Calibrated Clustering and Analogy-Based Expectation Equilibrium*

Philippe Jehiel[†] and Giacomo Weber[‡]

February 8, 2024

Abstract

Normal-form two-player games are categorized by players into K analogy classes so as to minimize the prediction error about the behavior of the opponent. This results in Calibrated Analogy-Based Expectation Equilibria in which strategies are analogy-based expectation equilibria given the analogy partitions and analogy partitions minimize the prediction errors given the strategies. We distinguish between environments with self-repelling analogy partitions in which some mixing over partitions is required and environments with self-attractive partitions in which several analogy partitions can arise, thereby suggesting new channels of belief heterogeneity and equilibrium multiplicity.

Keywords: Analogy-based Expectation Equilibrium, Prototype theory, K-mean clustering.

JEL Classification Numbers: D44, D82, D90

*We thank seminar audiences at the Parisian Seminar of Game Theory, the TOM Seminar at PSE, the CES Economic Theory Seminar at Paris 1, at the Bounded Rationality: Theory and Experiments workshop in Tel Aviv, Berlin, Vienna, ESSET 2023, AMES 2023, and the Summer School of Econometric Society in Economic Theory 2023 for comments. We thank Evan Friedman and Frédéric Koessler for useful suggestions. Jehiel thanks the European Research Council for funding (grant n° 742816).

[†]PSE, 48 boulevard Jourdan, 75014 Paris, France and University College London; jehiel@enpc.fr

[‡]PSE, 48 boulevard Jourdan, 75014 Paris, France; giacomo.weber@psemail.eu

1 Introduction

Many economists recognize that the rational expectation hypothesis that is central in solution concepts such as the Nash equilibrium is very demanding, especially in complex multi-agent environments involving lots of different situations (games, states or nodes, depending on the application). Several approaches have been proposed to relax it. When the concern with the hypothesis is that there are too many situations for players to fine tune a specific expectation for each such situation, a natural approach consists in allowing players to lump together situations into just a few categories, and only require that players form expectations about the aggregate play in each category.

The analogy-based expectation equilibrium (Jehiel, 2005) is a solution concept that has been proposed to deal with this. In addition to the usual primitives describing a game form, players are also endowed with analogy partitions, which are player-specific ways of partitioning situations or contingencies in the grand game. In equilibrium, the expectations in each analogy class correctly represent the aggregate behavior in the class, and players best-respond as if the behavior in every element of an analogy class matched the expectation about the aggregate play in the corresponding analogy class. This approach has been developed and applied to a variety of settings, but in almost all these developments, the analogy partitions are taken as exogenous (see Jehiel (2022) for a recent exposition of this strand of literature).

From a different perspective, psychologists have long recognized the use of categories to facilitate decision making and predictions (see, in particular, Anderson (1991) on predictions). For psychologists, a categorization bundles distinct objects into groups or categories, whose members are viewed as sufficiently similar to warrant a similar treatment. In some leading approaches, there is a prototype in each category that can be viewed as a representative -possibly fictitious- object in this category (say the mean or the mode of the objects in the category, see Rosch (1978) for an early account of the literature on prototype theory),[1] and categories are defined so that objects are assigned to the category with nearest prototype (see Posner and Keele (1968) or Reed (1972)).

From yet another perspective, the K-means clustering technique considered in Machine Learning has been widely used as a method to categorize exogenously given datapoints into a pre-specified number K of clusters (see Steinhaus (1957), Lloyd (1957) and MacQueen (1967)). From that perspective, datapoints are the primitives, and the clustering problem consists in partitioning the datapoints into K clusters with representative points for each cluster defined so that the original datapoints are best approximated by the representative points in their cluster. Solving the clustering problem (defined as deriving the variance-minimizing categorization, say) is hard (NP-complete), and practitioners most of the time rely on a simple algorithm to approximate its solution (see the end of Subsection 2.3 for a succinct exposition of the algorithm).

In this paper, we propose endogenizing the choice of partitions in the analogy-based expectation setting on the basis of the above basic principles. Specifically, we consider a strategic environment consisting of different normal form two-player games drawn by nature according to some prior distribution where we have in mind that the various games are played at many different times by many different subjects. In each of the normal form games $\omega \in \Omega$, player $j = 1, 2$ has the same action set A_j. An analogy partition for player i takes the form of a partition of the set of games Ω, which is used by player i to assess the behavior of player j in the various games. The data points accessible by players consist of the empirical frequencies of past play of the subjects assigned to the role of the opponent in the various games. That is, a typical data point for player i consists of an element of ΔA_j for each of the games ω. To make sense of these data points (and prior to knowing which specific game ω will

[1] An alternative approach in psychology is that of exemplar theory (Medin and Schaffer 1978, Nosofsky 1986)) in which only real existing objects are considered to describe the category. Such an alternative approach would introduce an element of stochasticity in the choice of representative exemplar that is somehow orthogonal to the focus of the present study, hence our primary reference to the prototype approach.

apply), a subject assigned to the role of player i is viewed as clustering these data points into an exogenously given (typically small) number K of categories where K can be related to the number of items human beings can remember in short-term memory (see Miller (1956) for pioneering research on this). While we are agnostic on how the clustering is achieved by a subject, we require in our approach that each game is assigned to the cluster with nearest representative behavior of the opponent and that the representative behavior in a cluster is identified with the mean behavior of the opponent across the games assigned to the cluster.[2] When called (later) to pick an action in a randomly selected game ω, this subject then identifies the behavior of his opponent in this game ω with the representative expectation that comes out from the clustering stage, and best-responds to it. This in turn generates new data points, and we are interested in the steady states - referred to as calibrated analogy-based expectation equilibria - generated by such dynamic processes.

Roughly, the calibrated analogy-based expectation equilibria (C-ABEE) can be described as profiles of analogy partitions and strategies such that i) given the analogy partitions, players' strategies form an analogy-based expectation equilibrium and ii) given the strategies, clustering leads players to adopt the analogy partitions considered in steady state.

Different formalizations of clustering can be considered whether we only require that games are assigned to categories such that the observed opponent's behavior is best approximated by the representative behavior in the category (local criterion, similarly as in the psychological perspective suggested above) or else whether we consider an exact resolution of the clustering problem (variance minimization, say). But, no matter what approach to clustering is adopted, a key observation is that it may not be possible in some cases to have a steady state with a single analogy partition for each player. This is so because unlike in

[2] In other words, the clustering is based on the attribute of the data consisting of the opponent's behavior as opposed to other aspects of the game ω. While other attributes of ω could be considered as well, we note that the opponent's behavior is the essential attribute needed to determine the best-response (and this is the only decision to be made by our subjects). Our main insights would not be qualitatively affected by the consideration of extra attributes in addition to the opponent's behavior.

the usual clustering problem, there is here an extra endogeneity of the dataset. A change in analogy partitions may affect the adopted strategies through the working of the analogy-based expectation equilibrium, which in turn may affect how clustering is done. This extra channel from the clustering to the dataset makes it sometimes impossible to have a calibrated analogy-based expectation equilibrium with a single analogy partition for each player. This will be first illustrated with a simple example involving three matching pennies games.

This observation leads us to extend the basic definition of C-ABEE to allow for distributions over analogy partitions defined so that that each analogy partition in the support is required to solve (either locally or globally) the clustering problem for that player and the strategies now also parameterized by the chosen analogy partition satisfy the requirements of the analogy-based expectation equilibrium appropriately extended to cope with distributions over analogy partitions. We refer to such an extension as a calibrated distributional analogy-based expectation equilibrium (CD-ABEE).

We show that in finite environments (i.e. environments such that there are finitely many normal form games and finitely many actions for each player), there always exists at least one CD-ABEE. We also sketch a dynamic learning model making more precise the learning environment that would give rise to CD-ABEE as steady states. The learning environment involves populations of subjects assigned to the roles of the various players. When several analogy partitions are required in steady state, it implies that different subjects assigned to the role of the same player would end up categorizing the various games differently, despite being exposed to the same objective datasets.[3]

We next consider various applications. A common theme that we consider throughout is whether the environment is such that starting from any candidate analogy partition profile, we obtain behaviors (through the machinery of ABEE) that would lead to the re-categorization of at least one game (through the

[3]We rationalize this heterogeneity of clustering by introducing subject-specific (measurement) errors in the observations made by the learning subjects, thereby introducing a stochastic element in the spirit of Harsanyi's purification argument, introduced to rationalize the mixing in Nash equilibria involving mixed strategies.

machinery of clustering) or else whether the environment is such that starting from several (sometimes many) analogy partition profiles, we obtain (ABEE) behaviors that in turn lead to the same partition profile from the clustering perspective. We say that analogy partitions are self-attractive in the former case and self-repelling in the latter.

We first provide an illustration of an environment with self-attractive analogy partitions in the context of beauty contest games in which players care both about being close to a fundamental value (which parametrizes the game) and being close to what the opponent is expected to be choosing. We show that when the concern for coordination with the opponent is sufficiently strong, virtually all analogy partitions of the fundamental value can be used to construct a CABEE, and as a result many different strategies can arise in a CABEE. This illustrates in a very extreme way a case of self-attractive analogy partitions that we believe provide a novel perspective on the phenomenon of echo chambers.

We next illustrate the possibility of self-repelling analogy partitions in the context of a monitoring game involving an employer who has to choose whether or not to exert some control and a worker who can be one of three possible types (which defines the game), and has to decide his effort attitude. We assume that one type of worker always exerts low effort, one type always exerts high effort and the last type exerts high effort only if he expects to be controlled with high enough probability. This defines an environment with self-repelling analogy partition when the employer can use only two categories and she finds control best only when the worker is expected to be shirking with high enough probability. We observe that when the type whose effort choice depends on the expectation is the least likely, it is always bundled with one of the other types in all CD-ABEE. Such an illustration can be used to shed new light on discrimination and why there may be heterogeneity in how the effort attitude of minority groups is assessed in societies.

Finally, we consider families of games with linear best-responses parameterized by the magnitude of the impact of opponent's action on the best-response (a one-dimensional parameter). We analyze separately the case of strategic complements and the case of strategic substitutes allowing us to cover applications such as Bertrand or Cournot duopoly with product differentiation, linear demand and constant marginal costs, or moral hazard in teams. We show that the strategic complements case corresponds to an environment with self-attractive analogy partitions and the strategic substitute case corresponds to an environment with self-repelling analogy partitions.

In the rest of the paper, we develop the framework (solution concepts, existence results, learning foundation) in Section 2. We develop applications in Section 3. We conclude in Section 4.

1.1 Related Literature

This paper belongs to a growing literature in behavioral game theory, proposing new forms of equilibrium designed to capture various aspects of misperceptions or cognitive limitations. While some papers in this strand posit some misperceptions of the players and propose a corresponding notion of equilibrium (see Eyster-Rabin (2005) on misperceptions about how private information affects behavior, Spiegler (2016) on misperceptions on the causality links between variables of interest or Esponda-Pouzo (2016) for a more abstract and general formulation of misspecifications), other papers motivate their equilibrium approach by the difficulty players may face when trying to understand or learn how their environment behaves (see Jehiel (1995) on limited horizon forecasts, Osborne-Rubinstein (1998) on sampling equilibrium, Jehiel (2005) on analogical reasoning or Jehiel-Samet (2007) on coarse reinforcement learning).[4] Our paper has a motivation more in line with the latter, but it adds structure on the coarsening of the learning based on insights or techniques borrowed from psychology and/or machine learning (which the previous literature just mentioned did not consider).

This paper also relates to papers dealing with coarse or categorical thinking

[4] See also Mullainathan et al. (2008) for a study of persuasion when listeners make inferences through coarse categories.

in decision-making settings (see, in particular, Fryer and Jackson (2008) for such a model used to analyze stereotypes or discrimination, Peski (2011) for establishing the optimality of categorical reasoning in symmetric settings or Al-Najjar and Pai (2014) and Mohlin (2014) for models establishing the superiority of using not too fine categories in an attempt to mitigate overfitting or balance the bias-variance trade-off). Our paper considers a clustering technique not discussed in those papers, and due to the strategic character of our environment the data-generating process in our setting is itself affected by the categorization, which these papers did not consider.

Finally, a contemporaneous and alternative approach to categorization in the context of the analogy-based expectation equilibrium is that introduced in Jehiel and Mohlin (2023) who propose putting structure on the analogy partitions based on the bias-variance trade-off where an exogenously given notion of distance between the various nodes or information sets is considered by the players. This is a different approach to categorization than the one considered in this paper, and in the context considered here with a set of normal form games (all arising with positive probability) it would lead to consider the finest analogy partition and accordingly the Nash Equilibrium, in sharp contrast with the findings obtained here.[5]

2 Theoretical setup

2.1 Strategic environment

We consider a finite number of normal form games indexed by $\omega \in \Omega$ where game ω is chosen (by Nature) with probability $p(\omega)$. To simplify the exposition, we restrict attention to games with two players $i = 1, 2$, and we refer to player

[5]Some other papers consider categorization in games (see in particular Samuelson (2001), Mengel (2012) or Gibbons et al. (2021)) with the view that the strategy should be measurable with respect to the categorization. This is somewhat different from the expectation perspective adopted here. It may be mentioned that Gibbons et al. (2021) discuss the possibility that some third party could influence how categorizations are chosen, which differs from our perspective that views the categories as being chosen by the players themselves.

j as the player other than i.[6] In every game ω, the action space of player i is the same and denoted by A_i. It is assumed in this part to be finite. The payoff of player i in game ω is described by a von Neumann-Morgenstern utility where $u_i(a_i, a_j, \omega)$ denotes the payoff obtained by player i in game ω if player i chooses $a_i \in A_i$ and player j chooses $a_j \in A_j$. Let $p_i \in \Delta A_i$ denote a probability distribution over A_i for $i = 1, 2$. With some abuse of notation, we let:

$$u_i(p_i, p_j, \omega) = \sum_{a_i, a_j} p_i(a_i) p_j(a_j) u_i(a_i, a_j, \omega)$$

denote the expected utility obtained by player i in game ω when players i and j play according to p_i and p_j, respectively.

We assume that players observe the game ω they are in when choosing their actions. A strategy for player i is denoted $\sigma_i = (\sigma_i(\omega))_{\omega \in \Omega}$ where $\sigma_i(\omega) \in \Delta A_i$ denotes the (possibly mixed) strategy employed by player i in game ω. The set of player i's strategies is denoted Σ_i, and we let $\Sigma = \Sigma_i \times \Sigma_j$.

A Nash equilibrium is a strategy profile $\sigma = (\sigma_i, \sigma_j) \in \Sigma$ such that for every player i, $\omega \in \Omega$, and $p_i \in \Delta A_i$,

$$u_i(\sigma_i(\omega), \sigma_j(\omega), \omega) \geq u_i(p_i, \sigma_j(\omega), \omega).$$

2.2 Analogy-based expectation equilibrium

Players are not viewed as being able to know or learn the strategy of their opponent for each game ω separately as implicitly required in Nash equilibrium. Maybe because there are too many games ω, they are assumed to learn the strategy of their opponent only in aggregate over collections of games, referred to as analogy classes. Throughout the paper, we impose that player i considers K_i different analogy classes, where K_i is kept fixed. Such a bound can be the result of constraints on memory (see the introduction), and we will have in mind that K_i is no greater and typically (possibly much) smaller than $|\Omega|$, the

[6] The framework, solution concept and analysis extend in a straightforward way to the case of more than two players.

number of possible normal form games. We refer to $\mathcal{K}_i$ as the set of partitions of Ω with K_i elements. Formally, considering for now the case of a single analogy partition for each player i, we let $An_i = \left\{\alpha_i^1, \ldots \alpha_i^{K_i}\right\}$ denote the analogy partition of player i. It is a partition of the set Ω of games with K_i classes, hence an element of $\mathcal{K}_i$. For each $\omega \in \Omega$, we let $\alpha_i(\omega)$ denote the (unique) analogy class to which ω belongs.

