Categorization in Games: A Bias-Variance Perspective

Jehiel, Philippe; Mohlin, Erik

2025

Document Version:
Other version

Working Paper 2025:7

Department of Economics
School of Economics and Management

Categorization in Games: A Bias-Variance Perspective

Philippe Jehiel
Erik Mohlin

Sep 2025

LUND UNIVERSITY

Categorization in Games: A Bias-Variance Perspective*

Philippe Jehiel† Erik Mohlin‡

August 13, 2025

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 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. We apply our framework to classic games including the chainstore game and adverse selection games, thereby suggesting novel predictions for these applications.

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

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 and Bar-Ilan University 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.

‡Swedish Defence University, 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.

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). A significant part of this literature has worked with exogenously given categories.[1] In this paper, we impose structure on the categories taking inspiration from the insights developed in relation to the bias-variance trade-off.

Our starting point is the analogy-based expectation equilibrium (Jehiel, 2005; Jehiel and Koessler, 2008) in which players use categories (analogy classes) to form predictions about opponents' play in a game. To fix ideas we interpret our model in terms of the following dynamic process. Time is discrete and in each period a single

[1] Exceptions include Dow (1991), Fryer and Jackson (2008), Mengel (2012), and Heller and Winter (2020), which to various extent let categories be endogenously determined.

cohort is active. Individuals from the active cohort are drawn to play in the different player roles. Many different groups of individuals (consisting of one individual for each player role in the game) play the same game at the same time.

In the first phase of a time period t, players receive some feedback about how the game has been played by all the groups in the previous cohort. Specifically feedback consists in the disclosure of behaviors in a subset of situations. In extensive form games, a situation corresponds to a node, and the feedback will typically consist of behaviors at on-path nodes. In Bayesian games situations represent types (product quality, in our leading example) and feedback may disclose type conditional on some event (such as trade) taking place.

In the second phase, individuals in a particular player role 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 a priori perception of how similar the various situations are to each other.[2] They form *endogenous* categories by bundling together situations they perceive 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. In particular, we formalize the latter by requiring 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. The categorization is then chosen so as to maximize the within-category similarities subject to this constraint.

In the third phase, players use the categories and the feedback to form expectations about what their opponents will do. A player's prediction about the play of the opponents in a given situation (i.e. play at a given node or play for a given type) is assumed to match the empirical distribution of the behaviors observed in the previous cohort in the category to which the situation has been assigned.

In the fourth and final phase, players best respond given their beliefs. In the the tradition of Selten's trembling hand idea (Selten, 1975) we assume that in every situation a player picks non-intended actions with probability ε.

These phases are implemented for every cohort over the various time periods generating a dynamic process which depends on the similarity functions, the specifications of ε and κ as well as the initial conditions. Steady states of the process

[2]These perceptions can be thought of as resulting from cultural and psychological factors, which are external to our model.

are referred to as (ε, κ)-categorization equilibria. While our approach allows for any specification of κ and ε we focus on the case in which κ and ε vanish at such a rate that ε is asymptotically not too large relative to κ. This implies that on-path situations (defined as situations for which feedback is obtained in the absence of trembles) can be distinguished perfectly but off-path situations have to be bundled (according to their similarity). In one application, instead of considering the steady states, we study the learning dynamic directly, as it gives rise to interesting cycling phenomena.

Our first main contribution is to provide a general framework that endogenizes the analogy partitions along the lines outlined above. Compared to the analogy-based expectation equilibrium setting, our framework adds an extra channel relating the categories or analogy classes to the strategy profile through the bias-variance trade-off principle, as explained above. Our second main contribution is to provide a series of applications, thereby highlighting how non-standard predictions can arise for what we believe are plausible specifications of the similarity functions.

Our main applications are as follows. We first 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 (perhaps because the monopolist reveals a form of weakness in one case but not in the other). We establish the existence of a categorization equilibrium with no entry except in the last few periods. We next 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. Assuming that feedback about quality is obtained only when there is trade, we show that the learning dynamics leads to cycles with bid prices always weakly above the Nash equilibrium price for the natural specification that considers nearby qualities as being similar to one another.[3]

In the last part of the paper, we provide a general discussion. We note that even when considering finite games, the existence of (ε, κ)-categorization equilibria may require extending our basic framework to allow for mixed distributions over analogy partitions. We also observe that in the context of extensive form games of complete information, categorization equilibria can be viewed as refinements of self-confirming

[3]In the Online Appendix S.1, we consider ultimatum games in which the responder has an outside option, and illustrate how our approach can predict that positive rents are left to the receiver. In the Online Appendix S.3, we consider public good games for which we derive insights related to those obtained in the chainstore game.

equilibria (Fudenberg and Levine, 1993, 1998) in which off-path beliefs are structured by the actual behaviors through the endogenously determined coarse analogy classes that apply there.[4]

Our paper is related to different branches of the literature. Regarding the aplications, the relevant literature will be mentioned in the respective sections. At a broader level, our paper can be related to a growing literature on misspecifications in games, which, in addition to the already mentioned analogy-based expectation equilibrium (Jehiel, 2005), include the cursed equilibrium (Eyster and Rabin, 2005), 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), 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.

Finally, a contemporaneous alternative approach to categorization in the context of the analogy-based expectation equilibrium is introduced in Jehiel and Weber (2024). They consider distributions over normal form games in which players select their analogy partitions so as to minimize the prediction error about opponents' play subject to using at most k analogy classes. Their approach relates to the k-means clustering technique, and differs sharply with the one considered here that relates to the bias-variance trade-off with no a priori constraint on the number of categories. In particular, an important aspect of our analysis relies on how the magnitude of trembling ε relates to the minimum mass κ requirement, which has no counterpart in the analysis of Jehiel and Weber.

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

[4]The obtained refinement should be contrasted with that proposed in Fudenberg and Levine (2006) in which for nodes that are two steps away from the path, the beliefs can be arbitrary.

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 situati ons 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 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 in the literature on learning in games, see Fudenberg and Levine, 1998). In Bayesian games, we will use this formulation to accommodate applications like trades in which the actions (bid or ask price) and types (determining the quality of the good) are disclosed only when the transaction takes place (as in Esponda, 2008).

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 $\mathcal{C}_i^k$. For every $x_j \in \mathcal{X}_j$, we let $\mathcal{C}_i(x_j)$ be the 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 $\beta_i(\mathcal{C}_i(x_j))$.[5]

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. 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,[6,7]

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.*

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 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.

[5]This can be interpreted as a form of correlation neglect within each category C_i^k, as player j's behavior may depend on x_j for the various $x_j \in C_i^k$, but player i thinks all these behaviors are the same.

[6]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.

[7]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 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 induces a probability $\mu^\sigma(a_j,x_j)$ that action a_j in situation x_j is disclosed.[8] We assume that ϕ is such that for every ε-perturbed strategy profile σ,[9] and for every analogy class $\mathcal{C}_i^k$, some behavior in $\mathcal{C}_i^k$ is disclosed with strictly positive probability. That is,[10]

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

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 $\beta_i(\mathcal{C}_i^k)$.

