UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for New Economic Thinking

Cambridge-INET Institute

Cambridge-INET Working Paper Series No: 2014/11

Cambridge Working Paper in Economics: 1460

CATEGORIZATION AND COORDINATION

Vessela Daskalova (University of Cambridge)

Nicolaas J. Vriend (Queen Mary University of London)

ABSTRACT

The use of coarse categories is prevalent in various situations and has been linked to biased economic outcomes, ranging from discrimination against minorities to empirical anomalies in financial markets. In this paper we study economic rationales for categorizing coarsely. We think of the way one categorizes one's past experiences as a model of the world that is used to make predictions about unobservable attributes in new situations. We first show that coarse categorization may be optimal for making predictions in stochastic environments in which an individual has a limited number of past experiences. Building on this result, and this is a key new insight from our paper, we show formally that cases in which people have a motive to coordinate their predictions with others may provide an economic rationale for categorizing coarsely. Our analysis explains the intuition behind this rationale.

Categorization and Coordination*

Vessela Daskalova[†] and Nicolaas J. Vriend[‡]

18 May 2014

Abstract

The use of coarse categories is prevalent in various situations and has been linked to biased economic outcomes, ranging from discrimination against minorities to empirical anomalies in financial markets. In this paper we study economic rationales for categorizing coarsely. We think of the way one categorizes one's past experiences as a model of the world that is used to make predictions about unobservable attributes in new situations. We first show that coarse categorization may be optimal for making predictions in stochastic environments in which an individual has a limited number of past experiences. Building on this result, and this is a key new insight from our paper, we show formally that cases in which people have a motive to coordinate their predictions with others may provide an economic rationale for categorizing coarsely. Our analysis explains the intuition behind this rationale.

JEL Classification: D83, C72.

Keywords: categorization, prediction, decision-making, coordination, learning.

1 Introduction

Casual observation as well as empirical research show that people tend to categorize. For example, when an employer encounters a new applicant, she may observe that the applicant is 'white, male, engineer', and she may use her past experience with other applicants in this category to make a prediction about the suitability of the applicant for the job. College admission officers may place pupils in different categories on the

*We thank Diego Cerdeiro, Jane Cooley-Fruehwirth, Edoardo Gallo, Sanjeev Goyal, Nobuyuki Hanaki, Penelope Hernandez, Matthew Jackson, David Levine, Bob Marks, Friederike Mengel, John Miller, Francesco Nava, Scott Page, Tom Palfrey, Sonje Reiche, Hamid Sabourian, Klaus Schmidt, Bob Sugden, Jan Tuinstra, and seminar and workshop participants at the University of Cambridge for comments. Earlier versions of the dynamic model were presented at the Santa Fe Institute, at Queen Mary University of London, and at the SCE conferences in London (2010) and San Francisco (2011). All errors are our own.

[†]Corresponding author: Cambridge-INET, University of Cambridge, <vrd22@cam.ac.uk>

[‡]School of Economics and Finance, Queen Mary University of London, <n.vriend@qmul.ac.uk>

basis of their grades in school, and make inferences about the ability of the pupil based on their experience with past students in the respective categories. Credit rating agencies sort countries and firms into different categories (e.g. A++ or C-) to convey an estimate of their probability of default on debt. Equity researchers place companies into the categories 'buy', 'sell', and 'neutral' to indicate their recommendations for traders. Customers or certification agencies group hotels and restaurants into categories such as $4^\star$ and $2^\star$, and we have certain expectations about the quality of a hotel or restaurant based on the rating it has received. Thus, people often group past experiences such as objects they have seen, other people they have met or situations they have experienced into categories, and they, then, use these categories as a tool on the basis of which to make predictions and decisions in new situations encountered. This widespread use of categories is a first empirical motivation for our paper. If categories are used in so many different spheres of life, it seems important to understand some basic properties of different ways of categorizing. This is what we aim to do here.

There are many alternative ways in which people can organize past experiences into categories. In particular, they could divide them into many fine categories, each category containing a small number of experiences. Or they could use just a few coarse categories instead, each category containing a large number of experiences. Whether they choose fine or coarse categories may matter for the predictions that they make. More specifically, coarse categorization has been linked to a number of biased economic outcomes ranging from discrimination against minorities (Fryer and Jackson, 2008) to empirical anomalies in financial markets (Barberis and Shleifer, 2003) to persuasion in advertising (Mullainathan et al., 2008). These potentially important economic implications of the use of coarse categories constitute a second empirical motivation for our paper, in which we will focus in particular on investigating economic rationales for choosing coarse categorizations.[1]

A third empirical motivation for our paper comes from observing that coarse categorization often occurs in situations in which people try to coordinate with each other. For example, people may use coarse categories and stereotypes when talking to others, even though they may individually not think in terms of such crude stereotypes.[2] Similarly, we often describe political beliefs in rather coarse terms, such as 'left' and 'right', even when our views are much more subtle. More generally, the language we use to interact with other people is often vague (Lipman, 2009), and within firms employees tend to use relatively generic jargon (Cremer et al., 2007). People also typically refer to colors using relatively coarse categories, such as e.g. 'red', 'green', 'blue', even though many well-defined, very fine color schemes exist (Steels et al., 2005; Komarova et al., 2007). Our model of categorization is meant as a tool for investigating the basic properties of different ways of categorizing in a relatively abstract setting, without aiming to capture

[1]We focus on deriving *economic* rationales for coarse categorization rather than on cognitive limitations. It may well be that people categorize coarsely because they are born with a limited number of 'boxes' in their brain, or because the more categories they use the higher the cognitive costs of accessing an individual category. Our analysis derives economic rationales that are complementary to such cognitive limitations factors.

[2]We thank Matt Jackson for suggesting this example.

any of these complex situations closely. The focus of our analysis is on gaining some more formal economic understanding of the questions why and when motives to coordinate with others may create incentives to categorize coarsely.

Starting point for our analysis is the following standard decision problem. A decision maker observes some characteristics of an object and has to predict its unobserved value.[3] There is an underlying function that relates the unobserved value to the observed characteristics, but the agent does not know this function. In our framework the tool for prediction available to the agent are her past experiences, which are sorted into categories. She puts the object into a particular category based on its observed characteristics. She makes a prediction that the object's unobserved value is equal to the average value of all objects in the respective category.[4]

We refer to a way of categorizing as a categorization. A categorization is a set of categories that the individual splits her past experiences into. Every category is associated with a belief about the unobserved value of objects placed in it. We thus think of a categorization as containing a set of estimators that an individual uses to make predictions. Our focus is on the properties of alternative sets of estimators the decision maker could use, and on the factors that affect their suitability for prediction in different environments.