$\beta_i(\alpha_i) \in \Delta A_j$ will refer to the (analogy-based) expectation of player i in the analogy class α_i. It represents the aggregate behavior of player j across the various games ω in α_i.

We say that β_i is *consistent* with σ_j whenever for all $\alpha_i \in An_i$,

$$\beta_i(\alpha_i) = \sum_{\omega \in \alpha_i} p(\omega)\sigma_j(\omega) / \sum_{\omega \in \alpha_i} p(\omega).$$

In other words, consistency means that the analogy-based expectations correctly represent the aggregate behaviors in each analogy class when the play is governed by σ.

We say that σ_i is a *best-response* to β_i whenever for all $\omega \in \Omega$ and all $p_i \in \Delta A_i$,

$$u_i(\sigma_i(\omega), \beta_i(\alpha_i(\omega)), \omega) \geq u_i(p_i, \beta_i(\alpha_i(\omega)), \omega).$$

In other words, player i best-responds in ω as if player j played according to $\beta_i(\alpha_i(\omega))$ in this game, which can be viewed as the simplest representation of player j's strategy given the coarse knowledge provided by β_i.

Definition 1 *Given the strategic environment and the profile of analogy partitions $An = (An_i, An_j)$, σ is an analogy-based expectation equilibrium (ABEE) if and only if there exists a profile of analogy-based expectations $\beta = (\beta_i, \beta_j)$ such that for each player i (i) σ_i is a best-response to β_i and (ii) β_i is consistent with σ_j.*

This concept has been introduced with greater generality in Jehiel (2005) (allowing for multiple stages and more than two players) and in Jehiel and

Koessler (2008) (allowing for private information).[7] We have chosen a simpler environment to focus on the choice of analogy partitions, which is the main concern of this paper.

2.3 Calibrated clustering

Psychologists have long recognized the use of categories to facilitate decision making and predictions (see, in particular, Anderson (1991) on predictions). A categorization bundles distinct objects into groups or categories, whose members are viewed as sufficiently similar to warrant a similar treatment. In each category, there is a prototype that can be viewed as a representative element in the category (say the mean or the mode, see Rosch 1978). And categories are defined so that objects are assigned to the category with nearest prototype (Posner and Keele (1968) or Reed (1972)). Another approach to categorization would require that objects are categorized so as to minimize some measure of dispersion such as the variance (or the relative entropy when objects can be identified with probability distributions, as in our environment).

In our framework, the objects considered by a subject assigned to the role of player i consist of the frequencies of choices of the subjects assigned to the role of player j in the various normal form games ω (this is made explicit when describing the learning environment, see section 2.6). That is, for player i, the objects are $\{(\omega, \sigma_j(\omega)), \omega \in \Omega\}$. An obvious attribute of $(\omega, \sigma_j(\omega))$ is the distribution of opponent's behavior as described by $\sigma_j(\omega)$, and we will assume that player i focuses on that attribute when choosing his categorization. This will in turn lead the chosen categorization to minimize (either locally or globally) the prediction error about the play of the opponent. This feature, we believe, would naturally be regarded as desirable by player i, as predicting the opponent play is the only thing player i cares about to determine his best-response.[8]

[7] See Jehiel (2022) for a definition in a setting covering both aspects and allowing for distributions over analogy partitions.

[8] Considering extra attributes in relation to ω would not raise conceptual difficulties, and

We first introduce a notion of approximation in the space of distributions over actions. Formally, for three distributions of player j's play p_j, p'_j and p''_j in ΔA_j, we say that p_j is less well approximated by p'_j than by p''_j whenever $d(p_j, p'_j) > d(p_j, p''_j)$ where d is a divergence function that will either be the square of the Euclidean distance $d(p_j, p'_j) = \left\|p_j - p'_j\right\|^2$ (as defined over $(\Delta A_j)^2$) or the Kullback-Leibler divergence applied to distributions $d(p_j, p'_j) = \sum_{a_j} p_j(a_j) \ln \frac{p_j(a_j)}{p'_j(a_j)}$ or in the applications to be developed later in which best-responses are linear, $d(p_j, p'_j) = \left(E(p_j) - E(p'_j)\right)^2$.

We next define two notions of calibrations relating the analogy partition of player i to the strategy of player j.

Definition 2 *A partition An_i of Ω is locally calibrated with respect to σ_j iff for every classes α_i, α'_i of An_i and every $\omega \in \alpha_i$,*

$$d(\sigma_j(\omega), \beta_i(\alpha_i)) \leq d(\sigma_j(\omega), \beta_i(\alpha'_i)).$$

It is globally calibrated with respect to σ_j iff An_i belongs to

$$\arg\min_{P_i \in \mathcal{K}_i} \sum_{c_i \in P_i} p(c_i) \sum_{\omega \in c_i} p(\omega \mid c_i) d(\sigma_j(\omega), \beta_i(c_i)) \tag{1}$$

where, for all $c \subseteq \Omega$, $\beta_i(c) = \sum_{\omega \in c} p(\omega \mid c)\sigma_j(\omega)$.

In the above definition, $\beta_i(c)$ is viewed as the prototype in category c and it is defined as the mean of the elements assigned to c. Local calibration retains the idea that objects should be assigned to the category with nearest prototype. Global calibration on the other hand requires that categorizations are chosen to minimize dispersion as measured by (1). When d is the square of the Euclidean distance, the measure of dispersion corresponds to variance. When d is the the Kullback-Leibler divergence, the measure of dispersion corresponds to relative entropy. We note that in both cases, choosing the prototype to be the mean of

our main insights would carry over as long as some positive weight is given to this attribute.

the elements assigned to the category is optimal (in the sense of minimizing the dispersion measure),[9] thereby giving a further argument for using the mean as the prototype in our context.

It should be stressed (and as formally established in Lemma 1 in the Online Appendix) that global calibration implies local calibration for the d as specified above. As it turns out, the local calibration conditions can be viewed as the first order conditions for the minimization problem (1).

Link to K means clustering.

In Machine Learning, a very popular way to categorize object is based on the K-means clustering algorithm (Steinhaus (1957), Lloyd (1957) and MacQueen (1967)). The objective of clustering is also to minimize variance (or relative entropy) as considered in our global calibration approach, but this is known to be NP hard in computer science. Practitioners instead use the following algorithm. Representative points are initially drawn, then at each subsequent iteration of the algorithm, first points are allocated to the cluster with closest representative, then, a new representative point, identified with the mean of the points allocated to the cluster, is determined in each cluster. Such an algorithm can be shown to converge, and it turns out that it necessarily converges to a local minimizer of the dispersion measure. Moreover, all local minimizers satisfy the local calibration condition requiring that points are assigned to the cluster with closest representative point. In light of this, one may interpret subjects who would have found analogy partitions that satisfy the local calibration requirements, as ones using the K-means clustering algorithm (but, note that we are agnostic as to whether real human subjects rely on such methods to find out their categorizations).

[9] Formally, for $i = 1, 2$ and any subset α_i of Ω, let d be either the square of the Euclidean distance or the Kullback-Leibler divergence, we have that

$$\sum_{\omega \in \alpha_i} p(\omega \mid \alpha_i)\sigma_j(\omega) = \arg\min_{q \in \Delta A_j} \sum_{\omega \in \alpha_i} p(\omega \mid \alpha_i) d(\sigma_j(\omega), q).$$

2.4 Calibrated analogy-based expectation equilibrium

Combining Definitions 1 and 2 yield:

Definition 3 *A pair (σ, An) of strategy profile $\sigma = (\sigma_i, \sigma_j)$ and analogy partition profile $An = (An_i, An_j) \in \mathcal{K}_i \times \mathcal{K}_j$ is a locally (resp. globally) calibrated analogy-based expectation equilibrium iff (i) σ is an analogy-based expectation equilibrium given An and (ii) for each player i, An_i is locally (resp. globally) calibrated with respect to σ_j.*

The more interesting and novel aspect in this definition is the fixed point element linking analogy partitions to strategies and vice versa. With respect to the previous papers (using the ABEE framework), it suggests a way to endogenize the analogy partitions (given the numbers K_i and K_j of allowed analogy classes). With respect to the clustering literature, the novel aspect is that the set of points to be clustered $(\sigma_j(\omega))_{\omega\in\Omega}$ for player i is itself possibly influenced by the shape of the clustering, as captured by the analogy-based expectation equilibrium.

When either player 1 or 2 has a dominant strategy in all games $\omega \in \Omega$, there always exists a (locally or globally) calibrated ABEE. To see this, suppose player i has a dominant strategy in all ω. The behavior of player i coincides with the dominant strategy irrespective of the profile of analogy partitions. This ensures that on player j's side, the analogy partition can simply be obtained by using the standard clustering techniques applied to the exogenous dataset given by player i's dominant strategy in the various games. Once such a clustering is derived, the rest of the construction of a calibrated ABEE is easily derived.

Also, if the number of games ω with different Nash equilibrium strategies is no larger than K_i for each player i , a (globally) calibrated ABEE is readily obtained by requiring that players play such a Nash equilibrium in each game and that they bundle games in the same analogy class, only when these games have the same Nash equilibrium. This is so because the opponent's behavior is then the same across the various games assigned to the same analogy class, and thus each player has correct expectations, thereby allowing to support the Nash equilibrium strategies in a calibrated ABEE.[10]

But, in general, a calibrated analogy-based expectation equilibrium may fail to exist. We illustrate this in a context with three matching penny games with different parameter values assuming that one of the players can use only two categories.

Example 1. Let $x = a, b, c$, with $0 < a < b < c < 2$. The following three games are played, each with probability $\frac{1}{3}$. The corresponding payoff matrices are given by:

G_x	L	R
U	$(1+x, 0)$	$(0, 1)$
D	$(0, 1)$	$(1, 0)$

Proposition 1 *Assume that $K_1 = 2$ and $K_2 = 3$, and d is the square of the Euclidean distance in the space of probability distributions on $\{L, R\}$. There is no Calibrated ABEE.*

Roughly, Proposition 1 can be understood as follows. Matching pennies games are such that the Nash equilibrium involves some mixing. When the Row player puts two games x and x' in the same analogy class, she can be mixing in at most one of these games (this is so because the incentives of Row are different in the two games and Row makes the same expectation about Column in both x and x' when these belong to the same analogy class). This in turn induces some polarization of the behaviors of both players in the two games x and x', which when a, b and c are not too far apart leads the Row player to re-categorize one of the two bundled games x or x' with the left alone game $x'' \neq x, x'$. Details appear in Appendix A.

[10] Observe that whenever all the objects in a category coincide exactly with the prototype, then the correct expectations would lead to play optimally. This would not be so in general if the clustering were based on alternative attributes (say, player i's own payoff structure), thereby giving a normative appeal to the use of opponent's behavior as the main attribute for clustering purpose.

2.5 Calibrated Distributional analogy-based expectation equilibrium

We address the existence issue by adopting a distributional approach (that will be interpreted from a learning perspective in the next subsection). Formally, we allow the analogy partition An_i of player i to take different realizations in $\mathcal{K}_i$, and we refer to λ_i as the distribution of An_i over $\mathcal{K}_i$. The distributions of analogy partitions of the two players are viewed as independent of one another (formalizing a random assignment assumption, see below the section on learning). We refer to $\lambda = (\lambda_i, \lambda_j)$ as the profile of these distributions, and we let $\Lambda = \Delta\mathcal{K}_i \times \Delta\mathcal{K}_j$ be the set of (λ_i, λ_j). For each analogy partition An_i of player i in the support of λ_i referred to as $Supp\lambda_i$, we let $\sigma_i(\cdot \mid An_i) : \Omega \to \Delta A_i$ refer to the mapping describing how player i with analogy partition An_i behaves in the various games $\omega \in \Omega$. We refer to $\sigma_i = (\sigma_i(\cdot \mid An_i))_{An_i \in Supp\lambda_i}$ as player i's strategy, and we let $\sigma = (\sigma_i, \sigma_j)$ denote the strategy profile, the set of which is still denoted Σ.

Given $\lambda \in \Lambda$ and $\sigma \in \Sigma$, we can define the aggregate behaviors of the two players in each game, as aggregated over the various realizations of analogy partitions. We have in mind that these aggregate behaviors in the various games $\omega \in \Omega$ constitute the only data accessible to players, thereby implying that only these aggregates are used to construct the analogy-based expectations and implement the clustering of the games. Formally, the aggregate strategy of player j in game ω is given by

$$\overline{\sigma}_j(\omega) = \sum_{An_j \in \mathcal{K}_j} \lambda_j(An_j)\sigma_j(\omega \mid An_j). \tag{2}$$

Let $\bar{\sigma} = (\bar{\sigma}_i, \bar{\sigma}_j)$ denote a profile of aggregate strategies and let $\overline{\Sigma}$ denote the set of such profiles.

The analogy-based expectation of player i defines for each analogy partition $An_i \in Supp\lambda_i$ and each analogy class $\alpha_i \in An_i$, the aggregate behavior of player j in α_i denoted by $\beta_i(\alpha_i \mid An_i) \in \Delta A_j$ (the dependence on An_i is here to stress that player i with analogy partition An_i considers only the aggregate behaviors in the various analogy classes in An_i). Similarly as above, $\beta_i(\cdot \mid An_i)$ is said to be consistent with $\overline{\sigma}_j$ iff, for all $\alpha_i \in An_i$,

$$\beta_i(\alpha_i \mid An_i) = \sum_{\omega \in \alpha_i} p(\omega)\overline{\sigma}_j(\omega) / \sum_{\omega \in \alpha_i} p(\omega). \tag{3}$$

We are now ready to propose the distributional extensions of our previous definitions.

Definition 4 *Given $\lambda = (\lambda_i, \lambda_j) \in \Lambda$, a strategy profile $\sigma = (\sigma_i, \sigma_j) \in \Sigma$ is a distributional analogy-based expectation equilibrium (ABEE) iff there exists $\beta = (\beta_i, \beta_j)$ such that for every player i and $An_i \in Supp\lambda_i$, we have that i) $\sigma_i(\cdot \mid An_i)$ is a best-response to $\beta_i(\cdot \mid An_i)$ and ii) $\beta_i(\cdot \mid An_i)$ is consistent with $\overline{\sigma}_j$ (where $\overline{\sigma}_j$ is derived from σ_j as in (2)).*

Definition 5 *A pair $(\sigma, \lambda) \in \Sigma \times \Lambda$ is a locally (resp. globally) calibrated distributional analogy-based expected equilibrium iff i) σ is a distributional ABEE given λ, and ii) for every player i and $An_i \in Supp\lambda_i$ (where $\lambda = (\lambda_i, \lambda_j)$), An_i is locally (resp. globally) calibrated with respect to $\overline{\sigma}_j$ (where $\overline{\sigma}_j$ is derived from σ_j as in (2)).*

Clearly, a calibrated distributional ABEE coincides with a calibrated ABEE if the distributions of analogy partitions assign probability 1 to a single analogy partition for both players i and j. Calibrated distributional ABEE are thus generalizations of calibrated ABEE. We now establish an existence result.

