A Theory of Deception*

David Ettinger† and Philippe Jehiel‡

5th January 2009

Abstract

This paper proposes an equilibrium approach to belief manipulation and deception in which agents only have coarse knowledge of their opponent's strategy. Equilibrium requires the coarse knowledge available to agents to be correct, and the inferences and optimizations to be made on the basis of the simplest theories compatible with the available knowledge. The approach can be viewed as formalizing into a game theoretic setting a well documented bias in social psychology, the Fundamental Attribution Error. It is applied to a bargaining problem, thereby revealing a deceptive tactic that is hard to explain in the full rationality paradigm.

Deception and belief manipulation are key aspects of many strategic interactions, including bargaining, poker games, military operations, politics and investment banking. Anecdotal evidences of belief manipulation and deception are very numerous, and Michael Lewis's (1990) best-seller "Liar's Poker" reports colorful illustrations of such strategic behaviors in the world of investment banking in the late 1980s. For example, Lewis explains how "he spent most of his working life inventing logical lies" that worked amazingly well (thanks to the logical appearance, see Lewis (1990) page 186). From the viewpoint of game theory,

*We would like to thank the editor and the referee for useful comments. We also thank K. Binmore, D. Fudenberg, D. Laibson, A. Newman, A. Rubinstein, the participants at ESSET 2004, Games 2004, ECCE 1, THEMA, Berkeley, Caltech, Institute for Advanced Study Jerusalem, the Harvard Behavioral/experimental seminar, Bonn University, the Game Theory Festival at Stony Brook 2005, and the conference in honor of Ken Binmore UCL 2005, for helpful comments. We are grateful to E. Kamenica for pointing out the literature on the Fundamental Attribution Error.

†Université de Cergy-Pontoise, THEMA, F-95000 Cergy-Pontoise, France

‡PSE and UCL; jehiel@enpc.fr

belief manipulation and deception are delicate to capture because traditional equilibrium approaches assume that players fully understand the strategy of their opponents.[1] We depart from this tradition by assuming that players may have a *partial* rather than *total* understanding of the strategy of their opponents. This in turn allows us to propose an equilibrium approach to deception, where deception is defined to be the process by which actions are chosen to manipulate beliefs so as to take advantage of the erroneous inferences.[2]

To illustrate the phenomenon of deception, we will consider and formalize the following bargaining situation. The owner of a house, Mrs A, wishes to sell her good at some price considered to be high (say above the market price as perceived by real estate agents). A potential buyer, Mr B, comes in. Mr B will accept paying the high price if he is afraid enough that another buyer may be interested in the house. Otherwise, he will prefer to continue bargaining in the hope of getting a lower price. The owner, Mrs A, after mentioning some slight problems with the heating system (thereby conceding a small discount in the price) tells Mr B that there is another potential buyer, and so she is not willing to discount the price any further. Mr B has no way to verify Mrs A's claims (in a reasonable amount of time). Should Mr B trust Mrs A when she says that there is another buyer, or is she bluffing?

In the theory to be developed below, mentioning that there are heating deficiencies will make it more likely in Mr B's eyes that Mrs A is an honest seller always telling the truth. As a result, Mr B will be convinced enough that there is indeed another buyer when Mrs A says so, and he will accept paying the high price (minus the small discount conceded for the reported heating deficiencies). By mentioning that there are deficiencies, Mrs A manipulates Mr B's belief about her true nature (whether she is an honest seller or an opportunist), and she exploits Mr B's inference error when she says that there is another buyer.

Such a deceptive tactic works in our theory in so far that mentioning small deficiencies is more representative of honest sellers than of opportunist sellers over all transaction situations (with high or low prices, say), and in forming his judgement about Mrs A's type, Mr B somehow only considers the general attitudes of the various types of sellers and does not distinguish how the various types of sellers behave in those various transaction situations

[1] As regards Lewis' deceptive tactic, it is not at all clear from a game theoretic perspective why the fact that the lie is logical (in a given instance) should increase the likelihood that it is believed. If liers always use logical lies, then logic should even heighten the listener's suspicion.

[2] From the perspective of this paper, logic may be viewed as more typical of true statements (over all possible statements), thereby making the use of logical lies more effective.

with high or low price.

We will present a detailed formalization of the above deceptive bargaining tactic in Section II, pointing out that it would not work if Mr B were fully rational.[3] Before developing that application, we present in Section I a general framework that allows us to model quite generally such inference errors as the one made by Mr B in a game theoretic equilibrium approach.

Specifically, the class of games considered in this paper are two-player multi-stage games with incomplete information and observable actions in which players may be of several types, past actions are assumed to be observable by everyone, and types may affect the preference relations of players. A key non-standard ingredient is that players are also parameterized by how finely they understand their opponent' s strategy. In addition to their preference and informational characteristics, players are endowed with cognitive types.

Following Jehiel (2005), cognitive types are modelled by assuming that players partition the decision nodes of their opponents into various sets referred to as analogy classes, and that players understand only the aggregate behavior of their opponent over the various decision nodes forming their analogy classes. Cognitive types are further differentiated according to whether or not the player distinguishes the behaviors of the various types of his opponent.

Thus, cognitive types may vary in two dimensions: a player may be more or less fine in the partition of the decision nodes of his opponent (what we call the analogy part), and a player may or may not distinguish the behaviors of the various types of his opponent (the sophistication part). In the above bargaining story, Mr B bundles the announcement nodes of sellers into one analogy class, whether the price is high or low, and he distinguishes the behaviors of honest and opportunist sellers. Thus, Mr B uses a coarse analogy partition, but he is sophisticated in the terminology just defined.

Given a strategic environment that includes the specification of players' cognitive types, we define an equilibrium concept that we refer to as the analogy-based sequential equilibrium. In equilibrium, players have correct expectations about the aggregate behavior of their opponents in their various analogy classes - these are referred to as analogy-based expectations. Whenever they move, players play best-responses to their analogy-based expectations

[3]Indeed, if Mr B were fully rational, he should understand that opportunist sellers more systematically concede that there are small deficiencies when the price is high, and thus Mr B should be even more cautious about the true presence of another buyer when told that there are heating problems.

and to their belief about the type of their opponent. As the game proceeds, players update their beliefs about the type of their opponent according to Bayes' rule as derived from their analogy-based expectations.[4]

In Section I we show that in finite environments (finite numbers of types, actions, and nodes), an analogy-based sequential equilibrium always exists. We also suggest how to interpret the solution concept from a learning perspective. Finally, we illustrate the working of the concept in a simple two-person two-period zero-sum game in which the payoff structure is commonly known to players but players may have cognitive types other than the fully rational one. The example serves to illustrate 1) why a player with non-fully rational cognitive ability cannot be viewed as a rational player who does not distinguish between some situations (a player with coarser information), 2) how, in a mixed population of rational and coarse players, a rational player always performs better, and 3) why, in our framework with incorrect inferences, there may be room for reputation building even in zero-sum games where there is no value to commitment.[5]

The framework of Section I is then used in Section II to formalize the above deceptive bargaining tactic. Section III concludes. We shall start, however, by situating our work in the perspective of various literatures.

Related literature

There have been many attempts to relax the rationality assumptions imposed on economic agents. These include relaxing the ability of agents to optimize their strategy given their beliefs (as in the Quantal Response Equilibrium, Richard McKelvey and Thomas Palfrey, 1995) or relaxing the ability of agents to form correct expectations. By maintaining the ability of agents to optimize their strategies given their beliefs, our paper contributes to the second form of departure from rationality, which we refer to as cognitive limitations.

[4]More precisely, we assume that players adopt the simplest representation of their opponent's strategy that is consistent with their knowledge (the analogy-based expectation). That is, the opponent's behavior in the various nodes bundled into one analogy class is assumed to be the same and in equilibrium it coincides with the aggregate distribution of the opponent's behavior over the set of nodes forming the analogy class. The evolution of the belief system is then similar to that in sequential equilibrium (David Kreps and Robert Wilson (1982a)) except that it is based on the conjecture about the opponent's strategy as just defined (rather than on the opponent's true strategy).