We present two complementary frameworks to analyze the properties of alternative possible categorizations - a static and a dynamic model. In the static model we assume that the agent has already accumulated a certain number of past experiences and faces a one-off prediction task. We analyze the factors that determine which ways of categorizing these past experiences will help her make the best prediction on the next object encountered. That is, comparing all possible ways of categorizing, we analyze which ways of categorizing will help her minimize the expected prediction error.

We consider several variants of the prediction task. First, we examine the properties of alternative ways of categorizing past experiences in the case of a decision maker who is only interested in predicting the true unobserved value of an object (i.e. the situation described above). Second, and this is a key feature of our paper, we examine the properties of different ways of categorizing in the case of decision makers who want to coordinate their predictions with each other. Reality abounds with cases in which people have some motive to coordinate their predictions with each other, and it thus seems important to study how this may affect the way individuals categorize. The desire to coordinate predictions may stem, on the one hand, from strategic considerations in situations in which people have to make a common decision later on. On the other hand, the desire

[3] We will use the term object broadly to denote any observation or any experience a decision maker may have. It may stand, for example, for an applicant faced by an employer. In that case observed characteristics may be education, age, race, gender, etc. The unobserved value that the agent is trying to predict may be the applicant's suitability for a particular job.

[4] Thus, we are assuming *that* decision makers categorize. For a discussion and some rationalization of why decision makers may want to use categories at all to make predictions, see Mohlin (2014) and Peski (2011). Using the *average* of the observations in a category as a prediction is a standard assumption in economic models of categorization; see e.g. Fryer and Jackson (2008), Al-Najjar and Pai (2014), Peski (2011), or Mohlin (2014).

to coordinate may be due to reputation concerns that lead people to conform.[5] To gain a basic understanding of how the attempt to coordinate predictions with others may affect the way an individual categorizes in a simple setting, we consider a benchmark case of two individuals, each of whom has accumulated a number of past experiences, and we assume that their goal is simply to coordinate predictions on the next object encountered and that they use categorizations to make predictions. We analyze this situation as a one-shot game in which each player chooses a categorization to use. We characterize some equilibrium properties of this game, and we explore the connection between a categorization's optimality for individual prediction and the possibility that the corresponding symmetric categorization profile constitutes a NE in the categorization game. Third, we look at the convex combination of the two benchmark cases, i.e. when decision makers are interested both in correctly predicting the true value of an object and in coordinating their prediction with another person.

In the complementary dynamic framework, an agent encounters a stream of objects, one in each period. Every period, from period one onwards, she has to make a prediction about the unobserved value of the object she encounters. We consider the learning dynamics, with the agent learning which categorization to use on the basis of her past experiences. In the dynamic model we relax some assumptions of the static model, and this allows us to analyze categorization also in a non-stationary environment. Just as in the static model, we consider the three variants of individual prediction, coordination, and the convex combination of the two.

Our main results can be summarized as follows. In the static analysis we show that the expected prediction error of a categorization in individual prediction can be decomposed into a bias and a variance component. Focusing on a symmetric setting to facilitate analytical tractability of the model, we show the following. In a deterministic environment, the best the decision maker can do is to use the finest possible categorization as its estimators are unbiased and there is no noise, hence no variance. In a stochastic environment, assuming the agent has a limited number of past experiences, comparative statics show that as the level of noise increases, or as the sample size available to the agent decreases, coarse categorizations become increasingly better compared to fine categorizations as their estimators are more consistent. Note that this is true without assuming any exogenous costs of using many fine categories and without assuming any bounded computational powers of the agent. Coarse categorizations may be optimal simply because they help the agent make better predictions. The intuition behind the benefit of using a coarser categorization is that in a coarser categorization there are on

[5] As an example of decision makers who need to coordinate predictions in order to make a common decision, consider members of a committee that have to decide whether an applicant is suitable for a particular job. Or firms that produce complementary products and want to coordinate predictions about the profitability of competing technologies in order to be able to establish future industry standards. The idea that professional forecasters (e.g. equity researchers) may want to have their predictions coordinated with others' due to reputation concerns has been put forward in various papers in finance (Scharfstein and Stein, 1990; Bizer et al., 2014). The same argument applies also to product reviewers, for example, on websites reviewing products for the entertainment industry; see e.g. Swisher (2013) for an analysis of reviewers on the Metacritic website. See also Morris and Shin (2002).

average more objects per category, and hence the expected variance of the category beliefs will be lower. That is, a coarser categorization has more consistent estimators compared to a finer categorization.[6]

In the case that two players simply want to coordinate their predictions, we find that in equilibrium they will not use categorizations at different levels of coarseness. In a deterministic environment all symmetric categorization profiles constitute a NE. What is interesting is that in a stochastic environment, finer categorization profiles may be NE only at sufficiently low levels of noise. As noise level increases, a player eventually has an incentive to deviate to a coarser categorization. Furthermore, both players using the coarsest possible categorization is always a Nash equilibrium, and it is the Pareto-superior one in any stochastic environment. Moreover, symmetric categorization profiles can be Pareto-ranked on their coarseness, with any coarser symmetric categorization profile always being more efficient than any finer symmetric categorization profile for any positive noise level. The intuition is that by using symmetric categorization profiles players have no bias from one another, and by using coarser categorizations, they reduce the variance in their predictions. These results suggest that in case people want to coordinate predictions with others in noisy environments, there is pressure to categorize more coarsely as this helps them coordinate more reliably.

Our analysis of the dynamic model shows that agents starting with a random set of categorizations can learn which categorizations to use depending on the environment. Comparing the coarseness of the categorizations an agent learns if she cares only about predicting the true object values correctly with the coarseness of the categorizations she learns if she (also) wants to coordinate predictions with others, we find that the higher the weight on coordination, the coarser the categorizations the individual learns. The properties of the categorizations that the agents learn in the dynamic model are in line with those predicted by our analysis of the static model for the different environments.

