).$$

Case 1 : $\mathcal{F}_1 \subseteq \widehat{\mathcal{F}}_1 \subseteq \mathcal{F}_2 \subseteq \widehat{\mathcal{F}}_2$

$$\Delta_2 = E\left[\sigma_2^2 - \hat{\sigma}_2^2 - \frac{1}{2}\left[(\omega_1 - \widehat{\omega}_2)^2 - (\omega_1 - \omega_2)^2\right]\right] = E\left[\sigma_2^2 - \widehat{\sigma}_2^2 - \frac{(\widehat{\omega}_2 - \omega_2)^2}{2}\right]$$

$$\widehat{\Delta}_2 = E\left[\sigma_2^2 - \widehat{\sigma}_2^2 - \frac{1}{2}\left[(\widehat{\omega}_1 - \widehat{\omega}_2)^2 - (\widehat{\omega}_1 - \omega_2)^2\right]\right]$$

$$\widehat{\Delta}_2 - \Delta_2 = 0$$

Intuitively, the coordination ability is the same for player 2 regardless of whether his information structure is $\mathcal{F}_2$ or $\widehat{\mathcal{F}}_2$. The only gain from being better informed comes from a better adjustment to the state of nature and is independent of player 1's information structure.

$$\Delta_1 = E\left[\sigma_1^2 - \hat{\sigma}_1^2 - \frac{1}{4}\left[(\omega_1 + \omega_2 - 2\widehat{\omega}_1)^2 - (\omega_1 - \omega_2)^2\right]\right] = E\left[\sigma_1^2 - \hat{\sigma}_1^2\right]$$

$$\widehat{\Delta}_1 = E\left[\sigma_1^2 - \hat{\sigma}_1^2 - \frac{1}{4}\left[(\widehat{\omega}_1 - \widehat{\omega}_2)^2 - (\widehat{\omega}_1 + \widehat{\omega}_2 - 2\omega_1)^2\right]\right] = E\left[\sigma_1^2 - \hat{\sigma}_1^2 + (\widehat{\omega}_1 - \omega_1)^2\right]$$

$$\widehat{\Delta}_1 - \Delta_1 = E(\omega_1 - \widehat{\omega}_1)^2$$

Case 2: $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \widehat{\mathcal{F}}_1 \subseteq \widehat{\mathcal{F}}_2$

$$\Delta_2 = E\left[\sigma_2^2 - \hat{\sigma}_2^2 - \frac{(\widehat{\omega}_2 - \omega_2)^2}{2}\right]$$

$$\widehat{\Delta}_2 = E\left[\sigma_2^2 - \hat{\sigma}_2^2 - \frac{1}{2}\left[(\widehat{\omega}_2 - \widehat{\omega}_1)^2 - 2(\omega_2 - \widehat{\omega}_1)^2\right]\right]$$

$$\widehat{\Delta}_2 - \Delta_2 = \frac{3}{2}E(\omega_2 - \widehat{\omega}_1)^2$$

$$\Delta_1 = E\left[\sigma_1^2 - \hat{\sigma}_1^2 + \frac{1}{4}(\omega_1 - \omega_2)^2 - \frac{1}{2}\left(\widehat{\omega}_1 - \frac{\widehat{\omega}_1 + \omega_2}{2}\right)^2\right]$$

$$\widehat{\Delta}_1 = E\left[\sigma_1^2 - \hat{\sigma}_1^2\right] + E\left(\widehat{\omega}_1 - \omega_1\right)^2$$

$$\widehat{\Delta}_1 - \Delta_1 = \frac{3}{4}E\left(\omega_1 - \omega_2\right)^2 + \frac{1}{2}E\left(\widehat{\omega}_1 - \frac{\omega_1 + \omega_2}{2}\right)^2 + E\left(\widehat{\omega}_1 - \omega_2\right)^2$$

Case 3: $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \widehat{\mathcal{F}}_2 \subseteq \widehat{\mathcal{F}}_1$

$$\Delta_2 = E\left[\sigma_2^2 - \hat{\sigma}_2^2 - \frac{(\widehat{\omega}_2 - \omega_2)^2}{2}\right]$$

$$\widehat{\Delta}_2 = E\left[\sigma_2^2 - \hat{\sigma}_2^2 + (\widehat{\omega}_2 - \omega_2)^2\right]$$

$$\widehat{\Delta}_2 - \Delta_2 = \frac{3}{2}E\left(\widehat{\omega}_2 - \omega_2\right)^2$$

$$\Delta_1 = E\left[\sigma_1^2 - \hat{\sigma}_1^2 + \frac{1}{4}(\omega_1 - \omega_2)^2 - \frac{1}{2}\left(\widehat{\omega}_1 - \frac{\omega_1 + \omega_2}{2}\right)^2\right]$$

$$\widehat{\Delta}_1 = E\left[\sigma_1^2 - \hat{\sigma}_1^2 + (\omega_1 - \widehat{\omega}_2)^2 - \frac{1}{2}\left(\widehat{\omega}_1 - \widehat{\omega}_2\right)^2\right]$$

$$\widehat{\Delta}_1 - \Delta_1 = E\left(\widehat{\omega}_2 - \omega_2\right)^2 + \frac{3}{4}E\left(\omega_2 - \omega_1\right)^2 + \frac{1}{2}E\left(\widehat{\omega}_2 - \frac{\omega_1 + \omega_2}{2}\right)^2$$