Theorem 1 *In finite environments, there always exists a locally (resp. globally) calibrated distributional ABEE when d is the square of the Euclidean distance or the Kullback-Leibler divergence.*

We prove this result by making use of Kakutani's fixed point theorem. Details appear in Appendix A.[11]

[11] We also provide in the online Appendix a description of the calibrated distributional ABEE in the context of the matching pennies environment of Example 1.

Comment. An implication of calibrated distributional ABEE that would involve several analogy partitions is that different subjects exposed to the same objective datasets (and the same memory constraints as summarized by the number of allowed categories) may end up choosing different analogy partitions. Such a motive for an heterogeneous way of processing the same objective dataset is a consequence of the link between the categorizations and the datasets (through ABEE) and it has no analog in the literature.

2.6 Learning foundation

Consider the following learning environment. There are populations of mass 1 of subjects assigned to the roles of player 1 and 2. At time 0, subjects are randomly matched, they are informed of the game ω they are in and they play. The play at time 0 is a parameter of the learning environment. At all other time periods t, there are two stages.

- In stage 1, subjects assigned to the role of i see the datasets $\left\{(\omega, \overline{\sigma}_j^{t-1}(\omega)), \omega \in \Omega\right\}$, where $\overline{\sigma}_j^{t-1}(\omega)$ denotes the aggregate frequencies of actions of subjects assigned to the role of j observed at $t-1$ when the game was ω. We consider the possibility of measurement error by which we mean that each observation $\overline{\sigma}_j^{t-1}(\omega)$ may be perturbed by a (small) subject-specific idiosyncratic term. Every subject i implements a clustering of his dataset into K_i categories. That is, he either implements a solution to (1) or he finds a local solution to this problem (say resulting from the implementation of the K-means algorithm). At the end of this, a subject is able to recognize to which category/analogy class $\alpha_i^t(\omega)$ a game ω is assigned as well as the representative point $\beta_i^t(\alpha_i^t(\omega))$ in $\alpha_i^t(\omega)$ defined as the mean of the elements assigned to $\alpha_i^t(\omega)$.

- In stage 2, subjects are randomly matched and each subject is informed of the game ω he is in. He then expects that the subject j he is matched with will play according to $\beta_i^t(\alpha_i^t(\omega))$, which is the representative behavior in the analogy class to which ω has been assigned. Subject i plays a best-response to this expectation. At the best-response stage, we consider the possibility of small perturbations in the payoff specifications as is commonly assumed in learning models (Fudenberg and Kreps, 1993).

In Appendix B, we establish that the dynamic model just proposed admits steady states. Moreover, when subjects solve the full clustering problem (i.e., solving (1) in stage 1), we show that the limit of these steady states as the measurement error and the payoff perturbations vanish correspond to the globally calibrated distributional ABEE. These results provide a learning foundation for the globally calibrated distributional ABEE.[12]

3 Applications

In this Section we consider various applications. We focus throughout on whether the environment has multiple self-attractive analogy partitions or only self-repelling analogy partitions. In the former case, several choices of analogy partitions may lead to behaviors (through the machinery of ABEE) that agree (form the clustering perspective) to the initial choice of analogy partitions. In the latter case, any choice of analogy partitions leads to behaviors (through the machinery of ABEE) in at least one game that would have to be reallocated to another analogy class (from the clustering perspective).[13]

Clearly, environments in which there is no pure CABEE are self-repelling ones. While the matching penny environment discussed above gives an illustration of this, we will cover more applications in which this arises. We will also illustrate that the polar case of self-attractive analogy partitions can arise in classic environments, thereby suggesting a novel source of multiplicity of equilibria in such cases.

[12]In Jehiel and Weber (2023) (the WP version), we discuss why more work would be needed to provide a learning foundation for the locally calibrated distributional ABEE.

[13]We use the labels attractive and repelling by analogy with their use in magnetic fields.

3.1 Self-Attractive Analogy Partitions: A Beauty Contest Game Illustration

We consider a family of games that induce strategic behavior in the spirit of the "beauty contest" example mentioned in Keynes's General Theory (1936). These games are similar to those introduced in Morris and Shin (2002) except that in our setting there is no private information and players form their expectations in a coarse way.

Formally, a fundamental value (playing the role of the state in the above construction) can take values θ in $\Theta \subset \mathbb{R}$ and θ is assumed to be distributed according to some smooth density $f(\cdot)$ on Θ. Player i has to choose an action $a_i \in \mathbb{R}$.

When the fundamental value is θ, player i 's utility is

$$U_i(a_i, a_{-i}; \theta) = -(1-r)(a_i - \theta)^2 - r(a_i - a_j)^2,$$

where $0 < r < 1$. In other words, players would like to choose an action that is close both to the fundamental value and to the action chosen by the opponent where r measures the weight attached to the latter and $1-r$ the weight attached to the former. It is the coordination aspect that gives to this game the flavor of the beauty contest game.[14]

As in the framework of Section 2, players are assumed to observe θ. The quadratic loss formulation implies that if player i expects j to play according to the distribution $\sigma_j \in \Delta\mathbb{R}$ the best-response of player i in game θ is

$$BR(\theta, \sigma_j) = (1-r)\theta + rE(a_j \mid \sigma_j).$$

[14] While these games are usually presented with more than two players, we note that the analysis to be presented now is the same whether there are two or more players (where in the latter case, one should require that a player wants to coordinate with the mean action of the others). Furthermore, while this environment has a continuum of games and a continuum of actions, extensions of the definitions provided above for finite environments are straightforward in this case.

The unique Nash Equilibrium in game θ requires then that

$$a_1^{NE}(\theta) = a_2^{NE}(\theta) = \theta.$$

Consider now this same environment assuming players use K categories as in Section 2. Given the symmetry of the problem, we focus on symmetric equilibria in which players 1 and 2 would both choose the same analogy partition $(\Theta_k)_{k=1}^K$ and we let for each k

$$\overline{\theta}_k = E(\theta \mid \theta \in \Theta_k)$$

denote the mean of the fundamental values conditional on θ lying in the analogy class Θ_k. Straightforward calculations (detailed for completeness in the online Appendix) show that the analogy-based expectation equilibrium requires that for $\theta \in \Theta_k$

$$a_1^{ABEE}(\theta) = a_2^{ABEE}(\theta) = (1-r)\theta + r\overline{\theta}_k. \tag{4}$$

In this equilibrium, players do not choose the fundamental value θ because their coarse expectation leads them to expect the mean action of the opponent to be $\overline{\theta}_k$ and not θ when $\theta \in \Theta_k$.

Assume that players use the square of the Euclidean distance applied to the mean of the distribution for calibration purposes. That is, for global calibration, given the behaviors $a_j^{ABEE}(\theta)$, player i seeks a partition $(\Theta'_k)_{k=1}^K$ that minimizes

$$\sum_{k=1}^K \int_{\Theta'_k} (a_j^{ABEE}(\theta) - E(a_j^{ABEE}(\theta') \mid \theta' \in \Theta'_k))^2 f(\theta) d\theta.$$

Our main insight is the observation that there are many possible Calibrated Analogy-based Expectation Equilibria when the concern for coordination is large enough. More precisely,

Proposition 2 *Take any partition $(\Theta_k)_{k=1}^K$ such that $\overline{\theta}_k = E(\theta \mid \theta \in \Theta_k)$ are all different. Then for r sufficiently close to* 1*, when both players use this analogy partition and play according to (4), we have a Calibrated Analogy-based Expectation Equilibrium.*

Proof. When r is close to 1, actions $a_j^{ABEE}(\theta)$ in Θ_k are all close to $\overline{\theta}_k$. When $\overline{\theta}_k$ are all distinct, the calibration (whether local or global) leads to $(\Theta_k)_{k=1}^K$. **Q.E.D.**

In other words, our beauty contest game illustrates in a stark way the possibility of self-attractive analogy partitions. When r is close to 1, virtually all analogy partitions can be sustained as part of a calibrated ABEE.[15] Or to put it differently: Once the analogy partition $(\Theta_k)_{k=1}^K$ is chosen and no matter what this partition is, players are led through the working of the ABEE to behave in a way that makes the clustering into $(\Theta_k)_{k=1}^K$ best for the purpose of minimizing prediction errors. Observe that the vast multiplicity of analogy partitions so derived results in a vast range of possible equilibrium behaviors as well, where behaviors are concentrated around the various $\bar{\theta}_k$.[16]

Of course, we should not expect to have such an extreme form of self-attraction for all parameter values of the beauty contest game. For example, when r is small (close to 0), then behaviors are not affected much by the analogy partition (see (4)), and in the limit as $r = 0$, the only globally calibrated ABEE would require choosing $(\Theta_k)_{k=1}^K$ so as to minimize

$$\sum_{k=1}^{K} \int_{\Theta_k} (\theta - \overline{\theta}_k)^2 f(\theta) d\theta.$$

In the case θ is uniformly distributed on $\left[\underline{\theta}, \overline{\theta}\right]$, this would lead to the equal splitting analogy partition (i.e., $\Theta_k = (\theta_{k-1}, \theta_k)$ where $\theta_0 = \underline{\theta}$ and $\theta_k = \underline{\theta} + \frac{k}{K}(\overline{\theta} - \underline{\theta})$).

For intermediate values of r, we can support a bigger range of analogy partitions as part of a CABEE but not as many as when r approaches 1. The next Proposition establishes that as r grows larger, more and more analogy

[15] It may be mentioned that for local (not global) calibration, we could dispense with the requirement that $\overline{\theta}_k$ are all different. This is so because, with local calibration, if $\overline{\theta}_k = \overline{\theta}_{k'}$, any game $\theta \in \Theta_k \cup \Theta_{k'}$ could be assigned to either Θ_k or $\Theta_{k'}$ as from (4) it is readily verified that $E(a_i^{ABEE}(\theta) \mid \theta \in \Theta_k)$ and $E(a_i^{ABEE}(\theta) \mid \theta \in \Theta_{k'})$ would both be equal to $\overline{\theta}_k = \overline{\theta}_{k'}$.

[16] Observe also that our construction would be robust to the introduction of any share of rational agents to the extent that in the limit as r tends to 1 players in our equilibrium are picking best-responses to their opponent's strategy.

partitions are self-attractive.[17]

Proposition 3 *Take a partition* $(\Theta_k)_{k=1}^K$ *and the corresponding ABEE. Suppose it is a locally CABEE for some* $r < 1$*. Then it is a locally CABEE for all* $r' > r$*.*

Proof. As r increases, for each $\theta \in \Theta_k$, $a(\theta)$ gets strictly closer to $\bar{\theta}_k$ and weakly further apart from any $\bar{\theta}_{k'}$, where $k' \neq k$. Thus, if $(\Theta_k)_k$ is locally calibrated with respect to $a(\theta)$, given r, then it also locally calibrated for any $r' > r$. **Q.E.D.**

The insight of this Proposition can possibly be related to some features of echo chambers. When the concern for coordination is big, the clustering of fundamental values into analogy classes may not be much related to the objective realization of the fundamental values, and different societies (which may settle down on different CABEE) may end up forming different beliefs and adopting different behaviors in objectively identical situations.

From a different perspective, even within a given CABEE, we observe that the categorization leads to more polarized behaviors as compared with the Nash Equilibrium case. In the Nash equilibrium, the distribution of actions is continuous in θ. This is not so in any CABEE when $r > 0$: behaviors in CABEE are more concentrated around the $\overline{\theta}_k$, and the polarization is stronger as r gets larger. At the limit when $r = 1$, the distribution of behaviors is simply K mass points each corresponding to one of the $\overline{\theta}_k$. In other words, even within a single society assumed to have settled down on one CABEE, we observe more polarization of behavior in any such CABEE than in the Nash equilibrium, and this is all the more pronounced that the motive for coordination (i.e., r) is larger.

The analysis presented here should be contrasted with insights obtained in the tradition of global games, suggesting a unique selection of equilibrium (see

[17] We conjecture that the same result holds for globally calibrated ABEE, even if we have not proven it. A weaker statement that can easily be established is that if a partition is part of a globally CABEE at r, then there exists r^* such that it is also part of a globally CABEE at all $r' \geq r^*$.

Morris and Shin (2002) for references). Of course, our setting is different (no private information), and our formalization of expectations through categories is also different, leading to alternative predictions and a novel perspective on the possibility of multiple equilibria. We believe that our finding that a larger variety of behaviors and larger departures from the fundamental can be obtained as agents are more concerned with coordinating with others is more in line with Keynes' intuition about beauty contests, even if in our analysis the beliefs of agents are not arbitrary but are pinned down by actual behaviors as in tradional game theoretic approaches. While rational expectations in the beauty contest game would lead agents to adopt behaviors coinciding with the fundamental value, the coarse categories used by agents to form their expectations allows for more heterogeneity of behaviors even when the categories are chosen to minimize the prediction error as in our approach.

3.2 Self-Repelling Analogy Partitions: A Monitoring Game Illustration

We consider the following Employer-Worker environment. There are three types of workers a, b, c. An employer is matched with one worker who can be either of type a, b, or c with probability p_a, p_b, p_c, respectively. The type $\omega = a, b, c$ of the worker is observed by the employer. It can be identified with the state in our general framework.

In each possible interaction ω, the employer and the worker make simultaneous decisions. The employer decides whether he will Control (C) the worker or not (D). The worker decides whether to exert low effort $e = 0$ or high effort $e = 1$.[18]

We assume that the employer cares about the type only through the effort attitude that she attributes to the type of worker. Specifically, we assume that the employer finds it best to choose C only if she expects the worker to choose

[18] A practical way to motivate the simultaneous move formulation is to view the C decision as requiring time before it is implemented.

low effort with probability no less than ν^*.

Workers' effort attitudes depend on their type and possibly on their expectation about whether they will be controlled or not. Type a of worker always chooses $e = 0$ (irrespective of his expectation about the Control probability). Type b always chooses $e = 1$. Type c of worker finds it best to exert low effort $e = 0$ when he expects C to be chosen with probability no more than μ^*.

In the unique Nash equilibrium, the employer would choose C when facing type a, D when facing type b and would mix between C with probability μ^* and D with probability $1 - \mu^*$ when facing type c. Type c of worker would choose $e = 0$ with probability ν^*.