Our paper makes several contributions to the literature. In terms of the questions we consider, we are to the best of our knowledge the first to analyze the properties of optimal categorization if players want to coordinate predictions. In terms of methodology, our contribution consists of developing a unified framework to analyze the properties of optimal categorization for the cases of individual prediction, coordination, and the convex combination of the two. We view categorizations as consisting of sets of estimators, and we use basic statistics to decompose the expected prediction error of alternative categorizations into a bias and a variance component, and basic game theory to characterize Nash equilibria in categorizations for the case when players want to coordinate predictions. In terms of new insights, the main message of our paper is that the attempt to coordinate predictions with others may be a further rationale for coarse categorization, additional and complementary to those hitherto considered in the literature.

The structure of the paper is the following: In section 2 we discuss this paper's relation to the literature. Section 3 describes our static model, and in section 4 we set out in detail

[6] This result is in line with Al-Najjar and Pai (2014) and Mohlin (2014).

our main analytical results. In section 5 we present our dynamic model, followed by its numerical analysis in section 6. Section 7 concludes.

2 Relation to the Literature

This paper is most closely related to some recent papers studying categorization as a model of individual decision making (Al-Najjar and Pai, 2014; Mohlin, 2014; Peski, 2011), and more distantly to some papers on categorization in games (Azrieli, 2009; Jehiel, 2005; Jehiel and Samet, 2007; Heller and Winter, 2014; Mengel, 2012a,b; Grimm and Mengel, 2012). Our work complements both lines of research by analyzing strategic considerations not studied in these strands of the literature. We start this section by discussing the literature that is most directly related to this paper, explaining our similarities and differences from it. We then place the paper in a broader context.

While coarse categorization has previously been viewed as a form of bounded rationality (resulting from innate limitations on the number of available categories or bounded computational power making fine categorization costly), both our and the above papers on categorization as a model of individual decision making, in particular Al-Najjar and Pai (2014) and Mohlin (2014), show that coarse categorization may result even without restrictions and costs on categorizing finely. An economic rationale to categorize coarsely is that while by using fine categories a decision maker may fit past observations more precisely, coarse categorization avoids overfitting in prediction in noisy environments when decision makers have a limited number of past observations. The paper most similar to ours in terms of analyzing categorization as a tool for individual prediction is Mohlin (2014). He also considers the question which categorizations minimize expected prediction error and independently from us formulates the costs and benefits of using fine and coarse categories in terms of bias and variance components of expected prediction error for a slightly different decision situation. Al-Najjar and Pai (2014) look at categorization as an example of coarse decision making and use different methodology (statistical learning theory) to show that coarse categorization is useful if the decision maker is operating in a noisy environment and has a limited number of observations. Mullainathan (2002) compares the accuracy of predictions of an agent who categorizes with the accuracy of predictions of a Bayesian decision maker. Peski (2011) shows that in a symmetric environment (if the agent has the same prior over all objects and over all properties he is making a prediction about) categorization is an optimal method for making predictions. Azrieli and Lehrer (2007) develop a complementary model in which a categorization is generated by extended prototypes.

This paper complements the papers discussed above by considering an additional class of situations. While they focus on properties of categorizations if an individual wants to make optimal predictions or decisions with respect to objects she encounters, we develop a unified framework under which we also consider the properties of categorizations if agents want to coordinate predictions about these objects with one another. We add to the above literature by showing that there may be another economic rationale for coarse categorization - the attempt to coordinate with others. Additionally, we consider making predictions on the basis of categories as a dynamic problem and we show that agents learn different ways of categorizing depending on the environment they are in and in line with the results from our static model. In our dynamic model, we are also able to relax a number of assumptions of the static model that are standard in the literature. For example, as we will explain in detail later on, we can analyze hierarchical and incomplete categorizations, and the model is applicable to a non-stationary environment.

While we also consider the role of categorization in strategic interactions, we pose a different question from those studied in the existing papers on categorization in games (Azrieli, 2009, 2010; Heller and Winter, 2014; Jehiel, 2005; Jehiel and Samet, 2007; Mengel, 2012b; Grimm and Mengel, 2012). We analyze a situation in which players use categories to make predictions about objects that they encounter and want to coordinate these predictions with one another as well as the situation in which they care both about predicting the true value correctly and about coordinating their predictions with the other. This complements the questions analyzed in other game-theoretic papers on categorization. Azrieli (2009) and Azrieli (2010) consider the use of categories to categorize opponents in games. Jehiel (2005) analyzes multi-stage games in which players bundle nodes at which other players move into analogy classes. In Jehiel and Samet (2007) each player partitions her own moves into similarity classes. Mengel (2012a) takes an evolutionary perspective and shows that populations relying on coarser partitions have higher fitness. Mengel (2012b) considers players who categorize games. Grimm and Mengel (2012) present an experiment in which they find that players learn to play games that are strategically equivalent (i.e. they fall in the same category) in the same way. Players categorizing games is also the underlying idea of Heller and Winter (2014).

Further economic papers related to categorization include Manzini and Mariotti (2012) who develop a model in which agents categorize alternatives before making a choice and Mandler et al. (2012) who analyze decision making on the basis of a checklist. More generally, categorization can be viewed as an example of decision making on the basis of heuristics as discussed by Gigerenzer and Brighton (2009), who also argue that less information can improve accuracy in prediction. Gilboa and Samuelson (2009) present a model of how a preference for simplicity can improve efficiency in inductive reasoning. The question of how many categories to use is related to the question of specifying the right number of parameters in econometric models for out-of-sample predictions. Making decisions on the basis of categories is also related to making decisions on the basis of analogies (Mitchell and Hofstadter, 1996; Hofstadter, 1996) and to decision making on the basis of similar past experiences as in Case Based Decision Theory (Gilboa and Schmeidler, 1995). All of these papers focus on models for individual decision making only. A crucial difference of our model with Case Based Decision Theory is that in their paper when making decisions on the basis of past cases, the decision maker considers how similar the current case is to each of the previous cases using a predefined measure of similarity. In our model of categorization, all observations in the same category are treated in the same way; as is the case in other economic models of categorization, i.e.

Mohlin (2014), Peski (2011), and Al-Najjar and Pai (2014). Thus, within a category there is no weighing of which observation is more similar to the new experience. Also, a basic assumption underlying decision making on the basis of categories is that experiences from other categories are not taken into account.

There have been some papers in the economic literature studying the implications of categorization. Fryer and Jackson (2008) show that discrimination against minorities may result if a decision maker has a limited number of categories available. Mullainathan et al. (2008) show that advertisers may use the tendency of people to think coarsely to persuade them to buy a certain product. Barberis and Shleifer (2003) show that style investing (the tendency to invest in classes of stocks rather than individual stocks) may be an explanation of some biases observed in financial markets. While we focus only on analyzing the basic theoretical properties of different ways of categorizing, we believe that our insight that links coordination and coarse categorization may have implications for real world situations that could be analyzed further.

