Categorization in Games: A Bias-Variance Perspective*

Philippe Jehiel† Erik Mohlin‡

January 30, 2024

Abstract

We develop a framework for categorization in games, applicable both to multi-stage games of complete information and static games of incomplete information. Players use categories to form coarse beliefs about their opponents' behavior. Players best-respond given these beliefs, as in analogy-based expectations equilibria. Categories are related to previously used strategies via the requirements that categories contain a sufficient amount of observations and exhibit sufficient within-category similarity, in line with the bias-variance trade-off. When applied to classic games including the ultimatum games, the chainstore game and adverse selection games our framework yields more intuitive predictions than those arising with standard solution concepts.

Keywords: Bounded rationality; Categorization; Bias-variance trade-off; Adverse selection; Chainstore paradox; Ultimatum bargaining.

JEL codes: C70, C73, D82, D83, D91.

*This paper has benefited from comments by Tore Ellingsen, Drew Fudenberg, Topi Miettinen, Alexandros Rigos, and Larry Samuelson. We also thank audiences at Lund University (Arne Ryde Workshop on Attention in Decision Making) and Bar-Ilan University (Learning Evolution Games 2019) for comments. Maria Juhlin provided excellent research assistance at an early stage of the project. Philippe Jehiel thanks the European Research Council (grant no. 742816) for funding. Erik Mohlin is grateful for financial support from the Swedish Research Council (grant no. 2015-01751 and 2019-02612) and the Knut and Alice Wallenberg Foundation (Wallenberg Academy Fellowship 2016-0156).

†Paris School of Economics and University College London. Address: PSE, 48 boulevard Jourdan, 75014 Paris, France. E-mail: jehiel@enpc.fr.

‡Lund University and the Institute for Futures Studies (Stockholm). Address: Lund University Department of Economics, Tycho Brahes väg 1, 220 07 Lund, Sweden. E-mail: erik.mohlin@nek.lu.se.

1 Introduction

Human decision-makers need to make simplifications in order to navigate social reality. We need to divide the complex web of interactions into manageable pieces to evaluate different courses of action. We need to extrapolate from past interactions to be able to predict what others will do. Categories serve these functions (Anderson, 1991; Laurence and Margolis, 1999; Gärdenfors, 2000; Murphy, 2002; Xu, 2007). A categorization bundles distinct objects or situations into groups or categories, whose members are viewed as sufficiently similar to warrant a similar treatment. As a result, categorical reasoning facilitates prediction: when a situation is classified as belonging to a category then by virtue of its similarity with other members of the category we expect similar behavior.

From the perspective of statistics and machine learning, categorizations should satisfy some properties to address the bias-variance trade-off (e.g. Geman et al., 1992). On the one hand, if categories are too coarse, bundling together situations that are too dissimilar, the resulting estimates are likely to be too biased. On the other hand, if categories are too narrow, bundling together too few data points, the resulting estimates will be unreliable, as they are plagued by high variance. Gigerenzer and Brighton (2009) discuss how simple heuristics typically used by humans can be viewed as devices inducing some bias in order to reduce variance. Mohlin (2014) derives properties of categorizations that solve the bias-variance trade-off optimally for the purpose of making predictions.[1]

In economics, a growing literature has introduced categorical thinking into game theory (Samuelson, 2001; Jehiel, 2005; Jehiel and Samet, 2007; Jehiel and Koessler, 2008; Azrieli, 2009; Mengel, 2012; Arad and Rubinstein, 2019). [2] A significant part of this literature has worked with exogenously given categories. While in some interactions it may be reasonable to assume that the categorizations are given exogenously, either by the framing of the game, or by players' culture or previous personal experience, in other interactions, it seems more appropriate to view the categories as being formed within the learning environment.[3]

Our starting point is the analogy-based expectation equilibrium (Jehiel, 2005; Jehiel and Koessler, 2008) in which categories (analogy classes) are used to form predictions about opponents' play. We endogenize the analogy partitions relying on principles inspired by the bias-variance trade-off. Specifically, we envision different cohorts of players. Within each cohort players are randomly matched to play a given game, and, depending on player roles,

[1] In hierarchical categorizations the location of the basic level (Rosch et al., 1976), which is neither the most fine-grained nor the most abstract level, may be influenced by bias-variance considerations. Similarly it may be responsible for why experts have more fine-grained categorizations than laymen (Tanaka and Taylor, 1991).

[2] See also Dow (1991), Rubinstein (1998), and Fryer and Jackson (2008).

[3] Some of the cited papers endogenize the categories assuming there is a fixed cost to adding a category and considering the best possible categorization in terms of the induced payoff consequence minus the cost associated with the categorization. Such approaches require a high degree of rationality (arguably at least as high as in standard economic models) while our interest lies in situations in which the rationality of subjects is more limited.

they are called to move in different situations (nodes or states depending on the application) in the game. After the play of a given game, new players receive some feedback about it, consisting in the disclosure of behaviors in a subset of situations. For example, in extensive form games, the feedback will typically consist of the on-path behaviors, and in Bayesian games we will consider cases in which behaviors are disclosed if some event (such as trade) occurs. We assume that in each situation, a player may pick a non-intended action with some exogenous probability ε, as in the trembling-hand formulation (Selten, 1975). This generates some observations for all situations.

When a new cohort arrives, the corresponding players categorize the situations in which their opponents have to make a move, using the available data from the games played by the previous cohort. Players are endowed with *exogenous* similarity functions, representing their perception of how similar various situations are to each other. They form *endogenous* categories by bundling together situations perceived to be as similar as possible, while respecting the desiderata that each category should contain enough data points, in line with the bias-variance trade-off. We formalize this by imposing that each analogy class should have a mass of observations no less than a threshold κ, unless doing so creates too high within-category dissimilarity. A player's prediction about the play of the opponent in a given situation is assumed to correspond to the empirical distribution of the behaviors observed in the previous cohort in the category to which the situation has been assigned. When a player does not tremble, she best-responds to such predictions. In order to the describe the steady-states of the induced dynamic system we define a notion of (ε, κ)-categorization equilibrium.

While our approach allows for any specification of κ and ε, we focus on the case in which κ and ε are small and vanish at such a rate that ε is asymptotically not too large relative to κ, implying that on-path situations can be distinguished perfectly but off-path situations have to be bundled (according to their similarity). When considering such limits and in some applications, we need to describe the learning dynamic more explicitly since the limit does not have a steady state representation. Throughout, both the feedback structure and the similarity functions are primitives of our model.

Our first main contribution is to provide a general framework that endogenizes the analogy partitions based on the similarity functions used by the players as well as the relative rate of the trembling behavior and the threshold mass used to implement the bias-variance trade-off. Our second main contribution is to provide a series of applications where in each case we motivate our choice of similarity functions based on our intuitive understanding of the interaction.

Our main applications are as follows. We first consider ultimatum games in which the responder has a fixed outside option. We illustrate that offers leaving positive surplus to the responder can be sustained as categorization equilibria using as similarity functions ones based on the Euclidean distance between offers. We next consider chainstore games

(Selten, 1978), and assume (for both the monopolist and the challengers) that histories in which there was some entry that was not immediately followed by a fight are treated as very dissimilar from other histories (in which either the challengers never entered or when there was entry it was always followed by a fight). We establish the existence of a categorization equilibrium with no entry except in the last few periods. Finally, we discuss adverse selection games of the Akerlof type, modeled as a Bayesian game between an informed seller and an uninformed buyer who values the good more than the seller (as in Esponda, 2008). Assuming that qualities (seller types) are considered more similar when they are close to each other, we show that the learning dynamics leads to cycles with intended bid prices always lying above the Nash equilibrium price.

Our paper can be related to several strands of the literature. First, since we assume that the threshold κ and the trembling probability ε are such that on-path nodes are treated as singleton analogy classes in extensive form games, our categorization equilibria in such games can be viewed as offering a selection device for self-confirming equilibria (Fudenberg and Levine (1993, 1998). Fudenberg and Levine (2006) propose a different selection referred to as subgame-confirmed equilibria in which strategy profiles are Nash equilibria on-path and self-confirming equilibria one step away from the path. They have provided a learning foundation for it in a class of extensive form games (in which players move once) when players are patient and experiment optimally.[4] Our approach differs from Fudenberg and Levine (2006) in several respects. Most importantly, in our case, when beliefs are incorrect, they are related to the true behaviors via a categorization which is itself structured by the similarity functions, the trembling probability, and the minimum size requirement imposed on categories.

Second, a number of approaches have been proposed to avoid the unintuitive predictions obtained in finite horizon interactions, including Kreps et al. (1982)'s crazy type approach and Neyman (1985)'s finite automaton approach (c.f. Rubinstein, 1998). We note that these approaches avoid the classical predictions in all versions of the finite horizon paradox (this is true also of Jehiel (2005)'s ABEE-approach). This is not the case in our setting as we discuss in the Section devoted to the chainstore game.

Third, a number of approaches have revisited the classic adverse selection games introduced by Akerlof (1970) and studied whether relaxations of the buyer's rationality could generate more trading activity. These include Eyster and Rabin (2005)'s cursed equilibrium, Jehiel and Koessler (2008)'s analogy-based expectation equilibrium, and Esponda (2008)'s behavioral equilibrium.[5] Our modeling of such interactions is inspired by Esponda (2008), in particular with respect to the feedback function. But, our derivation of categorization-based expectations based on that feedback is different, leading to more trade than in the rational case (in contrast to Esponda's finding), as well as cycling (which has no counter-

[4]See also Kalai and Neme (1992) who have proposed another notion of equilibrium for extensive form games in which strategy profiles are Nash equilibria up to p steps away from the path.

[5]See Miettinen (2009) on the relationship between these various approaches.

part in the other approaches).[6,7]

Fourth, our paper can be related to a growing literature on misspecifications in games, which, in addition to the already mentioned cursed equilibrium (Eyster and Rabin, 2005) and analogy-based expectation equilibrium (Jehiel, 2005), include the Berk-Nash equilibrium (Esponda and Pouzo, 2016) and the Bayesian Network Equilibrium (Spiegler, 2016). Some papers have suggested endogenizing the misspecifications based on evolutionary arguments (in particular He and Libgober, 2020; Fudenberg and Lanzani, 2023; Heller and Winter, 2020), but to the best of our knowledge, none of these papers have developed an approach based on the bias-variance trade-off to endogenize misspecifications.[8]

2 Framework

We present our approach within a unified setup covering both multi-stage games of complete information and (static) Bayesian games. Specifically, we consider games with two players $i \in I = \{1, 2\}$ such that player $i \in I$ faces various possible situations referred to as $x_i \in \mathcal{X}_i$, and in situation x_i player i has to choose an action $a_i \in A_i(x_i)$. Extension to more players is straightforward. In an extensive-form game with complete information, $\mathcal{X}_i$ will represent the nodes at which player i must move. In a Bayesian game, $\mathcal{X}_i$ will represent the set of types of player i. In the former case, the profile of actions chosen by the two players at the various nodes determines which nodes are visited. In the latter case, nature chooses the profile of types according to some probability assumed to be known by both players. For simplicity and mostly to avoid notational complexity dealing with densities instead of probabilities, we consider the finite case in which the set of situations and the sets of actions are all finite. In some of the applications developed next, we will consider straightforward extensions of the definitions to the case of a continuum of actions and situations.

A strategy for player i is defined by $\sigma_i = (\sigma_i(x_i))_{x_i \in \mathcal{X}_i}$ where $\sigma_i(x_i) \in \Delta A_i(x_i)$ describes the probability distribution over possible actions chosen by player i at x_i. A realized play of the game is described by the set of situations that occurred and the actions taken in

[6] We note that our predictions for this type of interactions are broadly in line with the experimental findings reported in Fudenberg and Peysakhovich (2016). They observe more trade than predicted by the Nash equilibrium and they suggest comparative statics with respect to the difference of valuation between the seller and the buyer that agree with our predictions.

[7] A few recent papers identify cycles of beliefs in the context of misspecified models. In Esponda et al. (2021) and Bohren and Hauser (2021) (see also Nyarko, 1991), the evidence accumulated while taking a particular action may push beliefs in a direction that makes another action seem optimal, and once this new action is taken the data that are being generated induce a belief that makes the previous action seem optimal again. In Fudenberg et al. (2017) cycles may arise from the fact that the learner never ceases to perceive an information value of experimenting with another action. None of these papers feature endogenous categorizations.

[8] In a contemporaneous paper, Jehiel and Weber (2023) endogenize the analogy partitions based on clustering techniques that are standard in machine learning. This is a different approach from the one pursued here.

those situations, as dictated by $\sigma = (\sigma_1, \sigma_2)$ and the strategy of nature. A realized play is denoted

$$(\hat{a}, \hat{x}) = \left\{ (\hat{a}_i, \hat{x}_i)_{i \in I} : \hat{x}_i \text{ occurred and } i \text{ chose } \hat{a}_i \text{ at } \hat{x}_i \right\}.$$

Regarding the feedback, we assume that after the play of a game only a subset of $(\hat{a}, \hat{x})$ is disclosed to outsiders (which will be used by new players to form expectations). We refer to such a disclosure as the feedback given the play and denote it by $\phi(\hat{a}, \hat{x})$.

In dynamic games, we assume that only the actions on the path of play are observed (as is commonly assumed, see Fudenberg and Levine, 1998). In Bayesian games, we will use this formulation to accommodate applications like trades in which the actions (bargaining offers) and types (determining the quality of the good) would be disclosed only when the transaction takes place (as in our adverse selection game).

2.1 Analogy-Based Expectations

Player i categorizes $\mathcal{X}_j$ (the set of player j's situations) into analogy classes $\mathcal{C}_i^1, ..., \mathcal{C}_i^K$ that constitute a partition $\mathcal{C}_i = \left\{\mathcal{C}_i^1, ..., \mathcal{C}_i^K\right\}$ of $\mathcal{X}_j$. An analogy class $\mathcal{C}_i^k \in \mathcal{C}_i$ of player i satisfies the requirement that if x_j and x_j' belong to the same analogy class $\mathcal{C}_i^k$, then the action spaces of player j at x_j and x_j' are the same. We let $\beta_i(\mathcal{C}_i^k)$ denote the analogy-based expectation of player i about the play of player j in $\mathcal{C}_i^k$. It is a probability distribution over the action space of player j in $\mathcal{C}_i^k$ meant to capture how player i views player j's representative behavior in $\mathcal{C}_i^k$.[9] For every $x_j \in \mathcal{X}_j$, we let $\mathcal{C}_i(x_j)$ be the unique analogy class $\mathcal{C}_i^k$ to which x_j belongs. We refer to $\beta_i = \left(\beta_i(\mathcal{C}_i^k)\right)_{k=1}^K$ as the analogy-based expectation of player i.

Given β_i, player i expects player j to behave according to the strategy defined by $\sigma_j^{\beta_i} = \left(\sigma_j^{\beta_i}(x_j)\right)_{x_j \in \mathcal{X}_j}$, with $\sigma_j^{\beta_i}(x_j) = \beta_i(\mathcal{C}_i(x_j))$. That is, player i expects player j in situation x_j to behave according to the representative behavior in the analogy class $\mathcal{C}_i(x_j)$ to which x_j belongs as defined by $\beta_i(\mathcal{C}_i(x_j))$.[10]

Most of the time player i plays a best-response to $\sigma_j^{\beta_i}$ (given his utility and information) and the rest of the time player i trembles and chooses any available action.[11] We require that the trembles occur independently at the various x_i. In other words, our treatment is similar to the extensive-form version of the trembling-hand equilibrium (Selten, 1975). Formally,

[9]In the case of n players $\mathcal{C}^i$ partitions $\times_{j \neq i} X_j$, with the requirement that if x_j and x_l' (possibly belonging to different players) belong to the same analogy class $C_k^i \in \mathcal{C}^i$, then the action spaces of player j at x_j and player l at x_l' are the same. Furthermore, $\beta_i(C_k^i)$ denotes the analogy-based expectation of player i about the play of all players acting at the situations belonging to C_k^i.

[10]The rationale for this is that this is the simplest representation of player j's strategy consistent with β_i (for elaboration see Jehiel, 2022).

[11]Such trembling can be viewed as reflecting exogenous experimentation at the learning stage (similar to Fudenberg and Kreps, 1993).

Definition 1 *σ_i is an ε_i-*perturbed best-response *to β_i if σ_i is a best-response to $\sigma_j^{\beta_i}$ subject to the constraint that at every x_i, $\sigma_i(x_i)$ assigns a probability no less than ε_i to every action at x_i and the probability distributions $\sigma_i(x_i)$ are independent across the various x_i.*

Remark 1 *(a) In the definition of ε_i-perturbed best-response, we implicitly assume that the probability of tremble is the same for all actions at x_i, and the same at all x_i. We could obviously extend this to allow for more general trembling strategies, but this would bring no additional insight. (b) The best-response is implicitly defined at the ex ante stage, but given that we consider games with perfect recall and all situations are reached with positive probability (due to trembling), the same choice of strategy would arise had we required an interim or sequential notion of best-response.*

In general, we allow for the possibility that players i and j have different probabilities of trembles, and we denote the profile of tremble probabilities by $\varepsilon = (\varepsilon_i, \varepsilon_j)$. This is to allow us to accommodate applications in which we believe one player is less likely to tremble than the other player (perhaps because the former but not the latter has a dominant strategy). The situations that are reached with positive probability in the absence of trembles ($\varepsilon = 0$) will be referred to as *on-path situations*. The remaining situations, which are reached with positive probability only when there are trembles (ε_i, $\varepsilon_j > 0$) are *off-path situations*. This distinction will play a role when we endogenize the analogy partitions.

In steady state, the analogy-based expectations are required to be related to the strategy profile and the feedback structure through a consistency requirement. Formally, a strategy profile σ together with a feedback structure ϕ and trembling behavior (as parameterized by ε), induces a probability $\mu^\sigma(a_j,x_j)$ that action a_j in situation x_j is disclosed.[12] We assume that ϕ is such that for every ε-perturbed strategy profile σ, and for every analogy class $\mathcal{C}_i^k$, some behavior in $\mathcal{C}_i^k$ is disclosed with strictly positive probability. That is,[13]

$$\mu^\sigma(\mathcal{C}_i^k) = \sum_{x'_j \in \mathcal{C}_i^k, a'_j \in \mathcal{A}_j(x'_j)} \mu^\sigma(a'_j,x'_j)$$

is strictly positive for every $\mathcal{C}_i^k$.