[5]The traditional approach to reputation pioneered by Thomas Schelling (1960) associates the idea of successful reputation building with the successful ability to commit to a particular behavior (which is of no use in a zero-sum game, due to the minmax theorem).

Several routes have been pursued to model cognitive limitations either introducing explicit biases in the inference process (see Daniel Khaneman et al., 1982 for an exposition of such biases as the gambler's fallacy, the base rate neglect, the conjunction fallacy etc...) or deriving the expectations from limited introspective reasoning (as in the level k approach, Dale Stahl, 1993) or deriving the expectations and inference process from the erroneous or coarse perception held by agents about their environment (approaches based on subjective prior or the self-confirming equilibrium and this paper, respectively). Our paper contributes to the last of these routes by further postulating that the coarse perception held by boundedly rational agents is the *simplest* representation -or model of others- that is consistent with their coarse statistical knowledge.

Such a line of research that views bounded rationality equilibrium concepts as a result of partial learning is the common theme of the limited foresight equilibrium (Jehiel, 1995), the analogy-based expectation equilibrium (Jehiel, 2005) and the valuation equilibrium (Jehiel and Samet, 2007).[6] Jehiel (2005) developed the analogy-based equilibrium concept to capture bounds on rationality that accommodate coarse perception but fully rational information processing, and extended to static games of incomplete information in Jehiel and Frederic Koessler (2008). Our aim in this paper is to extend this basic structure to extensive games with incomplete information, which is necessary to analyze the evolution of beliefs over time. The extension of these concepts to dynamic games allows us to examine the basic ideas of belief manipulation and deception. Connected to the analogy-based expectation equilibrium, Erik Eyster and Matthew Rabin (2005) have proposed a concept for static games of incomplete information, called cursed equilibrium, in which players do not fully take into account how other people's actions depend on their information.[7] In problems with interdependent preferences, the cursed equilibrium of Eyster and Rabin gives rise to erroneous equilibrium beliefs (as the analogy-based expectation equilibrium does) about the relation between the strategy and the signal of the opponent. Yet, by the very static nature

[6] Other approaches based on the idea that to facilitate learning agents do not consider the set of all possible strategies but only a subset are also available, see in particular Olivier Compte and Andrew Postlewaite (2008).

[7] The cursed equilibrium was developed independently of the analogy-based expectation equilibrium. The *fully* cursed equilibrium can be viewed as a special case of the analogy-based expectation equilibrium in which players' analogy partitions coincide with their own information partitions. The *partially* cursed equilibrium can be viewed as an alternative approach to the idea of partial sophistication to that captured by the analogy-based expectation equilibrium (see Jehiel and Koessler (2008) for further discussion).

of the games considered by Eyster and Rabin, no belief manipulation can be captured by their approach, which constitutes a key difference from the present framework.

Even though the starting point of our approach is about modeling the consequences of the coarse perception of agents with cognitive limitations as just explained, it turns out that our paper can also be viewed as formalizing a well studied bias in social psychology, e.g., the Fundamental Attribution Error (FAE) (see Edward Jones and Keith Davis (1965), Lee Ross (1977), Ross, Teresa Amabile and Julia Steinmetz (1977)). Roughly speaking, the FAE is "the tendency in forming one own's judgement about others to underestimate the importance of the specific situation in which the observed behavior is occurring" (Maureen O' Sullivan (2003)).[8] In the above bargaining story, Mr B is subject to the FAE. In forming his judgement about whether he is facing an honest seller after Mrs A has reported minor heating deficiencies, Mr B "ignores" that sellers' attitudes are not the same whether the price is high or low. Our model provides an explicit way to formalize such a neglect by Mr B.

There have been several earlier game theoretic attempts to capture the phenomenon of deception. These include the ideas of playing mixed strategy (to avoid being detected) in zero-sum interactions (John von Neuman and Oskar Morgenstern (1944)) and of playing a pooling or semi-pooling equilibrium (thereby not revealing one's own type) in signaling games (Michael Spence (1973)) or communication games (Joel Sobel (1985) and Vincent Crawford (2003)) or repeated games (Kreps and Wilson (1982b), Kreps et al. (1982), Drew Fudenberg and David Levine (1989)). Our approach to deception differs from these earlier approaches in that it is based on the idea of belief manipulation (by which we mean that some players end up having erroneous beliefs based on their observation), which cannot arise in the standard rationality paradigm considered in these earlier approaches. In our theory, deception can be viewed as the exploitation by rational players of the FAE made by other

[8] Ross et al. (1977) report a striking example in support of the FAE. In a pool of Stanford students from various fields, subjects were divided between questioners and answerers. The "questioners" were requested to ask the answerers difficult questions. Every questioner was matched to a single answerer who was almost always from a different field. After the quizz (answerers and questioners then knew how many correct answers were given in their match), it was observed that answerers consistently thought they were worse than questioners, thereby ignoring the fact that the pool of questions on which they performed relatively poorly was not generated at random but drawn from the esoteric knowledge of the questioner. Note that answerers were explicitly told before the quizz that questioners could freely choose the questions they liked best.

players, where FAE allows for belief manipulation.

Finally, it should be mentioned that our setup can be used to formalize a model of persuasion in the vein of the one developed independently of this paper by Sendhil Mullainathan et al. (2008), in which a persuader finds it advantageous to send (costly) messages even when they are not informative.[9]

I. A General Framework

A. *The class of games and the cognitive environment*

We consider multi-stage two-player games with observed actions and incomplete information. Extension to more than two players raises no conceptual difficulties. Each player $i = 1, 2$ can be one of finitely many types $\theta_i \in \Theta_i$. Player i knows his own type θ_i, but not that of player j, $j \neq i$. We assume that the distribution of types is independent across players, and we let $p_{\theta_i} > 0$ denote the prior probability that player i is of type θ_i. These prior probabilities $p_i = (p_{\theta_i})_{\theta_i}$ are assumed to be known to the players. Players observe past actions and earlier moves by nature except for the choice of their opponent's type. Moreover, there is a finite number of stages, and, at every stage and for every player including nature, the set of pure actions is finite.

Player i plays at the same set H_i of histories, whatever his type θ_i.[10] Moreover, the action space of player i at history $h \in H_i$ is common to all types θ_i, and is denoted by $A_i(h)$.

The set of all histories is denoted by H and the set of terminal histories is denoted by Z. The set of players who must move at history h is denoted by $I(h)$, and ha is the history starting with h and followed by a where $a \in \underset{i \in I(h)}{\times} A_i(h)$ is the action profile played by the players who must move at h.

Each player i is endowed with a VNM utility function defined on lotteries over terminal histories $h \in Z$. Player i's VNM utility is denoted by u_i and it may depend on the types of

[9] In their model, such an application requires nature in state $s = 1$ (or 2) to be identified with the strategic persuader in state $s = 0$. It also requires to assume that the listener pools the message moves in state $s = 1$ (or 2) and $s = 0$ into one analogy class (while distinguishing the persuader's behavior according to her private information). The analogy-based sequential equilibrium thus obtained corresponds to the more "Bayesian" approach they present in appendix II, thereby providing a learning justification to that approach rather than to the simpler one pursued in the body of their paper.

[10] A history refers to the earlier moves made by the players and possibly the earlier moves made by nature except for the choice of players' types which is not included in the history. Given our observability assumptions, histories are commonly known to the players.

players i and j together with the terminal history. That is, $u_i(h;\theta_i,\theta_j)$ is player i's payoff if the terminal history $h \in Z$ is reached, and players i and j are of type θ_i and θ_j, respectively. Each player i is assumed to know his own payoff structure (but not *a priori* that of his opponent).