There is an enormous number of empirical studies as well as procedural models by cognitive scientists and psychologists on how people categorize. For some overviews of cognitive science models, see Ashby and Maddox (2005), Zaki et al. (2003), and Smith and Medin (1981). The discussion is typically centered around the question which procedural model best describes the way humans categorize - comparing prototype (Rosch, 1975, 1978), exemplar (Nosofsky, 1986), and mixture models. As economists we are less interested in the exact physical and mental mechanisms of categorizing. Our focus is on explaining which ways of categorizing would be most useful in a given environment and why. The details of what actually goes on in the human brain in the process of categorizing is beyond the scope of our analysis.

Finally, categorization is also related to clustering and classification algorithms in machine learning; see e.g. Bishop (2007) or Murphy (2012). Clustering algorithms partition objects into groups in a way such that objects assigned to the same group are more similar to each other according to some distance metric than objects from different groups. Classification algorithms are used to predict which class a new observation belongs to. The algorithm is given a set of training data from which it learns. While the goal of clustering algorithms is to form the best clusters according to some measure of similarity, in economic analyses and in our model the driving force of categorization is the goal to make the best predictions on some unobserved attribute. This also differs from classification where the goal is to use past experiences to predict whether an object belongs to a certain class of objects or not.

An adaptive framework for learning categories to make predictions has also been developed by Anderson (1990, 1991). In contrast to our learning model in his framework there are exogenous costs of searching for the best categorization and the search is discontinued when the costs exceed some predefined threshold. We offer a more flexible model in which the agent is not constrained to neighboring categorizations as in Anderson (1990, 1991), but can search the entire space of possible categorizations.

3 Static Model

In our static model we consider the following situation. An individual who has already accumulated a number of past experiences faces a one-off prediction task. She encounters an object and has to make a prediction about its unobserved value. She places it in a category based on its observable characteristics and makes a prediction about its unobserved value equal to the average of the unobserved values of all objects in the respective category. Note that we assume that the unobserved values of all objects experienced in the past have been revealed to the agent.

There are many alternative ways in which the agent could divide her past experiences into categories. She could use just a few coarse categories, lumping many different past experiences together in one category. Or she could use many fine categories, each containing a small number of experiences. Since the new object will be treated according to the category average, her prediction would generally differ depending on which way of categorizing past experiences she chooses. Our focus is on studying the properties of different ways of categorizing. More precisely, we want to gain some understanding of the factors that determine which way of categorizing will help the agent minimize her expected prediction error on the next object depending on the environment she is in.

We also consider a second variant of this one-off prediction task. There are two agents who observe the same object and each of them independently makes a prediction about its unobserved characteristics. We assume that these agents are only interested in coordinating their predictions with each other. We formulate this situation as a one-off game in which each agent chooses a categorization to use and we analyze the equilibrium properties of this game.

Finally, we extend our analysis to consider a situation in which agents care both about predicting the true object value correctly and about coordinating their prediction with one another. But first we introduce some definitions and notation.

3.1 Objects and Categories

Individuals make predictions about objects. An object o consists of an l-dimensional vector of observed binary attributes x together with an unobserved real-valued feature y that the individual is trying to predict, that is $o = (x, y) \in X \times Y$ with $x \in \{0,1\}^l$ and $y \in \mathbb{R}$.[7] We denote the set of all objects the individual has experienced in the past by O. This is the set of past experiences she draws upon when making a prediction about a new object. An object's type is determined by its observable attributes, that is by x only. Let O^T denote the set of all possible object types. The number of all possible object types is $|O^T| = 2^l$, where l is the length of the vector of observable attributes of an object. Thus, for example, if $l = 1$, then $O^T = \{0,1\}$ and if $l = 2$, then

[7]Note that the fact that the x-vector is binary does not constitute a restriction as its length may always be increased to approximate an object with continuous attributes. One could also make the assumption that y is a real-valued vector instead of a real-valued scalar, but we prefer to keep the analysis of the static model as simple as possible. We relax this assumption in our dynamic model.

$O^T = \{11, 10, 01, 00\}$. Throughout the analysis we assume for simplicity that the agent has sampled an equal number of objects n from each possible object type.[8] Thus, the total number of experiences the individual has is $|O| = 2^l n$.

The unobserved feature y that the individual is trying to predict is a function of the observed attributes of the object, i.e. $y = f(x) + \epsilon$ with $\epsilon \sim N(0, \sigma^2)$. The noise term is drawn from the normal distribution and is i.i.d. for each object. For simplicity our analysis focuses on the case when the variance of the objects σ^2 is independent of the object type.[9] A deterministic environment is defined by $\sigma^2 = 0$, that is all objects that can be described by the same vector of observable attributes have the same unobserved value. In a stochastic environment $\sigma^2 > 0$, i.e. the unobserved value is a noisy function of the observed attributes. A stochastic environment represents the more realistic case in which observable characteristics do not completely reflect the unobserved value of an object (e.g. same gender, race, or having studied at the same institution do not mean people have the same suitability for a particular job) or in which some variable of interest is not observed by the decision maker (e.g. ability is not observed when the agent makes a prediction about someone's suitability for a particular job). We do not make assumptions on $f(x)$. The agent does not know the data generating process (the relation between observed and unobserved attributes) and predicts the unobserved value of the object by assigning it to a category. We assume that after the individual has made a prediction the true object value is revealed to her. For example, after having predicted how suitable a person is for a particular job, the individual observes the true productivity of this person. Thus, the agent knows the true unobserved values of all objects she has experienced in the past.

Each category has a category type $C^T \subseteq O^T$, which determines which objects can be put in it. The assignment of objects to categories is based on the objects' observed attributes. For example, for the one attribute case there are three category types: $\{0\}, \{1\}, \{1, 0\}$. The first category can contain only objects of type 0, the second category only objects of type 1, and the third category both objects of type 0 and of type 1. If there are two attributes, there are 15 category types: $\{11\}$, $\{10\}$, $\{01\}$, $\{00\}$, $\{11, 10\}$, $\{11, 01\}$, $\{11, 00\}$, $\{10, 01\}$, $\{10, 00\}$, $\{01, 00\}$, $\{11, 10, 01\}$ $\{11, 10, 00\}$, $\{11, 01, 00\}$, $\{10, 01, 00\}$, $\{11, 10, 01, 00\}$.[10] A category C is a collection of objects such that the x attributes of the objects in this category are covered by the category type, i.e. $C = \{(x_i, y_i) | x_i \in C^T\}$.