Combining Definition 1 and Definition 2 we propose a generalized version of analogy-based expectation equilibrium:[11]

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 $i = 1, 2$:*

[8] 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.

[9] That is, a strategy profile such that for every player i and at every x_i, $\sigma_i(x_i)$ assigns a probability no less than ε_i to every action at x_i.

[10] 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.

[11] 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.

(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 the feedback structure ϕ either.

2.2 Endogenous Categorizations

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. These functions ζ_i are left exogenous in our approach and should be thought of as being determined by pyschological and cultural factors that apply broadly, and thus should not be thought of as being primarily determined by the play in the specific interaction under study.[12] We make the natural assumption that a singleton set has maximum homogeneity, i.e. $\zeta_i(\{x_j\}) = 1$ for all $x_j \in \mathcal{X}_j$. Observe that 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).[13,14]

As a key step in our proposed approach, we introduce the following definition.

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 $\mu^\sigma(\{x\}) > \kappa$, there exists k such that $\mathcal{C}_i^k = \{x\}$.*

2. *For each $X \subseteq X_j$ with $\zeta_i(X) = 0$, there exists no k such that $\mathcal{C}_i^k = X$.*

[12] This is a different perspective from the one considered in Jehiel and Weber (2025) in which the similarity is purely based on the behaviors in the interactions under study.

[13] 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 (see definition 4) will never bundle nodes with different action sets.

[14] 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 the qualitative insights developed next.

3. *Let* $\mathcal{X}_j^{sing}$ *denote the set of situations put into singleton analogy classes in* $\mathcal{C}_i$*. If* C_i^k *is such that* $\mu^\sigma(\mathcal{C}_i^k) < \kappa$*, then for any* $X \subseteq \mathcal{X}_j \setminus (\mathcal{C}_i^k \cup \mathcal{X}_j^{sing})$*, 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* $\cup_{j=1}^N X^j = \cup_{j=1}^M \mathcal{C}_i^{k_j}$*,* $\mu^\sigma(X^j) > \kappa$ *for all* j*, and* $\min_{j=1}^N \zeta_i(X^j) > \min_{j=1}^M \zeta_i(\mathcal{C}_i^{k_j})$*.*

Roughly, Definition 4 can be interpreted as follows. The threshold parameter κ relates to the amount of data that the players consider necessary in order to obtain within-category estimates that are sufficiently reliable, in line with the bias-variance trade-off.[15] If a single situation is encountered with a frequency that exceeds the threshold κ the situation is placed in a singleton category (this is condition 1). By contrast, situations that are encountered with frequency that does not exceed κ, are bundled together so as to create categories that meet the threshold condition κ and maximize within-category homogeneity whenever possible. Different formalizations could be proposed to model the latter requirement, but conditions 2-4 provide a simple version of this desideratum. Specifically, the second condition requires that when a set of situations is considered to induce maximal heterogeneity, these situations cannot be bundled together into one analogy class, which seems natural given that the bias-variance trade-off would require that too dissimilar situations should not be bundled together.[16] 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 in agreement with the second condition. The fourth condition is a kind of local optimality requirement formulated in terms of the the minimum of homogeneities over the various analogy classes. It aims at capturing the desire of players to increase within-category similarities when the minimum mass criterion can be achieved.[17]

[15]Note that we could have, in principle, considered a different threshold parameter κ_i for each player i, but our applications will not make use of such an asymmetry.

[16]As an elaboration, 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 an extra 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.

[17]With no impact on the analysis, one could have generalized condition 4 to require that

The reduced-form properties in Definition 4 can be related to optimality properties obtained in simple prediction problems, as considered by Mohlin (2014).[18] 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 generated by a continuous and bounded joint probability density function f, such that $Y = m(x) + \varepsilon(x)$ where $m(x)$ denotes the 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 in Mohlin's approach, 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.

The comparative static results for the optimal categorizations derived in Mohlin (2014) have analogs in the conditions of Definition 4, relating the Euclidean distance in Mohlin's setting to the similarity function in our setting. Indeed, 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).

Several notable differences between our setting and the one studied in Mohlin (2014) are worth mentioning. In our approach, the homogeneity function used by an agent is subjective and viewed as a primitive (remember our view that homogeneity is determined at a broader level, not at the level of the interaction under study). This is

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.

[18] More general results can be found in Mohlin (2018).

[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.

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. 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.

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*.*

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 on situations that are observed without trembles are abundant enough to allow a fine-grained categorization. This is captured by the assumption that κ is small ($\kappa \to 0$). 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 between (a) cases in which κ and ε have the same order of magnitude leading to a notion of ρ-coarse categorization equilibrium and (b) cases in which ε is much smaller than κ, leading to the notion of coarse categorization equilibrium. Formally,

[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.

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* $\lim_{m\to\infty} \kappa^m/\varepsilon_i^m = \rho_i$ *for* $i = 1, 2$*, 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*.*

Part 1 of Definition 4 implies that expectations about opponent's behavior in situations that are observed without tremble are correct in a categorization equilibrium. This is analogous to the requirement in self-confirming equilibrium (Fudenberg and Levine, 1993) developed for extensive-form games (see Section 5.2 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 not too small. 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, in order to obtain new predictions as compared to the standard ones.

2.4 Dynamics

An (ε, κ)- categorization equilibrium can be understood as a steady state of a dynamic process of learning. Suppose the process has settled on $(\sigma, \mathcal{C})$. 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.[21]

When a steady state does not exist or when we want to study the stability of a steady state we use the following learning dynamic.[22] In period τ agents form a profile of analogy partitions $\mathcal{C}(\tau) = (\mathcal{C}_i(\tau), \mathcal{C}_j(\tau))$ that is κ-adjusted to behavior in the

[21] 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.

[22] While many variants could be considered (in particular related to the consideration of older cohorts when forming categories and expectations), we have chosen this one for its simplicity.

preceding period, denoted $\sigma^{\tau-1}$.[23] Expectations for period τ are based on $\sigma^{\tau-1}$ filtered through $\mathcal{C}(\tau)$. That is, the expectation in period τ about a situation assigned to $\mathcal{C}_i(\tau)$ is identified with the aggregate distribution observed in $\mathcal{C}_i(\tau)$ given the behaviors $\sigma^{\tau-1}$ observed in period $\tau-1$. These expectations induce behavior σ^{τ} in period τ (assuming that players best respond to their expectations when they do not tremble). At $\tau+1$, agents form a new profile of analogy partitions $\mathcal{C}(\tau+1) = (\mathcal{C}_i(\tau+1), \mathcal{C}_j(\tau+1))$ which is κ-adjusted to σ^{τ}. Expectations for period $\tau+1$ are based on σ^{τ} filtered through $\mathcal{C}(\tau+1)$, and so on. The dynamics is parameterized by the initial choice of strategies in period 1 and the tie-breaking rule in case of multiple best responses and/or multiple κ-adjusted partitions.

3 Chainstore Game

3.1 Set-Up

3.1.1 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 enters then the monopolist chooses whether to Accommodate (A) or Fight (F). The stage game payoffs of the monopolist and a generic 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, which can be verified by 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.