The non-standard aspect of our strategic environment Γ lies in the definition of the types θ_i. Types θ_i are made of two components $\theta_i = (t_i, c_i)$ where t_i is the preference type of player i that acts on players' preferences - this is the standard component in the type - and c_i is the cognitive type of player i, defining how finely player i understands the strategy of player j - this is the non-standard component in the type.

As common sense suggests, the cognitive type of players do not affect players' preferences over the various terminal nodes. That is, for every terminal history $h \in Z$, we have that $u_i(h;\theta_i,\theta_j) = u_i(h;\theta_i',\theta_j')$ whenever θ_i and θ_i' have the same preference type t_i, and θ_j and θ_j' have the same preference type t_j.

Cognitive types c_i are defined as follows. Each player i forms an expectation about the behavior of player j by pooling together several histories $h \in H_j$ at which player j must move, and each such *pool* is referred to as a *class of analogy*. Players are also differentiated according to whether or not they distinguish between the behaviors of the various types of their opponent.

Formally, a cognitive type c_i of player i is characterized by (An_i, δ_i), where An_i stands for player i's analogy partition and δ_i is a dummy variable that specifies whether or not type θ_i distinguishes between the behaviors of the various types θ_j of player j. We let $\delta_i = 1$ when type θ_i distinguishes between types θ_j's behaviors and $\delta_i = 0$ otherwise. As in Jehiel (2005), An_i is defined as a partition of the set H_j of histories at which player j must move into subsets or analogy classes α_i.[11] When h and h' are in the same analogy class α_i, it is required that $A_j(h) = A_j(h')$. That is, at two histories h and h' which player i pools together, the action space of player j should be the same, and $A(\alpha_i)$ denotes the common action space in α_i.

[11] A partition of a set X is a collection of subsets $x_k \subseteq X$ such that $\bigcup_k x_k = X$ and $x_k \cap x_{k'} = \emptyset$ for $k \neq k'$.

B. *Analogy-based sequential equilibrium*

Analogy-based expectations:

An analogy-based expectation for player i of type θ_i is denoted by β_{θ_i}. It specifies, for every analogy class α_i of player i of type θ_i, a probability measure over the action space $A(\alpha_i)$ of player j. Types θ_j of player j are distinguished or not by player i according to whether $\delta_i = 1$ or 0. If $\delta_i = 1$, β_{θ_i} is a function of θ_j and α_i, and $\beta_{\theta_i}(\theta_j, \alpha_i)$ is player i's expectation about the average behavior of player j with type θ_j in class α_i. If $\delta_i = 0$, player i merges the behaviors of all types θ_j of player j, and β_{θ_i} is a sole function of α_i: $\beta_{\theta_i}(\alpha_i)$ is then player i's expectation about the average behavior of player j in class α_i (where the average is taken over all possible types).[12] We let $\beta_i = (\beta_{\theta_i})_{\theta_i \in \Theta_i}$ denote the analogy-based expectation of player i for the various possible types $\theta_i \in \Theta_i$.

Strategy:

A behavioral strategy of player i is denoted by s_i. It is a mapping that assigns to every history $h \in H_i$ at which player i must move a distribution over player i's action space $A_i(h)$.[13] We let σ_{θ_i} denote the behavioral strategy of type θ_i, and for every $h \in H_i$ we let $\sigma_{\theta_i}(h) \in \Delta A_i(h)$ denote the distribution over $A_i(h)$ according to which player i of type θ_i selects actions in $A_i(h)$ when at h. We let $\sigma_{\theta_i}(h)[a_i]$ be the corresponding probability that type θ_i plays $a_i \in A_i(h)$ when at h, and we let $\sigma_i = (\sigma_{\theta_i})_{\theta_i}$ denote the strategy of player i for the various possible types θ_i; σ will denote the strategy profile of the two players.

Belief system:

When player i distinguishes the types of player j, i.e. $\delta_i = 1$, he holds a belief about the type of his opponent and this belief may typically change as time proceeds (and new observations become available). Formally, we let μ_{θ_i} denote the belief system of player i of type θ_i, where $\mu_{\theta_i}(h)[\theta_j]$ is the probability that player i of type θ_i assigns to the event "player j is of type θ_j" conditional on the history h being realized.

When player i does not distinguish the types of player j, no belief system is required. To

[12] We could more generally allow players to distinguish partially the types. This would lead to a partitional approach defining which of the types are being confused. The resulting presentation would however be more cumbersome without bringing additional insights.

[13] Mixed strategies and behavioral strategies are equivalent, since we consider games of perfect recall.

save on notation, we assume that in this case player i 's belief coincides with the prior p_j throughout the game. We call μ_i the belief system of player i for the various possible types θ_i, and we let μ be the profile of belief systems for the two players $i = 1, 2$.

Sequential rationality:

From his analogy-based expectation β_{θ_i}, player i of type θ_i derives the following representation of player j's strategy: Player i perceives player j to play at every history $h \in \alpha_i$ according to the average behavior in class α_i.[14] The induced strategy depends on the type θ_j of player j whenever $\delta_i = 1$ but not when $\delta_i = 0$. At every history $h \in H_i$ where he must play, player i is assumed to play a best-response to this perceived strategy of player j as weighted by his belief $\mu_{\theta_i}(h)$.

Formally, we define the β_{θ_i}-perceived strategy of player j, $\sigma_j^{\beta_{\theta_i}}$, as

$$\begin{aligned} \text{If } \delta_i &= 1 \quad \sigma_{\theta_j}^{\beta_{\theta_i}}(h) = \beta_{\theta_i}(\theta_j, \alpha_i) \quad \text{for every } h \in \alpha_i \text{ and } \theta_j \in \Theta_j \\ \text{If } \delta_i &= 0 \quad \sigma_{\theta_j}^{\beta_{\theta_i}}(h) = \beta_{\theta_i}(\alpha_i) \quad \text{for every } h \in \alpha_i \text{ and } \theta_j \in \Theta_j \end{aligned}$$

Given the strategy s_i of player i and given history h, we let $s_i \mid_h$ denote the continuation strategy of player i induced by s_i from history h onwards. We also let $u_i^h(s_i \mid_h, s_j \mid_h; \theta_i, \theta_j)$ denote the expected payoff obtained by player i when history h has been realized, the types of players i and j are given by θ_i and θ_j respectively, and players i and j behave according to s_i and s_j respectively.

Definition 1 *(Criterion) Player i's strategy σ_i is a sequential best-response to (β_i, μ_i) if and only if for all $\theta_i \in \Theta_i$, for all strategies s_i and all histories $h \in H_i$,*

$$\sum_{\theta_j \in \Theta_j} \mu_{\theta_i}(h)[\theta_j] u_i^h(\sigma_{\theta_i} \mid_h, \sigma_{\theta_j}^{\beta_{\theta_i}} \mid_h; \theta_i, \theta_j) \geq \sum_{\theta_j \in \Theta_j} \mu_{\theta_i}(h)[\theta_j] u_i^h(s_i \mid_h, \sigma_{\theta_j}^{\beta_{\theta_i}} \mid_h; \theta_i, \theta_j).$$

Consistency:

In equilibrium, two notions of consistency are required. First, analogy-based expectations

[14] This is the simplest representation compatible with type θ_i's knowledge.

are required to be consistent with the strategy profile. That is, they must coincide with the real average behaviors in every considered class and for every possible type (if types are differentiated), where the weight given to each element of an analogy class must itself be consistent with the real probability of visiting this element. A learning interpretation of this consistency requirement will be suggested. Second, the belief system held by players must be consistent with their expectations, as in Sequential Equilibrium.