Definition 2 *The analogy-based expectation β_i is* consistent *with the ε-perturbed strategy profile σ and the feedback ϕ if for every $\mathcal{C}_i^k$, and every action a_j in the action space of player j at $\mathcal{C}_i^k$,*

$$\beta_i(\mathcal{C}_i^k)[a_j] = \frac{1}{\mu^\sigma(\mathcal{C}_i^k)} \sum_{x_j \in \mathcal{C}_i^k} \mu^\sigma(a_j,x_j), \tag{1}$$

where $\beta_i(\mathcal{C}_i^k)[a_j]$ refers to the probability assigned to action a_j by $\beta_i(\mathcal{C}_i^k)$.

[12] We do not include a reference to ϕ in μ^σ since ϕ will be taken as fixed and exogenous throughout. We also do not include reference to ε as it will be clear from the context.

[13] Observe that $\mu^\sigma(\mathcal{C}_i^k)$ is not a probability as it could be greater than 1 in some cases. This reflects that in extensive-form games, a single play of the game typically allows one to reach more than one situation. Also note that μ is normalized so that there is a mass 1 of games being played.

Combining Definition 1 and Definition 2 we propose a generalized version of analogy-based expectation equilibrium:[14]

Definition 3 *Given a profile of analogy partitions* $\mathcal{C} = (\mathcal{C}_1, \mathcal{C}_2)$*, and a feedback structure* ϕ*, an* ε-perturbed analogy-based expectation equilibrium *is a strategy profile* $\sigma = (\sigma_1, \sigma_2)$ *such that there exists a profile of analogy-based expectations* $\beta = (\beta_1, \beta_2)$ *satisfying for* $i = 1, 2$*:*

(a) σ_i *is an* ε_i*-perturbed best-response to* β_i*,*

(b) β_i *is consistent with* (σ, ϕ) *as defined in (1).*

We have in mind that the knowledge of β_i is derived by player i through learning (and not by introspection). To the extent that player i bases his choice of strategy solely on β_i, it makes sense to assume that player i is unaware of the payoff, information and categorization structure of player j. Player i need not be aware of ϕ, the feedback structure either.

2.2 Endogenous Categorizations

Players put together situations according to their perceived similarity under the constraint that categories should contain a sufficient amount of data if possible.

Formally, each player i is endowed with a subjective homogeneity function $\zeta_i : 2^{\mathcal{X}_j} \to [0, 1]$ defined over subsets of $\mathcal{X}_j$ where for every $\mathcal{C}_i^k \subseteq \mathcal{X}_j$, $\zeta_i(\mathcal{C}_i^k) \in [0, 1]$ is a measure of how similar to one another the situations in the set $\mathcal{C}_i^k$ are perceived by player i to be. We assume that a singleton set has maximum homogeneity, i.e. $\zeta_i(\{x_j\}) = 1$ for all $x_j \in \mathcal{X}_j$, though we also allow for homogeneity functions such that $\zeta_i(X) = 1$ for some non-singleton X. We allow for homogeneity functions such that for some non-singleton subset $X \subseteq \mathcal{X}_j$ it holds that $\zeta_i(X) = 0$, in which case the set X is considered maximally heterogeneous (because the situations in X are considered very dissimilar).[15,16]

We relate the analogy partitions of the players to the strategy profile for any given threshold parameter κ through the following definition:

[14]When the feedback ϕ is complete (i.e. when it contains information about the entire profile (a, x) for all choices of action profiles) or when it contains information only about the equilibrium path in extensive form games of complete information, the above definition is equivalent to the one provided in Jehiel (2005) for extensive form games or Jehiel and Koessler (2008) for Bayesian games. For more general specifications of the feedback structure ϕ , our definition can be viewed as a natural generalization of the analogy-based expectation equilibrium as previously defined.

[15]If two situations $x_i, x_i' \in \mathcal{X}_i$ have different actions sets, i.e. $\mathcal{A}_i(x_i) \neq \mathcal{A}_i(x_i')$, we assume that any subset that contains both situations has maximal dissimilarity, which implies that an adjusted analogy partition will never bundle nodes with different action sets, as required in our construction.

[16]It would be natural to impose further extra properties, such that if $X \subseteq X'$ then $\zeta_i(X) > \zeta_i(X')$, but this will not matter for our analysis in this paper.

Definition 4 *Given σ and a threshold $\kappa > 0$, we say that $\mathcal{C} = (\mathcal{C}_i, \mathcal{C}_j)$ is κ-adjusted to σ if for each player i, her analogy partition $\mathcal{C}_i = \left\{\mathcal{C}_i^1, ..., \mathcal{C}_i^K\right\}$ satisfies the following criteria*

1. *For each $x \in \mathcal{X}_j$ with $\mu^\sigma(\{x\}) \geq \kappa$, there exists k such that $\mathcal{C}_i^k = \{x\}$.*
2. *If $X \subseteq \mathcal{X}_j$ and $\zeta_i(X) = 0$, there exists no k such that $\mathcal{C}_i^k = X$.*
3. *Let $\mathcal{X}_j^{sing}$ denote the set of situations put into singleton analogy classes in $\mathcal{C}_i$. If $\mathcal{C}_i^k$ is such that $\mu^\sigma(\mathcal{C}_i^k) < \kappa$, then for any $X \subseteq \mathcal{X}_j \backslash (\mathcal{C}_i^k \cup \mathcal{X}_j^{sing})$, it holds that $\zeta_i(\mathcal{C}_i^k \cup X) = 0$.*
4. *For any collection of non-singleton analogy classes $\left\{\mathcal{C}_i^{k_1}, ..., \mathcal{C}_i^{k_M}\right\}$ in $\mathcal{C}_i$, there is no collection $\left\{X^1, ..., X^N\right\}$ of pairwise disjoint sets, such that $\cup_{j=1}^N X^j = \cup_{j=1}^M \mathcal{C}_i^{k_j}$, $\mu^\sigma(X^j) \geq \kappa$ for all j, and $\min_{j=1}^N \zeta_i(X^j) > \min_{j=1}^M \zeta_i(\mathcal{C}_i^{k_j})$.*

The threshold parameter κ captures the amount of data that is considered necessary by the players to find the estimate in a category sufficiently reliable in line with the bias-variance trade-off. We could have considered a different threshold parameter κ_i for each player i, but our applications will not make use of such an asymmetry.

Roughly, the first condition says that if a situation is encountered enough times (as parameterized by κ), it is treated as a singleton analogy class (as there is no need to bundle it with other situations to meet the minimum mass criterion). In applications, we will have in mind that on-path situations satisfy this minimum mass requirement. The second condition requires that when a subset of situations is considered to induce maximal heterogeneity, the corresponding situations cannot be bundled together into one analogy class, which seems like a natural condition to impose.[17] The third condition says that the only reason for an analogy class not to meet the minimum mass condition is that adding other situations to the analogy class would induce maximum heterogeneity. The fourth condition requires a kind of local optimality requirement considering as criterion the infimum of homogeneities over the various analogy classes.[18]

The reduced-form properties in Definition 4 can be related to optimality properties obtained in simple prediction problems (as considered in Mohlin, 2014). In a prediction problem, one has to predict a random variable $Y \in \mathbb{R}$ associated with an observation $X = x \in \mathcal{X} \subseteq \mathbb{R}^n$. Pairs (X, Y) are independent draws from a continuous and bounded joint probability density function f, such that $Y = m(x) + \varepsilon(x)$ where $m(x)$ denotes the

[17] Instead of employing the notion of sets with maximal heterogeneity we could speak of sets whose homogeneity is below some threshold. For example part 2 of Definition 4 could be rephrased as follows: 'If $X \subseteq \mathcal{X}_j$ and $\zeta_i(X) \leq \delta$, there exists no k such that $\mathcal{C}_i^k = X$.' The threshold δ would be a primitive of the model in the same vein as κ. Sets with homogeneity below the threshold would serve the same function as sets with maximal heterogeneity in our current set-up.

[18] With no impact on the analysis, one could have generalized condition 4 to require that it is not the case that when $\cup_{j=1}^N X^j = \cup_{j=1}^M \mathcal{C}_i^{k_j}$, we have that $\mu^\sigma(X^j) \geq \kappa$ for all j, and $W(\zeta_i(X^1), ...\zeta_i(X^n)) > W(\zeta_i(\mathcal{C}_i^k), \zeta_i(\mathcal{C}_i^{k'}))$ for some given increasing and concave function W. We have chosen the infimum criterion mostly to avoid adding an extra less central notation.

conditional mean of Y at x and $\varepsilon(x)$ denotes an error term with variance σ_x^2 assumed to be independently drawn across observations. The agent partitions $\mathcal{X}$ into categories and upon observing x predicts that Y is equal to the empirical average associated with objects in the category x belongs to given the finite observed sample. A categorization is said to be optimal if it minimizes the expected squared prediction error. It turns out that (asymptotically as the sample size grows large) an optimal categorization features categories that are larger for parts of $\mathcal{X}$ where the variance σ_x^2 is high, the density f is low, and the conditional mean is rough (in the sense that the local variations of m are big).[19] Since f is continuous, Euclidean distance acts as a proxy for differences in conditional mean. When the conditional mean moves more relative to Euclidean distance (i.e. the derivative of m is larger) there is a greater need to reduce Euclidean distance within categories, i.e. to increase within-category homogeneity.

Relating this to our current framework, we believe that an agent's intuitive judgment of similarity among decision situations are responsive to cues that tend to proxy for differences in behavior, in the typical environment of the agent. Consequently, the agent prefers analogy classes that are homogeneous with respect to these intuitive similarity judgments. The comparative static results for the optimal categorizations in Mohlin (2014) have analogs in the conditions of Definition 4. The first condition incorporates the effect of the density. The second condition relates to the effect of the roughness of the conditional mean. The third condition relates to the interaction of density (in the form of the minimum mass condition) and the roughness of the conditional mean (in the form of the maximum heterogeneity condition).

There are however notable differences between our setting and the one studied in the prediction problem of Mohlin (2014). In our approach, the homogeneity function used by an agent is subjective and viewed as a primitive. This is to be contrasted with Mohlin's setup in which f and thus the notion of homogeneity (as induced by the Euclidean distance and the roughness of the conditional mean) are objective. Moreover, we do not consider samples of finite size in our approach, which allows us to eliminate estimation errors in each category (as reflected in the definition of consistent analogy-based expectations). This is to simplify matters and to focus on the non-random dimension of the bias induced by the categorical expectation formation. It also implies that our Definition 4 cannot incorporate a role for the variance of the data-generating process (unlike in Mohlin, 2014).[20] Given the subjective character of the prediction problem to be solved by players, we believe that our reduced-form approach as captured in Definition 4 is preferable to an exact optimization criterion, especially taking into account the potential difficulty players may face when solving such optimization problems.

[19] One may ask why agents use categorizations rather than other statistical methods, such as kernel-regression, to form predictions. We refer to section 5.1 of Mohlin (2014) for a discussion of this matter.

[20] Extending the model to allow for estimation errors as well as for the possibility that players subjectively consider the presence of aggregate shocks that apply to all data of a given situation is left for future research.

As an alternative to Definition 4, player i may consider solving an optimization problem consisting in the choice of an analogy partition $\mathcal{C}_i = \left\{\mathcal{C}_i^k\right\}_{k=1}^K$ that is a solution to

$$\max_{\mathcal{C}_i \in \mathcal{P}(\mathcal{X}_j)} W_i(\widetilde{\kappa}_i\left(\zeta_i\left(\mathcal{C}_i^1\right), \mu^\sigma(\mathcal{C}_i^1)\right), ..., \widetilde{\kappa}_i\left(\zeta_i\left(\mathcal{C}_i^K\right), \mu^\sigma(\mathcal{C}_i^K)\right), \tag{2}$$

for some functions W_i and $\widetilde{\kappa}_i$ assumed to be weakly increasing in their arguments where $\mathcal{P}\left(\mathcal{X}_j\right)$ denotes the set of partitions of $\mathcal{X}_j$. We note that such an approach to the link between the analogy partition of player i and the strategy profile σ would be less parsimonious than our proposed one without adding much insight to our subsequent applications. Moreover, such an optimization problem would generally be hard to solve for player i, hence our modeling choice of only imposing the desiderata shown in Definition 4.[21]

2.3 Categorization Equilibrium

For fixed $\varepsilon = (\varepsilon_1, \varepsilon_2)$ and κ, we define:

Definition 5 *A profile* $(\sigma, \mathcal{C})$ *is an* (ε, κ)-categorization equilibrium *if*

(a) σ *is an* ε*-perturbed analogy-based expectation equilibrium given* $\mathcal{C}$ *and*

(b) $\mathcal{C}$ *is* κ*-adjusted to* σ*.*

An $(\varepsilon,\ \kappa)$- categorization equilibrium can be understood as a steady state as follows. Assume the system has stabilized to $(\sigma, \mathcal{C})$. When looking at the data generated by previous matches, players would be led to choose analogy partitions $\mathcal{C}$ that are κ-adjusted to the strategy profile σ used in those matches. When trying next to form analogy-based expectations using such analogy partitions, they would be led to have beliefs as defined in (1) given that the play is governed by σ. They would then play as assumed in σ given that σ is an ε-perturbed analogy-based expectations equilibrium (ABEE) for C, thereby yielding the desired steady state property.[22]

Like in trembling-hand equilibrium (Selten, 1975), we focus on environments in which trembles are rare ($\varepsilon \to 0$). We also focus on environments in which data for situations that are observed without trembles are abundant, thereby leading us to assume that κ is small

[21] More formally, consider a more structured version of this problem, such as one requiring that situations are categorized so as to produce the largest overall homogeneity – for example, measured as the sum or the infimum of homogeneities $\zeta_i\left(\mathcal{C}_i^k\right)$ over the different analogy classes $\mathcal{C}_i^k$ – subject to the constraint that each analogy class should have total mass no smaller than some threshold $\kappa \in \mathbb{R}^+$. It is readily verified that this more structured problem would be harder to solve than the knapsack problem studied in computer science. But, the knapsack problem is known to be NP-hard, thereby formalizing the difficulty of solving our problem in general.

[22] We implicitly describe here the case in which all players assigned to the same role would end up with the same analogy partitions (requiring all subjects to sue the same categorization algorithm). Extensions to non-unitary versions (c.f. Fudenberg and Levine, 1993) are possible but bring no additional insights to the applications.

($\kappa \to 0$). Given that trembles are rare, it seems natural to allow for environments in which the data for off-path situations are scarce enough to require some coarse categorization. We distinguish then between cases in which κ and ε have the same order of magnitude leading to a notion of ρ-coarse categorization equilibrium and cases in which ε is much smaller than κ, leading to the notion of coarse categorization equilibrium. Formally,

Definition 6 *A profile* $(\sigma, \mathcal{C})$ *is a* categorization equilibrium *if there are sequences* $(\varepsilon^m)_m$ *and* $(\kappa^m)_m$ *converging to zero and a sequence* $(\sigma^m)_m$ *converging to* σ*, such that* (σ^m, C) *is an* $(\varepsilon^m,\ \kappa^m)$*-categorization equilibrium for all* m*. If* $\lim_{m\to\infty} \kappa^m/\varepsilon_i^m = \rho_i$*, then* $(\sigma, \mathcal{C})$ *is referred to as a* (ρ_1, ρ_2)-coarse categorization equilibrium*. If* $\lim_{m\to\infty} \kappa^m/\varepsilon_i^m = \infty$ *for* $i = 1, 2$*, then* $(\sigma, \mathcal{C})$ *is referred to as a* coarse categorization equilibrium*.*

Through part 1 of Definition 4, we have that expectations about opponent's behavior in situations that are observed without tremble are correct in a categorization equilibrium, which is analogous to the requirement in self-confirming equilibrium (Fudenberg and Levine, 1993) developed for extensive-form games (see Section 6.1 for elaboration). In a (ρ_1, ρ_2)-*coarse categorization equilibrium*, several off-path situations must be bundled together in coarse categories when ρ_1 and ρ_2 are big enough. In the subsequent analysis, we either consider coarse categorization equilibria or (ρ_1, ρ_2)-*coarse categorization equilibrium* with either ρ_1 or ρ_2 not too small so as to obtain new predictions as compared to the standard ones.

2.4 Dynamics

In some cases there will be no (ρ_1, ρ_2)-coarse categorization equilibrium. Hence we need to describe an explicit learning dynamic for such cases. In an attempt to illustrate the possibility of cycling that could emerge then, we will consider the following dynamics chosen for its simplicity. In period t agents form a profile of analogy partitions $\mathcal{C}(t) = (\mathcal{C}_i(t), \mathcal{C}_j(t))$ that is κ-adjusted to behavior in the preceding period, denoted σ^{t-1}. Expectations for period t are based on σ^{t-1} filtered through $\mathcal{C}(t)$. That is, the expectation in period t about a situation assigned to $\mathcal{C}_i(t)$ is identified with the aggregate distribution observed in $\mathcal{C}_i(t)$ given the behaviors σ^{t-1} observed in period $t-1$. These expectations induce behavior σ^t in period t (assuming that players best respond to their expectations when they do not tremble). At $t+1$, agents form a new profile of analogy partitions $\mathcal{C}(t+1) = (\mathcal{C}_i(t+1), \mathcal{C}_j(t+1))$ which is κ-adjusted to σ^t. Expectations for period $t+1$ are based on σ^t filtered through $\mathcal{C}(t+1)$, and so on. The dynamics is parameterized by the initial choice of analogy partitions $\mathcal{C}(1)$, as well as the the tie-breaking rule in case of multiple best responses. When such a dynamic learning model has a steady state it corresponds to a categorization equilibrium. However, we will consider the dynamics to cover cases in which there is no steady state and cycles emerge instead. Our main illustration of this will be the adverse selection game developed in Section 5.

3 Ultimatum/Bargaining Game

Consider an Ultimatum game where a proposer (first-mover) offers a share to a responder (second-mover) which he can either accept or reject. That is, the strategy of the proposer is a splitting share $s_P \in [0,1]$, and a strategy for the responder is an acceptance decision rule that maps the various offers onto an acceptance/rejection decision $s_R : [0,1] \to \{R, A\}$. The proposer's payoff is equal to $1 - s_P$ if the offer is accepted and zero otherwise. The responder's payoff is s_P if the offer is accepted and $v \geq 0$ otherwise. Here v may represent the responder's value of an outside option. The proposer has to predict the acceptance probability of the responder for the various proposer's offers that can be identified with the situations in the above abstract formulation. In doing so she may bundle several offers.