Assume the employer uses two categories $K = 2$ whereas the worker is rational.[19] We have:

Proposition 4 *Assume p_c is no larger than p_a and p_b, $\frac{p_c}{p_a+p_c} < \nu^* < \frac{p_b}{p_b+p_c}$, and $\nu^* \neq \frac{p_b}{p_a+p_b}$. There is no pure Calibrated Analogy-based expectation equilibrium in which a single analogy partition is used by the employer.*

Proof. This is proven by contradiction. Let $\beta_{e=1}$ denote the probability attached to $e = 1$ by the employer in her non-singleton analogy class. 1) If ac are put together, consistency implies that $\beta_{e=1}$ is at most $\frac{p_c}{p_a+p_c}$, which is smaller than ν^*. Best-response of the employer implies that C (Control) is chosen when facing worker c. Best-response of the c-worker leads to $e_c = 1$. The resulting profile of effort attitude is $(e_a = 0, e_b = 1, e_c = 1)$, which leads at the clustering stage to reassign c with b, thereby yielding a contradiction. 2) If bc are put together, $\beta_{e=1} > \nu^*$ so that D is chosen by the employer when facing the c-worker. This leads the c-worker to choose $e_c = 0$ and the resulting profile of effort attitudes $(e_a = 0, e_b = 1, e_c = 0)$ would lead to reassign c to a, thereby leading to a contradiction. 3) If ab are put together, $\beta_{e=1} = \frac{p_b}{p_a+p_b}$, and the probability that $e_c = 1$ is ν^* (as the mixed Nash equilibrium is then played

[19] One can possibly motivate this asymmetry by saying that each type of worker can safely focus on the situations in which similar types are involved whereas employers need to have a more complete understanding on how the type maps into an effort attitude.

in the interaction with c). This violates the condition for local calibration, for either a or b whenever $\nu^* \neq \frac{p_b}{p_a+p_b}$. **Q.E.D.**

In other words, in this monitoring environment, any categorization is self-repelling, and steady state requires some mixing. Relying on full clustering or global calibration, we have:

Proposition 5 *Assume p_c is no larger than p_a and p_b and $\frac{p_c}{p_a+p_c} < \nu^* < \frac{p_b}{p_b+p_c}$. There is a unique globally calibrated distributional ABEE in which the analogy partition putting ac (resp. bc) together is chosen with probability μ^* (resp. $1-\mu^*$), the c worker chooses $e = 0$ and 1 each with probability half, and the employer chooses C in her analogy class containing a and D in her analogy class containing b.*

This can easily be understood as follows. Given the choice of the Employer for each of her analogy partition, the distribution over analogy partitions is chosen so as to make the c worker indifferent between his two effort options. The mixing of the c worker is then chosen so as to make the two analogy partitions equally good for the purpose of minimizing the total variance. The rest of the construction follows easily.

When local calibration is considered instead, a wider range of equilibria can be sustained, but in all of them, the analogy partition putting ac (resp. bc) together must be chosen with probability μ^* (resp. $1-\mu^*$). The range of these depend on the probabilities p_ω of $\omega = a, b, c$ and to fix ideas, we consider in the next proposition that types a and b are equally likely.

Proposition 6 *Assume that $p_a = p_b = p > \frac{1}{3}$ and $\frac{p_c}{p_a+p_c} < \nu^* < \frac{p_b}{p_b+p_c}$. The following define the locally calibrated distributional ABEE. The analogy partition putting ac (resp. bc) together is chosen with probability μ^* (resp. $1-\mu^*$), the c worker chooses $e = 0$ with probability ζ in the range $(p, \frac{1-p}{2-3p})$, and the employer chooses C in her analogy class containing a and D in her analogy class containing b.*

The main difference between the above two Propositions is that for local calibration, any mixing of the c type that assigns probability $\zeta \in (p, \frac{1-p}{2-3p})$ on $e = 0$ makes the two analogy partitions locally calibrated with respect to the proposed strategy of the worker. Indeed, when the analogy partition putting ac (resp. bc) together is considered, the aggregate effort distribution in ac (resp. bc) assigns probability $\frac{p+(1-2p)\zeta}{1-p}$ (resp. $\frac{(1-2p)}{1-p}\zeta$) to $e = 0$ and the mixing ζ on $e = 0$ is then closer to $\frac{p+(1-2p)\zeta}{1-p}$ (resp. $\frac{(1-2p)}{1-p}\zeta$) than to 0 (resp. 1) since $\zeta > p$ (resp. $\zeta < \frac{1-p}{2-3p}$).

With the above learning interpretation that involves populations of employers and workers, the observation that several analogy partitions must arise in a distributional calibrated ABEE imply that not all employers categorize c workers in the same way. This implies that different employers may have have different beliefs about the working attitude of c workers. Some employers (in proportion μ^*) believe c workers choose $e = 0$ with probability $\frac{p+(1-2p)\zeta}{1-p}$ and others (in proportion $1 - \mu^*$) believe c workers choose $e = 0$ with probability $\frac{(1-2p)}{1-p}\zeta$. That is, some overestimate the working attitude of c workers and some underestimate it, leading in our model to polarized beliefs. Those employers underestimating the working attitude will treat the minority group (c is the least likely type) exactly like type a, while the others will treat them exactly like type b. Even if stylized, we believe our analysis may provide a novel argument as to how different beliefs about the working attitude of minority groups and different treatments of said groups may co-exist in a society.

3.3 Strategic interactions with linear best-replies

In this section we apply the notion of Calibrated ABEE to families of games with continuous action spaces parameterized by an interaction parameter μ, which takes values in an interval of the real line. This parameter is a determinant of the intensity of players' reactions to their opponent's behavior. Players have best-responses which are linear both in the strategy of the opponent and in μ.

Formally, we consider a family of games parameterized by $\mu \in [-1, 1]$, where

μ is distributed according to a continuous density function f with cumulative denoted by F. Players observe the realization of μ and player $i = 1, 2$ chooses action $a_i \in \mathbb{R}$.[20] In game μ, when player i expects player j to play according to $\sigma_j \in \Delta\mathbb{R}$, player i's best-response is:

$$BR_i(\mu, \sigma_j) = A + \mu B + \mu C E_{\sigma_j}(a_j),$$

where $E_{\sigma_j}(a_j)$ denotes the mean action derived from the distribution σ_j, and A, B and C are constants with $0 < C < 1$. We will analyze separately the cases in which $\mu \in [0, 1]$ and $\mu \in [-1, 0]$, and in each case we will assume that $f(\cdot)$ has full support. In the former case, the games exhibit strategic complementarity. In the latter, they exhibit strategic substitutability.

The restriction to linear best-replies while demanding allows us to accommodate classic applications.[21] In particular, consider the case of strategic complementarity ($\mu \geq 0$). Such games with linear best-responses arise in a duopoly with differentiated products in which firms have constant marginal costs, demand is linear, and firms compete in prices à la Bertrand (see Vives 1999 for a textbook formulation). Such games can also be viewed as capturing in a reduced form moral hazard in team problems (a specific formulation of the model introduced by Holmström 1982) in which the agents receive a bonus if the team is successful, agents simultaneously choose how much effort to exert, the probability of success depends on the profile of effort in a bilinear way and the cost of effort is quadratic.[22]

Consider next the case when μ is non-positive so that the game exhibits

[20] The environment we consider here has a continuum of games and a continuum of actions. While our general existence results do not apply to such environments, we will obtain the existence of locally calibrated ABEE in the case of strategic complements by construction. The case of strategic substitutes will be shown not to admit locally calibrated ABEE.

[21] It may be mentioned that in our formulation, we allow the actions to take any value (positive or negative) whereas in some of the applications mentioned below it would be natural to impose that the actions (quantities, prices or effort level) be non-negative. We do not impose non-negativity constraints to avoid dealing with corner solutions, but none of our qualitative insights would be affected with such additional constraints.

[22] Complementarity is obtained for positive coefficients applying to the product of effort levels in the probability of success.

strategic substitutability. A setting fitting our formulation is one of a duopoly with differentiated products with constant marginal costs and linear demands, but this time assuming firms compete in quantities à la Cournot (see again Vives 1999 for elaborations).

Regardless of the sign of μ, it is readily verified that there exists a unique Nash Equilibrium of the game with parameter μ. It is symmetric, it employs pure strategies and it is characterized by $a_1^{NE}(\mu) = a_2^{NE}(\mu) = \frac{A+\mu B}{1-\mu C}$. The function $a_i^{NE}(\mu)$ is continuous and monotone in μ. When $B = -AC$, the function $a_i^{NE}(\mu)$ is flat. The function is strictly increasing (decreasing) and convex (concave) in μ for B greater (smaller) than $-AC$, when $\mu \in [-1, 1]$.

In our family of games parameterized by μ, we will impose that if two games belong to the same analogy class, any game in between the two belongs to that class as well. Accordingly, we will be considering analogy partitions with the property that each analogy class is an interval of μ, and we will refer to these as *interval analogy partitions.*

Specifically, assume that players use (pure) symmetric interval analogy partitions, splitting the interval into K subintervals, so that

$$An_1 = An_2 = \{[\mu_0, \mu_1], (\mu_1, \mu_2], \ldots, (\mu_{K-1}, \mu_K]\}$$

where $\mu_0 = 0$, $\mu_K = 1$ in the case of strategic complements, and $\mu_0 = -1, \mu_K = 0$ in the case of strategic substitutes.[23]

Since analogy partitions are symmetric, we simplify notation by dropping the subscript that indicates whether player 1 or 2 is considered. We simply denote the interval $(\mu_{k-1}, \mu_k]$ by α_k for $k = 2, ...K - 1$ and $\alpha_1 = [\mu_0, \mu_1]$.

Given that players care only about the mean action of their opponent, we will assume that players focus only on this mean and accordingly compare the behaviors in different games using the Euclidean distance between the mean action these games induce. That is, we will consider the square of the Euclidean

[23]Whether μ_k is assigned to (μ_{k-1}, μ_k) or (μ_k, μ_{k+1}) plays no role in our setting with a continuum of μ.

distance in the space of these mean actions for clustering purposes.

Specifically, with some abuse of notation, we will refer to $\beta_i(\alpha_k)$ as the expected mean action of player j in the analogy class α_k. The consistency of β_i with σ_j imposes that $\beta_i(\alpha_k) = \frac{1}{F(\mu_k)-F(\mu_{k-1})}\int_{\mu_{k-1}}^{\mu_k}\sigma_j(\mu)f(\mu)d\mu$, where $\sigma_j(\mu)$ denotes the (mean) action chosen by player j in game μ. Moreover in each game $\mu \in \alpha_k$, best-response requires that player i chooses action $BR_i(\mu, \beta_i(\alpha_k)) = A + \mu(B + C\beta_i(\alpha_k))$ as given by μ and his analogy-based expectation $\beta_i(\alpha_k)$ about the mean action in α_k.

Given a (symmetric) interval analogy partition profile $An_1 = An_2 = \{[\mu_0, \mu_1], \ldots, (\mu_{K-1}, \mu_K]\}$, an ABEE is a strategy profile (σ_1, σ_2) such that for each player i, each class α_k and each game $\mu \in \alpha_k$, we have $\sigma_i(\mu) \in BR_i(\mu, \beta_i(\alpha_k))$ with the requirement thate β_i is consistent with σ_j. Exploiting the linearity of the best-response, it is easily established (through routine calculations provided in the online Appendix) that there exists a unique ABEE, which is symmetric.

Proposition 7 *Assume players use symmetric interval analogy partitions. There exists a unique ABEE where, for all* $k = 1, \ldots, K$, $\beta_1(\alpha_k) = \beta_2(\alpha_k) = \frac{A+B\mathbb{E}[\mu|\alpha_k]}{1-C\mathbb{E}[\mu|\alpha_k]}$ *and for* $\mu \in \alpha_k$, $\sigma_1(\mu) = \sigma_2(\mu) = A + \mu\frac{B+AC}{1-C\mathbb{E}[\mu|\alpha_k]}$.

Since under symmetric interval analogy partitions the ABEE is symmetric, we drop the subscript that refers to players and we write $\beta_1(\alpha_k) = \beta_2(\alpha_k) = \beta(\alpha_k)$. We also let $a(\mu|\alpha_k)$ refer to $A + \mu\frac{B+AC}{1-C\mathbb{E}[\mu|\alpha_k]}$ for the remainder of this section where as seen in Proposition 7, $a(\mu|\alpha_k) = A + \mu\frac{B+AC}{1-C\mathbb{E}[\mu|\alpha_k]}$ describes the ABEE strategies of players 1 and 2 in the analogy class α_k. The function $a(\mu|\alpha_k)$ is linear in μ. Similarly to the discussion of Nash Equilibrium above, when $B = -AC$, the function $a(\mu|\alpha_k)$ is flat, and it is strictly increasing (decreasing) in μ for B greater (smaller) than $-AC$, if $\mu \in [0, 1]$.

As far as clustering is concerned, and as already mentioned, we consider the square of the Euclidean distance in the mean actions. That is, considering a symmetric interval analogy partition given by $\{\alpha_k\}_{k=1}^K$ and the associated ABEE (as described by $\beta(\alpha_k)$ and $a(\mu \mid \alpha_k)$), $\{\alpha_k\}_{k=1}^K$ is locally calibrated, if for all $k = 1, \ldots, K$ and all $\mu \in \alpha_k$,[24]

$$(\beta(\alpha_k) - a(\mu|\alpha_k))^2 \leq (\beta(\alpha_{k'}) - a(\mu|\alpha_k))^2, \; \forall k' \neq k.$$

By the monotonicity of the function $a(\mu|\alpha_k)$ the problem of local calibration boils down to checking the above inequalities only at the extreme points of each analogy class. That is, the sequence $\{\mu_0, \mu_1, \ldots, \mu_K\}$ generates an interval analogy partition that is locally calibrated with respect to the corresponding ABEE, if and only if, for $k = 1, \ldots, K-1$,

$$(\beta(\alpha_k) - a(\mu_k|\alpha_k))^2 \leq (\beta(\alpha_{k+1}) - a(\mu_k|\alpha_k))^2 \; and$$

$$(\beta(\alpha_{k+1}) - a(\mu_k|\alpha_{k+1}))^2 \leq (\beta(\alpha_k) - a(\mu_k|\alpha_{k+1}))^2 \tag{5}$$

As far as global calibration is concerned, one has to check for a given candidate interval analogy partition $An = \{\alpha_k\}_{k=1}^K$ whether

$$An = \arg \min_{\{\alpha'_k\}_{k=1}^K} \sum_k \int_{\alpha'_k} \left[\beta(\alpha'_k) - a^{ABEE}(\mu)\right]^2 f(\mu) d\mu$$

where $a^{ABEE}(\mu) \equiv A + \mu \sum_{k=1}^K 1_{\{\mu_{k-1} < \mu \leq \mu_k]\}} \frac{B+AC}{1-C\mathbb{E}[\mu|\alpha_k]}$ is the ABEE strategy given An and $\beta(\alpha'_k) = E(a^{ABEE}(\mu) \mid \mu \in \alpha'_k)$.

3.3.1 Strategic Complements

In this part, we consider the strategic complements case, and we assume that μ is distributed according to a continuous density f with support on $[0, 1]$.

Given $(\mu_k)_{k=0}^K$, we wish to highlight that

$$a^{ABEE}(\mu) = A + \mu \sum_{k=1}^K 1_{\{\mu_{k-1} < \mu \leq \mu_k]\}} \frac{B + AC}{1 - C\mathbb{E}[\mu|\alpha_k]}$$

[24]Clearly, for local calibration, we could consider the Euclidean distance instead of the square of the distance. This is so because comparisons are only in terms of the mean action, which is one-dimensional.

has discontinuities at $\mu_1, \mu_2, \ldots, \mu_{K-1}$. If $B \geq -AC$, the function $a^{ABEE}(\mu)$ is increasing in μ and the discontinuities take the form of upward jumps. Similarly, if $B < -AC$, the function $a^{ABEE}(\mu)$ is decreasing in μ and the discontinuities take the form of downward jumps. The direction of the jumps is a consequence of the strategic complement aspect, and it will play a key role in the analysis of local calibration. Indeed assuming $B \geq -AC$, as one moves in the neighborhood of μ_k from the analogy class $(\mu_{k-1}, \mu_k]$ to the analogy class $(\mu_k, \mu_{k+1}]$, the perceived mean action of the opponent jumps upwards and this leads to an upward jump in the best-response due to strategic complementarity.