3.2 Category Belief and Categorizations

As the individual assigns a newly encountered object to a category, she predicts that the y-value of the object equals the average of the y-values of all past experiences in this category. We call this average the category belief. The category belief for category C

[8] The analysis can be extended to allow for objects of each type to be sampled from some probability distribution. We leave this extension for future research.

[9] The analysis could also be extended to allow for objects of different object types to have different variance.

[10] Overall, the number of category types is given by $|C^T| = \sum_{i=1}^{|O^T|} \binom{|O^T|}{i} = \sum_{i=1}^{|O^T|} \frac{|O^T|!}{i!(|O^T|-i)!}$ where $|O^T|$ is the number of different object types.

is defined as $\hat{Y}^C = \frac{1}{|C|}\sum_{(x_i,y_i)\in C} y_i$. Category beliefs are the estimators on the basis of which the individual makes predictions. Note that a category belief is capitalized as it is a random variable which depends on the exact realization of the y-values of the objects in the category.

A categorization is a decision maker's complete model of the world on the basis of which she makes predictions. A categorization P is a set of categories that form a disjoint partitioning of the object set (the set of the agent's past experiences), i.e. $P = \{C_1, C_2, ..., C_k\}$ such that $(\bigcup_i C_i = O)$ and $(C_i \cap C_j = \emptyset)$ for all $i \neq j$. No two categories in the same categorization can cover the same object type $C_i^T \cap C_j^T = \emptyset$ for all $i \neq j$.[11] Using a categorization to make predictions amounts to using a set of estimators, one for each category of objects.

There can be different possible ways to partition the object set into categories.[12] And these categorizations differ in coarseness. The coarseness of the categorization is determined by the number of categories k in which the set of objects O is partitioned. The smaller the number of categories for a given O, i.e. the smaller the k, the coarser the categorization. The larger the number of categories for a given O, i.e. the larger the k, the finer a categorization.

The finest possible categorization is a partitioning such that each object type has a separate category.[13] The coarsest possible categorization is a partitioning such that all object types are assigned to a single category. There is always one finest and one coarsest possible categorization. Depending on the number of different attributes describing an object and therefore on the number of different object types, there can be many categorizations between the finest and the coarsest possible.[14] These intermediate levels of coarseness of categorizations can be reached by (consecutive) coarsening of the finest possible categorization or by (consecutive) refining of the coarsest possible categorization. A coarsening of a categorization consists of merging two or more of its categories. A categorization formed through coarsening of a finer categorization will have a smaller number of categories and a greater number of object types in at least one category, compared to the categorization it was formed from. A refinement of an existing categorization consists of splitting one or more of its categories into two or more

[11] This assumption is relaxed in the dynamic model in which we also consider hierarchical categorizations.

[12] The set of possible categorizations $\mathcal{P}$ is equal to the set of all possible disjoint partitionings of the object set. Its cardinality will be given by the Bell number B_d with $d = |O^T|$ (see Rota (1964)).

[13] Note that using the finest possible categorization is not the same as not categorizing. If the individual uses the finest possible categorization, experiences with objects that have different observable attributes and are therefore placed in another category under the finest possible categorization will have no effect on her prediction. If the individual does not categorize, then such experiences may affect her prediction.

[14] For example, for the case of $l = 2$ (two observable attributes, four object types), we have the following categorizations. The finest possible categorization, with four categories, is $\big\{\{11\}\,\{10\}\,\{01\}\,\{00\}\big\}$. There are the following six categorizations with three categories: $\big\{\{11,10\}\,\{01\}\,\{00\}\big\}$, $\big\{\{11,01\}\,\{10\}\,\{00\}\big\}$, $\big\{\{11,00\}\,\{10\}\,\{01\}\big\}$, $\big\{\{10,01\}\,\{11\}\,\{00\}\big\}$, $\big\{\{10,00\}\,\{11\}\,\{01\}\big\}$, $\big\{\{01,00\}\,\{11\}\,\{10\}\big\}$. And the following seven categorizations with two categories: $\big\{\{11,10\}\,\{01,00\}\big\}$, $\big\{\{11,01\}\,\{10,00\}\big\}$, $\big\{\{11,00\}\,\{10,01\}\big\}$, $\big\{\{11,10,01\}\,\{00\}\big\}$, $\big\{\{11,10,00\}\,\{01\}\big\}$, $\big\{\{11,01,00\}\,\{10\}\big\}$, $\big\{\{10,01,00\}\,\{11\}\big\}$. The coarsest possible categorization is $\big\{\{11,10,01,00\}\big\}$. See also Figure 1.

categories with the restriction that all objects of the same type have to be assigned to the same category. A categorization formed through refinement of a coarser categorization will have a larger number of categories and a smaller number of object types in at least one category, compared to the categorization it was formed from.

Note that there may be different ways to coarsen or to refine an existing categorization. Thus, there may be different paths from the finest to the coarsest possible categorization. Figure 1 illustrates the different paths from the finest to the coarsest categorization in the case of four object types. We say that two categorizations are connected by a path if one may be reached through only coarsening (only refining) of the other. Depending on which categories of a finer categorization are merged together at the next level, one may not be able to reach a particular coarser categorization through combining of two or more existing categories only. For example, from the categorization $\{\{11, 10\}\{01\}\{00\}\}$ one can directly reach the coarser categorization $\{\{11, 10, 01\}\{00\}\}$ by merging the first and the second category. However, it is not possible to reach the coarser categorization $\{\{11, 01, 00\}\{10\}\}$ only by merging existing categories from $\{\{11, 10\}\{01\}\{00\}\}$. The latter is an example of categorizations that are not connected by a path. Note that in Figure 1 there is a line connecting $\{\{11, 10\}\{01\}\{00\}\}$ to $\{\{11, 10, 01\}\{00\}\}$, but there is no line connecting $\{\{11, 10\}\{01\}\{00\}\}$ to $\{\{11, 01, 00\}\{10\}\}$. Distinguishing categorizations that are connected by a path from categorizations that are not connected by a path will be useful for our proofs later on.