When assessing how the responder's acceptance probability depends on the offer s_P, it seems plausible that the the proposer would believe that the closer two offers are, the closer are their associated acceptance probabilities. This leads us to assume that the notion of similarity used by the proposer is based on the Euclidean distance in the space of offers $[0,1]$. More specifically, for any subset X of $[0,1]$, we assume that the homogeneity function used by the proposer is

$$\zeta_P(X) = 1 - \frac{1}{2}\left(\sup X - \inf X\right).$$

It follows that the homogeneity of a singleton analogy class is 1 and the homogeneity of the entire set $[0,1]$ of all offers is $1/2$.

Our ultimatum application has a continuum of actions for the proposer, but our general construction is easily adapted to this case. The proposer will use a pure strategy in our proposed categorization equilibrium. By part 1 of Definition 4, the corresponding (equilibrium) offer forms a singleton (on-path) analogy class in the proposer's analogy partition. By part 4 of Definition 4, if an off-path analogy class is not an interval then the union of this analogy class and the on-path singleton analogy class is an interval. Let K^{off} be the number of off-path analogy classes. In line with our general construction, we assume that trembles are uniform on $[0,1]$. By part 3 of Definition 4 each category must have a mass of at least κ (since under our assumptions no subset X of $[0,1]$ can have zero homogeneity). It follows that we need $\varepsilon/K^{off} > \kappa$. Additionally, to satisfy part 4 of Definition 4, we need the condition $\kappa > \varepsilon/\left(K^{off}+1\right)$ and that $\sup X - \inf X$ should also be the same for all off-path analogy classes.

In the next Proposition, we characterize the ρ-coarse categorization equilibria when $\frac{1}{2} < \rho < \frac{1}{3}$ ensuring that there are two off-path categories as just informally suggested.[23] We also characterize the coarse equilibrium (that can be viewed as a ρ-coarse categorization with $\rho < \frac{1}{2}$). Essential proofs not appearing in the main text are placed in the Appendix (with less essential aspects being relegated to the Online Supplement).

[23] ρ refers here only to the proposer, since for the responder the problem is a simple decision problem (with the no need to form expectation about the play of the opponent).

Proposition 1 *There is a unique coarse categorization equilibrium. It is such that the offer is $s_P^* = v$, and there is a single off-path analogy class . Assuming that $\frac{1}{2} < \rho < \frac{1}{3}$, ρ-coarse categorization equilibria have two off-path analogy classes. (a) If $v \geq 0.5$ then in any ρ-coarse categorization equilibrium $s_P^* = v$. (b) If $v \in (0.25, 0.5)$ then in any ρ-coarse categorization equilibrium $s_P^* \in [v, 0.5]$. (c) If $v \leq 0.25$ then in any ρ-coarse categorization equilibrium $s_P^* \in [v, 2v]$.*

A subgame perfect Nash equilibrium would require that the proposer offers $s_P^* = v$. In a categorization equilibrium with $\kappa^m/\varepsilon^m \to 0$ there would be arbitrarily many off-path analogy classes and so we would recover the subgame perfect Nash equilibrium, with $s_P^* = v$. Interestingly, this is also the prediction in a coarse categorization equilibrium (or more generally in a ρ-coarse categorization equilibrium with $\rho < 1/2$). In this case there is a single off-path analogy class. However, when $\rho > \frac{1}{2}$, ρ-coarse categorization equilibria allow for predictions away from the standard one. When $\frac{1}{2} < \rho < \frac{1}{3}$, there are ρ-coarse categorization equilibria in which (for some values of v) more equal splits may arise.[24]

As an alternative to the above homogeneity function, one could assume that $\zeta_P(X) = 0$ when X is not an interval, and that $\zeta_P(X) = 1 - \frac{1}{2}(\sup X - \inf X)$ otherwise. Such a modified notion of similarity and homogeneity may reflect a deeper understanding of the proposer that if two offers belong to the same analogy class, it would have to be that any intermediate offer also belongs to it. In this alternative, analogy classes would have to be intervals (as otherwise it would violate part 2 of Definition 4). Moreover, proposals away from the standard one could arise even in coarse categorization equilibria. Specifically, any offer $s_P^* \in [v, \sqrt{v}]$ could be sustained in a coarse categorization equilibrium in contrast to the finding of Proposition 1.[25]

4 Chainstore Game

In this Section, we apply our approach to the classic chainstore game, and illustrate how the monopolist may deter entry in most periods in a categorization equilibrium.

4.1 Set-Up

4.1.1 Game

In the finitely repeated chainstore game an *incumbent monopolist* faces a sequence of T *challengers*. Each challenger chooses to Enter (E) or to stay Out (O). If the challenger

[24] A similar qualitative insight would arise if there were N (instead of 2) off-path analogy classes, but as N grows large, the effect would eventually vanish yielding the standard SPNE prediction.

[25] When $s_P^* > v$, the perceived acceptance probability of offers $s_P \in [0, s_P^*)$ is $\frac{s_P^* - v}{s_P^*}$ (remember that trembling is uniform). Given the perception of the proposer, the best option in the range $[0, s_P^*)$ is $s_P = 0$ perceived to yield $\frac{s_P^* - v}{s_P^*}$. When $s_P^* < \sqrt{v}$, $\frac{s_P^* - v}{s_P^*} < 1 - s_P^*$ which is the correct perception of the payoff obtained by the proposer when proposing $s_P = s_P^*$.–

enters then the monopolist chooses whether to Accommodate (A) or Fight (F). The stage game payoffs of the monopolist and a challenger are denoted u_M and u_C, respectively, with $u_C(E,A) > u_C(O) > u_C(E,F)$ and $u_M(O) > u_M(E,A) > u_M(E,F)$. In words, the challenger prefers entering and facing an accommodating incumbent over not entering, and prefers not entering over entering and facing a fighting incumbent. The monopolist prefers the challenger to stay out over accommodating an entering challenger, and prefers the latter over fighting an entering challenger. Each challenger maximizes her payoff (in the stage at which she is present) and the monopolist maximizes the sum of stage game payoffs.

In the unique SPNE of this game, challengers choose E in every period and this is always followed by A, something that can be verified using backward induction. This prediction has been considered unintuitive, as the monopolist would seem to be able to deter early entry decisions by playing F in case of entry. While this kind of behavior cannot arise in a SPNE, we will establish that it can arise in a categorization equilibrium.

To make the chainstore game fit into our general two-player framework, we assume the challengers at the various time periods t form a single player, the challenger.[26] We also assume that the trembling probability is the same for the monopolist and the challenger.

4.1.2 Similarity and Homogeneity

A key modeling choice concerns the similarity between histories and the homogeneity of sets of histories. In the context of the chainstore game, we believe it plausible that players would consider that there is an important qualitative difference between histories in which there was a previous entry that was not immediately followed by a fight decision and other histories (either with no previous entry at all or with entries immediately followed by a fight decision). Accordingly, we will assume that subsets of histories that include both kinds of histories have minimal homogeneity. In effect, it will force us to have analogy classes that do not mix these two subsets of histories (according to part 2 of Definition 4). Other features can be incorporated into the homogeneity function, such as requiring that histories in nearby stages are more similar, but this will be play no role in our analysis of coarse categorization equilibria.

Formally, we first consider the nodes at which the challenger must make a decision and refer to the set of these nodes as $\mathcal{Q}_C$. We consider two subsets of $\mathcal{Q}_C$:

$$\begin{aligned} \mathcal{Q}_C^{Tough} &= \{q \in \mathcal{Q}_C : \text{No } E \text{ or all } E \text{ immediately followed by } F \text{ in history of } q\}; \\ \mathcal{Q}_C^{Soft} &= \{q \in \mathcal{Q}_C : \text{Some } E \text{ immediately followed by } A \text{ in history of } q\}. \end{aligned}$$

We require that for any $q^{Tough} \in \mathcal{Q}_C^{Tough}$ and $q^{Soft} \in \mathcal{Q}_C^{Soft}$, if q^{Tough} and q^{Soft} belong to X, then $\xi_M(X) = 0$. Any subset X containing only elements in $\mathcal{Q}_C^{Tough}$ or only elements

[26] This has no effect on the analysis of SPNE.

in $\mathcal{Q}_C^{Soft}$ is supposed to satisfy $\xi_M(X) > 0$.

Regarding the nodes at which the monopolist must make a decision, we denote the set of those corresponding to period t by $\mathcal{Q}_M^t$ and we distinguish in $\mathcal{Q}_M^t$ two subsets:

$$\mathcal{Q}_M^{t,Tough} = \left\{q \in \mathcal{Q}_M^t : \text{No } E \text{ or all } E \text{ immediately followed by } F \text{ in history of } q\right\};$$
$$\mathcal{Q}_M^{t,Soft} = \left\{q \in \mathcal{Q}_M^t : \text{Some } E \text{ immediately followed by } A \text{ in history of } q\right\}.$$

We require that if Y contains two nodes q and q' that either, (a) correspond to two different time periods, or (b) do not both belong to $\mathcal{Q}_M^{t,Tough}$, or (c) do not both belong to $\mathcal{Q}_M^{t,Soft}$ for some t, then $\xi_C(Y) = 0$. Any Y not containing two such elements is supposed to satisfy $\xi_C(Y) > 0$.

Observe that on the challenger's side, we do not allow histories at different calendar times to be bundled together, which may fit better with situations in which there is a different challenger at each time t, focusing on histories corresponding to that calendar time. We will later discuss what happens when histories with different calendar times are allowed to be bundled together also on the challenger's side.

4.2 Categorization Equilibrium

We will focus on coarse categorization equilibria and discuss later how ρ-coarse categorization equilibria look like for ρ large but not infinite (as implicitly required in coarse categorization equilibria).

4.2.1 Strategy profile

We define the threshold

$$k^* = \min\left\{k \in \mathbb{N} \text{ such that } u_M(E,F) + k u_M(O) \geq (k+1)\, u_M(E,A)\right\}. \tag{3}$$

Suppose that we are in the generic case where the inequality holds strictly for $k = k^*$. In this case we consider the following strategy profile σ_T:[27]

- Challenger $t \leq T - k^*$ strategy. If E was always matched with F in the past, or if there was no E in the past, play O. Otherwise play E.
- Challenger $t > T - k^*$ strategy. Play E.
- Monopolist strategy. At $t > T - k^*$, play A. At $t \leq T - k^*$; play F if E was always matched with F in the past, or if there was no E in the past; otherwise play A.

[27] When the condition in 3 holds with equality for $k = k^*$ we need to redefine the strategy profile so that entry and accommodation begins already in period $T - k^*$.

On the path of play induced by this strategy profile, the challenger enters only in the last k^* periods, and the monopolist accommodates those entries (while she would fight the challenger if entering in earlier periods).

4.2.2 Categorization profile

In a coarse categorization equilibrium, and given the strategy profile proposed above, the analogy partition profile $\mathcal{C}$ is characterized as follows.

- Each *on-path node* is in a separate analogy class.
- The monopolist categorizes *off-path challenger nodes* based on whether there was previously an act of E that was not met by F. The first analogy class bundles all off-path nodes with a history in which E was always met by F, and the second analogy class bundles all the remaining off-path nodes. Formally, let $\mathcal{Q}_C^{off}$ be the set of monopolist decision nodes that are located off the equilibrium path,

$$\mathcal{C}_M^1 = \left\{ q \in \mathcal{Q}_C^{off} \cap \mathcal{Q}_C^{Tough} \right\};$$
$$\mathcal{C}_M^2 = \left\{ q \in \mathcal{Q}_C^{off} \cap \mathcal{Q}_C^{Soft} \right\}.$$

- Challengers categorize *off-path monopolist nodes* based on the stage of the game only.[28] Formally, let $\mathcal{Q}_M^{off}$ be the set of off-path monopolist decision nodes. For each t let

$$\mathcal{C}_{Ct}^1 = \left\{ q \in \mathcal{Q}_M^{off} \cap \mathcal{Q}_M^{Tough} : q \text{ is in round } t \right\};$$
$$\mathcal{C}_{Ct}^2 = \left\{ q \in \mathcal{Q}_M^{off} \cap \mathcal{Q}_M^{Soft} : q \text{ is in round } t \right\}.$$

The above strategy profile and analogy partition profile together form a coarse categorization equilibrium when T is large enough.

Proposition 2 *There exists a T^* such that if $T > T^*$, then $(\sigma_T, \mathcal{C})$ is a coarse categorization equilibrium of the chainstore game with T periods, implying that in the absence of trembles the challenger enters only in the last k^* periods, and the monopolist fights the challenger in all but the last k^* periods.*

To emphasize the logic of the proposed equilibrium, observe that the only mistaken expectations are those of the monopolist regarding off-path nodes in $\mathcal{Q}_C^{Tough}$. In particular, if E occurs in period $t = T - k^*$ (i.e. the last period in which the challenger is supposed to

[28] We note that there are other categorizations that could be combined with σ^T to form a CE. For example we could let challengers bundle all monopolist nodes from the same period in a separate category for each time period. They would still have correct expectations.

stay out) then the monopolist mistakenly expects that by playing F, the challengers will be induced to stay out (with a probability roughly equal to $\frac{T-k^*-1}{T-k^*}$) from then on, whereas in reality, no matter what the monopolist does there will be entry in all remaining periods. This mistake is caused by the fact that there isn't enough mass of data on behavior at the subsequent challenger nodes (due to our assumption that $\lim_{m\to\infty} \kappa_T^m/\varepsilon_T^m = \infty$) so that they have to be bundled with many other nodes in $\mathcal{Q}_C^{Tough}$ at which indeed fighting after entry leads the challenger not to enter in the next period.

In a coarse categorization equilibrium, ε is supposed to be arbitrarily small compared to κ leading to bundle all off-path nodes in $\mathcal{Q}_C^{Tough}$ together. We note that the same strategy profile as the one considered in Proposition 2 could be used to support a ρ-coarse categorization equilibrium with $\rho > N$, as long as N and T are large enough.[29] This means that our construction only requires that κ be sufficiently (but not necessarily infinitely) large relative to ε.[30]

4.3 Discussion

4.3.1 Other coarse categorization equilibria

Are there other coarse categorization equilibria? It is clear that one cannot support coarse categorization equilibria with fewer periods of entry when the challenger is behaving optimally, as the challenger would always enter in the last k^* periods anticipating that the monopolist would find it optimal to play A (as implied by the definition of k^*). But, one can easily support equilibria with more periods of entry. In fact, take any $k^{**} > k^*$. It is readily verified that replacing k^* by k^{**} in the above strategy profile would be a categorization equilibrium for T large enough.

4.3.2 What if the challenger does not distinguish histories according to time?

Above we assumed a homogeneity function that implies that histories are distinguished according to time. What happens if we assume a homogeneity function which relaxes this while still keeping the idea that histories in which a previous entry was not immediately matched by a fight behavior are very dissimilar from others? This would fit with applications in which it is the same challenger who acts in the different time periods and the calendar time would not subjectively be considered by the challenger to affect dramatically the monopolist's behavior. In the Online Appendix S.2, we explore this alternative in detail. We demonstrate the existence of a coarse categorization equilibrium of the chainstore

[29]Here ρ refers to the Monopolist as the challenger behaves rationally in the proposed equilibrium.

[30]Indeed, in such a case, nodes in $\mathcal{Q}_C^{Tough}$ would have to be bundled in packages of at least N nodes, thereby leading to the belief that by playing F the challenger would stay out with probability no smaller than $\frac{N-1}{N}$. When N is large enough, this woudl give the same incentive to play the equilibrium as in the coarse categorization equilibrium considered in Proposition 2.

game where in the absence of mistakes there is no entry at all, and in case there is entry by mistake the monopolist fights the challenger in all but the last k^* periods.

4.3.3 Other finite horizon games

In the centipede game, a coarse categorization equilibrium would lead to immediate Take, as in the Subgame Perfect Nash Equilibrium. This is a corollary of a result we establish in section 6.1 that a coarse categorization is a self-confirming equilibrium (in the sense of Fudenberg and Levine, 1993).

In public good games, in each period agents privately decide on how much to contribute to the public good. The social benefit efficiency demands that everyone contributes in all periods, but the marginal cost of contribution is assumed to lie in between the private and social benefit so that in the unique subgame perfect equilibrium no one contributes in any period. In some variants, agents at the end of a period can also decide whether or not to punish other agents (after they have observed the profile of contributions in the current period). The subgame perfect Nash equilibrium still predicts no contribution, as well as no punishment in any period. This is in sharp contrast with behaviors experimentally observed in such games: Fehr and Gächter (2000) have documented significant levels of contribution, especially when agents have the possibility of punishing their peers, noting that contributions do not decrease over time in the presence of the punishment option.

In the Online Appendix S.2, we apply our framework to such games, assuming that histories in which agents have failed to contribute (in case there is no punishment stage), and histories in which agents failed to contribute and were not punished, or contributors were punished (in case there is a punishment stage) are very dissimilar from other histories where there was always contribution (in case there is no punishment stage), or non-contributors were always punished and contributors never punished (in case there is a punishment stage). This is analogous to our homogeneity assumption in the chainstore game. It leads to the conclusion that categorization equilibria with significant levels of contribution can be supported when agents have the opportunity to punish but not otherwise.

5 On Cycling in Adverse Selection Games

5.1 Set-Up

5.1.1 Market

Consider a market for trade of indivisible objects with random quality ω distributed on $\Omega = [0, 1]$ according to a continuous and differentiable density function g, with cumulative G. Sellers know the quality ω of their good. But buyers do not observe qualities; they only know the distribution of ω. The valuation of a given seller coincides with the quality ω of his good. The corresponding valuation of a buyer is $v = \omega + b$, where $b \in (0,1)$ represents gains from trade. We posit a one-to-one trading mechanisms between pairs consisting of one seller and one buyer drawn at random from their respective pools. In each pair, the seller and the buyer act simultaneously. The seller quotes an ask price $a(\omega)$ that depends on the quality ω. The buyer quotes a bid price $p \in [0,1]$. The market mechanism is such that if $p < a$ there is no trade, and if $p \geq a$ trade occurs at price p. Hence, if there is trade the buyer obtains utility $u(p) = v - p$, and the seller obtains utility p. If there is no trade, the seller gets ω and the buyer gets 0. This can be viewed as a Bayesian game between one seller informed of the state ω and one buyer not observing ω with action profiles and payoffs as just shown. This is the game considered in Esponda (2008).