Formally, letting $P^\sigma(\theta_i, \theta_j, h)$ denote the probability that history h is reached when players i and j are of types θ_i and θ_j respectively, and players play according to σ, the consistency of the analogy-based expectations is defined as:

Definition 2 *Player i's analogy-based expectation β_i is consistent with the strategy profile σ if and only if:*

- *For any $(\theta_i, \theta_j) \in \Theta$ such that $\delta_i = 1$, and for all $\alpha_i \in An_i$,*

$$\beta_{\theta_i}(\theta_j, \alpha_i) = \frac{\sum_{(\theta_i', h) \in \Theta_i \times \alpha_i} p_{\theta_i'} P^\sigma(\theta_i', \theta_j, h) \cdot \sigma_{\theta_j}(h)}{\sum_{(\theta_i', h) \in \Theta_i \times \alpha_i} p_{\theta_i'} P^\sigma(\theta_i', \theta_j, h)}$$

  whenever there exist θ_i' and $h \in \alpha_i$ such that $P^\sigma(\theta_i', \theta_j, h) > 0$.

- *For any $\theta_i \in \Theta$ such that $\delta_i = 0$, and for all $\alpha_i \in An_i$,*

$$\beta_{\theta_i}(\alpha_i) = \frac{\sum_{(\theta_i', \theta_j', h) \in \Theta \times \alpha_i} p_{\theta_i'} p_{\theta_j'} P^\sigma(\theta_i', \theta_j', h) \cdot \sigma_{\theta_j'}(h)}{\sum_{(\theta_i', \theta_j', h) \in \Theta \times \alpha_i} p_{\theta_i'} p_{\theta_j'} P^\sigma(\theta_i', \theta_j', h)}$$

  whenever there exist θ_i', θ_j' and $h \in \alpha_i$ such that $P^\sigma(\theta_i', \theta_j', h) > 0$.

The consistency of the belief system is defined as:

Definition 3 *Player i's belief system μ_i is consistent with the analogy-based expectation β_i if and only if for any $(\theta_i, \theta_j) \in \Theta$ such that $\delta_i = 1$*

$$\mu_{\theta_i}(\theta_j)(\emptyset) = p_{\theta_j}.$$

And for all histories h, ha

$$\begin{aligned}
\mu_{\theta_i}(ha)[\theta_j] &= \mu_{\theta_i}(h)[\theta_j] \text{ whenever } h \notin H_j \\
\mu_{\theta_i}(\theta_j)(ha) &= \frac{\mu_{\theta_i}(h)[\theta_j]\sigma_{\theta_j}^{\beta_{\theta_i}}(h)[a_j]}{\sum_{\theta'_j \in \Theta_j} \mu_{\theta_i}(h)[\theta'_j]\sigma_{\theta'_j}^{\beta_{\theta_i}}(h)[a_j]} \\
\text{whenever } h &\in H_j, \text{ there exists } \theta'_j \text{ s.t. } \sigma_{\theta'_j}^{\beta_{\theta_i}}(h)[a_j] > 0 \text{ and player } j \text{ plays } a_j \text{ at } h.
\end{aligned}$$

While the consistency of the analogy-based expectations (definition 2) should be thought of as the limiting outcome of a learning process, the consistency of the belief system μ_i (definition 3) should be thought of as an expression of player i 's inference process. Based on his representation of the strategy of the various types of his opponent, player i makes inferences using Bayes' law as to the likelihood of the various possible types he is facing.

The learning process we have in mind to justify the correctness of the analogy-based expectations involves populations of players i and j in which there is a constant share p_{θ_i} of players of type θ_i. In each round, players i and j are randomly matched. At the end of a round, the behaviors of the matched players and their types are revealed. These players exit the population, and they are replaced by new players with the same type.[15] All pieces of information are gathered in a general data set, and players have different access to this data set depending on their types.[16] At each round of the learning process, players choose their strategy as a best-response to the feedback they received (and the system of belief that derives from it), which in turn generates new data for the next round. If the pattern of behaviors adopted by the players stabilizes to some strategy profile σ, every player's analogy-based expectations should eventually converge to the ones that are consistent with σ given his cognitive type,[17] which motivates the solution concept defined below.

[15] The replacement scenario is reminiscent of the recurring game framework studied by Matthew Jackson and Ehud Kalai (1997), who assume that each individual player only plays once. This is to be contrasted with a recent paper by Ignacio Esponda (2008), who, in static games of incomplete information, elaborates on Eyster-Rabin's fully cursed equilibrium by assuming that players i have access both to the empirical distribution of actions of players j (but not to how these actions are related to j's private information) and to i's own distribution of payoffs.

[16] A player i with cognitive type $c_i = (An_i, \delta_i)$ such that $\delta_i = 0$ has access to the average empirical distribution of behavior in every analogy class $\alpha_i \in An_i$ where the average is taken over all histories $h \in \alpha_i$ and over the entire population of players j. A player with cognitive type $c_i = (An_i, \delta_i)$ such that $\delta_i = 1$ has access to the average empirical distribution of behavior in every $\alpha_i \in An_i$ for each subpopulation of types θ_j of players j.

[17] Observe that the average in the expression of $\beta_{\theta_i}(\theta_j, \alpha_i)$ is taken over all possible realizations of player

Equilibrium:

In equilibrium, both the analogy-based expectations and the belief systems are consistent, and players play best-responses to their analogy-based expectations at every history. In line with the Sequential Equilibrium (Kreps and Wilson (1982a)), we require the analogy-based expectations and belief systems to be consistent with respect to slight totally mixed perturbations of the strategy profile where a totally mixed strategy for player i is a strategy that assigns strictly positive probability to every action $a_i \in A_i(h)$ at every history $h \in H_i$. This in turn puts additional structure on the expectations and beliefs at histories that belong to analogy classes that are never reached in equilibrium.[18]

Definition 4 *A strategy profile σ is an Analogy-based Sequential Equilibrium if and only if there exist analogy-based expectations β_i, belief systems μ_i for $i = 1, 2$, and sequences $(\sigma^k)_k$, $(\beta_i^k)_k$, $(\mu_i^k)_k$ converging to σ, β, μ, respectively, such that each σ^k is a totally mixed strategy profile, and for every i and k:*

1. *σ_i is a **sequential best-response** to (β_i, μ_i)*
2. *β_i^k is **consistent** with σ^k and*
3. *μ_i^k is **consistent** with β_i^k.*

Compared to the sequential equilibrium, the main novelty lies in the introduction of cognitive types who may only know partial aspects of the strategy of their opponent. Compared to the analogy-based expectation equilibrium (Jehiel (2005)), the main novelty lies in the introduction of players' uncertainty about the type of their opponent and the possibility that a cognitive type may distinguish the behaviors of the various types of his opponent. It is the combination of these features that allows us to speak of deception as the exploitation of the FAE. More precisely, such a deception requires the presence of players who are both uncertain about their opponent's type (so that there is room for inference processes) and

i's types θ_i', hence the summation over θ_i' . That is, we are assuming that player i of type θ_i is informed of θ_j's behaviors whatever the type of player i they are matched with. The weight $p_{\theta_i'} P^\sigma(\theta_i', \theta_j, h)$ on $\sigma_{\theta_j}(h)$ simply reflects the relative frequency with which $\sigma_{\theta_j}(h)$ contributes to the aggregate behavior.

[18] For those readers who dislike trembles, one can offer a weaker notion of equilibrium without trembles, similar in spirit to the self-confirming equilibrium (see Drew Fudenberg and David Levine (1998)). Note, however, that trembles have less bite in our setup than in the standard framework because for an analogy class to be reached with positive probability it is enough that one of the histories in the analogy class is reached with positive probability - a requirement that is weaker when the analogy class is larger.

are partially knowledgeable of the strategy of their opponent, so that the inferences may be erroneous.

C. *Basic properties*

We note that in finite environments, an equilibrium always exists, no matter how cognitive types are specified and distributed.