We denote a categorization at some level of coarseness L by P^L, a categorization that is coarser than P^L by P^{L^+}, and a categorization that is finer than P^L by P^{L^-}. A categorization P^{L^+} is coarser than P^L if and only if the number of categories of P^{L^+} is smaller than the number of categories of P^L. We then write $P^{L^+} \succ P^L$. A categorization P^{L^-} is finer than P^L if and only if the number of categories of P^{L^-} is greater than the number of categories of P^L. We denote this case as $P^{L^-} \prec P^L$. Analogically, we use $P^{L^{++}}$ to indicate a categorization that is coarser than P^{L^+} and we use $P^{L^{--}}$ to indicate a categorization that is finer than P^{L^-}. Furthermore, where necessary we use an index in parenthesis to indicate a specific categorization. That is, $P^L(i)$ indicates a categorization i that has level of coarseness L. We denote the set of all categorizations that are connected by a path to categorization $P^L(i)$ as $R(P^L(i))$. Thus, a categorization $P^{L^+}(j) \in R(P^L(i))$ is a categorization that is coarser than $P^L(i)$ and is connected by a path to it. A categorization $P^{L^+}(k) \notin R(P^L(i))$ is a categorization that is coarser than $P^L(i)$ and is not connected by a path to it. To indicate that two different categorizations have the same level of coarseness, we write $P^L(i) \sim P^L(j)$.

3.3 Expected Prediction Error

The goal of the agent is to make the best possible prediction on the next object she encounters.[15] For simplicity, we assume that the next object is drawn from a discrete

[15]Note that considering prediction on the next object only is without loss of generality in this set up. We could also consider prediction on the next string of objects, but since the model is static the results

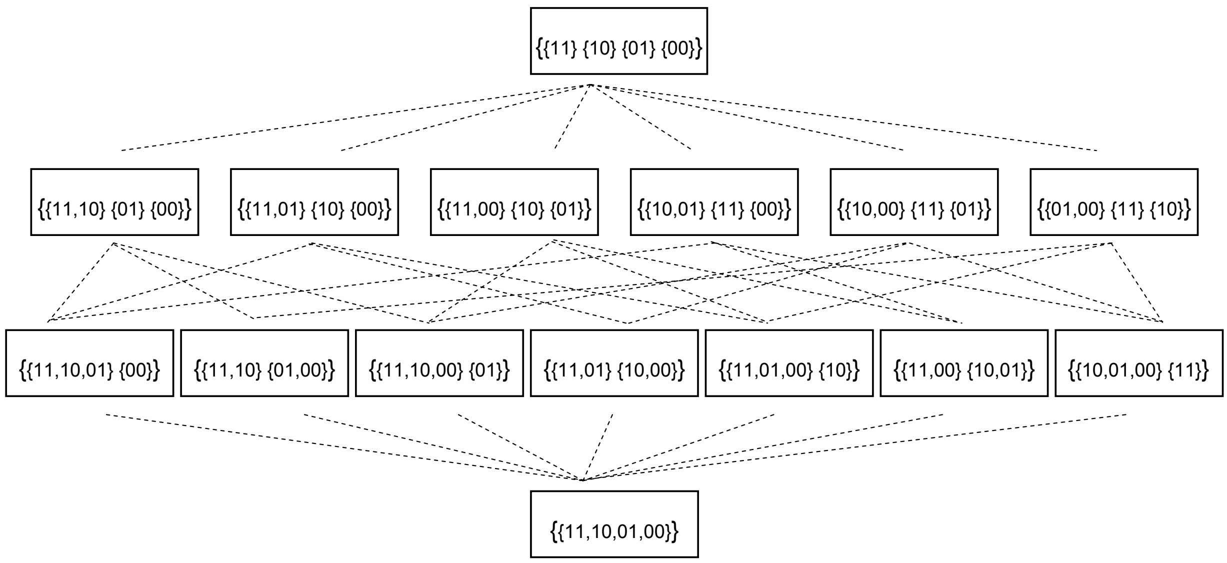


Figure 1: Paths connecting categorizations for the case of four object types

uniform distribution of all possible object types. The indicator we use to measure how good a categorization P is for individual prediction is the expected prediction error the individual would make on the next object she encounters by using this categorization: $EPE^{IP}(P)$. As we can see from Definition 1, the expected prediction error on an object type j is the expected mean squared error between the object's unobserved value and the category belief of the category the object is assigned to.[16] The expected prediction error of a categorization combines the expected prediction error on all object types weighing them by the probability that an object of a given type is observed. Note that $\hat{Y}^{C_k}$ and Y_j are capitalized as they denote random variables. The category belief $\hat{Y}^{C_k}$ depends on the random realization of the y-values of past experiences the individual has made in this category and Y_j depends on the realization of the y-value of the next object.

Definition 1. Expected Prediction Error Individual Prediction

$$EPE^{IP}(P) = \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j E[(\hat{Y}^{C_k} - Y_j)^2] \tag{1}$$

The indicator that we use to measure how good a given categorization profile is for the coordination of the two players' predictions on the next object is the expected prediction error from coordination of the two categorizations from each other. We denote it by $EPE^C(P_1, P_2)$ where P_1 denotes the categorization Player 1 uses and P_2 denotes the categorization Player 2 uses. As Definition 2 shows, the expected prediction error between the two players' predictions on an object type j is equal to the mean squared

would not change. In our dynamic model we consider an agent who experiences a sequence of objects and learns from her experience with each object.

[16] The mean squared error is a standard way of measuring how good an estimator is; see any statistics textbook, e.g. Berry and Lindgren (1996).

error between the category beliefs that the players use on object type j. The expected prediction error of two categorizations from each other combines the expected prediction errors on all object types, weighing them by the probability that an object of a given type is observed. As in this case both players only care about coordinating, the $EPE^C(P_1, P_2)$ is the same for both of them.

Definition 2. Expected Prediction Error Coordination

$$EPE^C(P_1, P_2) = \sum_{C_k \in P_1} \sum_{C_l \in P_2} \sum_{x_j \in (C_k^T, C_l^T)} p_j E[(\hat{Y}_1^{C_k} - \hat{Y}_2^{C_l})^2] \tag{2}$$

The expected prediction error for the case when decision makers care both about individually predicting the true object value correctly and about coordinating their prediction with each other is denoted by $EPE^{IP\&C}(P_1, P_2)$ and is given by the convex combination of the expected prediction errors from Definition 1 and from Definition 2. Let w denote the weight the individual places on making correct predictions about the true object value and $(1-w)$ denote the weight she places on coordinating her predictions with the other $(0 \leq w \leq 1)$. We assume that both players place the same weight w on individual prediction.[17] Note that if the two players use different categorizations then their $EPE^{IP\&C}(P_1, P_2)$ will be generally different, as although they would be making the same error with respect to each other, they would be generally making different mistakes with respect to the true object value.