In this modeling of the trading mechanism, setting the ask price equal to the quality $a(\omega) = \omega$ is a weakly dominant strategy for the seller (just as bidding one's own valuation is a weakly dominant strategy in the second-price auction), and from now on we will assume that the seller employs this strategy.

To make the analysis simple, we assume that $b < (g(1))^{-1}$ and that G has the *monotone reversed hazard rate property.* That is, for all p,

$$\frac{\partial}{\partial p}\left(\frac{g(p)}{G(p)}\right) < 0.$$

Moreover, we assume the following smoothness condition: $|g'(p)| < g(p)$ for all p.[31]

In a Nash equilibrium, the buyer quotes a bid price p so as to maximize:

$$\pi^{NE}(p) = \int_{\omega=0}^{p} (\omega + b - p)\, g(\omega)\, d\omega = G(p)\left(\mathbb{E}[\omega|\omega \leq p] + b - p\right).$$

It is readily verified (see Online Appendix) under our assumptions that there exists a unique Nash equilibrium in which the bid price p^{NE} of the buyer is uniquely defined by $\frac{g(p^{NE})}{G(p^{NE})} = \frac{1}{b}$.[32]

5.1.2 The Categorization Setup

To apply the general framework introduced above we identify Ω with $\mathcal{X}$, and we adopt straightforward extensions of our definitions to deal with the case of a continuum of states and a continuum of actions.

Feedback. Since the coarse categorization will only concern the buyer, it is enough to specify which profiles (ω, a) of quality ω and ask prices a are disclosed to new buyers. As seems natural in this application and in line with Esponda (2008), we posit that (ω, a)

[31] While not essential for our main conclusion regarding the presence of price cycles, these extra assumptions will simplify the analysis and ensure that there is a unique interior Nash equilibrium.

[32] In the case of a uniform quality distribution g this is $\pi^{NE}(p) = p\left(b - \frac{p}{2}\right)$, so $p^{NE} = b$.

appears in the feedback only when there is trade, i.e. when $a < p$. This defines the ϕ-function for the application.

Trembles. We will assume that only the buyer trembles. This is motivated on the ground that the seller, but not the buyer, has a weakly dominant strategy, thus making the discovery of the best strategy simpler for the seller. Specifically, with probability $1-\varepsilon$ the buyer picks a best response to her expectations and with probability ε she trembles. When trembling, we assume that the buyer choose bids according to a pdf f and cdf F with full support on $[0,1]$.[33] The seller always chooses his weakly dominant strategy.

Similarity and Homogeneity. Given that payoffs depend continuously on ω, it is natural to assume that when categorizing Ω, the buyer employs a homogeneity function that is decreasing in the Euclidean distances between the various elements in the considered set. For concreteness we let $\xi(C)$ be equal to the the difference between the supremum and infimum ω among the elements of C. Note that minimal homogeneity is obtained for $C = [0,1]$ and maximal homogeneity is achieved for intervals that vanish to points. This notion of homogeneity will (in line with part 4 of Definition 4) give rise to interval analogy partitions in which the set Ω is partitioned into various consecutive intervals.

Threshold Mass. In line with our general assumptions, we have in mind that for on-path qualities ω, i.e., ω such that (ω, a) is disclosed when the buyer does not tremble, there are enough data about the seller's ask price so that ω can be categorized finely. Since we are considering a setup with a continuum of ω, a strict application of part 1 of Definition 4 would not allow to categorize on-path qualities ω as singleton analogy classes, regardless of how small κ is. We nevertheless make this assumption so as to simplify the exposition, and we note that the assumption could be justified by our view of the continuum as an approximation of the discrete case. (See Jehiel and Mohlin (2021) for a fuller discussion.)

We consider the dynamic formulation sketched in Subsection 2.4. Denote by p^* the bid price chosen by non-trembling buyers in generation $t-1$. In generation t, all $\omega \leq p^*$, will be treated as singleton analogy classes so that buyers will understand that the ask price is $a = \omega$ for $\omega < p^*$. However, for $\omega > p^*$, buyers will be using a coarse analogy partition of $(p^*, 1]$ consisting of $K \geq 1$ analogy classes $\mathcal{C}^1, \mathcal{C}^2, ..., \mathcal{C}^K$ defined by $\mathcal{C}^k = (c_{k-1}, c_k]$ where

$$p^* = c_0 < c_1 < c_2 < ... < c_{K-1} < c_K = 1.$$

We will require that any $\mathcal{C}^k$ corresponds to a mass no less than κ. As in our general framework, we will be interested in the shape of categorizations and bidding strategies in the limiting case of $\kappa \to 0$ and $\varepsilon \to 0$.

[33] In line with our trembling formulation described in Section 2, we could impose that $f \equiv 1$ but our results apply to any f, hence our formulation.

5.1.3 Preliminary Analysis

Mass of Observations. The density of transactions conditional on trembling is $\tilde{g}(\omega) := g(\omega)(1 - F(\omega))$, and the density of transacted qualities in the dataset is thus given by

$$\mu_{p^*}^{\sigma,\varepsilon}(\omega) = \begin{cases} (1-\varepsilon)\, g(\omega) + \varepsilon \tilde{g}(\omega) & \text{if } \omega \leq p^*; \\ \varepsilon \tilde{g}(\omega) & \text{if } p^* < \omega. \end{cases}$$

In what follows we suppress the subscript reference to p^*, relying on the context to indicate the relevant p^*.

Adjustment of Categorizations to Observations. As already mentioned, we view our continuum model as a tractable approximation of a discrete model with finitely many prices and types. In a discrete model, as $\kappa \to 0$ and $\varepsilon \to 0$ each type $\omega \leq p^*$ is put in a singleton analogy class. This motivates our focus on the limiting case of an arbitrarily fine-grained categorization below p^* in the continuum model. For types above p^* the number of categories depend on κ and ε in a more complex way. Each analogy class above p^* should satisfy $\kappa \leq \int_{c_{k-1}}^{c_k} \mu_{p^*}^{\sigma,\varepsilon}(\omega)(s)\, ds$. Consequently, the number of categories above p^* (for any $p^* < 1$) is

$$K = \max\left\{1, \left\lfloor \frac{1}{\kappa} \int_{p^*}^{1} \mu(s)\, ds \right\rfloor \right\} = \max\left\{1, \left\lfloor \left(\tilde{G}(1) - \tilde{G}(p^*)\right) \frac{\varepsilon}{\kappa} \right\rfloor \right\} \leq \max\left\{1, \frac{\varepsilon}{\kappa}\right\}.$$

If $\kappa/\varepsilon \to \rho$ for some constant $\rho > 0$, then in the limit an adjusted categorization will have K analogy classes where K is bounded from above by $\max\left\{1, \frac{1}{\rho}\right\}$, which is finite, but possibly larger than one. If we impose $\kappa/\varepsilon \to 0$, as in the definition of coarse categorization equilibrium, then there is a single analogy class above p^*.

Analogy-Based Expectations. Buyers predict the distribution of ask price a of a type ω seller, knowing that trade occurs if $a \leq p$. For a quality $\omega \leq p^*$ the buyers understand that $a(\omega) = \omega$. Consequently, for a quality $\omega \leq p^*$ the buyer understands that the probability of trade is zero conditional on $\omega \leq p$ and one conditional on $p > \omega$, i.e

$$\widehat{\Pr(a \leq p|\omega)} = \Pr(a < p|\omega) = \Pr(\omega < p|\omega) = \mathbb{I}_{\{\omega \leq p\}}. \tag{4}$$

For a quality $\omega > p^*$ the buyer forms a prediction of the ask price distribution associated with qualities in analogy class $\mathcal{C}^k$ using the data generated under trembling. Using the fact that $a(\omega) = \omega$ we can write the probability density function (pdf) of ask prices conditional on a quality in $\mathcal{C}^k$ as

$$\tilde{g}\left(a|\omega \in \mathcal{C}^k\right) = \frac{\tilde{g}(a)}{\int_{\omega \in \mathcal{C}^k} \tilde{g}(\omega)\, d\omega} = \frac{\tilde{g}(a)}{\tilde{G}(c_k) - \tilde{G}(c_{k-1})}.$$

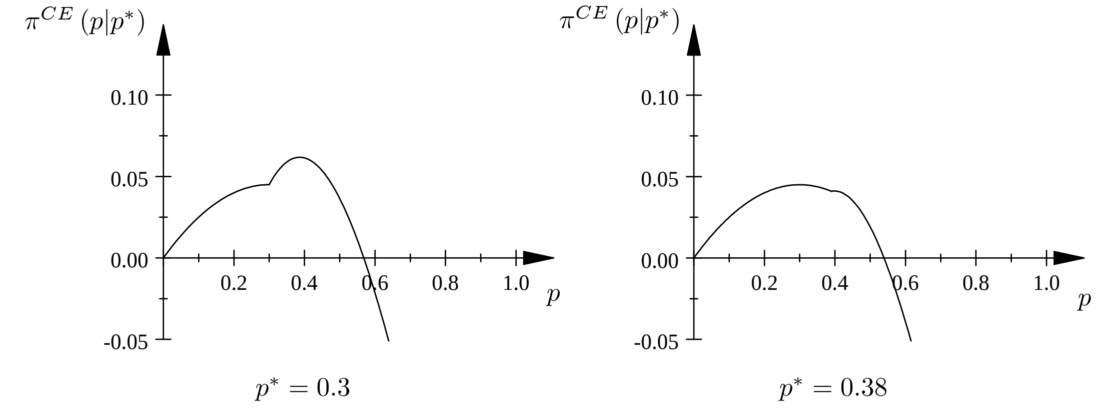


Figure 1: Expected utility for different values of p^*.

Thus, the buyer believes that the pdf of ask prices due to sellers with quality in $\mathcal{C}^k$ is

$$h_{\mathcal{C}^k}(a) = \begin{cases} \frac{\tilde{g}(a)}{\tilde{G}(c_k) - \tilde{G}(c_{k-1})} & \text{if } a \in \mathcal{C}^k; \\ 0 & \text{otherwise.} \end{cases}$$

This implies that, for a quality $\omega > p^*$ with $\omega \in \mathcal{C}^k$, the buyer perceives the probability of trade at price p to be

$$\Pr(a \widehat{\leq p | \omega} \in \mathcal{C}^k) = \int_{a=0}^{p} h_{\mathcal{C}^k}(a)\, da = \begin{cases} 1 & \text{if } c_k < p; \\ \frac{\tilde{G}(p) - \tilde{G}(c_{k-1})}{\tilde{G}(c_k) - \tilde{G}(c_{k-1})} & \text{if } c_{k-1} < p \leq c_k; \\ 0 & \text{if } p < c_{k-1}. \end{cases} \tag{5}$$

Using the perceived probability of trade as a function of price p, and letting $k(p)$ be such that $p \in (c_{k(p)-1}, c_{k(p)}]$ for $p > p^*$, the following lemma derives the perceived expected payoff as a function of p.

Lemma 1 *Let $v(\mathcal{C}_j) := \mathbb{E}[\omega | \omega \in \mathcal{C}_j] + b$. The perceived expected payoff is*

$$\pi^{CE}(p|p^*) = \begin{cases} G(p)\left(\mathbb{E}[\omega|\omega \leq p] + b - p\right) & \text{if } p \leq p^* \\ \\ G(p^*)\left(\mathbb{E}[\omega|\omega \leq p^*] + b - p\right) & \\ + \sum_{k=1}^{k(p)-1}\left(G(c_k) - G(c_{k-1})\right)\left(v\left(\mathcal{C}^k\right) - p\right) & \text{if } p > p^*. \\ + \left(\tilde{G}(p) - \tilde{G}\left(c_{k(p)-1}\right)\right) \frac{G\left(c_{k(p)}\right) - G\left(c_{k(p)-1}\right)}{\tilde{G}\left(c_{k(p)}\right) - \tilde{G}\left(c_{k(p)-1}\right)} \left(v\left(\mathcal{C}^{k(p)}\right) - p\right) & \end{cases}$$

As an illustration consider the case of a uniform quality distribution g, i.e., $G(p) = p$ for all $p \in [0, 1]$, a uniform mistake distribution f_{p^*} above p^*, i.e., $F_{p^*}(p) = p/(1 - p^*)$ for all $p \in [p^*, 1]$. We assume κ high enough to induce a single analogy class above p^*. Figure 1 illustrates the payoff function for $b = 0.3$ and two different values of p^*.

Dynamics. Letting p_t^* denote the price quoted by buyers of generation t when not trembling, our dynamic system is completely characterized by the initial value of this price p^0 and the recursive condition

$$p_{t+1}^* = \arg\max_{p\in[0,1]} \pi^{CE}(p \mid p_t^*).$$

5.2 Results

5.2.1 Learning and Cycling

In the following, we consider the case in which $\kappa/\varepsilon \to \rho$ for some constant ρ possibly equal to 0. Our main result is that the sequence of p_t^* in the dynamics just described has no rest point and must cycle over finitely many values $p^{(1)}, ..., p^{(m)}$, one of them being the Nash Equilibrium price p^{NE} as previously characterized, and the others being above p^{NE}. In order to establish this, we first derive three properties related to how p_{t+1}^* varies with p_t^* depending on whether p_t^* is below, above, or equal to p^{NE}. These properties are referred to as lemmata and are proven in the Appendix.

Lemma 2 *If $p_t^* = p^{NE}$ then $p_{t+1}^* > p^{NE}$.*

Lemma 3 *If $p_t^* > p^{NE}$, then either $p_{t+1}^* = p^{NE}$ or $p_{t+1}^* > p_t^*$.*

Lemma 4 *If $p_t^* < p^{NE}$, then $p_{t+1}^* > p_t^*$.*

Roughly, these three properties can be understood as follows. As already mentioned, categorical reasoning induces uninformed buyers to correctly infer that the quality corresponding to an ask price a below p^* is a. On the other hand, the coarse bundling for ask prices above p^* leads uninformed buyers to incorrectly infer that ask prices slightly above p^* are associated with an average quality that lies strictly above p^*. Thus, a buyer would choose a bid price strictly above p^* whenever $p^* \leq p^{NE}$ as she would incorrectly perceive a jump in quality when increasing slightly the bid price above p^* (and any bid price below p^* would rightly be perceived to be suboptimal). This is in essence the content of lemmata 4 and 2. By contrast, when $p^* > p^{NE}$, the best bid price below p^* is rightly perceived to be p^{NE} and the same logic leads the uninformed buyer to either choose p^{NE} or a bid price strictly above p^* with the aim of taking advantage of the jump in the perceived quality when the ask price lies above p^*.

The above properties immediately imply that the price dynamics has no rest point, i.e., there is no p_t^* such that $p_{t+1}^* = \arg\max_{p\in[0,1]} \pi^{CE}(p|p_t^*) = p_t^*$. To see this, assume by contradiction that p^* is a rest point. By Lemma 4, it cannot be that $p^* < p^{NE}$ since $p_t^* = p^* < p^{NE}$ would imply that $p_{t+1}^* > p_t^* = p^*$. By Lemma 2, it cannot be that $p^* = p^{NE}$ since $p_t^* = p^*$ would imply that $p_{t+1}^* > p^{NE}$. Finally, by Lemma 3, it cannot be that $p^* > p^{NE}$ since $p_t^* = p^*$ would imply either that $p_{t+1}^* > p_t^*$ or that $p_{t+1}^* = p^{NE}$ and thus

$p^*_{t+1} \neq p^*_t$ (given that $p^*_t = p^* \neq p^{NE}$). Even though there is no rest point, we can establish that there is a price cycle (making use of Lemmas 2-4), that consists of the Nash price and one or more prices above the Nash price.

Proposition 3 *There exists an increasing sequence* $(p^{(1)}, ..., p^{(\tau)})$ *with* $\tau \geq 2$ *and* $p^{(1)} = p^{NE}$ *such that if* $p^*_t = p^{(i)}$ *for* $i \in \{1, ..., \tau - 1\}$ *then* $p^*_{t+1} = p^{(i+1)}$*, and if* $p^*_t = p^{(\tau)}$ *then* $p^*_{t+1} = p^{(1)}$*. Moreover, the dynamic converges to the set* $\{(p^{(1)}, ..., p^{(\tau)})\}$ *from any initial price* $p_0 \in [0, 1]$.

In the above analysis, we have considered pure strategies on the buyer side. Could it be that by allowing mixing on the buyer side, we restore the existence of steady states? We suspect that if we stick to our assumption that (at least) every ω weakly below the support of the bid price strategy of the buyers would be treated as a singleton analogy class, then there is no such steady state. Suppose that the lower bound $\underline{p}$ of the support of buyer' strategy is strictly lower than p^{NE}. By the logic of lemma 4 the best response would be strictly above $\underline{p}$, so that $\underline{p}$ could not be part of the support in the steady state. Suppose instead that $\underline{p} > p^{NE}$. (a) If the support is coarsely categorized in a neighborhood of $\underline{p}$ then the best response is either strictly above $\underline{p}$, or equal to p^{NE}. (b) If the support is finely categorized in a neighborhood of $\underline{p}$ each $\omega \leq \underline{p} + \delta$ (for some $\delta > 0$) is perfectly distinguished, then the best response is either strictly above $\underline{p} + \delta$, or equal to p^{NE}. In either case (a) or (b) $\underline{p}$ cannot be the lower bound of a steady state support. Altogether, this is suggestive that it would not be possible to support a steady state even allowing for mixing on the bidding price.

6 Discussion

6.1 Relation to Other Solution Concepts

Focusing on extensive form games of complete information (i.e. allowing for simultaneous moves but no asymmetric information), and assuming that feedback consists in disclosing the played path, our notion of categorization equilibrium relates to self-confirming equilibrium (Fudenberg and Levine, 1993) and subgame perfect Nash equilibrium as follows:

Proposition 4 *Consider an extensive-form game of complete information and assume that the feedback consists of observing the path of play.*

(a) For any homogeneity function, if $(\sigma, \mathcal{C})$ *is a categorization equilibrium then* σ *is a (unitary) self-confirming equilibrium (Fudenberg and Levine, 1993, 1998).*

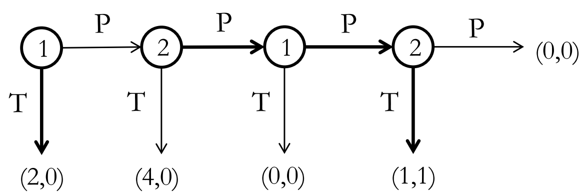
(b) For any homogeneity function, if σ *is a subgame perfect Nash equilibrium (SPNE) then there is a* $\mathcal{C}$ *such that* $(\sigma, \mathcal{C})$ *is a categorization equilibrium.*

(c) If σ is a subgame perfect Nash equilibrium (SPNE) then there may be no $\mathcal{C}$ such that $(\sigma, \mathcal{C})$ is a coarse categorization equilibrium.