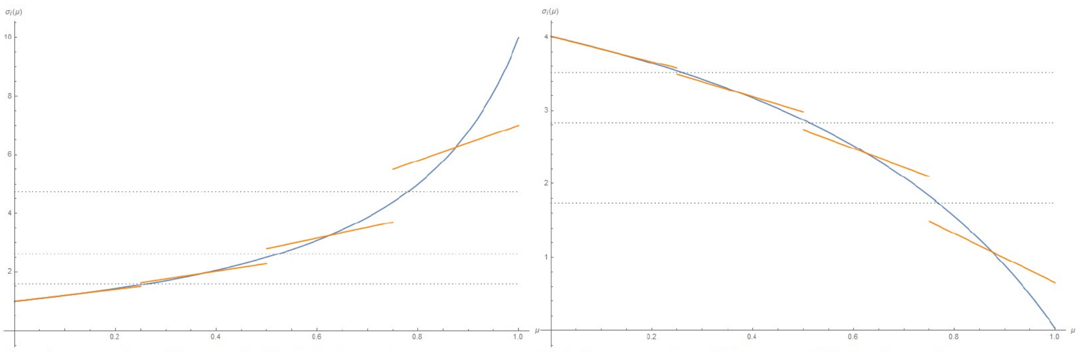
There is a simple geometric characterization of local calibration. Assuming $B \geq -AC$, we have that $\beta(\alpha_k) \leq \beta(\alpha_{k+1})$, for all k. The local calibration requirements summarized by inequalities in (5) are equivalent to the condition that the arithmetic average of the analogy-based expectations of two adjacent analogy classes should be between the largest action in the first and the smallest action in the second analogy class. That is,

$$a(\mu_k|\alpha_k) \leq \frac{\beta(\alpha_k) + \beta(\alpha_{k+1})}{2} \leq a(\mu_k|\alpha_{k+1}).$$

Similarly, when $B < -AC$, the ABEE function is strictly decreasing and (5) can be reduced to $a(\mu_k|\alpha_k) \geq \frac{\beta(\alpha_k)+\beta(\alpha_{k+1})}{2} \geq a(\mu_k|\alpha_{k+1})$. These inequalities can receive a simple graphical interpretation as illustrated in Figure 1 where the horizontal dashed lines in black represent the arithmetic average between the analogy-based expectations of two consecutive classes, and whenever $a^{ABEE}(\mu)$ does not cross any dashed line, the requirements for local calibration are satisfied by that analogy class.[25]

[25] Figure 1 shows how the (simplified) local calibration requirements would appear graphically. There are two graphs, one for $B > -AC$ on the left and one for $B < -AC$ on the right. Both graphs depict how the Nash Equilibrium function $a^{NE}(\mu) = \frac{A+\mu B}{1-\mu C}$ (in blue) and the ABEE function $a^{ABEE}(\mu)$ (in orange) change as μ varies. For these graphs we assume that μ is distributed uniformly over $[0, 1]$, we let $K = 4$, and we pick the interval analogy partition induced by the equal splitting sequence $\{0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1\}$. When $B > (<)AC$, $a^{NE}(\mu)$ and $a^{ABEE}(\mu)$ are strictly increasing (decreasing).

Figure 1: Graphical Illustration of Local Calibration Requirements for $K = 4$ and μ-sequence such that $\mu_k = \frac{k}{4}$, $k = 0, 1, \ldots, 4$



(a) Increasing NE and ABEE functions $A = 1$, $B = 1$ and $C = 0.8$

(b) Decreasing NE and ABEE functions $A = 4.01$, $B = -4$ and $C = 0.6$

One can easily see from Figure 1 that the analogy partitions depicted in the graphs are locally calibrated. As a matter of fact, and as we will show later, when μ is uniformly distributed between 0 and 1, an analogy partition that splits the interval into K subintervals of equal size leads to a locally calibrated ABEE. For more general distributions, we introduce the notion of *equidistant-expectations sequence* $\mu_0, \mu_1, \ldots, \mu_K$ defined so that for any μ_k, with $k \neq 0, 1$, the Euclidean distance between μ_k and the mean value of μ in $(\mu_{k-1}, \mu_k]$ is equal to the Euclidean distance between μ_k and the mean value of μ in $(\mu_k, \mu_{k+1}]$. That is, $\mu_k - \mathbb{E}[\mu|(\mu_{k-1}, \mu_k]] = \mathbb{E}[\mu|(\mu_k, \mu_{k+1}]] - \mu_k$. We refer to the corresponding interval partition $(\alpha_k)_{k=1}^K$ with $\alpha_k = (\mu_{k-1}, \mu_k]$ as the *equidistant-expectations partition.* We note that when μ is uniformly distributed on $[0, 1]$, the equidistant-expectations sequence is uniquely defined by $\mu_k = \frac{k}{K}$. For more general density functions f, it is readily verified (by repeated application of the intermediate value theorem) that there always exists at least one equidistant-expectations partition.[26]

Our main result in this application is:

Proposition 8 *In the environment with strategic complements, consider an equidistant-expectations partition* $(\mu_k^*)_{k=0}^K$. *There exist* $\left\{\underline{\mu}_k, \overline{\mu}_k\right\}_{k=0}^K$ *satisfying*

[26]See the online appendix for details.

$\underline{\mu}_k < \mu_k^* < \overline{\mu}_k$ *such that for any* $(\mu_k)_{k=0}^K$ *satisfying* $\mu_k \in (\underline{\mu}_k, \overline{\mu}_k)$, *the analogy partition* $An = (\alpha_k)_{k=1}^K$ *with* $\alpha_k = (\mu_{k-1}, \mu_k]$ *together with the corresponding ABEE is a locally calibrated ABEE.*

The rough intuition for this result can be understood as follows. First, start with an equidistant-expectations partition $(\mu_k^*)_{k=0}^K$. Suppose we were considering games with no interaction term, i.e., such that $C = 0$. Then in game μ, players would be picking their dominant strategy $a(\mu) = A + \mu B$ irrespective of the analogy partition, given that players would not care about the action chosen by their opponent. It is readily verified that the local calibration conditions for the points $a(\mu)$ would lead to pick an *equidistant-expectations partition* in this case, and this would force a locally calibrated ABEE when $C = 0$ to be relying on such *equidistant-expectations partitions.* Allowing for non-null interaction parameters C makes the problem of finding a locally calibrated ABEE a priori non-trivial due to the endogeneity of the data generated by the ABEE with respect to the chosen analogy classes, as emphasized above. However, what the Proposition implies is that using the same analogy classes as those obtained when $C = 0$ can be done to construct a locally calibrated ABEE. Intuitively, this is so because the strategic complement dimension makes the points obtained through ABEE in a given class of the equidistant-expectations partition look closer to one another relative to points outside a class, as compared with the case in which $C = 0$. As a result, the local calibration conditions which hold for the equidistant-expectation partition when $C = 0$ hold a fortiori when C is non-null. Now given the jumps, there is some slack in the conditions for local calibration, thereby ensuring that one can find open intervals of boundary points μ_k that satisfy the conditions of the Proposition.

Our environment with strategic complements can be viewed as one in which there is an element of self-attraction in the choice of analogy partitions. As can easily be inferred from the above discussion, the larger C is, the more analogy interval analogy partitions can be used to support locally calibrated ABEE, i.e., the more partitions are self-attractive. In Jehiel and Weber (2023) (the working paper version), we have characterized more fully the set of locally calibrated ABEE confirming that insight, and we have noted that when C is too large the equidistant-expectations partition need need not be compatible with the requirement for global calibration.[27]

3.3.2 Strategic Substitutes

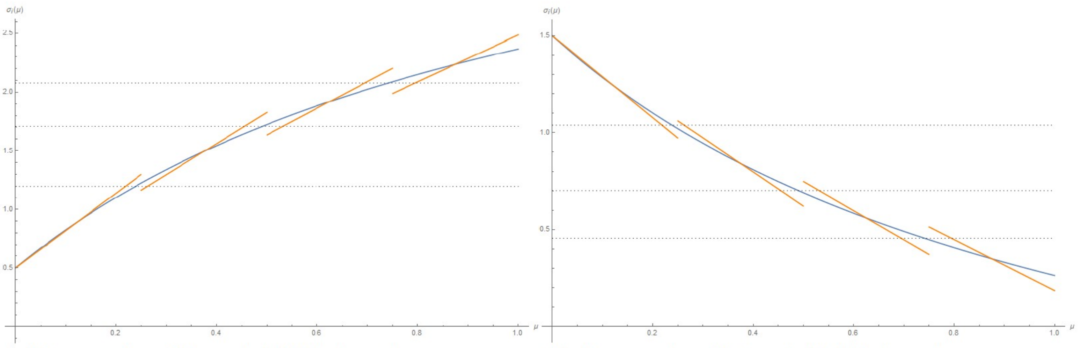
We consider now the strategic substitute case, and we assume that μ is distributed according to a continuous density function f with support $[-1, 0]$.

Note that, differently from the strategic complements environment, here the ABEE function $a^{ABEE}(\mu) = A + \sum_{k=1}^{K} \mathbf{1}_{\{\mu \in [\mu_{k-1}, \mu_k)\}} \mu \frac{B+AC}{1-C\mathbb{E}[\mu|\alpha_k]}$ is not monotone in μ. This is due to the fact that at the discontinuity points the jumps of the function are in the opposite direction with respect to the slope of $a(\mu|\alpha_k) = A + \mu \frac{B+AC}{1-C\mathbb{E}[\mu|\alpha_k]}$, and this is a fundamental difference induced by the change from strategic complements to strategic substitutes. To illustrate this, consider the case where $a(\mu|\alpha_k)$ has a positive slope, that is, $B < -AC$. Recall that $\beta(\alpha_k) = \frac{A+B\mathbb{E}[\mu|\alpha_k]}{1-C\mathbb{E}[\mu|\alpha_k]}$. Since μ is non-positive, $B < -AC$ and $\mathbb{E}[\mu|\alpha_k] < \mathbb{E}[\mu|\alpha_{k+1}]$ imply that $\beta(\alpha_k) < \beta(\alpha_{k+1})$. Since in the strategic substitutes environment the best-response is decreasing in the analogy-based expectations, at the adjacency point between two classes, the action played in equilibrium will be greater in the first of the two classes: $\beta(\alpha_k) < \beta(\alpha_{k+1})$ implies that $a(\mu_k|\alpha_k) > a(\mu_k|\alpha_{k+1})$. Hence, when $a(\mu|\alpha_k)$ is increasing in μ, the ABEE function jumps downwards at the discontinuity points. This is depicted in Figure 2.

The non-monotonicities that arise in the ABEE strategies with interval analogy partitions in turn make it impossible to satisfy the local calibration conditions in a neighborhood of $\mu = \mu_k$. This is so because it cannot be simultaneously the case that $\lim_{\mu \to \mu_k^+} a^{ABEE}(\mu)$ is closer to $\beta((\mu_k, \mu_{k+1}])$ and $a^{ABEE}(\mu_k)$

[27]While our analysis there suggests that intervals with larger μ should be smaller than in the equidistant-expectations partition, we have not been able to provide a general characterization of globally calibrated ABEE in this case.

Figure 2: Graphical Illustration of Local Calibration Requirements for $K = 4$ and μ-sequence such that $\mu_k = \frac{k}{4}$, $k = 0, 1, \ldots, 4$



(a) Increasing NE and ABEE functions $A = 0.5$, $B = -4$ and $C = 0.9$

(b) Decreasing NE and ABEE functions $A = 1.5$, $B = 1$ and $C = 0.9$

is closer to $\beta((\mu_{k-1}, \mu_k])$.[28] Formally, we have:

Proposition 9 *In the strategic substitutes environment, whenever $B \neq -AC$, there are no symmetric interval analogy partitions that are locally calibrated with respect to the induced ABEE.*

Our environment with strategic substitutes illustrates a case with self-repelling analogy partitions. In light of our general theoretical framework, it would then be natural to look for calibrated distributional ABEE. In our setting with a continuum of games, this is a challenging task, and in Jehiel and Weber (2023) (the working paper version), we characterize the calibrated distributional ABEE when μ can take only three values.

4 Conclusion

In this paper we have introduced the notion of Calibrated ABEE defined so that i) given the analogy partitions, players choose strategies following the ABEE machinery, and ii) given the raw data on the opponent's strategies, players select analogy partitions so as to minimize the prediction errors (either locally or

[28] For example when $B + AC > 0$, we would have $\beta((\mu_k, \mu_{k+1}]) > a^{ABEE}(\mu_k) > \lim_{\mu \to \mu_k^+} a^{ABEE}(\mu_k^+) > \beta((\mu_{k-1}, \mu_k])$, making it impossible to satisfy the local calibration conditions for $\mu = \mu_k$ and $\mu = \mu_k^+$ (i.e., μ slightly above μ_k).

globally). We have highlighted the existence of environments with self-repelling analogy partitions in which some mixing over analogy partitions must arise as well as environments with self-attractive analogy partitions for which multiple analogy partitions can arise in equilibrium. In the case of environments with self-repelling analogy partitions, an implication of our analysis is that faced with the same objective datasets and the same objective constraints (as measured by the number of classes), players must be processing information in a heterogeneous way in equilibrium. Our derivation of this insight follows from the strategic nature of the interaction, and it should be contrasted with other possible motives of heterogeneity, for example, based on the complexity of processing rich datasets.[29] The case of self-attractive analogy partitions suggests a novel source of multiplicity, not discussed in the previous literature. Our applications reveal among other things that environments with strategic complements have self-attractive analogy partitions and environments with strategic substitutes have self-repelling analogy partitions.

This study can be viewed as a first step toward a more complete understanding of the structure and impact of categorizations on expectation formation in strategic interactions. Among possible future research avenues, one could consider the effect of allowing the clustering to be based not only on the opponent's distribution of action but also on other characteristics of the interaction such as the own payoff structure. One could also consider the effect of allowing the pool of subjects assigned to the role of a given player to be heterogeneous in the number of categories that is considered.

[29]Such forms of heterogeneity are implicitly suggested in Aragones et al (2005) (when they highlight that finding regularities in complex datasets is NP-hard) or Sims (2003) (who develops a rational inattention perspective to model agents who would be exposed to complex environments).

Appendix A

Proof of Proposition 1

The column player partitions games finely. Let row player's analogy partition be $An_1 = \{\{x_1, x_2\}, \{x_3\}\}$ with $x_1 < x_2$, and let β_L denote the probability attached to L in the expectation $\beta_1(x_1, x_2)$.

In any ABEE, the Nash equilibrium strategies are played in x_3, thus the column player plays L with probability $\frac{1}{2+x_3}$ in x_3 to make the row player indifferent. The row player must be mixing in game x_1 too: if $\sigma_1(x_1) = D$, then $\sigma_2(x_1) = R$ which implies $\beta_L \geq \frac{1}{2}$, while the best response of the row player in x_1, given β_L, would be U; If $\sigma_1(x_1) = U$ is a best response for the row player, it has to be that $\beta_L \geq \frac{1}{2+x_1}$ which is greater than $\frac{1}{2+x_2}$ and so the row player would play U in x_2 as well. This in turn would imply that column player would play R in both games, which would violate the consistency requirement on β_L.