Definition 3. Expected Prediction Error Individual Prediction and Coordination

The $EPE_1^{IP\&C}(P_1, P_2)$ of Player 1 is:

$$EPE_1^{IP\&C}(P_1, P_2) = wEPE^{IP}(P_1) + (1-w)EPE^C(P_1, P_2)$$
$$= w \sum_{C_k \in P_1} \sum_{x_j \in C_k^T} p_j E[(\hat{Y}^{C_k} - Y_j)^2] + (1-w) \sum_{C_k \in P_1} \sum_{C_l \in P_2} \sum_{x_j \in (C_k^T, C_l^T)} p_j E[(\hat{Y}^{C_k} - \hat{Y}^{C_l})^2]$$

The $EPE_2^{IP\&C}(P_1, P_2)$ of Player 2 is:

$$EPE_2^{IP\&C}(P_1, P_2) = wEPE^{IP}(P_2) + (1-w)EPE^C(P_1, P_2)$$
$$= w \sum_{C_l \in P_2} \sum_{x_j \in C_l^T} p_j E[(\hat{Y}^{C_l} - Y_j)^2] + (1-w) \sum_{C_k \in P_1} \sum_{C_l \in P_2} \sum_{x_j \in (C_k^T, C_l^T)} p_j E[(\hat{Y}^{C_k} - \hat{Y}^{C_l})^2] \tag{3}$$

4 Analysis of the Static Model

In this section we present the analysis of the static model. We first analyze the properties of optimal ways of categorizing in the two benchmark cases - on the one hand, when an

[17] The analysis can be extended to allow for the players to place different weights on individual prediction.

individual cares only about predicting the true value correctly, and on the other hand, when agents care only about coordinating with each other. We then combine our insights from these cases in order to analyze their convex combination. All results and where relevant a sketch of the proof and some intuition are given in the text, while the technical proofs are in Appendix B.

4.1 Individual Prediction

We first consider the case of an individual who is only interested in predicting the true unobserved object value correctly. We develop some preliminary results in Lemma 1 and build upon them to derive our main results in this section in Proposition 1 and in Proposition 3. In Lemma 1 we show that the Expected Prediction Error of a categorization P in individual prediction, i.e. $EPE^{IP}(P)$ can be decomposed into two components - a bias component and a variance component.[18] These two components are crucial for understanding the costs and benefits of using fine and coarse categorizations, respectively.

Lemma 1. *Bias-Variance Decomposition of $EPE^{IP}(P)$*

$$EPE^{IP}(P) = \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j E[(\hat{Y}^{C_k} - Y_j)^2]$$

$$= \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j Var(\hat{Y}^{C_k}) + \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j Var(Y_j) + \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j (E[\hat{Y}^{C_k}] - \mu_j)^2 \qquad (4)$$

$$= Var(P) + Var(Y) + Bias^2(P)$$

Lemma 1 shows that the $EPE^{IP}(P)$ of a categorization can be decomposed into the expected variance of the category beliefs of the categorization $Var(P)$, the expected variance of the underlying object types in the population $Var(Y)$, and the expected bias of the category beliefs in the categorization $Bias^2(P)$. The $EPE^{IP}(P)$ will increase if any of these underlying components increases.

As a reminder note that the bias and the variance are standard properties describing how good an estimator is for prediction. The variance shows how sensitive the estimator is to the particular sample. The lower the variance of one estimator compared to another, the more consistent the first estimator is compared to the second. The bias is the difference between the expected value of an estimator and the true population mean of the object the prediction is about. Note that Lemma 1 is a bias-variance decomposition of the EPE of a categorization, i.e. of a set of estimators rather than of one estimator. Thus, the variance component is the expected variance of the category beliefs of all categories in the categorization and the expected variance of all object types in the population. The

[18] A bias-variance decomposition of the mean squared error of an estimator is shown in many statistical textbooks; e.g. Berry and Lindgren (1996). As we show, a bias-variance decomposition will be particularly useful to analyze the EPE of a categorization, when we view a categorization as a set of estimators.

bias component of a categorization is equal to the expected bias of its category beliefs. The bias of a category belief is the squared difference between its expected value $E[\hat{Y}^{C_k}]$ and the population mean of a particular object type j, μ_j, averaged over all object types that enter this category. Note that although the category belief is a sample average, in the case of a category that contains more than one object type, this category belief is not generally an unbiased estimator.

In Proposition 1 we derive results on the implications of the coarseness of a categorization for the expected bias and the expected variance components of its $EPE^{IP}(P)$. For expositional simplicity from now on we write bias and variance instead of expected bias and expected variance component of the expected prediction error.[19] We also show how the stochasticity of the environment and the sample size of the agent affect the variances of the category beliefs of the categorization.

Proposition 1. *Comparative Statics Bias and Variance Components* $EPE^{IP}(P)$

Part 1. Variances

The variance component of the $EPE^{IP}(P)$ *of a categorization* $Var(P) + Var(Y) = f(k, \sigma^2, n)$ *is increasing in the noise level* σ^2*, increasing in the number of categories k, and decreasing in the sample size n. The difference in variance between a finer and a coarser categorization* $Var(P^L) - Var(P^{L^+}) = g(m, \sigma^2, n)$ *is increasing in the noise level* σ^2*, increasing in the difference in the number of categories in the two categorizations m, and decreasing in the sample size n.*

Part 2. Biases

Let $P^L(i)$ *be a categorization of some level of coarseness L and* $P^{L^+}(j)$ *be any coarser categorization such that* $P^{L^+}(j) \in R(P^L(i))$*, i.e. such that* $P^{L^+}(j)$ *and* $P^L(i)$ *are connected by a path. Then it is always true that* $Bias^2(P^{L^+}(j)) \geq Bias^2(P^L(i))$.