[23] We do not explicitly describe how players arrive at these categorizations, but we believe that it could be thought of as the result of a trial and error process.

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.[24] We also assume that the trembling probability is the same for the monopolist and the challenger.

3.1.2 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 is natural 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 (as the former but not the latter suggests a form of weakness on the monopolist's side). 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 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 $\mathcal{Q}_C$:

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

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 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

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

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 the challenger can be thought of as a collection of different challengers, one for each of the calendar times t, so that the period t-challenger would naturally focus on histories corresponding to t. We will later discuss what happens when histories with different calendar times are allowed to be bundled together also on the challenger's side.

3.2 Categorization Equilibrium

We will focus on coarse categorization equilibria and discuss later what (ρ_M, ρ_C)-coarse categorization equilibria look like for finite ρ_M and ρ_C.

3.2.1 Strategy profile

We define the threshold

$$k^* = \min \{k \in \mathbb{N} \text{ such that } u_M(E,F) + k u_M(O) \geq (k+1)\, u_M(E,A)\}. \qquad (2)$$

Suppose that we are in the generic case where the above inequality holds strictly for $k = k^*$. In this case we consider the following strategy profile σ_T:[25]

- 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.

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).

[25] When the condition in 2 holds with equality for $k = k^*$ we need to redefine the strategy profile so that entry and accommodation begins already in period $T - k^*$.

3.2.2 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 as well as the distinction between $\mathcal{Q}^{Tough}$ and $\mathcal{Q}^{Soft}$.[26] Formally, let $\mathcal{Q}_M^{off}$ be the set of off-path monopolist decision nodes. For each t let

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

We have:

Proposition 1 *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

[26] 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.

particular, if E occurs in period $t = T-k^*$ (i.e. the last period in which the challenger is supposed to stay out) then the monopolist mistakenly expects that by playing F, the challenger(s) will be induced to stay out (with a probability roughly equal to $\frac{T-k^*-1}{T-k^*+1}$) from then on, whereas in reality, no matter what the monopolist does there will be entry in all remaining periods.[27] 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 all other nodes in $\mathcal{Q}_C^{Tough}$, at which it is indeed the case that fighting after entry leads the challenger not to enter in the next period**.**

We note that the same strategy profile as the one considered in Proposition 1 could be used to support a (ρ_M, ρ_C)-coarse categorization equilibrium with $\rho_M > N$, as long as N and T are large enough. This means that our construction only requires that κ be sufficiently (but not necessarily infinitely) large relative to ε.[28] By contrast, when κ is sufficiently small, we recover the usual Subgame Perfect Nash Equilibrium prediction with entry and no fight in every period, thereby illustrating in this application how increasing the κ threshold may enlarge the set of steady state behaviors.

3.3 Discussion

3.3.1 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 for the challenger. 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 dissimi-

[27]The reason for the expression for the probability is as follows. As mistakes become vanishingly rare, almost all data on off-path nodes come from nodes that are reached by a single mistake. The category $\mathcal{C}_M^1$ can be reached in a single mistake, either by mistaken E in period $t \leq T-k^*$, or following a mistaken F after non-mistaken E in period $T-k^*+1$. When it is the challenger's turn to act following a single mistaken E in period $t < T-k^*$ the equilibrium strategy requires the challenger to play O. However, if there is a mistaken E in period $t = T-k^*$ then the next time it is the challenger's turn to act it is period $T-k^*+1$ and according to the equilibrium strategy the challenger should then play E. The same is true after non-mistaken E in period $T-k^*+1$.

[28]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 would give the same incentive to play the equilibrium as in the coarse categorization equilibrium considered in Proposition 1.

lar 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 dramatically affect the monopolist's behavior. In the Online Appendix S.3, we explore this alternative in detail. We demonstrate the existence of a coarse categorization equilibrium such that 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.[29] In this alternative scenario, it is the challenger and not the monopolist who has mistaken expectations in contrast with the scenario described in Proposition 1. This difference illustrates the effect the homogeneity functions may have on the obtained categorization equilibria.

3.3.2 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 5.2 that a coarse categorization is a self-confirming equilibrium.

In the Online Appendix S.3, we study multi-stage contribution games in which in each stage, assumed to be finitely many, agents decide whether or not to contribute to a public good. We develop the analysis of coarse categorization equilibria in two variants, depending on whether or not the agents can punish their peers after observing their contributions at the end of each stage. In the latter case we assume (similarly as in the chainstore game) that histories in which someone fails to contribute but is not punished are very dissimilar from other histories. We observe that positive contributions can be supported in coarse categorization equilibria when there is a punishment option but not otherwise, in agreement with the qualitative findings reported in the experiment of Fehr and Gächter (2000).

[29] Such strategies cannot arise in a coarse categorization equilibrium with the previously considered homogeneity setting, as challengers would then find it best to enter in the last k^* periods.

4 On Cycling in Adverse Selection Games

4.1 Set-Up

4.1.1 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 mechanism between pairs consisting of one seller and one buyer drawn at random from their respective pools. In each pair, the seller (he) and the buyer (she) 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 to be 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. We restrict attention to pure strategies on the buyer side as well.

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, $\frac{d}{dp}\left(\frac{g(p)}{G(p)}\right) < 0$. Moreover, we assume the following smoothness condition: $|g'(p)| < g(p)$ for all p.[30]

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

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

[30] 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.

It is readily verified (see Online Appendix) that under our assumptions there exists a unique Nash equilibrium in which the bid price p^{NE} of the buyer is uniquely defined by $\frac{g\left(p^{NE}\right)}{G(p^{NE})} = \frac{1}{b}$.[31]

We model prices and qualities on a continuum for analytical convenience. We intend this as an approximation of what is in reality a finite set of nearby qualities and prices. This will motivate our treatment of categorization of qualities below. (See Jehiel and Mohlin (2021) for a fuller discussion.)

4.1.2 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) appears in the feedback only when there is trade, i.e. when $a \leq p$. 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. 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 chooses bids according to a pdf f and cdf F with full support on $[0, 1]$.[32] The seller always chooses his weakly dominant strategy.

Similarity and Homogeneity. Given that payoffs depend continuously on quality 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 qualities in the considered set. For concreteness we let $\xi(C)$ be equal to the difference between

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

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

the supremum and infimum ω among the qualities in 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 consecutive intervals.

Threshold Mass. In line with our general assumptions, we have in mind that for on-path qualities, 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. As mentioned above we think of this assumption as a limit case in which the continuum is viewed as an approximation of the fine grid case.

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 $\omega \leq p^*$. 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

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

In line with part 3 of Definition 4, we will require that any $\mathcal{C}^k$ corresponds to a mass no less than κ if possible (or else if κ is too large, the entire $(p^*, 1]$ will be one analogy class).