Proof. (a) Since $\kappa^m \to 0$ and $\varepsilon^m \to 0$ players must have correct expectations about behaviors on the path, given criterion 1 in definition 4. The result follows.

(b) Let L be the length of the longest path of play. This is the highest number of mistakes needed to reach any terminal node under any strategy profile. By choosing sequences $(\varepsilon^m)_m$ and $(\kappa^m)_m$ such that $\lim_{m\to\infty} \kappa^m / (\varepsilon^m)^L < 1$ we ensure that there is some M such that for any $m > M$ any (ε, κ)-categorization equilibrium will put all off-path nodes in singleton analogy classes. This implies that all players have correct expectations at all nodes. And since (for any finite m) all nodes are reached with positive probability all players will play ε-best responses at all nodes, converging to exact best responses as $m \to \infty$.

(c) Consider the following version of the centipede game where players 1 and 2 take turn choosing between Pass and Take. The unique SPNE is TP for Player 1 and PT for



Player 2 (indicated by the fat arrows). Both of Player 2's nodes are off-path and reached by a single mistake (by Player 1 at the first node). If $\lim_{m\to\infty} \kappa^m/\varepsilon^m = \infty$ then Player 1 will bundle these two nodes together (assuming Player 1 does not perceive them as maximally dissimilar) and form the expectation that Player 2 passes with probability 1/2. Given this belief, Player 1 perceives the expected utility of passing at both of her nodes to be 2.5 making it seem optimal to deviate from the strategy SPNE. ■

Part (a) of Proposition 4 establishes that categorization equilibrium refines (unitary) self-confirming equilibrium, and hence coarse categorization equilibrium refines self-confirming equilibrium. This happens because categorization equilibrium (compared to self-confirming equilibrium) puts more structure on the admissible off-path beliefs, while perfectly distinguishing on-path nodes, thereby inducing correct on-path beliefs.

Part (b) says that subgame perfect Nash equilibrium (SPNE) is a refinement of categorization equilibrium. The reason is that with complete freedom on how to choose sequences $(\varepsilon^m)_m$ and $(\kappa^m)_m$, we can always ensure that all nodes are put in singleton analogy classes (this requires that ε is high enough relative to κ), thereby inducing best-responses in all subgames.[34] However part (c) tells us that this is not true for coarse categorization equilibrium: there are SPNE that cannot be supported as a coarse categorization equilibrium.

[34]The fact that the homogeneity function does not matter in part (a) is simply a consequence of the

The reason is that in a coarse categorization equilibrium one is *not* free choose sequences $(\varepsilon^m)_m$ and $(\kappa^m)_m$ such that there are enough mistakes to put all nodes in singleton analogy classes. In general we note that if $\kappa^m / (\varepsilon^m)^l < 1 < \kappa^m / (\varepsilon^m)^{l+1}$ then any node that is at most l steps off the equilibrium path will be placed in a category of its own under any (ε^m, κ^m)-categorization equilibrium, whereas nodes that are further away from the equilibrium path may be bundled more coarsely.

Remark 2 *(1) Part (a) of Proposition 4 states that there are SPNE strategy profiles that are not supported by any coarse categorization equilibrium. The example invoked to prove this actually shows something slightly stronger, that there may be no coarse categorization equilibrium that supports a strategy profile that is outcome equivalent to the SPNE.*

(2) When considering other classes of games such as Bayesian games and other feedback structures ϕ such as the one considered in Section 5, one may wonder how categorization equilibria relate to self-confirming equilibria defined in the sense of Battigalli (1987) or Dekel et al. (2004). We note that if players are aware of the tremble structure as well as ϕ, categorization equilibria need not be self-confirming equilibria given ϕ, even assuming that κ is set so that on the path situations are treated as singleton analogy classes.[35]

(3) In the Online Appendix we provide two examples in which (σ, C) is a categorization equilibrium but σ is not a Nash equilibrium. Constructing such examples either require that the feedback differs from the path of play (in which case a normal form game with just two players can be used to illustrate the claim) or (if the feedback is the path of play) that one considers games with at least three players and some asymmetric information. In the latter case we adapt an example from Fudenberg and Levine (1993) used to illustrate that a self-confirming equilibrium may differ from a Nash equilibrium.

6.2 On the existence of steady state

When does a coarse (or ρ-coarse) categorization equilibrium exist? Suppose that in our approach to endogenizing the categorizations, we had required that the categorization chosen by player i should be a solution to the maximization problem (2) for some functions W_i and $\widetilde{\kappa}_i$. A generalized notion of steady state could then be defined to include distributions over analogy partitions all solving the above maximization problem as well as strategies and analogy-based expectations that would be indexed by the analogy partition in the support of such distributions that would satisfy the consistency and best-response properties as defined in Section 2. We conjecture that in finite environments (i.e., environments with

on-path nodes being distinguished perfectly, so that homogeneity is maximal in each singleton on-path analogy class. In part (b) the irrelelvance of the homogeneity function stems from choosing sequences $(\varepsilon^m)_m$ and $(\kappa^m)_m$ such that all off-path nodes are put in singleton analogy classes.

[35] This is so because, the conjecture $\sigma_j^{\beta_i}$ may be at odds with the observations when the trembling structure as well as ϕ is known (for example, this is the case in the trade application developed below when the trembles are not concentrated on bid prices above the maximum value of the seller).

finitely many situations and actions), such a notion of steady state (allowing for a mixed extension both to strategies as in Nash and to analogy partitions as in some recent work by Jehiel and Weber (2023) would always exist.

Our approach uses Definition 4 instead of the optimization problem (2). However, we conjecture that for fixed (ε, κ), such a notion of steady state might still exist in finite environments, even if requiring in general a mixed extension, as just noted.[36] Considering the limit of such steady states as (ε, κ) go to 0 and κ does not get small relative to ε as we do in the coarse categorization equilibrium of Definition 6 would complicate matters, and as our analysis of the adverse selection market suggests, it is not clear we could support a steady state in this limit. A more complete investigation of this requires further work.

6.3 On the similarity/homogeneity function

In this paper, the similarity between situations as well as the homogeneity function that derives from it were left exogenous. While we believe the choice of similarity and homogeneity functions we have made in each application is intuitively appealing, it may be desirable in future work to think of principles that could be used to guide this choice.

A perspective that we feel could be fruitful is the following. One may think of situations as being parameterized by a vector of attributes. The similarity and homogeneity functions could then be thought of as relying only on a subset of these attributes (for example, those attributes appearing more frequently across the various games or environments faced by the subject). One approach could then consist in choosing the similarity and homogeneity functions based only on those attributes in an attempt to maximize the similarity and homogeneity of opponent's behavior across the various situations (as measured by some notion of correlation). For example, in the bargaining application, we could consider an environment consisting of games obtained by varying the outside option v of the responder. In our language, a situation as viewed by the proposer would be parameterized by the offer s_P as well as v. Assuming that the only attribute used to form the similarity and homogeneity function is s_P would lead to adopt some homogeneity function that depends on the euclidean distance in the s_P space, as considered above in Section 3.

Obviously, more work is needed to develop such a perspective, but we hope the framework introduced in this paper will be a useful step in pursuing this task left for future research.

[36] One approach to prove this would be to propose W_i and $\widetilde{\kappa}_i$ functions such that any solution to optimization problem (2) would satisfy the conditions of Definition 4.

Appendix

A.1 Ultimatum Game Application

Proof of Proposition 1. Suppose that $\kappa^m > \varepsilon^m/2$ as $m \to \infty$ (which must hold in a coarse categorization equilibrium). It implies that there is a single off-path analogy class for all m. As $\varepsilon^m \to 0$ the following holds. The responder rejects if $s_P < v$ and accepts if $s_P > v$ and at $s_P = v$ she is indifferent between accepting and rejecting. Hence, the proposer believes that the acceptance probability is $1 - v$ for an off-path offer, and consequently believes that the expected utility of making an off-path offer s_P is $(1 - s_P)(1 - v)$. Note that $(1 - s_P)(1 - v)$ is decreasing in s_P and approaches $1-v$ (from below) as s_P approaches 0. Thus in categorization equilibrium the proposer must get at least $1 - v$, meaning that we need $s_P^* \leq v$. Suppose $v > 0$. If $s_P^* \in (0, v)$ then the proposer earns 0 in categorization equilibrium meaning that a deviation to off-path $s_P = 0$ appears profitable. Thus, if $v > 0$ then $s_P^* = v$ in a categorization equilibrium. Suppose $v = 0$. If $s_P^* > 0$ then the proposer earns less than 1 in categorization equilibrium meaning that a deviation to off-path $s_P = 0$ appears profitable. Thus, if $v = 0$ then $s_P^* = 0$ in a categorization equilibrium.

Now suppose that $\varepsilon^m/2 > \kappa^m > \varepsilon^m/3$ as $m \to \infty$, implying that there are two off-path analogy classes for all m. As $\varepsilon^m \to 0$ the following holds.

For (a), consider the case of $v \geq 0.5$. The responder rejects if $s_P < v$ and accepts if $s_P > v$ and at $s_P = v$ she is indifferent between accepting and rejecting. Hence, the proposer believes that the acceptance probability is 0 for an off-path offer $s_P < 0.5$, and believes that the acceptance probability is $2(1 - v)$ for an off-path offer $s_P > 0.5$. It follows that the proposer perceives the expected utility of offering an off-path $s_P < 0.5$ to be 0 and perceives the expected utility of offering an off-path $s_P > 0.5$ to be $(1 - s_P)\, 2\,(1 - v)$. Note that $(1 - s_P)\, 2\,(1 - v)$ is decreasing in s_P and approaches $1 - v$ (from below) as s_P approaches 0.5. Thus in categorization equilibrium the proposer must get at least $1 - v$, meaning that we need $s_P^* \leq v$. If $s_P^* < v$ then the proposer earns 0 in categorization equilibrium meaning that a deviation to off-path $s_P \in (0.5, 1)$ appears profitable.

For (b) and (c), consider the case of $v \in (0, 0.5)$. The proposer believes that the acceptance probability is 1 for an off-path offer $s_P > 0.5$, and believes that the acceptance probability is $2\left(\frac{1}{2} - v\right) = 1 - 2v$ for an off-path offer $s_P < 0.5$. It follows that the proposer perceives the expected utility of offering an off-path $s_P > 0.5$ to be $1 - s_P$ and perceives the expected utility of offering an off-path $s_P < 0.5$ to be $(1 - s_P)(1 - 2v)$. Thus by deviating to $s_P > 0.5$ she perceives that she can get an amount that approaches 0.5 from below and by deviating to $s_P = 0 < 0.5$ she perceives that she can get exactly $1 - 2v$. Deviation to $s_P = 0$ is perceived more profitable than deviation to $s_P > 0.5$ if and only if $v \leq 0.25$. Naturally, in categorization equilibrium we must have $s_P \geq v$, as otherwise the responder rejects and the proposer would perceive it profitable to deviate to $s_P > 0.5$. Combining this we see that if $v > 0.25$ then any $s_P \in [v, 0.5]$ is part of a categorization equilibrium, and if $v \leq 0.25$ then any $s_P \in [v, 2v]$ is part of a categorization equilibrium. ■

A.2 Chainstore Application

Proof of Proposition 2. We need to show that for $T > T^*$ there is a sequence $(\sigma_T^m)_m$ converging to σ_T, such that $(\sigma_T^m, \mathcal{C})$ is an $(\varepsilon^m, \kappa^m)$-categorization equilibrium for all m. We define σ_T^m as the strategy profile which at each node puts probability ε^m on the action that σ_T puts zero probability on. Since there are only two actions at each node this is enough to specify σ_T^m. Since the starting point of $(\varepsilon^m, \kappa^m)$ is arbitrary it is sufficient to show the following: There exists a T^* such that for any $T > T^*$ there is exists an m^* such that if $T > T^*$ and $m > m^*$ then σ_T^m is an $(\varepsilon_T^m, \kappa_T^m)$-categorization equilibrium of the chainstore game with T periods.

1. First we explain why $\mathcal{C}$ is adjusted to σ_T^m for all $m > m^*$ (and all T).

   (a) For any T, if m is large enough, then $\kappa_T^m < (1 - \varepsilon_T^m)^T$, ensuring that on-path nodes have a mass exceeding the threshold κ_T^m and thus are treated as singleton analogy classes, by point 1 of Definition 4.

   (b) For off-path nodes following histories in which there was some E not matched with F, our homogeneity assumptions imply that nodes in $\mathcal{Q}_C^{Soft}$ cannot be bundled with nodes that are not in $\mathcal{Q}_C^{Soft}$, and nodes in $\mathcal{Q}_M^{t,Soft}$ cannot be bundled with nodes that are not in $\mathcal{Q}_M^{t,Soft}$, according to point 2 of Definition 4. (The total mass of such histories would typically fall short of the κ_T^m threshold, but the dissimilarity with other histories would not allow further bundling.)

   (c) Furthermore, all off-path nodes in $\mathcal{Q}_C^{Soft}$ have to be bundled together and all off-path nodes in $\mathcal{Q}_M^{t,Soft}$ have to be bundled together (but separately for each t) according to point 3 of Definition 4. This follows from the assumption that $\lim_{m\to\infty} \kappa_T^m / \varepsilon_T^m = \infty$, which implies that the total mass of the off-path nodes vanishes relative to the threshold κ.

   (d) The situation is analogous for off-path nodes following histories in which there was no E or any E was immediately followed by an F. The off-path nodes of the challenger $\mathcal{Q}_C^{off}$ have to be partitioned into $\mathcal{C}_M^1$ and $\mathcal{C}_M^2$, and the off-path nodes of the monopolist have to be partitioned, for each t, into $\mathcal{C}_{Ct}^1$ and $\mathcal{C}_{Ct}^2$.

2. Second we examine the analogy-based expectations

   (a) Players have correct expectations at on-path nodes.

   (b) Players also have correct expectations at nodes following off-path histories in which there was some E not matched with F, i.e. at off-path nodes in $\mathcal{Q}_C^{Soft}$ and $\mathcal{Q}_M^{t,Soft}$. This is so because after such histories, the challenger consistently chooses E and the monopolist consistently chooses A after E.

(c) Next consider off-path monopolist nodes following histories in which there was no E or any E was immediately followed by an F, i.e. off-path nodes in $\mathcal{Q}_M^{t,Tough}$ for some t. (Such a node is only reached when the challenger plays E before $t \leq T - k^*$.) Challengers have correct expectations since they do not bundle together nodes from different time periods. (Indeed this would be true even if challengers did not distinguish between $\mathcal{Q}_M^{t,Tough}$ and $\mathcal{Q}_M^{t,Soft}$.)

(d) It only remains to check the monopolist's expectations at off-path nodes in Q_C^{Tough}. As $\varepsilon^m \to 0$ the expectations here are determined by behavior at nodes with histories containing a single mistake. The fraction of such nodes at which the challenger chooses E vanishes as $T \to \infty$. It follows that as T gets large, the monopolist will expect that O is chosen with a probability close to 1.

3. Third and finally we verify that σ_T^m induces a ε_T^m-best-responses given the analogy-based expectations. We have found that the challengers have correct expectations and it is easy to see that they best-responds to the monopolist's strategy, so we focus on the monopolist.

(a) Monopolist in period $t \leq T$ at an off-path node in Q_M^{Tough}. By playing F, the monopolist expects that with a probability close to 1, a string of O occur from then on until the end of the game. By playing A, the monopolist correctly expects a string of (E, A) until the end of the game. The former is at least as good as the latter if $u_M(E,F) + (T-t)\, u_M(O) \geq (T-1+1)\, u_M(E,A)$. For $t \leq T - k^*$ this is satisfied, but for $t > T - k$ it is not satisfied, by the definition of k^*.

(b) Monopolist at the on-path node in period $t = T - k^* + 1$. This node is in Q_M^{Tough}, immediately preceded by the first instance of E. By deviating from σ^T and playing F, the monopolist expects that with a probability close to 1, a string of O occur from the next period until the end of the game. By complying with σ^T and playing A, the monopolist correctly expects a string of (E, A) until the end of the game. Deviation is then perceived unprofitable by the same condition as before.

(c) Monopolist at an off-path node in Q_M^{Soft}. Regardless of what happens in the current period, the monopolist (correctly) expects E in all subsequent periods. The best response is to play A from now until the end of the game.

(d) Monopolist at an on-path node in period $t > T - k^* + 1$. In the history of such a node there has been at least one instance of E that was not immediately followed by A, i.e. the node is in Q_M^{Tough}. The monopolist (correctly) expects E in all subsequent periods. The best response is to play A until the end of the game.