Proposition 1 *In finite environments, there always exists at least one Analogy-based Sequential Equilibrium.*

Proof: The proof follows standard methods, first noting the existence of equilibria in which each player i is constrained to play any action $a_i \in A_i(h)$ at any history $h \in H_i$ with a probability no less than ε, and then showing that the limit as ε tends to 0 of such strategy profiles is an Analogy-based Sequential Equilibrium. **Q. E. D.**

We next observe that if every player i is rational (in the sense that for all types $\theta_i = (t_i, c_i)$ of player i, the cognitive type $c_i = (An_i, \delta_i)$ is such that An_i is the finest analogy partition $\bigcup_{h \in H_j} \{h\}$, and player i distinguishes between player j's types, $\delta_i = 1$), then an analogy-based sequential equilibrium coincides with a sequential equilibrium of the game in which every type $\theta_i = (t_i, c_i)$ of player i is identified with her preference type t_i. Thus, our framework can be viewed as providing a generalization of the sequential equilibrium that allows us to cope with situations in which the cognitive abilities of players need not be perfect.

D. *A simple illustration*

In this part, we construct an analogy-based sequential equilibrium in a simple two-person two-period zero-sum game. This example serves to illustrate the working of the concept in a simple scenario. Specifically, consider the two-period-repetition of the following zero-sum stage game G. In stage game G the Row player chooses an action U or D, the Column player chooses an action L or R, and stage game payoffs are as represented in Figure 4. The overall payoff obtained by the players is the sum of the payoffs obtained in the two periods. That is, there is no discount between period 1 and period 2 payoffs.

	L	R
U	5, -5	3, -3
D	0, 0	7, -7

Figure 4. The stage game G

We assume that there are two types of Row players, the *Rational* type and the *Coarse* type, where both types are assumed to be equally likely. The *Rational* Row player has a perfect understanding of the strategy of the Column player, as in the standard case. The *Coarse* Row player only knows the average behavioral strategy of the Column player over the two time periods (i.e., he bundles period 1 and the possible histories in period 2 into one analogy class).

There is one type for the Column player. The Column player is *Sophisticated* in the sense that he distinguishes between the behaviors of the *Rational* Row player and the *Coarse* Row player. But, he is assumed to be *Coarse* in the sense that for each type of the Row player he only knows the average behavior of this type over the two time periods, i.e. he bundles all histories into one analogy class.

Proposition 2 *The following strategy profile is an Analogy-based Expectation Sequential Equilibrium. 1) Rational Row Player: Play* U *in period* 1*. Play* D *in period* 2 *if* U *was played in period* 1*, and* U *otherwise. 2) Coarse Row Player: Play* U *both in periods* 1 *and* 2*. 3) Column Player (Sophisticated Coarse): Play* L *in period* 1*. Play* R *in period* 2 *if the Row player played* U *in period* 1*. Play* L *in period* 2 *if the Row player played* D *in period* 1*.*

In equilibrium, (U, L) is played in period 1 and then (D, R) in period 2 whenever the Row player is rational, and (U, L) is played in period 1 and then (U, R) in period 2 whenever the Row player is coarse. The Column player gets an expected payoff of -10 that is less than her value $-70/9$. The Rational Row player gets an overall payoff of $5 + 7 = 12$ and the Coarse Row player gets an overall payoff of $5 + 3 = 8$.

A key aspect of this equilibrium involves understanding the inference process of the *Sophisticated Coarse* Column player. The *Coarse* Row player always plays U, and the *Rational* Row player plays U and D with an equal frequency on average. These (average) behaviors of the two types of Row players define the analogy-based expectations of the

Column player. Given these expectations, the Column player updates her belief about the type of the Row player as follows: when action D is being played in period 1, the Column player believes that she faces the *Rational* Row player for sure. When action U is being played in period 1, the Column player believes that she faces the *Coarse* Row player with probability $\frac{1/2}{1/2+1/2\times 1/2} = \frac{2}{3}$. Accordingly, the Column player plays R in period 2 because given her belief, this looks like the smartest decision, even though in reality it is not. Thus, by playing U in period 1, the Rational Row player builds a false reputation for being more likely to be a Coarse Row player, which he later exploits in period 2 by getting the high payoff of 7.[19]

We make several comments about the equilibrium shown in Proposition 2.

First, the Column player gets an expected payoff that is less than her value, $-70/9$, even though, by the very property of the value, the Column player could very well guarantee $-70/9$ - no matter what the Row player does - by playing the maximin strategy (i.e., play L with probability $4/9$ and R with probability $5/9$ in both periods). The Column player chooses not to follow the maximin strategy because she thinks that she can do better, given her understanding of the strategy of Row players. Such a feature would, of course, not arise in a standard rationality framework in which the Column player should obtain, in equilibrium, at least what she can secure irrespective of other players' strategies. This helps to clarify the difference from Vincent Crawford (2003), who assumes in a zero-sum pre-play communication game that those agents whose behaviors are not exogenously specified are fully rational and are thus bound to get at least their value in equilibrium.[20] It also helps to explain why it is not possible to interpret the analogy-based sequential equilibrium as a sequential equilibrium that would obtain in the full rationality paradigm under alternative informational assumptions.[21]

[19] The rest of the argument to establish Proposition 2 goes as follows. It is readily verified that the Rational Row player plays a best-response to the Column player's strategy. (He gets an overall payoff of $5 + 7 = 12$ and would only get an overall payoff of $0 + 11/2$ at best if he were to play D in period 1, and he would obviously get a lower payoff by playing U in period 2.) The Coarse Row player finds it optimal to play U whenever he has to move, because he perceives the Column Player to play L and R with an equal frequency on average over the two time periods, and $\frac{1}{2}(5+3) > \frac{1}{2}(0+7)$.

[20] Vincent Crawford (2003) captures the idea of lying for strategic advantage in a zero-sum pre-play communication game that is populated by sufficiently many mechanical types. But in Crawford's model, the belief of rational players cannot be manipulated, as equilibrium requires that rational players are not mistaken about either the distribution of types or about their strategies. This is a key difference from our approach.

[21] Even if the Column player were assumed not to remember whether she is in stage 1 or 2, she could still

Second, in the equilibrium of Proposition 2, the Rational Row player obtains a larger payoff than the Coarse Row player. This is no coincidence, as the Rational Row player always has the option to mimic other types' strategies and Rational players assess correctly the payoff attached to any strategy.

Finally, it should be noted that it would be impossible to reproduce the behavioral strategies described in Proposition 2 if there were only one type for each player, who would be characterized solely by his analogy partition as in Jehiel (2005).[22]

II. Deception as a Bargaining Tactic

A. *The basic setup*

The owner of a house, Mrs A, wishes to sell her good. The initial price has already been publicly announced. It is either $\underline{p}$ or $\overline{p}$ where $\overline{p} > \underline{p}$ and $\underline{p}$ may be thought of as being the "market price" of the house as perceived by real estate agents.

A potential buyer, Mr B, comes in, and the following interaction between Mrs A and Mr B takes place. Mrs A tells Mr B whether or not some small repairs (say for heating deficiencies) are needed in the house.[23] If minor deficiencies are announced, the price drops by an amount Δ. That is, the new price is $p - \Delta$ where p was the originally announced price (Δ should be thought of as being small relative to $\overline{p} - \underline{p}$). Then Mrs A tells Mr B whether or not there is another buyer who has expressed interest in the house. When the initial price was $p = \underline{p}$ and Mrs A says that another buyer has expressed interest, the price increases by a very small amount, say ε. No such price increase occurs when the initial price is $p = \overline{p}$.[24]

Only Mrs A knows whether indeed there are small repairs needed and whether there is another potential buyer. After the announcements are made, Mr B has to decide whether or not to accept the offer (before he can verify the correctness of Mrs A's announcements).

secure the value, given that the maxmin strategy does not require any recall (it is stationary). See Jehiel (2005) and Jehiel and Koessler (2008) for further examples illustrating why the analogy-based expectation equilibrium cannot be interpreted as a standard equilibrium of a different game with modified information structure.

[22]For the Column player to play a different action in periods 1 and 2, she should either be indifferent between playing L or R (which cannot be the case here, since the Row player does not play U with probability 7/9 on average) or treat separately the behavior of the Row in the two time periods, but then in period 1 she could not find it optimal to play L given that the Row player always plays U.