Therefore, any ABEE requires $\beta_L = \frac{1}{2+x_1}$. Given β_L, row plays $\sigma_1(x_2) = U$ leading column to play $\sigma_2(x_2) = R$. By consistency, the column player must be playing L with probability $\frac{2}{2+x_1}$ in game x_1.

Note that the strategy profile σ_2 just obtained is such that An_1 is not locally calibrated. If $x_1 = b$, the probability of playing L in x_1 is greater than $\frac{1}{2}$, while $\beta_L = \frac{1}{2+b} < \frac{1}{2+a} < \frac{1}{2}$, where $\frac{1}{2+a}$ is the probability assigned to L by row's expectations in $x_3 = a$: game x_1 should be reassigned to x_3. Whenever $x_1 = a$, regardless of x_2 being b or c, x_2 should be reassigned to x_3 because the probability of L being played in x_2 is zero, which is closer to $\frac{1}{2+x_3}$ than to $\beta_L = \frac{1}{2+a}$. **Q.E.D**

Proof of Theorem 1.

Compared to classic existence results in game theory, the main novelty is to show that the global calibration correspondence has properties that allow to apply Kakutani fixed point theorem to a grand mapping $M : \bar{\Sigma} \times \Lambda \rightrightarrows \bar{\Sigma} \times \Lambda$, which is a composition of the following functions and correspondences. Given $(\bar{\sigma}, \lambda)$ we compute the analogy-based expectations β through consistency and we call this function C. Given (β, λ), the Best Response correspondence (BR)

yields the optimal strategies for each analogy partition in the support of λ. We aggregate the strategies following (2) obtaining $\bar{\sigma}'$ and define β' to be consistent with $\bar{\sigma}'$. We denote this function AG. We perform global calibration (GC) on $(\bar{\sigma}', \beta')$. Then, we obtain the following composition:

$$(\bar{\sigma}, \lambda) \mapsto_C (\beta, \lambda) \mapsto_{BR} (\sigma', \lambda) \mapsto_{AG} (\bar{\sigma}', \beta') \mapsto_{GC} (\bar{\sigma}', \lambda')$$

where $M(\bar{\sigma}, \lambda)$ denotes the set of $(\bar{\sigma}', \lambda')$ that can be obtained through this composition.

Note that C and AG are continuous functions, while BR and GC are correspondences. The mapping BR is upper-hemicontinuous (uhc) with non-empty, convex and compact values by standard arguments. Since, as we will prove later, GC is also uhc with non-empty, convex and compact values, it follows that:

(i) M is nonempty;

(ii) M is uhc as a composition of uhc mappings;

(iii) M is convex-valued since BR and GC are convex-valued;

(iv) M is compact-valued because BR being compact-valued and uhc implies that $BR(\beta, \lambda)$ is compact. Also, since, AG is single-valued and continuous, and GC is compact-valued and uhc, then $GC \circ AG \circ BR \circ C(\bar{\sigma}, \lambda)$ is compact, for all $(\bar{\sigma}, \lambda) \in \bar{\Sigma} \times \Lambda$.

Since $\bar{\Sigma} \times \Lambda$ is a compact and convex set, properties (i) to (iv) ensure that M has a fixed point by Kakutani's theorem.

To conclude the proof we need to show the properties of the GC correspondence. GC maps $\bar{\Sigma} \times \bar{\Sigma}$ into $\bar{\Sigma} \times \Lambda$, where both $\bar{\Sigma}$ and Λ are convex and compact. The image of the correspondence is defined as follows:

$$GC(\bar{\sigma}, \beta) = \{\bar{\sigma}\} \cup \{\lambda \in \Lambda | \lambda_i(An_i) > 0 \iff An_i \in \arg \min_{An_i' \in \mathcal{K}_i} V(\bar{\sigma}_j, \beta_i')\}$$

where $V_i(\bar{\sigma}_j, \beta_i') = \sum_{\alpha_i \in An_i'} p(\alpha_i) \sum_{\omega \in \alpha_i} p(\omega|\alpha_i) d\left(\bar{\sigma}_j(\omega), \beta_i(\alpha_i|An_i')\right)$.

For ease of exposition, let us denote the latter set in the union above as $G_i(\bar{\sigma}_j) \cup G_j(\bar{\sigma}_i)$. Note that G_i is nonempty because $\mathcal{K}_i$ is finite, thereby implying that there is always a solution to the minimization problem. Also, G_i is a simplex hence it is convex and compact. Thus, GC is nonempty, convex and compact valued. We check that $GC(\bar{\sigma}, \beta)$ is uhc by showing that it has a closed graph.

We first establish the continuity of V_i by verifying that d is a continuous function. When d is the squared Euclidean distance, d is clearly continuous in $\bar{\sigma}$ and in β. Instead, when d represents the KL divergence, it is not generally continuous because whenever there is $\omega \in \alpha_i$ such that $supp[\bar{\sigma}_j(\omega)] \not\subset supp[\beta_i(\alpha_i|An_i)]$, then $d(\bar{\sigma}_j, \beta_i)$ goes to infinity. However, the consistency requirements impose $supp[\bar{\sigma}_j(\omega)] \subseteq supp[\beta_i(\alpha_i|An_i)]$. Since global calibration imposes for both players that β_i is consistent with $\bar{\sigma}$, for all ω, α_i and An_i, then $d(\bar{\sigma}_j, \beta_i)$ is finite. Recall that, $d(x, y) = \sum_a (x_a \ln x_a - x_a \ln y_a)$. Since $x_a, y_a \in [0, 1]$ and $x_a > 0$ implies $y_a > 0$, under the convention that $0 \ln 0 = 0$, d is continuous when it represents the KL divergence. Hence, V_i is continuous, if β_i is consistent with $\bar{\sigma}_j$ for both players.

We can now proceed to establish that GC is uhc. Assume by contradiction that GC does not have a closed graph. That is, $\bar{\sigma}_j^n \to \bar{\sigma}_j$, $\lambda_i^n \to \lambda_i$ and $\lambda_i^n \in G_i(\bar{\sigma}_j^n)$, but $\lambda_i \notin G_i(\bar{\sigma}_j)$. Note that $\lambda_i \notin G_i(\bar{\sigma}_j)$ implies that there is some $\tilde{An}_i \in supp[\lambda_i]$ and $\varepsilon', \varepsilon > 0$ such that

$$+\infty > V_i(\bar{\sigma}_j, \tilde{\beta}_i) \geq V_i(\bar{\sigma}_j, \beta_i) + \varepsilon + \varepsilon',$$

where $\tilde{\beta}_i$ is consistent with $\bar{\sigma}_j$ according to $\tilde{An}_i$. Also, let $\tilde{\beta}_i^n$ be consistent with $\bar{\sigma}_j^n$, according to $\tilde{An}_i$. We want to show that for some n, $\lambda_i^n(\tilde{An}_i) > 0$ and $V_i(\bar{\sigma}_j^n, \tilde{\beta}_i^n) > V_i(\bar{\sigma}_j^n, \beta_i^n)$. For $\lambda_i^n \to \lambda_i$ and $\lambda_i(\tilde{An}_i) > 0$, for any n large enough, $\lambda_i^n(\tilde{An}_i) > 0$. As $\bar{\sigma}_j^n \to \bar{\sigma}_j$, for n large enough, by continuity of V_i, $V_i(\bar{\sigma}_j^n, \beta_i^n)$ is in a neighborhood of $V_i(\bar{\sigma}_j, \beta_i)$ so we can write: $V_i(\bar{\sigma}_j, \beta_i) > V_i(\bar{\sigma}_j^n, \beta_i^n) - \varepsilon$. Then, $V_i(\bar{\sigma}_j, \tilde{\beta}_i) \geq V_i(\bar{\sigma}_j, \beta_i) + \varepsilon + \varepsilon' > V_i(\bar{\sigma}_j^n, \beta_i^n) + \varepsilon'$. Similarly, $\bar{\sigma}_j^n \to \bar{\sigma}_j$ implies that, for any n large enough, $V_i(\bar{\sigma}_j^n, \tilde{\beta}_i^n) > V_i(\bar{\sigma}_j, \tilde{\beta}_i) - \varepsilon'$. Thus,

$$V_i(\bar{\sigma}_j^n, \tilde{\beta}_i^n) > V_i(\bar{\sigma}_j, \tilde{\beta}_i) - \varepsilon' \geq V_i(\bar{\sigma}_j, \beta_i) + \varepsilon > V_i(\bar{\sigma}_j^n, \beta_i^n)$$

We get $V_i(\bar{\sigma}_j^n, \tilde{\beta}_i^n) > V_i(\bar{\sigma}_j^n, \beta_i^n)$ and $\lambda_i^n(\tilde{An}_i) > 0$, which contradicts $\lambda_i^n \in G_i(\bar{\sigma}_j^n)$. It follows that GC is uhc. **Q.E.D.**

Proof of Propositions 8

Consider the increasing sequence $\{\mu_k\}_{k=0}^K$ with $\mu_0 = 0$, $\mu_K = 1$ and $\mu_k = \frac{\mathbb{E}[\mu|(\mu_{k-1},\mu_k]]+\mathbb{E}[\mu|(\mu_k,\mu_{k+1}]]}{2}$.[30] We simply check that the sequence we propose satisfies the conditions for local calibration.

That is, An is locally calibrated if, for all $k = 1, \ldots, K-1$, (i) $(\beta(\alpha_k) - a(\mu_k|\alpha_k))^2 \leq (\beta(\alpha_{k+1}) - a(\mu_k|\alpha_k))^2$, and (ii) $(\beta(\alpha_{k+1}) - a(\mu_k|\alpha_{k+1}))^2 \leq \big(\beta(\alpha_k) - a(\mu_|\alpha_{k+1})\big)^2$.

It is readily verified that condition (i) reduces to $\mathbb{E}[\mu|(\mu_k, \mu_{k+1}]] \geq \frac{2\mu_k - \mathbb{E}[\mu|(\mu_{k-1},\mu_k]]}{1+2C(\mu_k - \mathbb{E}[\mu|(\mu_{k-1},\mu_k]])}$. By substituting μ_k in the inequality with $\frac{\mathbb{E}[\mu|(\mu_{k-1},\mu_k]]+\mathbb{E}[\mu|(\mu_k,\mu_{k+1}]]}{2}$ we obtain:

$$\mathbb{E}[\mu|(\mu_k, \mu_{k+1}]] > \frac{\mathbb{E}[\mu|(\mu_k, \mu_{k+1}]]}{1 + 2C\mathbb{E}[\mu|(\mu_k, \mu_{k+1}]]}$$

which is true because the denominator is greater than 1.

Condition (ii) is equivalent to requiring that, if $\mathbb{E}[\mu|(\mu_{k-1}, \mu_k]] < \frac{1}{2C}$, the following inequality holds: $\mathbb{E}[\mu|(\mu_k, \mu_{k+1}]] \leq \mu_k + \frac{\mu_k - \mathbb{E}[\mu|(\mu_{k-1},\mu_k]]}{1-2C\mathbb{E}[\mu|(\mu_{k-1},\mu_k]]}$. Recalling that $\mu_k = \frac{\mathbb{E}[\mu|(\mu_{k-1},\mu_k]]+\mathbb{E}[\mu|(\mu_k,\mu_{k+1}]]}{2}$, we obtain

$$\frac{\mathbb{E}[\mu|(\mu_k, \mu_{k+1}]] - \mathbb{E}[\mu|(\mu_{k-1}, \mu_k]]}{2} < \frac{\frac{\mathbb{E}[\mu|(\mu_k,\mu_{k+1}]] - \mathbb{E}[\mu|(\mu_{k-1},\mu_k]]}{2}}{1 - 2C\mathbb{E}[\mu|(\mu_{k-1}, \mu_k]]}$$

which holds because $0 < 1 - 2C\mathbb{E}[\mu|(\mu_{k-1}, \mu_k]] < 1$ when $\mathbb{E}[\mu|(\mu_{k-1}, \mu_k]] < \frac{1}{2C}$.

Since both conditions hold strictly, sequences in the neighborhood of $\{\mu_k\}_{k=0}^K$ would also satisfy the conditions for local calibration. **Q.E.D.**

Appendix B: Learning Environment

In this part of the appendix we formalize the learning environment introduced in section 2.6.

Learning dynamics. There is a continuum of mass 1 of subjects assigned to the role of player $i = 1, 2$.

[30] We prove the existence of such a sequence in Lemma 3 in the Online Appendix.

We introduce two perturbations that are used to deal with possible indifferences. First, when playing a game, we assume the payoffs are slightly perturbed (see Fudenberg and Kreps (1993) or Esponda and Pouzo (2016)). Specifically, let $\widetilde{\rho}_i$ be a random variable with a continuous density g_i on $[0,1]$ and $\varepsilon > 0$ a number that should be thought of as small (ε measures the degree of perturbation)and we assume that the distribution of realizations in the population matches the densities and probabilities induced by p and $\widetilde{\rho}_i$. Player i in game ω (with draws $\rho_i(a_i,\omega), \rho_i(a_i',\omega)$) picks action a_i whenever for all $a_i' \neq a_i$,[31]

$$u_i(a_i,\beta_i,\omega) + \varepsilon\rho_i(a_i,\omega) > u_i(a_i',\beta_i,\omega) + \varepsilon\rho_i(a_i',\omega)$$

where β_i refers to player i's expectation about player j's behavior in ω.

The second perturbation concerns how clustering is implemented. Suppose in the previous period $\overline{\sigma}_j(\omega)$ represents the aggregate play of population j when playing game ω. Let $\widetilde{\eta}_i$ be a random vector with continuous density h_i over the interior of ΔA_j. We assume that a given player of population i implements a (global) clustering into K_i classes of $\left(\overline{s}_j(\omega) \equiv \frac{\overline{\sigma}_j(\omega)+\varepsilon\eta_i(\omega)}{1+\varepsilon}\right)_{\omega\in\Omega}$ where $\eta_i(\omega)$ is a realization drawn from $\widetilde{\eta}_i$ and the draws are assumed to be independent across games ω. As before, we assume that the distributions of realizations in the population match the density $\widetilde{\eta}_i$. To be more specific, player i (with draws $\eta_i(\omega)$) picks the partitioning into K_i classes so as to solve[32]

$$\arg\min_{P_i\in\mathcal{K}_i} \sum_{c_i\in P_i} p(c_i) \sum_{\omega\in c_i} p(\omega \mid c_i) d(\overline{s}_j(\omega), \beta_i(c_i)),$$

note that in cluster c_i, the belief β_i is identified with the mean of $\overline{s}_j(\omega)$ conditional on $\omega \in c_i$.

The learning dynamics is as described in the main text and it is fully pinned down by the initial values of $\overline{\sigma}^0(\omega)$ used in period 1 (as well as ε, g_i, h_i).

[31] Cases of indifference are insignificant whenever $\widetilde{\rho}$ is distributed in the continuum as assumed here.