4.1.3 Preliminary Analysis

Mass of Observations. The density of transactions of quality ω conditional on trembling is $\tilde{g}(\omega) := g(\omega)(1 - F(\omega))$, and the density of transacted quality ω 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, each type $\omega \leq p^*$ is put in a singleton analogy class, because it is on path and thus disclosed, even in the absence of trembles. For types above p^* the number of categories depends on κ and ε in a more complex way. Each analogy class above p^* should satisfy $\kappa \leq \int_{c_{k-1}}^{c_k} \mu_{p^*}^{\sigma,\varepsilon}(\omega)(s)\,ds$ if possible. 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 $\omega > p$ and one conditional on $p \geq \omega$, i.e

$$\Pr\widehat{(a \leq p|\omega)} = \Pr(a \leq p|\omega) = \Pr(\omega \leq p|\omega) = \mathbb{I}_{\{\omega \leq p\}}. \tag{3}$$

For a quality $\omega > p^*$ the buyer forms a prediction about the distribution of ask prices 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

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

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

$$\Pr\left(a \leq \widehat{p|\omega} \in \mathcal{C}^k\right) = \int_{a=0}^{p} h_{\mathcal{C}^k}(a)\, da = \begin{cases} 1 & \text{if } c_k < p; \\ \frac{\tilde{G}(p)-\tilde{G}(c_{k-1})}{\tilde{G}(c_k)-\tilde{G}(c_{k-1})} & \text{if } c_{k-1} < p \leq c_k; \\ 0 & \text{if } p < c_{k-1}. \end{cases} \tag{4}$$

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*

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

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

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

4.2 Results

4.2.1 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^*_τ 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 $p^*_{\tau+1}$ varies with p^*_τ depending on whether p^*_τ is below, above, or equal to p^{NE}. These properties are referred to as lemmata and are proven in the Appendix.

Lemma 2 *If* $p^*_\tau = p^{NE}$ *then* $p^*_{\tau+1} > p^{NE}$.

Lemma 3 *If* $p^*_\tau > p^{NE}$, *then either* $p^*_{\tau+1} = p^{NE}$ *or* $p^*_{\tau+1} > p^*_\tau$.

Lemma 4 *If* $p^*_\tau < p^{NE}$, *then* $p^*_{\tau+1} > p^*_\tau$.

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^*. This is the content of lemmata 3.

The above properties immediately imply that the price dynamics has no rest point, i.e., there is no p^*_τ such that $p^*_{\tau+1} = \arg\max_{p\in[0,1]} \pi^{CE}(p|p^*_\tau) = p^*_\tau$. To see this, assume by contradiction that p^* is a rest point. By Lemma 4, it cannot be that $p^* < p^{NE}$ since $p^*_\tau = p^* < p^{NE}$ would imply that $p^*_{\tau+1} > p^*_\tau = p^*$. By Lemma 2, it cannot be that $p^* = p^{NE}$ since $p^*_\tau = p^*$ would imply that $p^*_{\tau+1} > p^{NE}$. Finally, by Lemma 3, it cannot be that $p^* > p^{NE}$ since $p^*_\tau = p^*$ would imply either that $p^*_{\tau+1} > p^*_\tau$ or that $p^*_{\tau+1} = p^{NE}$ and thus $p^*_{\tau+1} \neq p^*_\tau$ (given that $p^*_\tau = p^* \neq p^{NE}$). Even though there is no rest point, we can establish that there is a price cycle that consists of the Nash price and one or more prices above the Nash price.[33]

[33]Note that the exact cycle will depend on, among other things, the threshold κ. When κ converges to 0 the interval $\left[p^{NE}, 1\right]$ will be will become ever more finely categorised and the cycle will be close to the usual Nash prediction.

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

The result obtained here should be put in the perspective of the literature which has 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.[34] As already mentioned, 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 counterpart in the other approaches).[35,36]

5 Discussion

5.1 On the existence of categorization equilibria

When player are required to use a single analogy partition, the existence of a (ε, κ)-categorization equilibrium is not guaranteed, even in finite environments. To see this consider the following two-stage game. First, Player 1 chooses an action $a_1 \in \{A, B, C\}$, and then, upon observing player 1's choice, player 2 chooses an action

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

[35] 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.

[36] 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.

$a_2 \in \{L, M, R\}$. The payoffs of player 1 as a function of the profile of actions are.

	L	M	R
A	2	2	2
B	4	1	4
C	4	4	1

The payoffs of Player 2 are such that it is strictly dominant for player 2 to choose L after A, M after B, and R after C. Suppose that the homogeneity function does not rule out any bundling of A, B, and C. We assume that only Player 1 trembles, and does so uniformly with probability ε.[37] Moreover, we assume that κ is large relative to ε so that if Player 1 plays a pure action then the nodes following the two remaining actions have to be bundled together (as in the Coarse Categorization Equilibrium). Under these circumstances the game does not have any (ε, κ)-categorization equilibrium for small ε, or more precisely:

Claim 1 *There is no pure (ε, κ)-categorization equilibrium in which $\frac{1}{3} > \kappa > 2\varepsilon$.*

A proof of the claim appear in the Appendix, but we now sketch the main steps. If the strategy of player 1 were pure, then the nodes following the other two actions would have to be bundled into one analogy class (because $\kappa > 2\varepsilon$), thereby leading to the contradiction that player 1 would find one of the non-played actions preferable.[38] Having established that player 1 should be mixing, it follows that player 1 cannot be using the fine analogy partition (as A would then be the pure strategy chosen by player 1). Player 1 also cannot use the coarse analogy partition that puts all decision nodes of player 2 into a single analogy class, as it must be that at least one of player 1's actions is played with probability no less than κ (given that $\kappa < \frac{1}{3}$). The remaining three two-class analogy partitions can then be ruled out using the property that player 1 must perceive that at least two actions are equally good, as required for mixing.

The inexistence of a pure (ε, κ)-categorization equilibrium suggests allowing for mixed partitions. In the context of the example above, this would require Player

[37] Player 2 does not tremble similarly to our assumption that the buyer does not tremble in the adverse selection game (because this player has a weakly dominant strategy).

[38] For example, if A were the pure choice, then both B and C would be perceived to yield $\frac{4+1}{2}$ which is more than 2, what A would be (rightly) perceived to yield.

1 to mix between two partitions that would both be κ adjusted to the strategies. Specically, let Player 1 mix between the two analogy partitions, (A, B, C), which is the the maximally fine partition, and (A, BC), which bundles B and C separately from A. Under the former partition, action A is perceived as optimal and is thus played with probability one unless the player trembles. Under the latter partition, Player 1 plays B and C with probability 0.5 each, unless the player trembles. This induces Player 1 to expect M and R to be played with equal probability (each close to 0.5) following either B or C and hence Player 1 is indifferent between B and C. Moreover, since the probability of L is close to zero following either B or C, Player 1 strictly prefers B and C to A. Hence, it is a best-response to mix between B and C when using partition (A, BC).

In order for (A, B, C) to be adjusted all actions need to occur with at least probability κ. In order for (A, BC) to be adjusted, actions B and C need to occur with at most probability κ each. Thus, actions B and C need to occur with exactly probability κ each. Suppose partition (A, B, C) is played with probability p and partition (A, BC) is played with probability $1 - p$. The probability that action A is played is $p(1 - 2\varepsilon) + (1 - p)\varepsilon$. The probability that action B is played is $p\varepsilon + (1 - p)\left(\frac{1}{2} - \varepsilon\right)$. This is also the probability that action C is played. Hence, for given κ and ε we need to set p such that $p\varepsilon + (1 - p)\left(\frac{1}{2} - \varepsilon\right) = \kappa$. Note that this requires that p vanishes to 0 as ε as κ tends to 0.