[23]It is assumed that Mr B cannot verify the nature of these repairs within a reasonable amount of time.

[24]We assume this only for plausibility. The analysis is unaffected if we assume that there is also a price increase when $p = \overline{p}$ (this is because ε is assumed to be small in comparison with $\overline{p} - \underline{p}$).

If Mr B says yes, the transaction takes place at the agreed price (i.e., p if no deficiencies were announced and $p - \Delta$ if deficiencies were announced). We let V_B^{yes} denote Mr B' s payoff when the original price was $\overline{p}$ and deficiencies were announced (so that the final price is $\overline{p} - \Delta$).

If Mr B says no, there are several cases. When the original price was the "market price" $\underline{p}$, no transaction takes place between Mrs A and Mr B, as we assume that Mrs A expects to sell her house at a price close to $\underline{p}$ and Mr B expects to buy a similar house at a price close to $\underline{p}$ (both Mrs A and Mr B would be slightly better off making the transaction now even at prices $\underline{p} - \Delta$, $\underline{p} + \varepsilon$, respectively, due to extra delays imposed by the transaction not being made now).

When the original price was $\overline{p}$ and there is effectively another buyer, no transaction between Mrs A and Mr B takes place. Mrs A gets a payoff that is less than $\overline{p}$, due to the risk that the other buyer does not confirm his interest, but significantly larger than $\underline{p}$, and Mr B gets a payoff of V_B^{out} (corresponding to the outcome of a search for another house).

When the original price was $\overline{p}$ and there is no other buyer,[25] bargaining between Mrs A and Mr B goes on. We do not model this extra piece of bargaining explicitly, but we assume that a transaction eventually takes place at a price significantly lower than $\overline{p} - \Delta$ (say not too far from $\underline{p}$).[26] We denote by V_B^{no} the payoff obtained by Mr B in this case.[27]

On top of the above specifications, we assume that there are two categories of sellers, those who always tell the truth (whom we call honest sellers) and those who do what serves their interest best (whom we call opportunists). Mrs A can belong to either of these categories, but there is no way for Mr B to know which, except by making inferences from how she behaves (here, what she says in the announcement stage).

Finally, we describe the probabilities of the various events, which are assumed to be known to both Mrs A and Mr B. We assume that the probability of the seller being honest is $\mu = \Pr(\text{Mrs A is honest})$ independently of the other random variables. We assume that the probability of a deficiency is $\lambda_d = \Pr(\text{deficiency})$ independently of the other random variables. For plausibility reasons, we allow the probabilities that the price is $\overline{p}$ and that there is

[25] Presumably Mr B gets further signals (not under Mrs A' s control) about this.

[26] This may be because Mr B values this specific house more than the average buyer and everyone is aware of this at this bargaining stage.

[27] More precisely, V_B^{no} will denote Mr B's payoff assuming a deficiency has been announced.

another buyer to be (presumably negatively) correlated. We let $\lambda_b = \Pr(\text{other buyer})$ be the probability that there is another buyer, $\lambda^* = \Pr(p = \overline{p} \,|\text{other buyer})$ and $\lambda^{**} = \Pr(p = \overline{p} \,|\text{no other buyer})$ be the probabilities that the initial price is $\overline{p}$ conditional on having another buyer or not having one, respectively. We also let $\overline{\lambda} = \lambda_b\lambda^* + (1-\lambda_b)\lambda^{**} = \Pr(p = \overline{p})$ denote the unconditional probability that the initial price is $\overline{p}$.

B. *Cognitive environment*

In this bargaining problem, a key strategic aspect is the judgement Mr B makes as to the likelihood that there is another buyer as a function of the announcements made by Mrs A. When Mr B is told that there is another buyer, should Mr B trust Mrs A? And how is Mr B's judgement affected by the announcement (or the non-announcement) of minor deficiencies?

We wish to analyze a situation in which Mr B somehow confuses the two price scenarios $p = \underline{p}, \overline{p}$ when assessing the announcement strategies of sellers such as Mrs A (whereas he distinguishes the strategies of sellers in all other respects). We will also assume that opportunist sellers are fully rational, and we will show how Mrs A, when opportunist, deceives Mr B by mentioning minor deficiencies (whether or not there are any) so as to increase his belief that there is indeed another buyer when she says there is. From the viewpoint of social psychology, Mr B is victim of the fundamental attribution error, and when she is opportunist, Mrs A exploits this. This is the essence of the deceptive tactic that we wish to highlight here. We will also illustrate later on how deception would be unsuccessful if Mr B were assumed to be fully rational instead.

To cast the above strategic environment into the framework of Section I, we let t_A be the preference type of Mrs A, where $t_A = (\tau, d, b)$, $\tau = h,\ o$ indicates whether A is honest or opportunist, $d = 1, 0$ indicates whether there are (minor) deficiencies or not and $b = 1, 0$ indicates whether there is another buyer or not. Mrs A is assumed to be fully rational. Thus, her cognitive type corresponds to the standard situation in which every decision node of Mr B constitutes a singleton analogy class (since there is only one type of Mr B - see below - no distinction of the various types of Mr B is relevant here). We will identify Mrs A's type θ_A with her preference type t_A.

Mr B can only be of one type. While his preferences have already been described, his cognitive type is described as follows. Mr B puts the "deficiency announcement nodes" of

Mrs A in the same analogy class, whether $p = \underline{p}$ or $\overline{p}$. Similarly, he puts the "other buyer announcement nodes" of Mrs A in the same analogy class, whether $p = \underline{p}$ or $\overline{p}$. In addition, he differentiates between the behaviors of the various types of sellers (i.e., $\delta_B = 1$ in the language of Section I).

Finally, before the interaction between Mrs A and Mr B starts, Nature chooses the type θ_A of Mrs A and then the initial price $p = \underline{p}, \overline{p}$ according to the distribution described above.

C. *Analysis*

When Mrs A is honest, her strategy is imposed by the definition of her type. She always tells the truth. That is, she mentions the deficiency if there is one and she mentions the existence of another buyer if there is one, whether the price is $p = \underline{p}$ or $\overline{p}$. Moreover, when the price is $p = \underline{p}$ and Mrs A is opportunist, she never mentions any deficiency (whether or not there is one), so as to save on the discount Δ, and she always says that there is another buyer (whether or not there is one), so as to get the extra ε in the final price.[28] Transaction always takes place between Mrs A and Mr B when $p = \underline{p}$.

It only remains to determine the announcement strategy of Mrs A when she is opportunist and the price is $\overline{p}$, and also the acceptance strategy of Mr B in this case.

We will test when the following strategies constitute an analogy-based sequential equilibrium. When she is opportunist and the price is $p = \overline{p}$, Mrs A always reports that there are deficiencies and that there is another buyer, no matter what the truth is. Mr B says "yes" to the offer after such announcements and "no" after any other announcement (i.e., when Mrs A says that there is no other buyer, or that there is another buyer and no deficiencies in the house).

A key variable in the analysis is the belief that Mr B attaches to the existence of another buyer after Mrs A has made her announcements (and the initial price was $p = \overline{p}$). Call γ such a belief. Given the above definitions of V_B^{out}, V_B^{no} and V_B^{yes}, Mr B would accept the deal (after $d = 1$ and $b = 1$ were announced) if $\gamma V_B^{out} + (1 - \gamma) V_B^{no} < V_B^{yes}$, and he would reject it otherwise. Accordingly, we let

$$\overline{\gamma} = \frac{V_B^{no} - V_B^{yes}}{V_B^{no} - V_B^{out}}$$

[28] Mrs A knows that the transaction will be approved by Mr B in any event, even if the final price is $\underline{p} + \varepsilon$ (rather than $\underline{p} - \Delta$ or $\underline{p}$).

denote the threshold belief such that Mr B would say "yes" if $\gamma > \overline{\gamma}$ and "no" if $\gamma < \overline{\gamma}$.