■

A.3 Adverse Selection Application

A.3.1 Deriving Perceived Expected Payoff

Proof of Lemma 1. The perceived expected payoff is

$$\begin{aligned}\pi^{CE}(p|p^*) = &\int_0^{p*} \Pr\left(\widehat{a \leq p}|\omega\right)(\omega + b - p)\, g(\omega)\, d\omega \\ &+ \int_{p*}^1 \Pr\left(a \widehat{\leq p|\omega} \in \mathcal{C}^k\right)(\omega + b - p)\, g(\omega)\, d\omega,\end{aligned}$$

where, using (4) we obtain

$$\int_0^{p*} \Pr\left(\widehat{a \leq p}|\omega\right)(\omega + b - p)\, g(\omega)\, d\omega = \begin{cases} G(p)\left(\mathbb{E}[\omega|\omega \leq p] + b - p\right) & \text{if } p < p^* \\ G(p^*)\left(\mathbb{E}[\omega|\omega \leq p^*] + b - p\right) & \text{if } p \geq p^* \end{cases}$$

and, writing $i(p)$ for the analogy class that contains $\omega = p$, using (5) we obtain,

$$\begin{aligned}&\int_{p*}^1 \Pr\left(a \widehat{\leq p|\omega} \in \mathcal{C}^k\right)(\omega + b - p)\, g(\omega)\, d\omega \\ &= \sum_{k=1}^{k(p)-1} \left(\left(G(c_k) - G(c_{k-1})\right)\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^k\right] + b - p\right)\right) \\ &+ \left(\tilde{G}(p) - \tilde{G}\left(c_{k(p)-1}\right)\right) \frac{G\left(c_{k(p)}\right) - G\left(c_{k(p)-1}\right)}{\tilde{G}\left(c_{k(p)}\right) - \tilde{G}\left(c_{k(p)-1}\right)} \left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^{k(p)}\right] + b - p\right)\end{aligned}$$

■

A.3.2 Preliminary Observations

Note that $\lim_{p\uparrow c_k} \pi^{CE}(p|p^*) = \lim_{p\downarrow c_k} \pi^{CE}(p|p^*)$, for all $i \in \{1, ..., K-1\}$, implying that $\pi^{CE}(p|p^*)$ is continuous everywhere. Moreover, $\pi^{CE}(p|p^*)$ is piecewise differentiable with points of non-differentiability only at category boundaries. The first derivative at $p \in \left(c_{k(p)-1}, c_{k(p)}\right)$ is

$$\begin{aligned}\frac{\partial \pi^{CE}(p|p^*)}{\partial p} = &-G\left(c_{k(p)-1}\right) - \left(\tilde{G}(p) - \tilde{G}\left(c_{k(p)-1}\right)\right) \frac{G\left(c_{k(p)}\right) - G\left(c_{k(p)-1}\right)}{\tilde{G}\left(c_{k(p)}\right) - \tilde{G}\left(c_{k(p)-1}\right)} \\ &+ \tilde{g}(p) \frac{G\left(c_{k(p)}\right) - G\left(c_{k(p)-1}\right)}{\tilde{G}\left(c_{k(p)}\right) - \tilde{G}\left(c_{k(p)-1}\right)} \left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^{k(p)}\right] + b - p\right).\end{aligned} \tag{A1}$$

One can show (Online Appendix S.3.1) that

$$\frac{\partial \pi^{CE}(p|p^*)}{\partial p} \geq \tilde{g}(p)\left(\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] + b - p\right) - \frac{\tilde{G}(p)}{\tilde{g}(p)}\right). \tag{A2}$$

Letting $p \downarrow p^* = c_{k(p)-1}$ we obtain

$$\left.\frac{\partial \pi^{CE}(p|p^*)}{\partial p}\right|_{p\downarrow p^*} = \tilde{g}(p^*)\frac{G(c_1) - G(p^*)}{\tilde{G}(c_1) - \tilde{G}(p^*)}\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] + b - p^*\right) - G(p^*).$$

One can show (Online Appendix S.3.1) that

$$\left.\frac{\partial \pi^{CE}(p|p^*)}{\partial p}\right|_{p\downarrow p^*} > g(p^*)\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] + b - p^*\right) - G(p^*). \tag{A3}$$

Finally, we can find a lower bound on the second derivative of $\pi^{CE}(p|p^*)$ with respect to p (see Online Appendix S.3.1). For $p \in (p_t^*, c_1)$ we have

$$\frac{\partial^2 \pi^{CE}(p|p^*)}{\partial p^2} \geq \tilde{g}'(p)\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] + b - p\right) - 2\tilde{g}(p). \tag{A4}$$

A.3.3 Proof of Lemmata 2-4

Proof of Lemma 2. Since $\pi^{CE}\left(p|p^{NE}\right)$ coincides with $\pi^{NE}(p)$ on $[0, p^*] = \left[0, p^{NE}\right]$, the constrained optimal $p \in [0, p^*]$ is at $p = p^* = p^{NE}$. Differentiating π^{CE} at $p \in \mathcal{C}^1 = \left(p^{NE}, c_1\right]$, and letting p go to p^{NE}, we obtain, using (A3),

$$\begin{aligned}
\left.\frac{\partial \pi^{CE}\left(p|p^{NE}\right)}{\partial p}\right|_{p\downarrow p^{NE}} &> g\left(p^{NE}\right)\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] + b - p^{NE}\right) - G\left(p^{NE}\right) \\
&= G\left(p^{NE}\right)\left(\frac{g\left(p^{NE}\right)}{G\left(p^{NE}\right)}\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] + b - p^{NE}\right) - 1\right) \\
&= G\left(p^{NE}\right)\left(\frac{1}{b}\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] + b - p^{NE}\right) - 1\right) \\
&= \frac{G\left(p^{NE}\right)}{b}\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] - p^{NE}\right) \\
&= g\left(p^{NE}\right)\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] - p^{NE}\right) > 0.
\end{aligned}$$

Here, the third and fifth equalities use the fact that $g\left(p^{NE}\right)/G\left(p^{NE}\right) = 1/b$. Since $\pi^{NE}(p)$ is continuous, the desired result is implied. ■

Proof of Lemma 3. Since $\pi^{CE}(p|p_t^*)$ coincides with $\pi^{NE}(p)$ on $[0, p_t^*]$, the constrained optimal $p \in [0, p_t^*]$ is at $p = p^{NE} < p_t^*$. Suppose that $\arg\max_{p\in[p_t^*,1]} \pi^{CE}(p|p_t^*) = p_t^*$ (requiring $\left.\frac{\partial \pi^{CE}(p|p_t^*)}{\partial p}\right|_{p\downarrow p_t^*} \leq 0$). By continuity of $\pi^{CE}(p|p_t^*)$, we have $\arg\max_{p\in[0,1]} \pi^{CE}(p|p_t^*) = p^{NE} < p_t^*$. ■

Proof of Lemma 4. Suppose, $p_t^* < p^{NE}$. Then the constrained optimal $p \in [0, p_t^*]$ is at p_t^*. Differentiating π^{CE} at $p \in \mathcal{C}^1 = (p_t^*, c_1]$, and letting p go to p_t^*, we obtain, using (A3),

$$\begin{aligned}
\left.\frac{\partial \pi^{CE}(p|p^*)}{\partial p}\right|_{p\downarrow p_t^*} &> g(p_t^*)\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] + b - p_t^*\right) - G(p_t^*) \\
&= G(p_t^*)\left(\frac{g(p_t^*)}{G(p_t^*)}\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] + b - p_t^*\right) - 1\right) \\
&\geq G(p_t^*)\left(\frac{g\left(p^{NE}\right)}{G\left(p^{NE}\right)}\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] + b - p_t^*\right) - 1\right) \\
&= G(p_t^*)\left(\frac{1}{b}\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] + b - p_t^*\right) - 1\right) \\
&= g(p_t^*)\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] - p_t^*\right) > 0.
\end{aligned}$$

Hence $\left.\frac{\partial \pi^{CE}(p|p_t^*)}{\partial p}\right|_{p\downarrow p_t^*} > 0$. By continuity of π^{CE} note $\arg\max_{p\in[0,1]} \pi^{CE}(p|p_t^*) > p_t^*$. ■

A.3.4 Proof of Convergence to Cycle in Proposition 3

Lemma A1 *There is some $\delta > 0$ such that if $p^* \leq p^{NE}$ then $\mathbb{E}[\omega|\omega \in C_1] > p^* + \delta$.*

Proof of Lemma A1. We only sketch the proof here. For details see Online Appendix S.3. Assume $p^* \leq p^{NE}$. The mass in each analogy class (above p^*) is at least κ. Let $g^{\min} = \min_{\omega\in[0,1]} g(\omega)$ and $g^{\max} = \max_{\omega\in[0,1]} g(\omega)$. By the full-support assumption we have $g^{\min} > 0$. It can then be shown that $c_1 - p^* \geq \frac{\kappa}{\varepsilon g^{\max}}$. Using this we can establish a lower bound on the expected quality in analogy class $\mathcal{C}^1$.

$$\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] \geq p^* + \frac{1}{2}\left(c_1^*(p^*) - p^*\right)^2 g^{\min}\left(1 - F\left(\frac{1}{2}\left(p^{NE} + 1\right)\right)\right)$$

■ We use Lemma A1 to establish convergence to the cycle from initial prices below p^{NE}.

Lemma A2 *Starting at $p_1^* < p^{NE}$ there is convergence to the set $\left[p^{NE}, 1\right]$.*

Proof of Lemma A2. Consider $p_t^* < p^{NE}$. By Lemma 4 we know that $p_{t+1}^* > p_t^*$. Using Lemma A1 in the proof of Lemma 4 we find that the first derivative of $\pi^{CE}(p|p_t^*)$ wrt to p, is bounded above zero as p goes to p_t^* (from above)

$$\left.\frac{\partial \pi^{CE}(p|p_t^*)}{\partial p}\right|_{p\downarrow p_t^*} > g(p_t^*)\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] - p_t^*\right) > g(p_t^*)\,\delta > \delta g^{\min} > 0. \tag{A5}$$

Here $g^{\min} = \min_{p\in[0,1]} g(p) > 0$ by the full support assumption. We can also find a lower bound for the second derivative of $\pi^{CE}(p|p_t^*)$ wrt to p. From equation (A4) we have

$$\begin{aligned}\frac{\partial^2 \pi^{CE}(p|p_t^*)}{\partial p^2} &\geq \tilde{g}'(p)\left(\mathbb{E}\left[\omega|\omega\in\mathcal{C}^1\right]+b-p\right)-2\tilde{g}(p)\\ &\geq \left(\min_{p\in[0,1]}\tilde{g}'(p)\right)(p_t^*+\delta+b-p)-2\left(\min_{p\in[0,1]}\tilde{g}(p)\right).\end{aligned} \tag{A6}$$

Note that

$$p_{t+1}^* \geq \min\left\{p\in[p_t^*,1]:\frac{\partial\pi^{CE}(p|p_t^*)}{\partial p}\leq 0\right\} \tag{A7}$$

The bounds in (A5) and (A6) imply that the left hand side of (A7) is bounded above p_t^*. ■

Proof of Proposition 3. Assume, to derive a contradiction, that the sequence p_t^* is monotonic. Lemmata 2–4 imply that $p_{t+1}^* > p_t^*$ for all t. Since $p_t^* \leq 1$ for all t, it follows that $p_t^* \to \bar{p}$ for some $\bar{p} > p^{NE}$ as $t \to \infty$. (To see that there is a $\bar{p} > p^{NE}$ note that if $p_1^* \geq p^{NE}$ then $p_t^* \geq p^{NE}$ for all t.) This implies $\left|p_{t+1}^*-p_t^*\right| \to 0$, which, by continuity of $\pi^{CE}(p|p_t^*)$, implies $\left|\pi^{CE}\left(p_{t+1}^*|p_t^*\right)-\pi^{CE}(p_t^*|p_t^*)\right| \to 0$. Since $\pi^{CE}(p|p_t^*)=\pi^{NE}(p)$ for $p\in[0,p_t^*]$, we have $\left|\pi^{CE}\left(p_{t+1}^*|p_t^*\right)-\pi^{NE}(p_t^*)\right| \to 0$, and consequently $\pi^{CE}\left(p_{t+1}^*|p_t^*\right) \to \pi^{NE}(\bar{p})$. Since the Nash equilibrium p^{NE} is unique it holds that $\pi^{NE}\left(p^{NE}\right) > \pi^{NE}(\bar{p})$, and since $\pi^{CE}(p|p_t^*)=\pi^{NE}(p)$ for $p\in[0,p_t^*]$ we get

$$\pi^{CE}\left(p_{t+1}^*|p_t^*\right) \to \pi^{NE}(\bar{p}) < \pi^{NE}\left(p^{NE}\right) = \pi^{CE}\left(p^{NE}|p_t^*\right).$$

This is in contradiction to $p_{t+1}^* = \arg\max_{p\in[0,1]}\pi^{CE}(p|p_t^*)$. We conclude that the sequence p_t^* is not monotonic. Lemmata 2–4 imply that it must be cyclical, consisting of cycles with p^{NE} and one or more price above p^{NE}.

Note that the preceding argument can be used to show, that starting at $p_1^* \geq p^{NE}$ there is convergence to the cycle, from which there is no escape. To see this, suppose (to obtain a contradiction) that there is some $p_1^* > p^{NE}$ that does not belong to the cycle (i.e., $p_1^* \neq p^{(1)}$ for all $i \in \{1,...,\tau\}$), from which there is no convergence to the cycle. This means that $p_{t+1}^* > p_t^*$ for all t and $p_t^* \to \bar{p}$ for some $\bar{p} \in \left[p^{NE}, p^{(\tau)}\right]$ as $t \to \infty$. It remains to show that starting at $p_1^* < p^{NE}$ there is convergence to the set $\left[p^{NE},1\right]$, which is established in Lemma A2 in the Online Appendix. ■

References

Akerlof, George A (1970): "The Market for" Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 488–500.

Anderson, J. R (1991): "The Adaptive Nature of Human Categorization," *Psychological Review*, 98 (3), 409–429.

Arad, Ayala and Ariel Rubinstein (2019): "Multidimensional reasoning in games: framework, equilibrium, and applications," *American Economic Journal: Microeconomics*, 11 (3), 285–318.

Azrieli, Yaron (2009): "Categorizing others in a large game," *Games and Economic Behavior*, 67 (2), 351–362.

Battigalli, Pierpaolo (1987): "Comportamento razionale ed equilibrio nei giochi e nelle situazioni sociali," *unpublished undergraduate dissertation, Bocconi University, Milano.*

Bohren, J Aislinn and Daniel N Hauser (2021): "Learning with heterogeneous misspecified models: Characterization and robustness," *Econometrica*, 89 (6), 3025–3077.

Dekel, Eddie, Drew Fudenberg, and David K Levine (2004): "Learning to play Bayesian games," *Games and Economic Behavior*, 46 (2), 282–303.

Dow, James (1991): "Search Decisions with Limited Memory," *Review of Economic Studies*, 58, 1–14.

Esponda, Ignacio (2008): "Behavioral equilibrium in economies with adverse selection," *American Economic Review*, 98 (4), 1269–1291.

Esponda, Ignacio and Demian Pouzo (2016): "Berk–Nash equilibrium: A framework for modeling agents with misspecified models," *Econometrica*, 84 (3), 1093–1130.

Esponda, Ignacio, Demian Pouzo, and Yuichi Yamamoto (2021): "Asymptotic behavior of Bayesian learners with misspecified models," *Journal of Economic Theory*, 195, 105260.

Eyster, Erik and Matthew Rabin (2005): "Cursed equilibrium," *Econometrica*, 73 (5), 1623–1672.

Fehr, Ernst and Simon Gächter (2000): "Cooperation and punishment in public goods experiments," *American Economic Review*, 90 (4), 980–994.

Fryer, Roland and Matthew O. Jackson (2008): "A Categorical Model of Cognition and Biased Decision Making," *The B.E. Journal of Theoretical Economics (Contributions)*, 8 (1), 1–42.

Fudenberg, Drew and David M Kreps (1993): "Learning mixed equilibria," *Games and economic behavior*, 5 (3), 320–367.

Fudenberg, Drew and Giacomo Lanzani (2023): "Which misspecifications persist?" *Theoretical Economics*, 18 (3), 1271–1315.

Fudenberg, Drew and David K Levine (1993): "Self-confirming equilibrium," *Econometrica: Journal of the Econometric Society*, 523–545.

——— (1998): *The theory of learning in games*, Cambridge, MA: MIT press.

——— (2006): "Superstition and rational learning," *American Economic Review*, 96 (3), 630–651.

Fudenberg, Drew and Alexander Peysakhovich (2016): "Recency, records, and recaps: Learning and nonequilibrium behavior in a simple decision problem," *ACM Transactions on Economics and Computation (TEAC)*, 4 (4), 1–18.

Fudenberg, Drew, Gleb Romanyuk, and Philipp Strack (2017): "Active learning with a misspecified prior," *Theoretical Economics*, 12 (3), 1155–1189.

Gärdenfors, Peter (2000): *Conceptual Spaces: The Geometry of Thought*, Cambridge, MA: MIT Press.

Geman, Stuart, Elie Bienenstock, and René Doursat (1992): "Neural networks and the bias/variance dilemma," *Neural computation*, 4 (1), 1–58.

Gigerenzer, Gerd and Henry Brighton (2009): "Homo heuristicus: Why biased minds make better inferences," *Topics in cognitive science*, 1 (1), 107–143.

He, Kevin and Jonathan Libgober (2020): "Evolutionarily stable (mis) specifications: Theory and applications," *arXiv preprint arXiv:2012.15007*.

Heller, Yuval and Eyal Winter (2020): "Biased-belief equilibrium," *American Economic Journal: Microeconomics*, 12 (2), 1–40.

Jehiel, Philippe (2005): "Analogy-Based Expectation Equilibrium," *Journal of Economic Theory*, 123, 81–104.

——— (2022): "Analogy-Based Expectation Equilibrium and Related Concepts: Theory, Applications, and Beyond," Manuscript prepared for the World Congress of the Econometric Society 2020.

Jehiel, Philippe and Frédéric Koessler (2008): "Revisiting games of incomplete information with analogy-based expectations," *Games and Economic Behavior*, 62 (2), 533–557.

Jehiel, Philippe and Erik Mohlin (2021): "Cycling and Categorical Learning in Decentralized Adverse Selection Economies," .

Jehiel, Philippe and Dov Samet (2007): "Valuation Equilibrium," *Theoretical Economics*, 2, 163–185.

Jehiel, Philippe and Giacomo Weber (2023): "Calibrated Clustering and Analogy-Based Expectation Equilibrium," .

Kalai, Ehud and Alejandro Neme (1992): "The Strength of a Little Perfection," *International Journal of Game Theory*, 20 (4), 335–55.

Kreps, David M, Paul Milgrom, John Roberts, and Robert Wilson (1982): "Rational cooperation in the finitely repeated prisoners' dilemma," *Journal of Economic theory*, 27 (2), 245–252.

Laurence, S. and E. Margolis (1999): "Concepts and Cognitive Science," in *Concepts: Core Readings*, ed. by E. Margolis and S. Laurence, Cambridge, MA: MIT Press, 3–81.

Mengel, F. (2012): "Learning across games," *Games and Economic Behavior*, 74 (2), 601–619.

Miettinen, Topi (2009): "The partially cursed and the analogy-based expectation equilibrium," *Economics Letters*, 105 (2), 162–164.

Mohlin, Erik (2014): "Optimal categorization," *Journal of Economic Theory*, 152, 356–381.

Murphy, G. L. (2002): *The Big Book of Concepts*, Cambridge, MA: MIT Press.

Neyman, Abraham (1985): "Bounded complexity justifies cooperation in the finitely repeated prisoners' dilemma," *Economics letters*, 19 (3), 227–229.

Nyarko, Yaw (1991): "Learning in mis-specified models and the possibility of cycles," *Journal of Economic Theory*, 55 (2), 416–427.

Rosch, E., C. B. Mervis, W. Gray, D. Johnson, and P. Boyles-Brian (1976): "Basic Objects in Natural Categories," *Cognitive Psychology*, 8, 382–439.