The insight from this example is general. If we allow for mixed analogy partitions then we can ensure the existence of (ε, κ)-categorization equilibrium for any (ε, κ), at least in finite environments. A proof of this, which is somewhat similar to the existence proof of Jehiel and Weber (2024), can be found in the Online Appendix. From the existence of such (ε, κ)- categorization equilibria with mixed partitions, one could consider the accumulation points of sequences of such equilibria as $\varepsilon, \kappa \to 0$ and $\frac{\varepsilon}{\kappa} \to 0$ (resp. $\frac{\varepsilon}{\kappa} \to \rho$) to parallel the definitions of coarse categorization equilibria (resp. ρ-coarse categorization equilibria) for this mixed analogy partition extension. In the context of the adverse selection game studied in Section 4, while such equilibria could arise, we suspect they would be unstable with respect to the learning dynamics studied there and that instead only the cycles analyzed in Proposition 2 would emerge for generic choices of initial beliefs in the dynamic process. The precise study of such stability considerations is left for future research.

5.2 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 3 *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^{SPNE} is a subgame perfect Nash equilibrium (SPNE) then there may be no coarse categorization equilibrium $(\sigma', \mathcal{C}')$ that supports a strategy profile that is outcome equivalent to σ^{SPNE}.

Part (a) of Proposition 3 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.[39] However part (c) tells us that this is not true for coarse categorization equilibrium: there are SPNE that cannot be supported

[39]The fact that the homogeneity function does not matter in part (a) is simply a consequence of the on-path nodes being distinguished perfectly, so that homogeneity is maximal in each singleton on-path analogy class. In part (b) the irrelevance 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.

as a coarse categorization equilibrium. The reason is that in a coarse categorization equilibrium one is *not* free to 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 $(\varepsilon^m, \kappa^m)$-categorization equilibrium, whereas nodes that are further away from the equilibrium path may be bundled more coarsely.

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.

In some cases our notion of categorization equilibrium seems to impose reasonable restrictions on off-path beliefs, allowing us to rule out implausible beliefs allowed by self-confirming equilibrium. Consider an extensive-form game with a pure strategy SPNE such that for each player behaviour is the same at all off-path nodes. For any value of κ the SPNE can be implemented in a categorization equilibrium, since any bundling of off-path nodes will induce correct expectations. This is not so in self-confirming equilibrium, as the beliefs off-path can be arbitrary and unrelated to the actual behaviors arising at those nodes.

6 Conclusion

Our paper has proposed a novel perspective on categorization, using the the bias-variance trade-off to think about how analogy partitions should be chosen in the analogy-based expectation equilibrium. In this construction, we have assumed that players are endowed with a pre-conceived perception about the similarity of the different situations in which the opponent is supposed to play. We have motivated this on the ground that sociological and psychological factors may be driving these perceptions. A further step of endogenization would require making progress on how the similarity perceptions should be determined.

Appendix

A.1 Chainstore Application

Proof of Proposition 1. 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 (ε^m, κ^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 (ε^m, κ^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 (ε_T^m, κ_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.

■

A.2 Adverse Selection Application

A.2.1 Deriving Perceived Expected Payoff

Proof of Lemma 1. The perceived expected payoff is

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

where, using (3) we obtain

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

and, writing $i(p)$ for the analogy class that contains $\omega = p$, using (4) we obtain,

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

■

A.2.2 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 piece-wise 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.4.1) 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.4.1) 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.4.1). 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.2.3 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 $\pi^{CE}(p|p_\tau^*)$ coincides with $\pi^{NE}(p)$ on $[0, p_\tau^*]$, the constrained optimal $p \in [0, p_\tau^*]$ is at $p = p^{NE} < p_\tau^*$. Suppose that $\arg\max_{p \in [p_\tau^*, 1]} \pi^{CE}(p|p_\tau^*) = p_\tau^*$ (requiring $\left.\frac{\partial \pi^{CE}(p|p_\tau^*)}{\partial p}\right|_{p \downarrow p_\tau^*} \leq 0$). By continuity of $\pi^{CE}(p|p_\tau^*)$, we have $\arg\max_{p \in [0,1]} \pi^{CE}(p|p_\tau^*) = p^{NE} < p_\tau^*$. ■

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

using (A3),

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

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

■

A.2.4 Proof of Convergence to Cycle in Proposition 2

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

Proof of Lemma A1. We only sketch the proof here. For details see Online Appendix S.4. Assume $p^*\leq p^{NE}$. The mass in each analogy class (above p^*) is at least κ. Let $g^{\min}=\min_{\omega\in[0,1]}g\left(\omega\right)$ and $g^{\max}=\max_{\omega\in[0,1]}g\left(\omega\right)$. 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 $\mathcal{C}^1$,

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

■

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

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

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

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 $\pi^{CE}(p|p^*_\tau)$ wrt to p. From equation (A4) we have

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

Note that

$$p^*_{\tau+1} \geq \min\left\{p\in[p^*_\tau,1] : \frac{\partial \pi^{CE}(p|p^*_\tau)}{\partial p}\leq 0\right\} \qquad \text{(A7)}$$

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

Proof of Proposition 2. Assume, to derive a contradiction, that the sequence p^*_τ is monotonic. Lemmata 2–4 imply that $p^*_{\tau+1} > p^*_\tau$ for all τ. Since $p^*_\tau \leq 1$ for all τ, it follows that $p^*_\tau \to \bar{p}$ for some $\bar{p} > p^{NE}$ as $\tau\to\infty$. (To see that there is a $\bar{p} > p^{NE}$ note that if $p^*_1 \geq p^{NE}$ then $p^*_\tau \geq p^{NE}$ for all τ.) This implies $\left|p^*_{\tau+1}-p^*_\tau\right|\to 0$, which, by continuity of $\pi^{CE}(p|p^*_\tau)$, implies $\left|\pi^{CE}\left(p^*_{\tau+1}|p^*_\tau\right)-\pi^{CE}(p^*_\tau|p^*_\tau)\right|\to 0$. Since $\pi^{CE}(p|p^*_\tau)=\pi^{NE}(p)$ for $p\in[0,p^*_\tau]$, we have $\left|\pi^{CE}\left(p^*_{\tau+1}|p^*_\tau\right)-\pi^{NE}(p^*_\tau)\right|\to 0$, and consequently $\pi^{CE}\left(p^*_{\tau+1}|p^*_\tau\right)\to\pi^{NE}(\bar{p})$. Since the Nash equilibrium p^{NE} is unique it holds that $\pi^{NE}\left(p^{NE}\right)>\pi^{NE}(\bar{p})$, and since $\pi^{CE}(p|p^*_\tau)=\pi^{NE}(p)$ for $p\in[0,p^*_\tau]$ we get

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

This is in contradiction to $p^*_{\tau+1}=\arg\max_{p\in[0,1]}\pi^{CE}(p|p^*_\tau)$. We conclude that the sequence p^*_τ 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 $i\in\{1,...,\bar{\tau}\}$), from which there is no convergence to the cycle. This means that $p^*_{\tau+1}>p^*_\tau$ for all τ and $p^*_\tau\to\bar{p}$ for some $\bar{p}\in\left[p^{NE},p^{(\bar{\tau})}\right]$ 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 by Lemma A2. ■

A.3 Proof of Non-Existence for Example with Pure Partitions

Proof of Claim 1. First we consider pure strategies.