Assuming Mrs A follows the above strategy and given Mr B's cognitive type, the consistency of Mr B's analogy-based expectations implies that he should expect honest sellers to always report the truth, and opportunist sellers either to say that $d = 1$ and $b = 1$ with probability $\overline{\lambda} = \Pr(p = \overline{p})$ or to say that $d = 0$ and $b = 1$ with probability $1 - \overline{\lambda}$, independently of p, d, b. From such a perception, the distribution of (p, d, b), and Bayes' law, one can now compute Mr B's posterior belief denoted $\mu^{post}(\tau = h)$ that Mrs A is honest when she says that $d = 1$, $b = 1$ and $p = \overline{p}$.

$$\mu^{post}(\tau = h) = \frac{\mu \lambda_d \lambda_b \lambda^*}{\mu \lambda_d \lambda_b \lambda^* + (1 - \mu)(\overline{\lambda})^2}$$

This follows from noting that Mr B's perceived probability that ($\tau = h$ and $p = \overline{p}$, Mrs A says $d = 1$, $b = 1$) is $\mu \lambda_d \lambda_b \lambda^*$ and Mr B's perceived probability that ($\tau = o$ and $p = \overline{p}$, Mrs A says $d = 1$, $b = 1$) is $(1 - \mu)(\overline{\lambda})^2$.[29]

After observing that $p = \overline{p}$ and hearing that $d = 1$ and $b = 1$ from Mrs A, Mr B assesses the probability that there is another buyer based on his prior belief that Mrs A is honest and his perceived informativeness of A's messages whether $\tau = h$ or o. Since Mr B knows that when Mrs A is honest she sends truthful messages, and since he correctly perceives that the message of opportunist sellers is not informative, we obtain that Mr B's posterior belief that $b = 1$ is:

$$\gamma^{post} = \mu^{post} + (1 - \mu^{post}) \Pr(b = 1 \mid p = \overline{p})$$

where $\Pr(b = 1 \mid p = \overline{p}) = \lambda^* \lambda_b / \overline{\lambda}$. We will also define the corresponding probability when the probability that Mrs A is honest coincides with the prior μ instead of μ^{post}. That is,

$$\gamma^{prior} = \mu + (1 - \mu) \Pr(b = 1 \mid p = \overline{p}).$$

We have:

Proposition 3 *Assume that $\gamma^{post} > \overline{\gamma} > \gamma^{prior}$. Then, the above strategy profile is an*

[29] The latter probability requires that $\tau = o$ (which has probability $1-\mu$), that $p = \overline{p}$ (which has probability $\overline{\lambda}$), and that the announcement "$d = 1$, $b = 1$" is picked (which is perceived to have probability $\overline{\lambda}$) while these three events are all perceived to be independent.

analogy-based sequential equilibrium.

Proof: Given that $\gamma^{post} > \overline{\gamma}$ and the above derivations, after observing that $p = \overline{p}$ and hearing from Mrs A that $d = 1$, $b = 1$, Mr B optimally says "yes" to the offer because he infers that the chance that there is indeed another buyer γ^{post} is high enough (larger than $\overline{\gamma}$). Clearly, if Mr B is told that there is no other buyer $b = 0$, he will infer that Mrs A is honest (because opportunist sellers never say that), and accordingly he will know for sure that there is no other buyer and will say "no" when the price is $p = \overline{p}$. Finally, after observing that $p = \overline{p}$ and hearing from Mrs A that $d = 0$, $b = 1$, Mr B will believe according to his cognitive perception that Mrs A is honest with some probability $\widehat{\mu}$.[30] Because $\mu^{post}(\tau = h) > \mu$ (as otherwise one could not have $\gamma^{post} > \gamma^{prior}$) and $\mu(\tau = h \mid p = \overline{p}$, A says $d = 0$ or 1, $b = 0) = 1$, we can infer that $\widehat{\mu} < \mu$ (the evolution of μ as perceived by Mr B should be a martingale). Given that $\gamma^{prior} < \overline{\gamma}$ (and that the offer when no discount is proposed is worse, everything else being equal), we conclude that Mr B finds it optimal to reject the offer after observing $p = \overline{p}$ and hearing from Mrs A that $d = 0$, $b = 1$. This also ensures that when Mrs A is opportunist and $p = \overline{p}$, she does not find it desirable to say that $d = 0$, $b = 1$ because she rightly anticipates that the offer would be rejected in such a case. **Q. E. D.**

D. *Discussion*

Observe that the condition $\overline{\gamma} > \gamma^{prior}$ is independent of $\Pr(p = \overline{p})$ whereas $\gamma^{post} > \overline{\gamma}$ is automatically satisfied when $\Pr(p = \overline{p})$ is sufficiently small.[31] Thus, the conditions of Proposition 3 require that the prior probability μ that Mrs A is honest be not too large, so that $\overline{\gamma} > \gamma^{prior}$, and that the probability $\overline{\lambda}$ that $p = \overline{p}$ be not too large, so that $\gamma^{post} > \overline{\gamma}$. The effect of $\Pr(p = \overline{p})$ on the equilibrium analysis is, of course, due to the erroneous inference process of Mr B, caused by his bundling of the decision nodes of Mrs A into a single analogy

[30] This $\widehat{\mu}$ is derived analogously to μ^{post} and can be expressed as

$$\widehat{\mu} = \frac{\mu(1-\lambda_d)\lambda_b\lambda^*}{\mu(1-\lambda_d)\lambda_b\lambda^* + (1-\mu)(1-\overline{\lambda})^2}$$

[31] To illustrate this, consider the case in which the probability that $p = \overline{p}$ is independent of whether $b = 0, 1$ so that $\lambda^* = \lambda^{**} = \overline{\lambda}$. Then $\mu^{post}(\tau = h)$ simplifies into $\frac{\mu\lambda_d\lambda_b}{\mu\lambda_d\lambda_b + (1-\mu)\overline{\lambda}}$, which converges to 1 as $\overline{\lambda}$ approaches 0, thereby implying that γ^{post} approaches $1 > \overline{\gamma}$ when $\overline{\lambda} \simeq 0$.

class, whether $p = \overline{p}$ or $\underline{p}$. Increasing $\Pr(p = \underline{p})$ makes the deceptive tactic of Mrs A when $p = \overline{p}$ and $\tau = o$ more effective, as reporting that there are deficiencies becomes more representative of the overall attitude of honest sellers than of opportunist sellers. Ironically, had we assumed that Mr B were fully rational instead, he should have inferred, when $p = \overline{p}$ and Mrs A reports deficiencies, that Mrs A is more likely to be an opportunist, since when $p = \overline{p}$ mentioning deficiencies is more typical of opportunists than of honest sellers. But Mr B is unaware of how the behaviors of sellers differ in situations $p = \overline{p}$ and $\underline{p}$, thereby explaining his erroneous judgement.

We note that in the equilibrium shown above, when $p = \overline{p}$, and Mrs A honestly reports that $d = 0$ and $b = 1$, Mr B does not trust Mrs A and he rejects her offer. So we see here that when Mrs A is opportunist, her deceptive tactic imposes a cost on honest sellers, who are no longer trusted in some scenarios.

It should also be mentioned that under the conditions of Proposition 3 , we could not sustain an equilibrium in which opportunist sellers always say that $d = 0$ and $b = 1$ irrespective of $p = \overline{p}$ and $\underline{p}$. Indeed, if this were so, after hearing that $d = 0$ and $b = 1$, Mr B would believe that he is facing an honest seller with a probability smaller than the prior μ (since saying $d = 0$ and $b = 1$ would then be more typical of opportunists than of honest sellers: - remember that opportunists say $d = 0$ and $b = 1$ when $p = \underline{p}$). Since $\overline{\gamma} > \gamma^{prior}$, Mr B would then reject Mrs A's offer. By contrast, if Mrs A were to say that $d = 1$ and $b = 1$ when $p = \overline{p}$, she would convince Mr B that she is an honest seller with probability 1 and Mr B would then accept Mrs A's offer, thereby implying that the assumed strategy of Mrs A when $p = \overline{p}$ and $\tau = o$ is not optimal.