Rubinstein, Ariel (1998): *Modeling bounded rationality*, Cambridge, MA: MIT press.

Samuelson, Larry (2001): "Analogies, adaptation, and anomalies," *Journal of Economic Theory*, 97 (2), 320–366.

Selten, Reinhard (1975): "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 25–55.

——— (1978): "The chain store paradox," *Theory and decision*, 9 (2), 127–159.

Spiegler, Ran (2016): "Bayesian networks and boundedly rational expectations," *The Quarterly Journal of Economics*, 131 (3), 1243–1290.

Tanaka, James. W. and Marjorie Taylor (1991): "Object Categories and Expertise: Is the Basic Level in the Eye of the Beholder?" *Cognitive Psychology*, 23, 457–482.

Xu, Fei (2007): "Sortal concepts, object individuation, and language," *Trends in Cognitive Sciences*, 11 (9), 400–406.

ONLINE APPENDIX
Categorization in Games: A Bias-Variance Perspective

Philippe Jehiel and Erik Mohlin

S.1 Chainstore Application

So far, in our analysis of the chainstore game we assumed a homogeneity function that implies that histories are distinguished according to time. What happens if we assume a homogeneity function which relaxes this while still keeping the idea that histories in which a previous entry was not immediately matched by a fight behavior are very dissimilar from other? Compared to the above setting, the only difference is that for the challenger we now consider $\mathcal{Q}_M^{Tough} = \cup_t \mathcal{Q}_M^{t,Tough}$ and $\mathcal{Q}_M^{soft} = \cup_t \mathcal{Q}_M^{t,Soft}$ and we require that if Y contains two nodes q and q' that do not both belong to $\mathcal{Q}_M^{Tough}$ nor both belong to $\mathcal{Q}_M^{soft}$, then $\tilde{\xi}_C(Y) = 0$ (while any set Y not having this property satisfies $\tilde{\xi}_C(Y) > 0$). We define a corresponding categorization profile $\tilde{\mathcal{C}}$ which only differs from $\mathcal{C}$ in that the challengers' categorizations do not differentiate periods, i.e. $\tilde{\mathcal{C}}_C^1 = \cup_t \mathcal{C}_{Ct}^1$ and $\tilde{\mathcal{C}}_C^2 = \cup_t \mathcal{C}_{Ct}^2$.

We now observe that in this alternative setting, there is a coarse categorization equilibrium, this time relying on erroneous expectations of the challengers. Still defining k^* as above, we consider the following strategy profile $\tilde{\sigma}_T$:

- Challenger strategy. If E was always matched with F in the past, or if there was no E in the past, play O. Otherwise play E.
- Monopolist strategy. At $t > T - k^*$, play A. At $t \leq T - k^*$; play F if E was always matched with F in the past, or if there was no E in the past; otherwise play A.

Proposition S1 *There exists a T^* such that if $T > T^*$, then $\left(\tilde{\sigma}_T, \tilde{\mathcal{C}}\right)$ is a coarse categorization equilibrium of the chainstore game with T periods, implying that in the absence of mistakes there is not entry, and the monopolist fights the challenger in all but the last k^* periods.*

On the path of play induced by this strategy profile the challenger never enters. In case there is entry the monopolist fights the challenger in all but the last k^* periods. In this construction, the monopolist plays a best-response to the challenger's strategy and the mistaken belief concerns the challenger who refrains from entering in all periods. She stays out at histories with no earlier (E, A) because she fears the monopolist would fight with a large probability in case of entry. This expectation arises due the bundling of many histories in $\mathcal{Q}_M^{Tough}$ and the observation that according to $\tilde{\sigma}_T$ the monopolist would play F at such histories in all but the last k^* periods. [1]

Proof of Proposition S1. The proof is similar to that of proposition 2. We focus on the differences.

1. Why $\tilde{\mathcal{C}}$ is adjusted to $\tilde{\sigma}_T^m$ for all $m > m^*$ (and all T). Our revised homogeneity assumptions imply that nodes in $\mathcal{Q}_M^{t,Soft}$ should be bundled with nodes in $\mathcal{Q}_M^{t',Soft}$, and nodes in $\mathcal{Q}_M^{t,Tough}$ should be bundled with nodes in $\mathcal{Q}_M^{t',Tough}$ for $t \neq t'$.

2. Analogy-based expectations.

   (a) Players have correct expectations at on-path nodes, as in the proof of proposition 2.

   (b) Players also have correct expectations at off-path nodes in $\mathcal{Q}_C^{Soft}$ and $\mathcal{Q}_M^{Soft}$, as in the proof of proposition 2.

   (c) Next consider off-path monopolist nodes in $\mathcal{Q}_M^{Tough}$. Challengers have erroneous expectations since they bundle together nodes from different time periods. As $\varepsilon^m \to 0$ the expectations here are determined by behavior at nodes with histories containing a single mistake with E. The fraction of such nodes at which the monopolist chooses A vanishes as $T \to \infty$. It follows that as T gets large, the challenger will expect that F is chosen with a probability close to 1.

   (d) It remains to check the monopolist's expectations at off-path nodes in Q_C^{Tough}. At all such nodes the challenger plays O unless trembling. Hence the monopolist has correct expectations.

3. Verify that $\tilde{\sigma}_T^m$ induces a ε_T^m-best-responsese given the analogy-based expectations. We have found that the challengers have correct expectations and it is easy to see that they best-responds to the monopolist's strategy, so we focus on the monopolist.

   (a) Monopolist at an off-path node in Q_M^{Tough}. By playing F, the monopolist correctly expects that with a probability close to 1, a string of O occur from then on until the end of the game. (Same belief as in the proof of proposition 2 but now it is a correct belief.) By playing A, the monopolist correctly expects a string of (E, A) until the end of the game (as in the proof of proposition 2). The

[1] It should be noted that $\tilde{\sigma}_T$ cannot part of a categorization equilibrium when using the homogeneity assumptions of Proposition 2, i.e. when the challenger is induced to categorize different time periods separately. This is so because the challenger would then have to expect that in the last k^* period histories in $\mathcal{Q}_M^{Tough}$ the monopolist would play A after entry, thereby leading challengers to choose E in those events in contrast to the prescription of $\tilde{\sigma}_T$. We see here the effect of the homogeneity functions in shaping the categorization equilibria.

time period $t \leq T - k^*$ where the incentive to take F is weakest is $t = T - k^*$. Taking F not unprofitable if

$$u_M(E,F) + k^* u_M(O) \geq (k^* + 1)\, u_M(E,A),$$

which is satisfied by the definition of k^*. At later time periods taking A is strictly profitable.

(b) Monopolist at an off-path node in Q_M^{Soft}. The monopolist (correctly) expects the challengers to play E in all subsequent periods and best-responds by playing A from now until the end of the game, as in the proof of proposition 2.

(c) Challenger at an off-path node in Q_M^{Tough}. Here, the challenger will expect that E is met by F with a probability close to 1 (as T gets large), hence plays O.

(d) Challenger at an off-path node in Q_M^{Soft}. Here the challenger has correct expectations, hence plays E.

■

S.2 Public Goods Game

S.2.1 The Game With or Without Punishment

We now apply our approach to public good games. The game has more than two players. So far we have only considered two-player games but it is straightforward to extend our basic definitions to the multi-player case. We consider a finitely repeated n-player linear public good game with punishment. The game is repeated T times and players maximize the sum of payoffs. Each round consists of a contribution stage and a punishment stage . Each player holds an endowment of e units. We focus on the simplified case where i can either contribute her entire endowment to the public good or not contribute at all, $g_i \in G = \{0, e\}$. The payoff of player i from the contribution stage is

$$u_i^{Cont}(g) = \alpha \sum_{j=1}^{n} g_j + (e - g_i),$$

where α, with $\frac{1}{n} < \alpha < 1$, captures the marginal per capita return from contributing to the public good. The contribution stage is followed by a punishment stage: each player i can decide whether to punish another player or not. In particular, each player i can subtract punishment points $p_{ij} \in P = \{0, p\}$ from each other player j. For each punishment point a cost of $\beta > 0$ is incurred. This gives rise to the following payoff function,

$$u_i^{Pun}(g,p) = \alpha \sum_{j=1}^{n} g_j + (e - g_i) - \sum_{j \neq i}^{n} p_{ji} - \beta \sum_{j \neq i}^{n} p_{ij}.$$

In the unique SPNE of this game no player contributes, and no player punishes, yielding payoffs of e to everybody. Total payoff is maximized when everyone contributes e, resulting in payoffs of αne.

S.2.2 Zero Contributions without Punishment Stage

We first examine the game without the punishment stage. In this game the stage game payoff of player i is given by u_i^{Cont}. All categorization equilibria are based on the same strategy profile, which coincides with the SPNE, implying that no one contributes. To see why note that in the last round no player contributes, since there is no punishment stage. Suppose there is an equilibrium with full contribution in the second to last round. In this case players on the equilibrium path in the second to last round have a correct belief that no one will contribute in the next round, despite everyone contributing in the second to last round. Thus not contributing in the second to last round is perceived to give a higher payoff, no matter what the off-path expectations about the last round are. Extending this reasoning, we get:

Proposition S2 *Every categorization equilibrium prescribes non-contribution by all players in all rounds.*

Proof. We prove this by induction using the following base case and induction step.

Base case: All categorization equilibria prescribe no-contribution by all players at all information sets in round T.

Induction step: If a categorization equilibrium prescribes no-contribution by all players on the equilibrium path in rounds $\{t+1, ...T\}$ then the categorization equilibrium also prescribes non-contribution by all players on the equilibrium path in round t.

To establish the base case, consider a player i in period T at an information set at which the her strategy prescribes contribution. Regardless of what she expects the other players to do, no-contribution yields a higher payoff.

To establish the induction step, consider a categorization equilibrium that prescribes no-contribution by all players on the equilibrium path in rounds $\{t+1, ...T\}$. Consider player i in period t at an information set H_t on the equilibrium path (there is only one unless non-degenerate mixed strategies are used). Suppose the strategy prescribes contribution by player i. All on-path nodes are singleton categories. Hence, player i has a correct belief that compliance, i.e. contribution in the current round and no-contribution in the following round yields $\alpha\left(e + \sum_{j \neq i} g_j\left(H_t\right)\right) + e\left(T-t\right)$. Deviation is expected to yield at least $\alpha\left(\sum_{j \neq i} g_j\left(H_t\right)\right) + e + e\left(T-t\right)$. The latter is larger than the former. ■

S.2.3 Positive Contributions with Punishment Stage

Our assumption regarding similarity and homogeneity is that players distinguish sharply between two kinds of histories: (i) histories in which all acts of non-contributions where punished (by all those who contributed) and no act of contribution was punished, and (ii) all other histories. A history of either kind is never bundled with a history of the other kind. We also assume that $(n-1)\,p \geq e\,(1-\alpha)$, meaning that the cost of being punished is high enough relative to the benefit of not contributing. Under these assumptions we can show, that for sufficiently long games (sufficiently large T) there is a categorization equilibrium with contribution in every round, and (off-path) punishment in a no-contribution event except in the last few periods. The construction is similar to the one underlying Proposition S1 for the chainstore game. In the first kind of histories (i) the strategy prescribes contribution and punishment of non-contributors (and only non-contributors), except in the last few rounds in which non-punishment is prescribed. In the second kind of history (ii) the strategy prescribes non-contribution and no punishment. The threat of punishment off-path would not be credible in a standard SPNE. The reason players contribute throughout the interaction in our categorization equilibrium is that the bundling of all off-path histories of the first kind induce players to believe that they will be punished with probability approaching one (as $T \to \infty$) if they fail to contribute, even towards the end of the game where in reality they would not be punished. In what follows we provide a detailed description of our construction

Similarity and Homogeneity In general it is natural to assume that if two situations $x_i, x_i' \in X_i$ have different actions sets, i.e. $A_i(x_i) \neq A_i(x_i')$, then any analogy class that contains both situations has minimal homogeneity. This implies that an adjusted analogy partition will never bundle nodes with different action sets, as in Jehiel (2005). Since contribution decision information sets and punishment information sets have different actions sets any analogy class that contains both kinds of information sets have minimal homogeneity. Let H^{Con} denote the sets of contribution decision information sets, and let H^{Pun} denote the set of punishment decision information sets. Since the action sets are different any set that bundles information sets from $\mathcal{H}^{Con}$ and $\mathcal{H}^{Pun}$ have minimal homogeneity. For both $\mathcal{H}^{Con}$ and $\mathcal{H}^{Pun}$ we assume that homogeneity is mainly determined by whether non-contributors, but not contributors, were punished. Let $\mathcal{H}^{Fair}$ denote the set of information sets with a history such that in each previous round all non-contributors were punished by all contributors, and no contributors were punished.

$$\mathcal{H}^{Fair} = \left\{ \begin{array}{l} \text{In each previous round in the history of } H\text{, for all } j\text{:} \\ \;\; g_j = 0 \Rightarrow p_{lj} = p \text{ for all } l \text{ with } g_l = e\text{, and} \\ \;\; g_j = 1 \Rightarrow p_{lj} = 0 \text{ for all } l. \end{array} \right\}$$

Let $\mathcal{H}^{Unfair}$ denote the complement, i.e. information sets with a history such that in at least one previous round there was a non-contributor who was not punished by all contributors, or there was a contributor who was punished. We assume if H and H' belong to X but

$H \in \mathcal{H}^{Fair}$ and $H' \in \mathcal{H}^{Unfair}$, then $\xi(X) = 0$. Let

$$\begin{aligned}
\mathcal{H}^{Con-Fair} &= \mathcal{H}^{Con} \cap \mathcal{H}^{Fair} \\
\mathcal{H}^{Con-Unfair} &= \mathcal{H}^{Con} \cap \mathcal{H}^{Unfair} \\
\mathcal{H}^{Pun-Fair} &= \mathcal{H}^{Pun} \cap \mathcal{H}^{Fair} \\
\mathcal{H}^{Pun-Unfair} &= \mathcal{H}^{Pun} \cap \mathcal{H}^{Unfair}
\end{aligned}$$

Any subset X containing only elements in $\mathcal{H}^{Con-Fair}$ or only elements in $\mathcal{H}^{Con-Unfair}$ satisfies $\xi(X) > 0$. Likewise, any subset X containing only elements in $\mathcal{H}^{Pun-Fair}$ or only elements in $\mathcal{H}^{Pun-Unfair}$ satisfies $\xi(X) > 0$.

Strategy profile We assume

$$(n-1)\,p \geq e\,(1-\alpha)\,. \tag{S1}$$

For each $\bar{n} \in \{1, ..., n-1\}$ let

$$k^*_{\bar{n}} = \min\left\{k \in \mathbb{N} \text{ such that } \left(\alpha n + 1\right) ek \geq \beta p \bar{n}\right\}. \tag{S2}$$

Consider the strategy profile $\hat{\sigma}$, where each individual i plays the following strategy:

- At $H \in \mathcal{H}^{Con-Fair}$, contribute e.
- At $H \in \mathcal{H}^{Con-Unfair}$, do not contribute.
- At $H \in \mathcal{H}^{Pun-Fair}$, where in the immediately preceding contribution stage,
  - i contributed and $\bar{n} \in \{1, ..., n-1\}$ other players did not contribute: punish if $t \leq T - k^*_{\bar{n}}$, otherwise do not punish.
  - i contributed and all other players contributed: do not punish.
  - i did not contribute: do not punish.
- At $H \in \mathcal{H}^{Pun-Unfair}$, do not punish.

On the path of play induced by this strategy profile everyone contributes in all rounds. In case there is non-contribution all contributors punish, except the last period.

Categorization profile Under the categorization profile $\widehat{\mathcal{C}}$, each on-path information set is in a separate analogy class, as usual. Off-path information sets are categorized based on the type of decision (contribution or punishment) and on whether the history was in

$\mathcal{H}^{Fair}$ or $\mathcal{H}^{Unfair}$. Formally, let $\mathcal{H}_{-i}^{off}$ denote the off-path information sets at which players other than i move, define

$$\begin{aligned}
\mathcal{C}_{-i}^{Con-Fair} &= \left\{H \in \mathcal{H}_{-i}^{off} : H \in \mathcal{H}^{Con-Fair}\right\}; \\
\mathcal{C}_{-i}^{Con-Unfair} &= \left\{H \in \mathcal{H}_{-i}^{off} : H \in \mathcal{H}^{Con-Unfair}\right\}; \\
\mathcal{C}_{-i}^{Pun-Fair} &= \left\{H \in \mathcal{H}_{-i}^{off} : H \in \mathcal{H}^{Pun-Fair}\right\}; \\
\mathcal{C}_{-i}^{Pun-Unfair} &= \left\{H \in \mathcal{H}_{-i}^{off} : H \in \mathcal{H}^{Pun-Unfair}\right\}.
\end{aligned}$$

Proposition S3 *If (S1) then there exists a T^* such that if $T > T^*$, then $\left(\hat{\sigma}_T, \widehat{\mathcal{C}}\right)$ is a coarse categorization equilibrium of the chainstore game with T periods, implying that in the absence of mistakes everyone contributes in all rounds.*

Remark S1 *Condition S1 requires that the cost of being punished is high enough relative to the benefit of not contributing, and the definition of $k_{\bar{n}}^*$ in (S2) implies that in period $t \leq T - k^*$ the cost of punishing is lower that the loss from others not contributing (in response to non-punishment), whereas in in period $t \leq T - k^*$ the cost of punishing is higher than the loss from others not contributing.*

Proof of Proposition S3. We need to show that for $T > T^*$ there is a sequence $(\hat{\sigma}_T^m)_m$ converging to $\hat{\sigma}_T$, such that $(\hat{\sigma}_T^m, \widehat{\mathcal{C}})$ is an $(\varepsilon^m, \kappa^m)$- categorization equilibrium for all m. We define $\hat{\sigma}_T^m$ as the strategy profile which at each node puts probability ε^m on the action that $\hat{\sigma}_T$ puts zero probability on. Since there are only two actions at each node this is enough to specify $\hat{\sigma}_T^m$. Since the starting point of $(\varepsilon^m, \kappa^m)$ is arbitrary it is sufficient to show the following: There exists a T^* such that for any $T > T^*$there is exists an m^* such that if $T > T^*$ and $m > m^*$ then $\hat{\sigma}_T^m$ is an $(\varepsilon_T^m, \kappa_T^m)$-categorization equilibrium of the chainstore game with T periods.