(A) If A is played, then the nodes following B and C are put together in the same analogy class BC. The belief for BC is $\beta_1(M) = \beta_1(R) = 0.5$, implying that B is viewed as better than A since $(4+1)/2 > 2$.

(B) If B is played, then the nodes following A and C are put together in the same analogy class AC. The belief for AC is $\beta_1(L) = \beta_1(R) = 0.5$, implying that A is viewed as better than B since $(2+2)/2 > 1$

(C) If C is played, then the nodes following A and B are put together in the same analogy class AB. The belief for AB is $\beta_1(L) = \beta_1(M) = 0.5$, implying that B is viewed as better than C since $(4+1)/2 > 1$.

Next we consider mixed strategies. We consider each of the different possible analogy partitions.

1. Suppose Player 1 uses the analogy partition (A, B, C) that places each of Player 2's nodes in a separate class. She will then have correct expectations so that her unique best response is action A. But then the weight put on either of B and C will not be high enough to make the analogy partition (A, B, C) adjusted, since by assumption $\kappa > \varepsilon$.

2. Suppose Player 1 uses the analogy partition (A, BC). In order for this to be adjusted, Player 1 needs to play each of B and C with probability less than κ. In order for Player 1 to be indifferent between A and B we need $2 = \beta_1(M|BC) + 4\beta_1(R|BC)$. Since $\beta_1(L|BC) = 0$ it must be that $\beta_1(M|BC) = 2/3$.[1] The analogy partition (A, BC) is adjusted provided that $\kappa > 2\varepsilon$. The resulting analogy based expectations satisfy $\beta_1(M|BC) = 2/3$, $\beta_1(R|BC) = 1/3$. However, if $\beta_1(M|BC) = 2/3$ then the perceived expected payoff from C is $4 \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} = \frac{9}{3} > 2$ so that action C is perceived to be optimal. Similarly in

[1] We can achieve this as follows. With probability 3ε Player 1 trembles and takes one of her actions, each with probaiblity ε. When she does not tremble she puts probability $1 - \frac{\varepsilon}{(1-3\varepsilon)}$ on A, probability $\frac{\varepsilon}{(1-3\varepsilon)}$ on B, and probability 0 on C. Thus, total mass put on B (adding trembles and non-trembles) is 2ε, and likewise the total mass put on C is ε.

order for Player 1 to be indifferent between A and C we need $\beta_1(M|BC) = 1/3$ but this makes action B to be considered optimal.

3. Suppose Player 1 uses the analogy partition (AB, C). In order for this to be adjusted, Player 1 needs to play each of A and B with probability less than κ. In order for her to be indifferent between C and A we need $1 = 2\beta_1(L|AB) + 2\beta_1(M|AB)$, which is impossible. In order for her to be indifferent between C and B we need $1 = 4\beta_1(L|AB) + 1\beta_1(M|AB)$, which is also impossible.

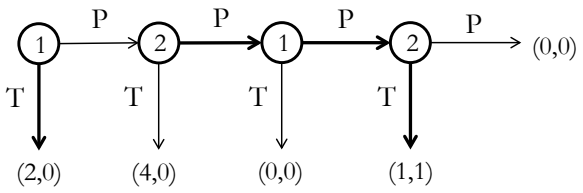
4. Suppose Player 1 uses the analogy partition (AC, B). This case is parallel to the previous one.

5. The analogy partition that bundles all actions is not adjusted since at least one action receives a weight of at least κ.

■

A.4 Relation to Other Solution Concepts

Proof of Proposition 3. (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. ■



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.

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

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 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.

Fudenberg, Drew and David K Levine (1998), *The theory of learning in games*. MIT press, Cambridge, MA.

Fudenberg, Drew and David K Levine (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*. MIT Press, Cambridge, MA.

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.

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 (2024), "Endogenous clustering and analogy-based expectation equilibrium."

Laurence, S. and E. Margolis (1999), "Concepts and cognitive science." In *Concepts: Core Readings* (E. Margolis and S. Laurence, eds.), 3–81, MIT Press, Cambridge, MA.

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.

Mohlin, Erik (2018), "Asymptotically optimal regression trees." Technical report, Lund University, Department of Economics, Working Papers; No. 2018:12.

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

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

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.

Selten, Reinhard (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.

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 Illustration: Ultimatum/Bargaining Game Application

To provide a first simple illustration of how our approach works, we consider the following ultimatum-like environment. 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 $s_P \in [0,1]$ that offers $1-s_P$ to the responder, 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. The proposer has to predict the acceptance probability for different offers. Thus, the set of offers can be identified with the set of situations in the above abstract formulation. When predicting acceptance as a function of offers the proposer may bundle several offers together.

When assessing how the responder's acceptance probability depends on the offer s_P, it seems plausible that the the proposer would subjectively assess 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 (with finite sets of actions and situations) 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 $\varepsilon/K^{off} \geq \kappa$. 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 informally suggested above.[1] We also characterize the coarse categorization equilibrium (that can be viewed as a ρ-coarse categorization equilibrium 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 Appendix.

Proposition S1 *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 $s_P^* \in [v, 2v]$.*

Proof of Proposition S1. Suppose that $\kappa^m > \varepsilon^m/2$ as $m \to \infty$ (which must hold in a coarse categorization equilibrium). This 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

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

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 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.[2,]

S.2 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 S2 *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*

[2]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 S1.

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.[3]

Proof of Proposition S2. The proof is similar to that of proposition 1. 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 1.

   (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 1.

   (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

[3]It should be noted that $\tilde{\sigma}_T$ cannot part of a categorization equilibrium when using the homogeneity assumptions of Proposition 1, 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.

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-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 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 1 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 1). The 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 1.

   (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.3 Public Goods Game

S.3.1 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.3.2 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 S3 *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.3.3 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 S2 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

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

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 S4 *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 S4. We need to show that for $T > T^*$ there is a sequence $\left(\hat{\sigma}_T^m\right)_m$ converging to $\hat{\sigma}_T$, such that $(\hat{\sigma}_T^m, \widehat{\mathcal{C}})$ is an (ε^m, κ^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 (ε^m, κ^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 (ε_T^m, κ_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[4]

   (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.

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

3. Verify that $\hat{\sigma}_T^m$ induces a ε_T^m-best-response 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 $t \leq T$.