[32] For generic $\eta_i(\omega)$, there is a unique solution, thus the handling of indifferences is inconsequential when $\widetilde{\eta}_i$ has a density with no atom, as assumed here.

Steady state. In this part, we show that for a fixed ε, there always exists a steady state of the learning dynamics just described. We then show that the limits of such steady states as ε converges to 0 correspond to the globally calibrated distributional ABEE.

Proposition 10 *For a fixed ε, there always exists a steady state of the learning dynamics.*

Proof. The proof shares similarities with the purification techniques introduced by Harsanyi (1973). We consider the same grand mapping M that we introduced in the proof of Theorem 1, but now in the perturbed environment. The perturbations of the payoffs make best-responses single-valued as commonly observed in the previous learning literature. The main novelty here is that the perturbations on the strategies at the clustering stage make the calibration mapping single-valued too. The argument to show this is a bit more involved than for the payoff perturbation part because the perturbations at the clustering stage do not allow for additive separability.

More precisely, consider the compound mapping $(\bar{\sigma}, \lambda) \mapsto_C (\beta, \lambda) \mapsto_{BR} (\sigma', \lambda) \mapsto_{AG} (\bar{\sigma}', \lambda)$ where $\bar{\sigma}'$ is the profile of aggregate best-responses, given λ.

Fix the probability distributions over analogy partitions λ. From the profile of aggregate strategies $\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2)$ we can compute the corresponding analogy-based expectations $(\beta_i(\cdot|An_i))_{An_i \in supp\lambda_i}$ that are consistent with $\bar{\sigma}$. The mapping $(\bar{\sigma}, \lambda) \mapsto (\beta, \lambda)$ is continuous, single-valued and defined over convex and compact sets.

We consider best-responses in the perturbed environment. Let us order the actions in A_i, so that a_i^z is the z-th element in A_i. We denote by $a_i^*(\omega|An_i)(\cdot)$ the function that maps each realization of the profile of random variables $\tilde{\rho}_i$ for each action $a_i \in A_i$ in game ω into a best-response, and we write $a_i = a_i^*(\omega|An_i)(\rho_i)$ to indicate that a_i is played when the profile of realizations is $\rho_i(\omega) = (\rho_i(a_i', \omega))_{a_i'}$ where $\rho_i = (\rho_i(\omega))_\omega$. We denote by $X_i^z(a_i^*(\omega|An_i))$ the set of perturbations under which the action a_i^z is chosen according to $a_i^*(\omega|An_i)$. That is:

$$X_i^z(a_i^*(\omega|An_i)) = \{\rho_i | a_i^z = a_i^*(\omega|An_i)(\rho_i)\}. \tag{6}$$

The mixed strategy played by player i, under the analogy partition An_i in game ω is induced by $a_i^*(\omega|An_i)$ if and only if $\sigma_i(\omega|An_i)$ assigns probability $p_i(a_i^z;\omega, An_i)$ to a_i^z where

$$p_i(a_i^z;\omega, An_i) = \int \cdots \int_{\rho_i \in X_i^z(a_i^*(\omega|An_i))} d\rho_i(a_i^1,\omega)\dots d\rho_i(a_i^{|A_i|},\omega) g_i(\rho_i(a_i^1,\omega))\dots g_i(\rho_i(a_i^{|A_i|},\omega)) \tag{7}$$

and $g_i(\rho_i)$ is the continuously differentiable pdf of $\tilde{\rho}_i$.

Consider first the mapping $BR : (\beta, \lambda) \mapsto (\sigma', \lambda)$, where σ' is a profile of mixed strategies that is a best-response to β. Let $a_i^*(\omega|An_i)$ prescribe actions that are best responses to $\beta_i(\cdot|An_i)$, given the perturbed payoffs. Then BR is single-valued (because the set of realizations of the perturbations under which there are indifferences has measure zero), and it is readily verified that BR is a continuous function over convex and compact sets.

Consider the AG function, $(\sigma', \lambda) \mapsto (\bar{\sigma}', \beta')$ which aggregates the strategies over games and computes consistent expectations. AG is single-valued and continuous. Thus, the compound mapping $(\bar{\sigma}, \lambda) \mapsto_C (\beta, \lambda) \mapsto_{BR} (\sigma', \lambda) \mapsto_{AG} (\bar{\sigma}', \beta')$ is also continuous and single valued over convex and compact sets.

For the clustering part, the argument is somewhat similar to the best response part. Consider the calibration mapping $GC : (\bar{\sigma}', \beta') \mapsto (\bar{\sigma}', \lambda')$, where $\lambda' = (\lambda_i', \lambda_j')$ is such that λ_i' solves the global clustering problem for player i.

Consider the perturbed strategies $\bar{s}$, where $\bar{s}_j(\omega) = \frac{\bar{\sigma}_j(\omega)+\varepsilon\eta_i(\omega)}{1+\varepsilon}$ and impose that β_i is consistent with $\bar{s}_j$. As established in Theorem 1, for each $\omega \in \alpha$, the function $d(\bar{s}_j(\omega), \beta_i(\alpha|An_i))$ is continuous in $\bar{s}_j$.

We define $An_i^*(\eta_i)$ as the function mapping the realization of the perturbation $\tilde{\eta}_i$ to an analogy partition An_i that solves the clustering problem. As before, we denote by $X_i^k(An_i^*) = \{\eta_i | An_i^k = An_i^*(\eta_i)\}$ the set of realizations such that the k-th analogy partition is prescribed by An_i^*.

The mixture of analogy partitions λ_i is induced by $An_i^*(\cdot)$ iff λ_i assigns probability $q(An_i^k)$ to An_i^k where $q(An_i^k) = \int \cdots \int_{\eta_i \in X_i^k(An_i^*)} d\eta_i(\omega_1) \ldots d\eta_i(\omega_N) h_i(\eta_i(\omega_1)) \ldots . h_i(\eta_i(\omega_N))$, where h_i is the continuously differentiable pdf of η_i. We show now that the calibration mapping is single-valued. To establish this, we rely on results from chapter 2 in Milnor (1965). More precisely, we show that if An_i and An_i' yield the same V value (the criterion used for the clustering problem), then the set of realizations of $\tilde{\eta}_i$ that allow this has measure zero. Given, $\bar{\sigma}_j$ we can define the function $h(\eta_i) = V_i(\bar{s}_j, \beta_i(\cdot|An_i)) - V_i(\bar{s}_j, \beta_i(\cdot|An_i'))$, which is a mapping $h : U \to R$, where $\eta_i \in U$.[33] The function h is smooth (all partial derivatives exist and are continuous). Since $\eta_i(a_i, \omega) > 0$, for all ω and all a_i, then U is an open set. As $h(\eta_i) = 0$ is a regular value,[34] then the set $\{\hat{\eta}_i | h(\hat{\eta}_i) = 0\}$ is a smooth manifold of dimension $dim(U) - 1 = (|A_i| - 1) \cdot |\Omega| - 1$, which has measure zero in U. Then, the argument for C being single valued and continuous are the same as those used for BR.

Thus, the compound mapping

$$M : (\bar{\sigma}, \lambda) \mapsto_C (\beta, \lambda) \mapsto_{BR} (\sigma', \lambda) \mapsto_{AG} (\bar{\sigma}', \beta') \mapsto_{GC} (\bar{\sigma}', \lambda')$$

is single-valued and continuous, and it maps $\bar{\Sigma} \times \Lambda$ into $\bar{\Sigma} \times \Lambda$, which are convex and compact sets. By Brouwer's fixed point theorem, this mapping has a fixed point.

It is then readily verified that the fixed point (σ, λ) is a steady state of the learning dynamics. **Q.E.D.**

Proposition 11 *Consider a sequence of steady states $(\sigma^{(\varepsilon)}, \lambda^{(\varepsilon)})$ of the learning dynamics induced by ε where $\sigma^{(\varepsilon)}$ denotes the ex ante strategy (prior to the realizations of the perturbations ρ) and $\lambda^{(\varepsilon)}$ denotes the distribution of the profile of analogy partitions.*[35] *Consider an accumulation point (σ, λ) of $(\sigma^{(\varepsilon)}, \lambda^{(\varepsilon)})$ as ε tends to 0. (σ, λ) is a globally calibrated distributional ABEE.*

[33] V_i is defined as in the proof of Theorem 1.

[34] To show this, we note that the first derivatives of $h(\cdot)$ wrt to $\eta_i(a_i, \omega)$ are linearly independent as one varies a_i and ω.

[35] $\sigma_i^{(\varepsilon)}$ is a strategy of player i that depends on the game ω and the analogy partition An_i of player i as in the general construction above.

Proof. If $\lim_{\varepsilon\to 0}(\sigma^{(\varepsilon)}, \lambda^{(\varepsilon)}) = (\sigma, \lambda)$, then for ε small enough $supp[\sigma_i(\omega)] \subseteq supp[\sigma_i^{(\varepsilon)}(\omega)]$ and $supp[\lambda_i] \subseteq supp[\lambda_i^{(\varepsilon)}]$, for $i = 1, 2$ and $\omega \in \Omega$.

Since $(\sigma^{(\varepsilon)}, \lambda^{(\varepsilon)})$ is a steady state, any $An_i \in supp[\lambda_i^{(\varepsilon)}]$ solves the clustering problem for player i in the perturbed environment. Thus, for $\varepsilon = 0$, $An_i \in supp[\lambda_i]$ solves the clustering problem because d is continuous in ε (as established in Theorem 1, imposing consistency on β suffices to guarantee continuity in the case of KL divergence). The same argument can be made to show that σ is a best-response to λ. Thus, (σ, λ) is a steady state of the learning dynamics when $\varepsilon = 0$.

It follows that (σ, λ) is a globally calibrated ABEE because the requirements for the equilibrium and the steady states coincide when $\varepsilon = 0$ and the independence of the random draws ensures that $\sigma \in \Sigma_1 \times \Sigma_2$ and $\lambda \in \Lambda_1 \times \Lambda_2$. **Q.E.D.**

References

[1] Al-Najjar N. and M. Pai (2014): "Coarse Decision Making and Overfitting," *Journal of Economic Theory*, 150, 467–86.

[2] Anderson, J. R. (1991): "The adaptive nature of human categorization," *Psychological Review*, 98(3), 409–429.

[3] Aragones E., I. Gilboa, A. Postlewaite and D. Schmeidler (2005): "Fact-Free Learning," *American Economic Review*, 95(5), 1355–1368.

[4] Esponda I. and D. Pouzo (2016): "Berk-Nash Equilibrium: A Framework for Modeling Agents With Misspecified Models," *Econometrica*, 84 (3), 1093–1130.

[5] Eyster E. and M. Rabin (2005): "Cursed Equilibrium," *Econometrica*, 73(5), 1623–1672.

[6] Fryer R. and M. Jackson (2008): "A Categorical Model of Cognition and Biased Decision-Making," *Contributions in Theoretical Economics*, 8(1).

[7] Fudenberg D. and D. Kreps (1993): "Learning Mixed Equilibria," *Games and Economic Behavior*, 5 (3), 320–367.

[8] Gibbons R., M LiCalzi, M Warglien (2021): "What situation is this? Shared frames and collective performance," *Strategy Science* 6 (2), 124–140

[9] Holmström B. (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13(2), 324–340.

[10] Jehiel P. (1995): "Limited horizon forecast in repeated alternate games," *Journal of Economic Theory*, 67 (2), 497–519.

[11] Jehiel P. (2005): "Analogy-based expectation equilibrium," *Journal of Economic Theory* 123, 81–104.

[12] Jehiel P. (2022): "Analogy-based expectation equilibrium and related concepts: Theory, applications, and beyond," Prepared for the twelfth World Congress of the Econometric Society.

[13] Jehiel P. and F. Koessler (2008): "Revisiting Bayesian games with analogy-based expectations," *Games and Economic Behavior*, 62, 533–557.

[14] Jehiel P. and E. Mohlin (2023): "Categorization in Games: A Bias-Variance Perspective," WP ffhalshs04154272f

[15] Jehiel P. and D. Samet (2007): "Valuation equilibrium," *Theoretical Economics*, 2 (2), 163–185.

[16] Jehiel P. and G. Weber (2023): "Calibrated Clustering and Analogy-Based Expectation Equilibrium," PSE Working Papers halshs-04154234, HAL.

[17] Keynes J.M. (1936): *'The General Theory of Employment, Interest and Money,* Macmillan, Cambridge University Press.

[18] Lloyd S.P. (1957): "Least squares quantization in PCM," *Bell Laboratories Technical Memorandum*, later published as: Lloyd, S.P. (1982): "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, 28(2), 129–137.

[19] MacQueen J. (1967): "Classification and analysis of multivariate observations," *In 5th Berkeley Symp. Math. Statist. Probability*, 281-297.

[20] Mengel, F. (2012): "Learning Across Games," *Games and Economic Behavior*, 74(2), 601–619.

[21] Medin, D. L. and M. M. Schaffer (1978): "Context theory of classification learning," *Psychological Review*, 85(3), 207–238.

[22] Miller G. A. (1956): "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychological Review*, 63(2), 81–97.

[23] Milnor J. (1965): *'Topology from the Differential Viewpoint'*, University Press of Virginia

[24] Mohlin E. (2014): "Optimal Categorization," *Journal of Economic Theory*, 152, 356–381.

[25] Mullainathan S., J. Schwartzstein and A. Shleifer (2008): "Coarse Thinking and Persuasion," *The Quarterly Journal of Economics,* 123(2), pp. 577–619

[26] Myerson, R. and P. Reny (2020), "Perfect Conditional ε-Equilibria of Multi-Stage Games With Infinite Sets of Signals and Actions," *Econometrica*, 88 (2), 495–531.

[27] Nosofsky, R. M. (1986): "Attention, similarity, and the identification-categorization relationship," *Journal of Experimental Psychology: General*, 115(1), 39–57.

[28] Osborne M. and A. Rubinstein (1998): "Games with Procedurally Rational Players," *American Economic Review,*, 88(4), 834–47.

[29] Pęski M. (2011): "Prior symmetry, categorization and similarity-based reasoning," *Journal of Economic Theory*, 146(1), 111–140

[30] Posner, M. I. and S.W. Keele (1968): "On the genesis of abstract ideas," *Journal of Experimental Psychology*, 77(3, Pt.1), 353–363.

[31] Reed, S. K. (1972): "Pattern recognition and categorization," *Cognitive Psychology*, 3(3), 382–407.

[32] Rosch, E. and B.B. Lloyd (Eds.). (1978): Cognition and categorization. Lawrence Erlbaum.

[33] Samuelson, L. (2001). "Analogies, adaptation, and anomalies," *Journal of Economic Theory*, 97(2), 320–366.

[34] Sims C.A. (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50(3), 665–690.

[35] Spiegler R. (2016): "Bayesian Networks and Boundedly Rational Expectations," *Quarterly Journal of Economics*, 131, 1243–1290.

[36] Steinhaus H. (1957): "Sur la division des corps materiels en parties," *Bulletin L'Academie Polonaise des Science*, 4, 801–804.

[37] Vives X. (1999): *Oligopoly pricing: old ideas and new tools*, MIT press.

For Online Publication: Online Appendix

Globally CD-ABEE in Example 1

We also provide a description of the globally calibrated distributional ABEE.

Let $An_x = \{\{x\}, \{x', x''\}\}$ denote an analogy partition of player 1 and $\lambda_x = Pr(An_x)$. Let p_x denote the probability that L is played in game x by player 2.