1. Why $\widehat{\mathcal{C}}$ is adjusted to $\hat{\sigma}_T^m$ for all $m > m^*$ (and all T).

   (a) For any T, if m is large enough, then $\kappa_T^m < (1 - \varepsilon_T^m)^{2nT}$, ensuring that on-path nodes have a mass exceeding the threshold κ_T^m and thus are treated as singleton analogy classes.

   (b) For off-path nodes our homogeneity assumptions imply that information sets in $\mathcal{H}^{Fair}$ and $\mathcal{H}^{Unfair}$ have to be separated. Likewise, information sets in $\mathcal{H}^{Con}$ and $\mathcal{H}^{Pun}$ have to be separated. No further refinement is allowed (for m large enough).

2. Analogy-based expectations[2]

[2]In a game with more than two players there are at least two options for how to specify analogy

(a) Players have correct expectations at on-path information sets.

(b) Players also have correct expectations at off-path information sets in $\mathcal{H}^{Unfair}$, since after the corresponding histories no one contributes at any information set and no one punishes at any information set.

(c) Next consider off-path information sets in $\mathcal{H}^{Con-Fair}$. At all such nodes everyone contributes, resulting in correct expectations.

(d) Finally consider off-path information sets in $\mathcal{H}^{Pun-Fair}$. As $\varepsilon^m \to 0$ the expectations here are determined by behavior at information sets with histories containing a single act of non-contribution (due to a mistake) in the present round. The fraction of such nodes at which not everyone punishes vanishes as $T \to \infty$. It follows that as T gets large, expects everyone except the non-contributor to punish with a probability close to 1.

3. Verify that $\hat{\sigma}^m_T$ induces a ε^m_T-best-responsese given the analogy-based expectations.

(a) First consider player i at an information set $H \in \mathcal{H}^{Con-Fair}$ (on-path or off-path) in round $t \leq T$. Complying with the proposed strategy profile yields for the continuation
$$EU_i\left(g_i = e|t\right) = \alpha ne\left(T - t + 1\right).$$
The player believes that if she makes a one-shot deviation then with probability approaching 1 (as $T \to \infty$) everyone else punishes her, and play remains in $\mathcal{H}^{Con-Fair}$. Hence, a one-shot deviation yields
$$EU_i\left(g_i = 0|t\right) = \alpha ne + e\left(1 - \alpha\right) + \left(-\left(n-1\right)p + \alpha ne\left(T - t\right)\right)$$
The difference is
$$EU_i\left(g_i = e|t\right) - EU_i\left(g_i = 0|t\right) = \left(n-1\right)p - e\left(1 - \alpha\right).$$
If (S1) holds then deviation is not profitable.

(b) Second, consider player i at information set $H \in \mathcal{H}^{Con-Unfair}$ in round $t \leq T$. Complying with the proposed strategy profile yields $EU_i\left(g_i = 0|t\right) = \left(T - t + 1\right)e$. A one-shot deviation yields $EU_i\left(g_i = 0|t\right) = \alpha e + \left(T - t\right)e$. The former is larger than the latter since $\alpha < 1$.

(c) Third, consider player i at an information set $H \in \mathcal{H}^{Pun-Fair}$ (on-path or off-path) in round $t \leq T$.

based expectations at off-path information sets. Players may ignore correlation across the other players' actions and form expectations about individual actions (here contributions), or they may form expectations about the distribution of actions (contributions). Here we present results derived for expectations about individual contributions. We can confirm that the results are essentially the same under expectations about the distribution of contributions.

i. If everyone complied in the contribution stage then (clearly) not punishing is perceived to be optimal.

ii. If player i was the only one not to contribute, then (clearly) not punishing is perceived to be optimal.

iii. If i contributed and $\bar{n}$ other players did not contribute then i believes that with probability approaching 1 (as $T \to \infty$) all other contributors will punish the non-contributors, so that punishing yields

$$EU_i\left(p_{il} = p|t\right) = -\beta p\bar{n} + \alpha ne\left(T - t\right).$$

not punishing leads to $\mathcal{H}^{Unfair}$, hence yields $EU_i\left(p_{il} = 0|t\right) = e\left(T - t\right)$. The difference is

$$EU_i\left(p_{il} = p|t\right) - EU_i\left(p_{il} = 0|t\right) = -\beta p\bar{n} + \left(\alpha n + 1\right)e\left(T - t\right).$$

This is decreasing in t. For $t = T - k^*_{\bar{n}}$ the difference is

$$EU_i\left(p_{il} = p|t\right) - EU_i\left(p_{il} = 0|t\right) = -\beta p\bar{n} + \left(\alpha n + 1\right)ek^*_{\bar{n}}.$$

By the definition of $k^*_{\bar{n}}$ this non-negative, hence punishing is profitable for $t \leq T - k^*$. For $t > T - k^*$ it is strictly negative so punishing is not profitable.

(d) Fourth, consider player i at information set $H \in \mathcal{H}^{Pun-Unfair}$ in round $t \leq T$. Clearly, punishing is not perceived as optimal.

■

S.3 Adverse Selection Application

S.3.1 Preliminary Observations

To demonstrate (A2) we rewrite (A1) as follows

$$\begin{aligned}\frac{\partial \pi^{CE}\left(p|p^*\right)}{\partial p} &= \frac{\tilde{G}\left(p^*_t\right)G\left(p^*_t\right)}{\tilde{G}\left(c_1\right) - \tilde{G}\left(p^*_t\right)}\left(\frac{G\left(c_1\right) - G\left(p^*_t\right)}{G\left(p^*_t\right)} - \frac{\left(\tilde{G}\left(c_1\right) - \tilde{G}\left(p^*_t\right)\right)}{\tilde{G}\left(p^*_t\right)}\right) \\ &+ \frac{G\left(c_1\right) - G\left(p^*_t\right)}{\tilde{G}\left(c_1\right) - \tilde{G}\left(p^*_t\right)}\tilde{g}\left(p\right)\left(\left(\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] + b - p\right) - \frac{\tilde{G}\left(p\right)}{\tilde{g}\left(p\right)}\right).\end{aligned}$$

We note that

$$\frac{\left(\tilde{G}\left(c_1\right)-\tilde{G}\left(p_t^*\right)\right)}{\tilde{G}\left(p_t^*\right)}=\frac{\int_{p_t^*}^{c_1}\left(1-F_{p^*}\left(\omega\right)\right)g\left(\omega\right)d\omega}{\left(\int_0^{p_t^*}\left(1-F_{p^*}\left(\omega\right)\right)g\left(\omega\right)d\omega\right)}<\frac{\left(1-F_{p^*}\left(p_t^*\right)\right)\int_{p_t^*}^{c_1}g\left(\omega\right)d\omega}{\left(1-F_{p^*}\left(p_t^*\right)\right)\int_0^{p_t^*}g\left(\omega\right)d\omega}$$
$$=\frac{\int_{p_t^*}^{c_1}g\left(\omega\right)d\omega}{\int_0^{p_t^*}g\left(\omega\right)d\omega}=\frac{G\left(c_1\right)-G\left(p_t^*\right)}{G\left(p_t^*\right)}.$$

Moreover, we note that

$$\frac{G\left(c_1\right)-G\left(p_t^*\right)}{\tilde{G}\left(c_1\right)-\tilde{G}\left(p_t^*\right)}=\frac{\int_{p_t^*}^{c_1}g\left(\omega\right)d\omega}{\int_{p_t^*}^{c_1}\left(1-F_{p^*}\left(\omega\right)\right)g\left(\omega\right)d\omega}\geq 1,$$

Thus (A2) is implied. To demonstrate (A3) note that

$$\tilde{g}\left(p^*\right)\frac{G\left(c_1\right)-G\left(p^*\right)}{\tilde{G}\left(c_1\right)-\tilde{G}\left(p^*\right)}=\frac{\left(1-F_{p^*}\left(p^*\right)\right)\left(G\left(c_1\right)-G\left(p^*\right)\right)}{\int_{p^*}^{c_1}g\left(\omega\right)\left(1-F_{p^*}\left(\omega\right)\right)d\omega}g\left(p^*\right)$$
$$=\frac{\left(1-F_{p^*}\left(p^*\right)\right)\int_{p^*}^{c_1}g\left(\omega\right)d\omega}{\int_{p^*}^{c_1}\left(1-F_{p^*}\left(\omega\right)\right)g\left(\omega\right)d\omega}g\left(p^*\right)>g\left(p^*\right).$$

Finally, to demonstrate (A4) we note that for $p\in\left(p_t^*,c_1\right)$

$$\frac{\partial^2\pi^{CE}\left(p|p^*\right)}{\partial p^2}=\tilde{g}'\left(p\right)\left(\left(\mathbb{E}\left[\omega|\omega\in\mathcal{C}^1\right]+b-p\right)-2\frac{\tilde{g}\left(p\right)}{\tilde{g}'\left(p\right)}\right)\frac{G\left(c_1\right)-G\left(p_t^*\right)}{\tilde{G}\left(c_1\right)-\tilde{G}\left(p_t^*\right)}.$$

Using (S.3.1) we obtain (A4)

S.3.2 Nash equilibrium

Proposition S4 *There exists a unique Nash equilibrium in which the bid price p^{NE} of uninformed buyers is uniquely defined by*

$$\frac{g\left(p^{NE}\right)}{G\left(p^{NE}\right)}=\frac{1}{b}.$$

Proof of Proposition S4. Note that

$$\frac{\partial}{\partial p}\left(\mathbb{E}\left[\omega|\omega\leq p\right]\right)=\frac{1}{G\left(p\right)}pg\left(p\right)-\left(\int_{\omega=0}^{p}\omega g\left(\omega\right)d\omega\right)\frac{g\left(p\right)}{G\left(p\right)^2}$$
$$=\frac{g\left(p\right)}{G\left(p\right)}\left(p-\int_{\omega=0}^{p}\omega\frac{g\left(\omega\right)}{G\left(p\right)}d\omega\right)$$
$$=\frac{g\left(p\right)}{G\left(p\right)}\left(p-\mathbb{E}\left[\omega|\omega\leq p\right]\right).$$

Thus

$$\begin{aligned}
\frac{\partial}{\partial p}\pi^{NE}(p) &= g(p)\left(\mathbb{E}[\omega|\omega \le p] + b - p\right) + G(p)\left(\frac{\partial}{\partial p}\left(\mathbb{E}[\omega|\omega \le p]\right) - 1\right) \\
&= g(p)\left(\mathbb{E}[\omega|\omega \le p] + b - p\right) + G(p)\left(\frac{g(p)}{G(p)}\left(p - \mathbb{E}[\omega|\omega \le p]\right) - 1\right) \\
&= g(p)\left(\mathbb{E}[\omega|\omega \le p] + b - p\right) + g(p)\left(p - \mathbb{E}[\omega|\omega \le p]\right) - G(p) \\
&= g(p)\, b - G(p),
\end{aligned}$$

and so the first-order condition of $\max_p \pi^{NE}(p)$ is

$$\frac{g(p)}{G(p)} = \frac{1}{b},$$

and the second-order condition is satisfied in virtue of the assumption that $|g'(p)| < g(p)$. Notice that $\lim_{p\to 0} \frac{g(p)}{G(p)} = \infty$ and $\frac{g(1)}{G(1)} = g(1)$. Hence, by the assumption that $g(1) < 1/b$ and $\frac{\partial}{\partial p}\left(\frac{g(p)}{G(p)}\right) < 0$, the first-order condition has a unique solution that is interior. ■

S.3.3 Lemma for Proof of Convergence to Cycle

Proof of Lemma A1. Assume $p^* \le p^{NE}$. The mass in each analogy class (above p^*) is at least κ. We establish a lower bound on the width of analogy class $\mathcal{C}^1$. Let $g^{\min} = \min_{\omega\in[0,1]} g(\omega)$ and $g^{\max} = \max_{\omega\in[0,1]} g(\omega)$. By the full-support assumption we have $g^{\min} > 0$. Note that

$$\int_{\omega\in\mathcal{C}^1} \mu(\omega)\, d\omega = \varepsilon \int_{\omega\in\mathcal{C}^1} \tilde{g}(\omega)\, d\omega \le \varepsilon \int_{\omega\in\mathcal{C}^1} g^{\max} d\omega = \varepsilon\left(c_1 - p^*\right) g^{\max} \Rightarrow c_1 - p^* \ge \frac{\kappa}{\varepsilon g^{\max}}.$$

Using this we can establish a lower bound on the expected quality in analogy class $\mathcal{C}^1$. Define

$$c_1^*(p^*) = \min\left\{p^* + \frac{\kappa}{\varepsilon g^{\max}}, \frac{1}{2}\left(p^{NE} + 1\right)\right\} \le c_1,$$

implying that

$$c_1^*(p^*) - p^* \ge \min\left\{\frac{\min\left\{\kappa, \varepsilon\left(\tilde{G}(1) - \tilde{G}\left(p^{NE}\right)\right)\right\}}{\varepsilon g^{\max}}, \frac{1 - p^{NE}}{2}\right\} := M_1.$$

Note that

$$\begin{aligned}
\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] \ge{} & \left(1 - \frac{1}{\mu(\mathcal{C}^1)}\int_{\omega=p^*}^{c_1^*(p^*)} g^{\min}\left(1 - F\left(c_1^*(p^*)\right)\right) d\omega\right)\cdot p^* \\
& + \frac{1}{\mu(\mathcal{C}^1)}\int_{\omega=p^*}^{c_1^*(p^*)} g^{\min}\left(1 - F\left(c_1^*(p^*)\right)\right) d\omega \cdot \left(p^* + \frac{c_1^*(p^*) - p^*}{2}\right).
\end{aligned}$$

Moreover,

$$\int_{\omega=p^*}^{c_1^*(p^*)} g^{\min}\left(1-F\left(c_1^*\left(p^*\right)\right)\right) d\omega \geq \left(c_1^*\left(p^*\right)-p^*\right) g^{\min}\left(1-F\left(\frac{1}{2}\left(p^{NE}+1\right)\right)\right)$$
$$\geq M_1 \cdot g^{\min}\left(1-F\left(\frac{1}{2}\left(p^{NE}+1\right)\right)\right) := M_2.$$

Thus we have

$$\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] \geq \left(1-\frac{M_2}{\mu\left(\mathcal{C}^1\right)}\right) p^* + \frac{M_2}{\mu\left(\mathcal{C}^1\right)}\left(p^* + \frac{c_1^*\left(p^*\right)-p^*}{2}\right)$$
$$= p^* + \frac{M_2}{\mu\left(\mathcal{C}^1\right)}\left(\frac{c_1^*\left(p^*\right)-p^*}{2}\right) \geq p^* + \frac{M_2}{2} M_1,$$

or

$$\mathbb{E}\left[\omega|\omega \in \mathcal{C}^1\right] \geq p^* + \frac{1}{2}\left(c_1^*\left(p^*\right)-p^*\right)^2 g^{\min}\left(1-F\left(\frac{1}{2}\left(p^{NE}+1\right)\right)\right)$$

■

S.4 Categorization Equilibrium and Nash Equilibrium

Here we present examples demonstrating that, CE may not be not outcome equivalent to any NE, for the reason that this would require inconsistent beliefs, as mentioned in Section 6.1.

S.4.1 Example where Feedback Differs from the Path of Play

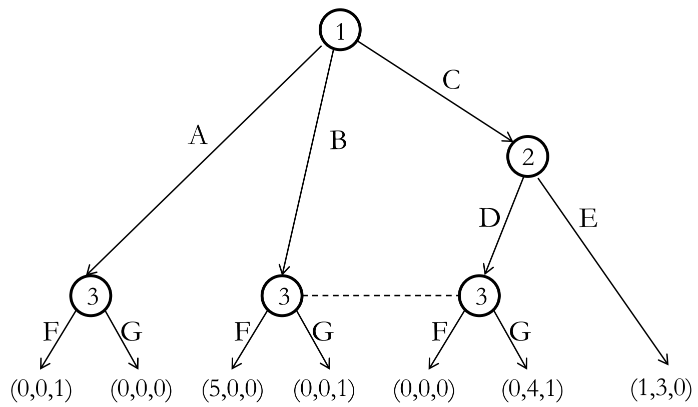
Consider the following game. Player 1 (row) and Player 2 (column) simultaneously choose between actions A and B, with the following outcomes.

	A	B
A	$0,1$	$1,0$
B	$1,1$	$0,0$

The unique Nash equilibrium is (B, A). Note that B is dominated for Player 2 so we can ignore her belief formation. Suppose that the feedback is such that the outcome of the game is reported if and only if it is (B, B). This means that an entering cohort will see a record consisting entirely of (B, B) outcomes, and those acting as Player 1 will form the belief that Player 2 plays action B with probability 1. The best response is action A. Thus the unique Categorization equilibrium outcome is (A, A).

S.4.2 Example where Feedback Coincides with the Path of Play

We now turn to an example where the feedback is the path of play. We need to assume that there are three players so that two of them can disagree about what the remaining player does off the path. Consider the following game.



There is a categorization equilibrium involving the strategy profile (C, E, FG), according to which Player 1 plays C, Player 2 plays E, and Player 3 plays F at the node following A and plays G at the information set following B and D. Only the root node and the node following C are on the path of play. Suppose that Player 1 deems all of Player 3's nodes sufficiently similar to be bundled together in a single analogy class, whereas Player 1 perceives them sufficiently dissimilar to put each of them in a separate category.

To see that this constitutes a categorization equilibrium note that F is dominant for Player 3 at the node following A, and G is dominant for Player 3 at the information set following B and D. Since Player 2 has correct beliefs about the behavior of Player 3 it follows that E is optimal for Player 2. All of Player 3's nodes are reached by a single mistake. Hence Player 1 believes that Player 3 plays F with probability $1/3$ at all of Player 3's nodes (since Player 1 bundles them all together). Player 1 has a correct belief about Player 2's behavior at the on-path node following C. Under these beliefs Player 1 optimally plays C.

In order for Player 2 to take action E she needs to believe that player 3 plays F with at least probability $1/4$ at the information set following B and D. Hence, in a Nash equilibrium implementing the outcome (C, E) Player 3 must follow a strategy that puts at least probability $1/4$ on F at the information set following B and D. In order for Player 1 to take action C rather than action B she needs to believe that player 3 plays F with at most probability $1/5$ at the node following B. Hence in a Nash equilibrium implementing the outcome (C, E) Player 3 must follow a strategy that puts at most probability $1/5$ on F at the information set following B and D. Thus the beliefs required for Players 1 and 2 are inconsistent.