      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.4 Adverse Selection Application

S.4.1 Preliminary Observations

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

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

We note that

$$\begin{aligned}\frac{\left(\tilde{G}\left(c_1\right)-\tilde{G}\left(p_{\tau}^{*}\right)\right)}{\tilde{G}\left(p_{\tau}^{*}\right)}&=\frac{\int_{p_{\tau}^{*}}^{c_1}\left(1-F_{p^{*}}\left(\omega\right)\right)g\left(\omega\right)d\omega}{\left(\int_{0}^{p_{\tau}^{*}}\left(1-F_{p^{*}}\left(\omega\right)\right)g\left(\omega\right)d\omega\right)}<\frac{\left(1-F_{p^{*}}\left(p_{\tau}^{*}\right)\right)\int_{p_{\tau}^{*}}^{c_1}g\left(\omega\right)d\omega}{\left(1-F_{p^{*}}\left(p_{\tau}^{*}\right)\right)\int_{0}^{p_{\tau}^{*}}g\left(\omega\right)d\omega}\\&=\frac{\int_{p_{\tau}^{*}}^{c_1}g\left(\omega\right)d\omega}{\int_{0}^{p_{\tau}^{*}}g\left(\omega\right)d\omega}=\frac{G\left(c_1\right)-G\left(p_{\tau}^{*}\right)}{G\left(p_{\tau}^{*}\right)}.\end{aligned}$$

Moreover, we note that

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

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

$$\begin{aligned}\tilde{g}(p^*) \frac{G(c_1) - G(p^*)}{\tilde{G}(c_1) - \tilde{G}(p^*)} &= \frac{(1 - F_{p^*}(p^*))(G(c_1) - G(p^*))}{\int_{p^*}^{c_1} g(\omega)(1 - F_{p^*}(\omega))\, d\omega} g(p^*) \\ &= \frac{(1 - F_{p^*}(p^*)) \int_{p^*}^{c_1} g(\omega)\, d\omega}{\int_{p^*}^{c_1} (1 - F_{p^*}(\omega))\, g(\omega)\, d\omega} g(p^*) > g(p^*).\end{aligned}$$

Finally, to demonstrate (A4) we note that for $p \in (p_\tau^*, c_1)$

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

Using (S.4.1) we obtain (A4)

S.4.2 Nash equilibrium

Proposition S5 *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 S5. Note that

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

Thus

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

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.4.3 Extended Proof of Lemma for Convergence to Cycle

Proof of Lemma A1. Assume $p^*\leq p^{NE}$. 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

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

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

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

implying that

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

Note that

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

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

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

■

S.5 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 5.2.

S.5.1 Example where Feedback Differs from the Path of Play

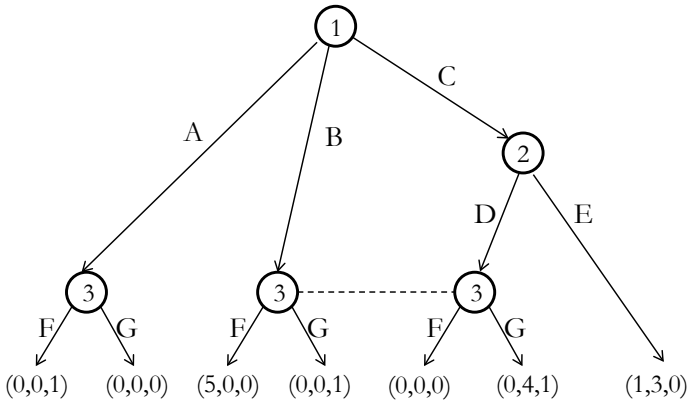
Consider the following game. Player 1 (row) and Player 2 (column) simultaneously choose between actions A and B, with the following outcomes.

	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.5.2 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.

S.6 Proof of Existence

Fix positive $\varepsilon = (\varepsilon_1, \varepsilon_2)$ and κ. Let $\mathcal{P}_i$ denote the set of partitions of $\mathcal{X}_i$ and let Π_i denote the set of mixed partitions of player i (i.e. the set of probability distributions on $\mathcal{P}_i$). Let $\mathcal{P} = \mathcal{P}_1 \times \mathcal{P}_2$ denote the set of pure partition profiles and let $\Pi = \Pi_1 \times \Pi_2$ denote the set of mixed partition profiles. A mixed partition of player i is denoted $\pi_i \in \Pi_i$ and a mixed partition profile is denoted $\pi \in \Pi$. As before, a pure partition profile is denoted $\mathcal{C}$. Let Σ_i denote the mixed strategy set of player i and let $\Sigma = \Sigma_1 \times \Sigma_2$ denote the set of mixed strategy profiles with a typical element denoted σ. A function from the set of pure partition profiles will be called a *partition-dependent strategy*.[5] Let $\Lambda := \Sigma^{|\mathcal{P}|}$ denote the set of partition-dependent strategy profiles, with

[5]This construction is similar to that of Jehiel and Weber (2024).

a typical element denoted λ.

Let $\lambda(\mathcal{C})$ denote the strategy profile induced by the partition-dependent strategy profile λ under the partition profile $\mathcal{C}$. A mixed partition profile π and a partition-dependent strategy profile λ induces behavior, in the form of a mixed strategy, $\mu(\lambda, \pi) := \sum_{\mathcal{C}' \in \mathcal{P}} \pi(\mathcal{C}') \lambda(\mathcal{C})$, thereby defining a function

$$\mu : \Lambda \times \Pi \to \Sigma,$$

which is continous in both arguments.

In line with Definition 2 let

$$\beta_i : \Sigma \times \mathcal{P}_i \to \Sigma,$$

denote the *analogy-based expectations function* which for a player i and any pure partition $\mathcal{C}_i$ assigns the analogy based expectation $\beta_i(\sigma, \mathcal{C}_i)$. This function is well-defined since ε and κ are kept positive throughout. Moreover, it is continuous in both arguments. In line with Definition 1 let

$$\xi_i : \Sigma \twoheadrightarrow \Sigma_i$$

be an *ε-perturbed best-response strategy scorrespondence for player i*, so that $\xi_i(\sigma)$ is the set of player i's ε-perturbed best-responses to σ_{-i}. We use β and ξ to define an *ε-best-response partition-dependent strategy correspondence*

$$\varphi : \Sigma \twoheadrightarrow \Lambda,$$

which for each player i and each mixed strategy profile σ assigns a partition-dependent strategy as follow

$$\varphi_i(\sigma) := \{\xi_i(\beta_i(\sigma, \mathcal{C}_i))\}_{\mathcal{C}_i \in \mathcal{P}_i}.$$

In line with Definition 4 we define a *κ-adjusted-pure-partition correspondence*

$$\psi^{\mathcal{P}} : \Sigma \twoheadrightarrow \mathcal{P},$$

which for each profile of mixed strategies σ assigns the set of profiles of κ-adjusted pure partitions $\psi^{\mathcal{P}}(\sigma)$. That is, for each player i the partition $\psi_i^{\mathcal{P}}(\sigma)$ of player i is

adjusted to σ_{-i}. Based on this we define a *κ-adjusted-mixed-partition correspondence*

$$\psi : \Sigma \twoheadrightarrow \Pi,$$

which for each profile of mixed strategies σ assigns the set of profiles of mixed partitions such that each pure partition is κ-adjusted to σ, i.e. $\psi(\sigma) \subseteq \Delta\left(\psi^{\mathcal{P}}(\sigma)\right)$.