In a different vein, it is also worth noting that if Mr B had not distinguished the different categories of sellers (i.e., $\delta_B = 0$), then the deceptive tactic of the opportunist Mrs A would have been pointless, since Mr B would have kept believing that he is facing an honest seller with probability μ whatever Mrs A's announcements, and he would have rejected the deal when $p = \overline{p}$ (given that $\overline{\gamma} > \gamma^{prior}$). Thus, we see that the deceptive tactic of Mrs A requires that Mr B be not too irrational, in the sense that Mr B's cognitive type should allow him to make some inferences from what he observes.

Finally, it is instructive to contrast the insights obtained in our cognitive environment with those that would arise had we assumed that Mr B were fully rational. An important

observation is that when Mr B is rational it can never be optimal for Mrs A, when she is opportunist, to say that $d = 1$ with probability 1 when $p = \overline{p}$. Indeed, if she did so, then Mr B - assumed to be rational - would infer that Mrs A is an honest seller with a probability smaller than the prior μ, and it is not hard to see that this cannot be beneficial to Mrs A (given the extra cost due to the discount imposed by the announcement of $d = 1$).[32] As it turns out, when $\overline{\gamma} > \gamma^{prior}$ and Mr B is rational, Mrs A, when she is opportunist and $p = \overline{p}$, will mix between announcing $d = 0$ or 1 and will announce $b = 1$. Mr B's posterior belief that Mrs A is honest after either ($p = \overline{p}$ and Mrs A says $d = 0$ and $b = 1$) or ($p = \overline{p}$ and Mrs A says $d = 1$ and $b = 1$) must be smaller than μ (due to the martingale property of beliefs), thereby implying that Mr B rejects Mrs A's offer in one of these two cases. Due to the required indifference of Mrs A between announcing ($d = 0$ and $b = 1$) or announcing ($d = 1$ and $b = 1$) we conclude that Mrs A cannot obtain that a transaction with Mr B takes place with probability 1 when $p = \overline{p}$ and $\overline{\gamma} > \gamma^{prior}$. This is, of course, in sharp contrast with what happens in the above cognitive environment as analyzed in Proposition 3.

III. Conclusion

What are the lessons to be drawn from our approach? Firstly, a description of the prototype of a deceptive tactic. In the above bargaining story (as in many real life situations), the first stage of the deception involves building a relation of confidence with the victim (even if this has some cost, as illustrated by the announcement of the deficiencies in the bargaining story) so as to better exploit it at a later stage. Thus, from a practical viewpoint, one is more likely to discover a deceptive tactic when one sees a party making an initial sacrifice that subsequently turns out to be of great benefit to this same party. Secondly, according to our theory, deception requires the presence of agents who are neither fully rational (otherwise, their beliefs could not be manipulated) nor fully irrational, in the sense of not distinguishing the various types of the opponent (otherwise, there could be no inference process as the interaction proceeds). More precisely, our theory of deception requires the presence of agents who somehow have a stereotypical understanding of others' attitudes. Our

[32] Under the condition of Proposition 3 ($\overline{\gamma} > \gamma^{prior}$), Mr B would reject Mrs A's offer. Even if $\overline{\gamma} < \gamma^{prior}$, when Mrs A is opportunist she will prefer to say that $d = 0$ and $b = 1$ so as to save on the discount Δ.

theory thus provides some content to the common-sense idea that the best candidates for belief manipulation and deception are individuals who are neither too smart nor too dumb.

References

[1] Compte, Olivier and Andrew Postlewaite (2008): 'Repeated Relationships with Limits on Information Processing', mimeo.

[2] Crawford, Vincent P. (2003): 'Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions', *American Economic Review*, **93**, 133-149.

[3] Esponda, Ignacio (2008): 'Behavioral Equilibrium in Economies with Adverse Selection', forthcoming *American Economic Review*.

[4] Eyster, Erik and Rabin, Matthew (2005): 'Cursed Equilibrium', *Econometrica*, **73**, 1623-1672.

[5] Fudenberg, Drew and Levine, David K. (1989): 'Reputation and Equilibrium Selection in Games with a Patient Player', *Econometrica*, **57**, 759-778.

[6] Jackson, Matthew O. and Kalai, Ehud (1997): 'Social Learning in Recurring Games,' *Games and Economic Behavior*, **21**, 102-134.

[7] Jehiel, Philippe (1995): 'Limited Horizon Forecast in Repeated Alternate Games', *Journal of Economic Theory*, **67**, 497-519.

[8] Jehiel, Philippe (2005): 'Analogy-based Expectation Equilibrium', *Journal of Economic Theory*, **123**, 81-104.

[9] Jehiel, Philippe and Koessler, Frédéric (2008): 'Revisiting Games of Incomplete Information with Analogy-based Expectations', *Games and Economic Behavior*, **62**, 533-557.

[10] Jehiel, Philippe and Dov Samet (2007): 'Valuation Equilibrium', *Theoretical Economics*, **2**, 163-185.

[11] Jones, Edward E. and Davis Keith E. (1965): 'From Acts to Dispositions: The Attribution Process in Person Perception' in Berkowitz L. (ed), *Advances in Experimental Social Psychology* (Vol 2, 219-266), New York: Academic.

[12] Kahnemann, Daniel., Slovic, P. and Tversky, Amos., eds (1982):'Judgement Under Uncertainty, Heuristic and Biases', Cambridge University Press.

[13] Kreps, David , Milgrom, Paul, Roberts, John and Wilson, Robert (1982): 'Rational cooperation in the finitely repeated prisoners' dilemma', *Journal of Economic Theory*, **27**, 245-252.

[14] Kreps, David and Wilson, Robert (1982): 'Sequential Equilibria', *Econometrica*, **50**, 863-894.

[15] Kreps, David and Wilson, Robert (1982b): 'Reputation and Imperfect Information', *Journal of Economic Theory*, **27**, 253-279.

[16] Lewis, Michael (1990): *Liar's Poker*, Penguin Books.

[17] Mullainathan, Sendhil, Schwartzstein, Joshua and Shleifer, Andrei (2008): 'Coarse Thinking and Persuasion', *Quarterly Journal of Economics*, **123**, 577-619.

[18] McKelvey, Richard and Thomas Palfrey (1995): 'Quantal Response Equilibrium,' *Games and Economic Behavior*, **10**, 6-38

[19] O'Sullivan, Maureen (2003): 'The Fundamental Attribution Error in Detecting Deception: The Boy-Who-Cried-Wolf Effect', *Personality and Social Psychology Bulletin*, **29**, 1316-1327.

[20] Ross, Lee (1977): 'The Intuitive Psychologists and his Shortcomings' in Berkowitz L. (ed), *Advances in Experimental Social Psychology* (Vol 10, 173-220), New York: Academic.

[21] Ross, Lee, Amabile, Teresa M. and Steinmetz, Julia L.(1977):'Social Roles. Social Control and Biases in Social-Perception Processes', *Journal of Personality and Social Psychology*, **35**, 485-494.

[22] Schelling, Thomas (1960): *The Strategy of Conflict.* Harvard University Press.

[23] Sobel, Joel (1985): 'A theory of Credibility', *Review of Economic Studies*, **52**, 557-573.

[24] Spence, Michael A. (1973): 'Job Market Signaling', *Quarterly Journal of Economics*, **87**, 357-374.

[25] Stahl, Dale O. (1993): Evolution of Smart$_n$ players,' *Games and Economic Behavior*, **5**, 604-17'

[26] Von Neuman, John and Morgenstern, Oskar (1944): *Theory of Games and Economic Behavior*, Princeton University Press.

[27] Vrij, Aldert (2001): *Detecting Lies and Deceit.* Wiley, New York.