Let $p_x \leq p_{x'} \leq p_{x''}$. For the purposes of global calibration it must be the case that $p_{x'} = \frac{p_x + p_{x''}}{2}$, and also $\lambda_{x'} = 0$ unless $p_x = p_{x'} = p_{x''}$, but it is readily verified that there is no ABEE such that $p_x = p_{x'} = p_{x''}$.

Then, in order to ensure global calibration, player 2's strategies must be such that $p_x < p_{x'} = \frac{p_x + p_{x''}}{2} < p_{x''}$. It is easily verified that there is no distributional ABEE where player 2 is playing pure strategies in game x or x''. Thus, player 2 is mixing in both games and, clearly, also in x'. For player 2's indifference, the aggregate strategy of player 1 $\bar{\sigma}_1(x)$ must be to play U with probability $\frac{1}{2}$ in all games x. In order to sustain such aggregate strategies for player 1, whenever $x < x' < x''$ player 1 must be indifferent in game x'' when using An_x and in game x when using $An_{x''}$.

To illustrate this, take $x < x' < x''$. Player 1 expects L to be played with probability $\frac{p_x + 3p_{x''}}{4}$ in games $\{x', x''\}$ when using An_x, and with probability $\frac{3p_x + p_{x''}}{4}$ in games $\{x, x'\}$ when using $An_{x''}$. By setting the former probability equal to $\frac{1}{2+x}$ and the latter equal to $\frac{1}{2+x''}$, we can always sustain $\bar{\sigma}_1(x)$ and $\bar{\sigma}_1(x'')$ where U is played with probability $\frac{1}{2}$. Also, since $\frac{p_x + 3p_{x''}}{4} < \frac{1}{2+x'} < \frac{3p_x + p_{x''}}{4}$, player 1 plays U in x' when using An_x and D when using $An_{x''}$. Therefore, U is played in aggregate in x' with probability $\frac{1}{2}$ if and only if $\lambda_x = \lambda_{x''} = \frac{1}{2}$.

ABEE in Beauty-Contest game, section 3.1

We provide here the details needed to compute the ABEE in the framework introduced in section 3.1, in the context of the Beauty-Contest game illustration. We assume players use the same analogy partition $(\Theta_k)_{k=1}^{K}$. For $\theta \in \Theta_k$, player i's best response is: $a_i(\theta) = (1-r)\theta + rE[a_j(\theta)|\Theta_k]$.

Then, the average best response of i in θ is:

$$E[a_i(\theta)|\Theta_k] = \int_{\Theta_k} a_i(\theta) f(\theta|\Theta_k) d(\theta) = (1-r)E[\theta|\Theta_k] + rE[a_j(\theta)|\Theta_k] \iff$$

$$E[a_i(\theta)|\Theta_k] = (1-r)E[\theta|\Theta_k] + r\left((1-r)E[\theta|\Theta_k] + rE[a_i(\theta)|\Theta_k]\right)$$

and so both players expects their opponent to play the average θ in the analogy class: $E[a_1(\theta)|\Theta_k] = E[a_2(\theta)|\Theta_k] = E[\theta|\Theta_k]$ and they play:

$$a_1^{ABEE}(\theta) = a_2^{ABEE}(\theta) = (1-r)\theta + rE[\theta|\Theta_k]$$

Proof of Proposition 7.

By definition, $\sigma_j(\mu) \in BR_j(\mu, \beta_i(\alpha_k))$ implies that the following equations must hold in equilibrium:

$$\sigma_j(\mu) = A + \mu B + \mu C \int_{\mu_{k-1}}^{\mu_k} \frac{f(\mu)}{F(\mu_k) - F(\mu_{k-1})} \sigma_i(\mu) d\mu$$

$$= A + \mu \left(B + AC + BC\mathbb{E}[\mu|\alpha_k] + C^2\mathbb{E}[\mu|\alpha_k] \int_{\mu_{k-1}}^{\mu_k} \frac{f(\nu)}{F(\mu_k) - F(\mu_{k-1})} \sigma_j(\nu) d\nu \right)$$

Taking the weighted average of $\sigma_j(\mu)$ over the interval $[\mu_{k-1}, \mu_k]$, according to the distribution of μ, yields the following equation:

$$\int_{\mu_{k-1}}^{\mu_k} \frac{f(\mu)}{F(\mu_k) - F(\mu_{k-1})} \sigma_j(\mu) d\mu =$$

$$= A + \mathbb{E}[\mu|\alpha_k] \left(B + AC + BC\mathbb{E}[\mu|\alpha_k] + C^2\mathbb{E}[\mu|\alpha_k] \int_{\mu_{k-1}}^{\mu_k} \frac{f(\nu)}{F(\mu_k) - F(\mu_{k-1})} \sigma_j(\nu) d\nu \right)$$

By consistency of $\beta_i(\alpha_k)$, the equation above simplifies into $\beta_i(\alpha_k) = \frac{A+B\mathbb{E}[\mu|\alpha_k]}{1-C\mathbb{E}[\mu|\alpha_k]}$.

In equilibrium, both players have the same expectations $\beta_1(\alpha_k) = \beta_2(\alpha_k)$. Substituting the expression of $\beta_i(\alpha_k)$ into the best-responses yields the following equilibrium (pure) strategies: for all $\alpha_k \in An_1$ (and An_2) and for all $\mu \in \alpha_k$,

$$\sigma_1(\mu) = \sigma_2(\mu) = A + \mu \frac{B + AC}{1 - C\mathbb{E}[\mu|\alpha_k]}$$

This is the unique ABEE and it is symmetric. **Q.E.D.**

Lemma 1 *Let σ be some strategy profile and d be either the squared Euclidean distance or the KL divergence. If $An_i \in \mathcal{K}_i$ is a globally calibrated analogy partition for player i with respect to σ, then An_i is a locally calibrated analogy partition for player i with respect to σ.*

Proof. Let An_i be a globally calibrated analogy partition with respect to σ. Let β_i be consistent with σ. Assume by contradiction that An_i is not locally calibrated. Then, $\exists \alpha_i, \alpha\prime_i \in An_i \wedge \hat{\omega} \in \alpha_i$ s.t. $d(\sigma_j(\hat{\omega}), \beta_i(\alpha_i)) > d(\sigma_j(\hat{\omega}), \beta_i(\alpha\prime_i))$.
Let $\widehat{\alpha_i} = \alpha_i \setminus \{\hat{\omega}\}$ and $\widehat{\alpha\prime_i} = \alpha\prime_i \cup \{\hat{\omega}\}$. Then,

$$\sum_{\omega \in \alpha_i} p(\omega) d(\sigma_j(\omega), \beta_i(\alpha_i)) + \sum_{\omega \in \alpha\prime_i} p(\omega) d(\sigma_j(\omega), \beta_i(\alpha\prime_i))$$

$$> \sum_{\omega \in \widehat{\alpha_i}} p(\omega) d(\sigma_j(\omega), \beta_i(\alpha_i)) + \sum_{\omega \in \widehat{\alpha\prime_i}} p(\omega) d(\sigma_j(\omega), \beta_i(\alpha\prime_i))$$

$$= p(\widehat{\alpha_i}) \sum_{\omega \in \widehat{\alpha_i}} p(\omega|\widehat{\alpha_i}) d(\sigma_j(\omega), \beta_i(\alpha_i)) + p(\widehat{\alpha\prime_i}) \sum_{\omega \in \widehat{\alpha\prime_i}} p(\omega|\widehat{\alpha\prime_i}) d(\sigma_j(\omega), \beta_i(\alpha\prime_i))$$

$$> p(\widehat{\alpha_i}) \sum_{\omega \in \widehat{\alpha_i}} p(\omega|\widehat{\alpha_i}) d(\sigma_j(\omega), \beta_i(\widehat{\alpha_i})) + p(\widehat{\alpha\prime_i}) \sum_{\omega \in \widehat{\alpha\prime_i}} p(\omega|\widehat{\alpha\prime_i}) d(\sigma_j(\omega), \beta_i(\widehat{\alpha\prime_i}))$$

where the second inequality holds because, when β_i is consistent with the strategies played in the games in a given analogy class, this is the representative object that minimizes the sum of prediction errors in that analogy class among all possible representative objects in ΔA_j.
Let $\widehat{An_i} = \widehat{\alpha_i} \cup \widehat{\alpha\prime_i} \cup \{An_i \setminus \{\alpha_i, \alpha\prime_i\}\}$, then:

$$\sum_{\alpha_i \in An_i} p(\alpha_i) \sum_{\omega \in \alpha_i} p(\omega|\alpha) d(\sigma_j(\omega), \beta_i(\alpha_i)) > \sum_{\alpha_i \in \widehat{An_i}} p(\alpha_i) \sum_{\omega \in \alpha_i} p(\omega|\alpha_i) d(\sigma_j(\omega), \beta_i(\alpha_i))$$

which contradicts An_i being a globally calibrated analogy partition. **Q.E.D.**

Lemma 2 *(Reverse Truncation) Let X be a RV on $[0,1]$ with continuous pdf f_X and cdf F_X. There exists a continuous pdf g over $[0,+\infty)$, such that:*

$$g_{X|[0,1]}(x) = f_X(x), \text{ where } g_{X|[0,1]}(x) = \frac{g_X(x)}{Pr_g[0 \leq X \leq 1]}$$

Proof. We want to find a continuous function g_X such that:

$$g_X(x) = \begin{cases} Pr_g[0 \leq X \leq 1] f_X(x) & 0 \leq x \leq 1 \\ v(x) & 1 < x < \infty \end{cases}$$

Note that g_X is continuous if $v(\cdot)$ is continuous and $v(1) = Pr_g[0 \leq X \leq 1] f_X(1)$. Also, g_X must be a pdf, so it must be the case that $\int_0^{+\infty} g_X(x)dx = 1$. That is, $\int_0^1 Pr_g[0 \leq X \leq 1] f_X(x)dx + \int_1^{+\infty} v(x)dx = Pr_g[0 \leq X \leq 1] + (1 - Pr_g[0 \leq X \leq 1]) = 1$.

Let us pick the right function $v(.)$. This function must be continuous and satisfy two conditions: (i) $\int_1^{+\infty} v(x)dx = 1 - T$, and (ii) $v(1) = T f_X(1)$, where $T \equiv Pr_g[0 \leq X \leq 1]$. Let $v()$ be defined as $v(x) = Tf(1)e^{\frac{Tf(1)}{1-T}(1-x)}$ which is continuous since, for $a, b \in \mathbb{R}$, the function ae^{bx} is continuous in x. Also, $v(1) = Tf(1)e^{\frac{Tf(1)}{1-T}(0)} = Tf(1)$.

And finally: $\int_1^{+\infty} v(x)dx = Tf(1)e^{\frac{Tf(1)}{1-T}} \int_1^{+\infty} e^{-\frac{Tf(1)}{1-T}x}dx = 1 - T$. **Q.E.D.**

Lemma 3 *There always exists an equidistant-expectations partition.*

Proof. We show that there always exists a sequence $\{\mu_k\}_{k=0}^{k}$ with $\mu_0 = 0$, some $\mu_1 \in [0,1]$ and, for $k = 2, \ldots, K$, μ_k defined so that $\mathbb{E}[\mu|(\mu_{k-1}, \mu_k]] = 2\mu_{k-1} - \mathbb{E}[\mu|(\mu_{k-2}, \mu_{k-1}]]$. We note that if μ_1 is too large, μ_K might be above 1. Lemma 2 allows us to consider μ to be a random variable from $[0, +\infty)$ distributed according to a continuous strictly positive pdf g, and cdf G, with $g(\mu) = f(\mu)G(1)$, for $0 \leq \mu \leq 1$.

Let $\mu_k \geq \mu_{k-1} \geq 0$. Since $g(\mu)$ is striclty positive and continuous in μ, then $\mathbb{E}[\mu|(\mu_{k-1}, \mu_k]] = \frac{1}{G(\mu_k) - G(\mu_{k-1})} \int_{\mu_{k-1}}^{\mu_k} \mu g(\mu)d\mu$ is continuous and strictly decreasing in μ_{k-1}, and it is continuous and strictly increasing in μ_k. Moreover, if $\mu_k \leq 1$, we have:

$$\frac{\int_{\mu_{k-1}}^{\mu_k} \mu g(\mu) d\mu}{G(\mu_k) - G(\mu_{k-1})} = \frac{\int_{\mu_{k-1}}^{\mu_k} \mu G(1) f(\mu) d\mu}{G(1)(F(\mu_k) - F(\mu_{k-1}))}$$

so the term $G(1)$ cancels out and we are back to the original distribution F.

Fix μ_{k-1}. The function $m(\mu_k) \equiv \mathbb{E}[\mu|[\mu_{k-1}, \mu_k]]$ is continuous and strictly increasing over $(\mu_{k-1}, +\infty)$, with image $(\mu_{k-1}, +\infty)$. Then the inverse function m^{-1} exists over $(\mu_{k-1}, +\infty)$ and it is continuous and strictly increasing over $(\mu_{k-1}, +\infty)$.

Given μ_{k-2}, μ_{k-1}, we use the inverse function to retrieve μ_k from the equation $\mathbb{E}[\mu|(\mu_{k-1}, \mu_k]] = 2\mu_{k-1} - \mathbb{E}[\mu|(\mu_{k-2}, \mu_{k-1}]]$. Let $h(\mu_{k-2}, \mu_{k-1}) \equiv 2\mu_{k-1} - \mathbb{E}[\mu|(\mu_{k-2}, \mu_{k-1}]]$. Note that $h(\cdot)$ is a continuous function and $h(\mu_{k-2}, \mu_{k-1}) \geq \mu_{k-1}$.

Starting from $\mu_0 = 0$ and some $\mu_1(\mu_1) \equiv \mu_1$, we recursively define μ_k as a function of μ_1 as follows: $\mu_2(\mu_1) = m^{-1}(h(\mu_0, \mu_1))$ and

$$\mu_k(\mu_1) = m^{-1}(h(\mu_{k-2}(\mu_1), \mu_{k-1}(\mu_1)))$$

for $k = 3, \ldots, K$. Note that, for each k, the function $\mu_k(\mu_1)$ is well defined. Since $h(\mu_{k-2}, \mu_{k-1}) \geq \mu_{k-1}$, then the inverse function exists at the point $h(\mu_{k-2}, \mu_{k-1})$.

Since $m^{-1} : (\mu_{k-1}, +\infty) \to (\mu_{k-1}, +\infty)$ is also strictly increasing and continuous, then $\mu_k(\mu_1)$ is continuous, being a composition of continuous functions, and $\mu_k(\mu_1) \geq \mu_{k-1}$, with equality if and only if $h(\mu_{k-2}, \mu_{k-1}) = \mu_{k-1} \iff \mu_{k-1} = \mathbb{E}[\mu|(\mu_{k-2}, \mu_{k-1}]] \iff \mu_{k-1} = \mu_{k-2}$. So, either we get the sequence with 0 everywhere, or a strictly increasing sequence.

Let $\mu_1 = 0$, then $\mu_K(0) = 0$. Let $\mu_1 = 1$, then $\mu_K(1) > 1$. Then, by the intermediate value theorem, there must exist $0 < \mu_1^* < 1$ such that $\mu_K = 1$. **Q.E.D.**