We combine all of the above to a *(ε, κ)-best-response-and-adjusted-partition correspondence*

$$\eta : \Lambda \times \Pi \twoheadrightarrow \Lambda \times \Pi,$$

defined by

$$\eta(\lambda, \pi) := \left\{ \left(\tilde{\lambda}, \tilde{\pi}\right) \in \Lambda \times \Pi : \tilde{\pi} \in \psi(\mu(\lambda, \pi)) \wedge \tilde{\lambda} \in \varphi(\mu(\lambda, \pi), \tilde{\pi}) \right\}.$$

The set $\Lambda \times \Pi$ is non-empty, compact and convex. Note that $\psi(\sigma)$ and is non-empty, closed, and convex (being a probability distribution over a positive and finite number of adjusted analogy partitions). Moreover, for standard reasons $\varphi(\sigma, \pi)$ is also non-empty, closed, and convex for all σ. It follow that $\eta(\lambda, \pi)$ is non-empty, closed, and convex, for all (λ, π). It remains to show that $\eta(\lambda, \pi)$ is upper hemi-continuous (u.h.c.) in $(\lambda, \pi) \in \Lambda \times \Pi$. Since $\Lambda \times \Pi$ is compact and η is closed for all (λ, π) this is equivalent to showing that η has the closed graph property, i.e. that

$$graph(\eta) = \left\{ \left((\lambda, \pi), \left(\tilde{\lambda}, \tilde{\pi}\right)\right) \in (\Lambda \times \Pi) \times (\Lambda \times \Pi) : \left(\tilde{\lambda}, \tilde{\pi}\right) \in \eta(\lambda, \pi) \right\}$$

is closed. Consider a sequence $\left((\lambda^t, \pi^t), \left(\tilde{\lambda}^t, \tilde{\pi}^t\right)\right)$ with $\left(\tilde{\lambda}^t, \tilde{\pi}^t\right) \in \eta(\lambda^t, \pi^t)$ for all t. We need to show that if $\left((\lambda^t, \pi^t), \left(\tilde{\lambda}^t, \tilde{\pi}^t\right)\right)$ converges to $\left((\lambda^*, \pi^*), \left(\tilde{\lambda}^*, \tilde{\pi}^*\right)\right)$ then $\left(\tilde{\lambda}^*, \tilde{\pi}^*\right) \in \eta(\lambda^*, \pi^*)$. If ψ and φ are u.h.c. then this is satisfied. It follows from standards arguments that φ is u.h.c. It remains to show that ψ is u.h.c.

Note that the image of ψ_i is a simplex, hence compact and and convex for all σ. Thus, to demonstrate that ψ is u.h.c. we only need to show that ψ has the closed graph property. Consider a sequence $(\sigma^t, \tilde{\pi}^t)$ with $\tilde{\pi}^t \in \psi(\sigma^t)$ for all t. Suppose that $(\sigma^t, \tilde{\pi}^t)$ converges to $(\sigma^*, \tilde{\pi}^*)$. We need to show that $\tilde{\pi}^* \in \psi^{\Pi}(\sigma^*)$. To obtain a contradiction suppose there is some i such that $\tilde{\pi}_i^* \notin \psi_i(\sigma^*)$, meaning that there is some $\mathcal{C}_i^*$ in the support of $\tilde{\pi}_i^*$ which is not κ-adjusted to σ_{-i}^*. This means that at least one of the following four statements must be true. (i) There is $x \in \mathcal{X}_j$

with $\mu^{\sigma^*}(\{x\}) > \kappa$ which is not in a singleton analogy class in $\mathcal{C}_i^*$. (ii) There is an analogy class $\mathcal{C}_i^{*k} = X$ with $\zeta_i(X) = 0$. (iii) There is an analogy class C_i^{*k} and a set $X \subseteq \mathcal{X}_j \setminus (\mathcal{C}_i^{*k} \cup \mathcal{X}_j^{sing})$ such that $\mu^{\sigma^*}(\mathcal{C}_i^{*k}) < \kappa$ and $\zeta_i(\mathcal{C}_i^{*k} \cup X) > 0$. (iv) There is a collection of non-singleton analogy classes $\left\{\mathcal{C}_i^{*k_1}, ..., \mathcal{C}_i^{*k_M}\right\}$ in $\mathcal{C}_i^*$, and a collection $\left\{X^1, ..., X^N\right\}$ of pairwise disjoint sets, such that $\cup_{j=1}^N X^j = \cup_{j=1}^M \mathcal{C}_i^{*k_j}$, $\mu^{\sigma^*}(X^j) > \kappa$ for all j, and $\min_{j=1}^N \zeta_i(X^j) > \min_{j=1}^M \zeta_i(\mathcal{C}_i^{*k_j})$. Now we show that each of these statements imply a contradiction. If (i) holds then there is a neighborhood B of σ^* such that $\mu^{\sigma^t}(\{x\}) > \kappa$ for all $\sigma^t \in B$, and hence it cannot be that σ^* is a limit point of σ^t. If (ii) holds so that $\zeta_i(X) = 0$ then regardless of behavior no κ-adjusted categorization $\mathcal{C}^t$ contains a category $\mathcal{C}_i^{tk} = X$ and hence $\mathcal{C}_i^*$ cannot be in the support of a limit point $\tilde{\pi}_i^*$. If (iii) holds then there is a neighborhood B of σ^* such that $\mu^{\sigma^t}((\mathcal{C}_i^{*k})) < \kappa$ for all $\sigma^t \in B$. Together with the fact $\zeta_i(\mathcal{C}_i^{*k} \cup X) > 0$ this implies that for any $\sigma^t \in B$ no κ-adjusted analogy partition can have $\mathcal{C}_i^{*k}$ as an analogy class. Hence it cannot be that $\mathcal{C}_i^*$ is in the support of a limit point $\tilde{\pi}_i^*$. If (iv) holds then there is a neighborhood B of σ^* such that for all $\sigma^t \in B$ it holds that $\mu^{\sigma^t}(X^j) > \kappa$ for all j. Together with the fact $\cup_{j=1}^N X^j = \cup_{j=1}^M \mathcal{C}_i^{*k_j}$ and $\min_{j=1}^N \zeta_i(X^j) > \min_{j=1}^M \zeta_i(\mathcal{C}_i^{*k_j})$ this implies that $\mathcal{C}_i^*$ cannot be in the support of a limit point $\tilde{\pi}_i^*$. We conclude that $\tilde{\pi}^* \in \psi(\sigma^*)$ and hence that ψ is u.h.c.

The above argument establishes that for given $\varepsilon^n = (\varepsilon_1^n, \varepsilon_2^n)$ and κ^n there exists an $(\varepsilon^n, \kappa^n)$-categorization equilibrium for any n. Since $\Lambda \times \Pi$ is compact any sequence $(\varepsilon^n, \kappa^n)_{n \in \mathbb{N}}$ has a convergent sub-sequence with limit in $\Lambda \times \Sigma$. Hence a categorization equilibrium exists.