We now explain the intuition behind Part 1. Lemma 1 showed that the variance component of EPE^{IP} is equal to the variance of the category beliefs in the categorization plus the variance of the object types in the population $Var(P) + Var(Y)$. Under the assumption of an equal variance of all object types in the population and an equal number of objects from each type in the agent's sample, we show in Appendix B that the latter term is equal to: $Var(Y) = \sigma^2$. We thus focus our attention on the first term, i.e. on the expected variance of the category beliefs in a categorization $Var(P)$. Each category belief $\hat{Y}^{C_k}$ is the average of the y-values of the objects in the respective category C_k. The y-values of these objects are normally distributed and thus the category belief is also normally distributed, with a variance that will be equal to the population variance

[19]Note, however, that the variance that we refer to is always the variance component of the EPE of a categorization, which is based on the EPE of its category beliefs. This is not to be confused with the variance of a category, i.e. with the variance of a sample. As we are interested in characterizing how good estimators are for prediction, our focus is on the bias and variance components of the EPE of the estimators, i.e. of the category beliefs.

divided by the number of objects in the category.[20] Thus, the intuition behind the benefit of using a coarser categorization is that in a coarser categorization there are on average more objects per category and the variance of the category beliefs will be lower. That is, a coarser categorization has more consistent estimators compared to a finer categorization. Note also that the variance of all categorizations of equal coarseness is the same.[21]

Proposition 1 Part 1 implies that the more noise there is in the environment and/or the smaller the sample size of an agent, the greater the difference in variance between a finer and a coarser categorization and hence the greater the benefits of using a coarse rather than a fine categorization. Table 2 in Appendix A provides a numerical illustration of this part of the proposition.

We now look at Part 2. The bias component of $EPE^{IP}(P)$ of a categorization used for individual prediction is equal to the bias of its category beliefs $Bias^2(P) = \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j (E[\hat{Y}^{C_k}] - \mu_j)^2$. The bias of each category belief is the expected mean squared error between the expected value of the category belief (the estimator $E[\hat{Y}^{C_k}]$) and the true population mean of each object type in the category. Note that the category belief of a category in which there is only one object type will be an unbiased estimator of the true value of this object type in the population, as the expected value of the estimator will be equal to the true population mean of this object type. However, the category belief of a category which contains object types with different underlying population means will be generally a biased estimator for a particular object type as its expected value will be equal to the average of the population means and thus the estimator will be making a mistake towards at least some of the population means.[22] In the Proof of Proposition 1 Part 2 in Appendix B we show that any coarser category formed through merging of two finer categories will have a higher bias than the sum of the biases of the finer categories that were merged to form it. As a result, a coarser categorization will always have a higher bias than a finer categorization that is connected to it by a path.[23] Table 2 in Appendix A provides a numerical illustration of the bias of finer and coarser categorizations.

To further illuminate the underlying relationship between categorization coarseness and $EPE^{IP}(P)$ consider the following two cases. Assume the individual uses the finest possible categorization, i.e. each category contains only one type of objects. In that case, the expected value of category belief for each category will be equal to the true mean of the respective type of objects in the population. Therefore, each category belief is an unbiased estimator and the bias term in $EPE^{IP}(P)$ will be zero. Unless the environment is deterministic ($\sigma^2 = 0$), however, the finest categorization will have a positive $EPE^{IP}(P)$ due to the variance component. Assume to the contrary that the

[20]Let $Y_1, Y_2, ...Y_m$ be normal random variables which are independent and for which $Y_i \sim N(\mu_i, \sigma^2)$. We know from the basic properties of normal random variables that $\sum_i Y_i \sim N(\sum_i \mu_i, \sum_i \sigma^2)$. Thus, the average of these normal random variables will also be normally distributed with $\frac{\sum_i Y_i}{m} \sim N(\frac{\sum \mu_i}{m}, \frac{\sigma^2}{m})$.

[21]As we are interested in deriving economic rationales for coarse categorization, our focus is not on the asymptotics but on the finite sample properties of different ways of categorizing.

[22]Apart from the trivial case when the population means of all object types are equal.

[23]The biases of categorizations that are not connected by paths are not directly comparable without making additional assumptions.

individual uses the coarsest possible categorization, i.e. all types of objects are put in one category. In that case, if the different types of objects have different population means, the prediction made using the category belief will be biased. However, $EPE^{IP}(P)$ may be smaller in the case of the coarsest possible categorization as Proposition 1 Part 1 tells us that the variance component of $EPE^{IP}(P)$ for the coarsest categorization will be smaller than the variance component of $EPE^{IP}(P)$ for the finest categorization for any positive noise level. While the category beliefs of the finest categorization are the unbiased estimators, the category belief of the coarsest categorization will be the more consistent estimator, as it will have the smallest variance due to the fact that it contains the largest number of objects.

As we assume that the goal of an economic agent is to make optimal predictions, we next characterize what an optimal categorization means.

Definition 4. Characterization of Optimal Categorization for IP

A categorization $P^L(i)$ is optimal for individual prediction if and only if the following three conditions hold: i) its EPE^{IP} is smaller or equal to the EPE^{IP} of all categorizations that are finer than it; ii) its EPE^{IP} is smaller or equal to the EPE^{IP} of all categorizations that are equally coarse; iii) its EPE^{IP} is smaller or equal to the EPE^{IP} of all categorizations that are coarser than it. Using the bias-variance decomposition from Lemma 1, this is equivalent to:

i) For all $P^{L^-} \prec P^L(i)$ it has to hold that:

$$\begin{aligned}&EPE^{IP}(P^L(i)) \leq EPE^{IP}(P^{L^-})\\ &\Leftrightarrow Var(P^L(i)) + Var(Y) + Bias^2(P^L(i)) \leq Var(P^{L^-}) + Var(Y) + Bias^2(P^{L^-})\\ &\Leftrightarrow Var(P^L(i)) - Var(P^{L^-}) \leq Bias^2(P^{L^-}) - Bias^2(P^L(i))\end{aligned} \tag{5}$$

ii) For all $P^L(j) \sim P^L(i)$ it has to hold that:

$$\begin{aligned}&EPE^{IP}(P^L(i)) \leq EPE^{IP}(P^L(j))\\ &\Leftrightarrow Var(P^L(i)) + Var(Y) + Bias^2(P^L(i)) \leq Var(P^L(j)) + Var(Y) + Bias^2(P^L(j))\\ &\Leftrightarrow Var(P^L(i)) - Var(P^L(j)) \leq Bias^2(P^L(j)) - Bias^2(P^L(i))\end{aligned} \tag{6}$$

iii) For all $P^{L^+} \succ P^L(i)$ it has to hold that:

$$\begin{aligned}&EPE^{IP}(P^L(i)) \leq EPE^{IP}(P^{L^+})\\ &\Leftrightarrow Var(P^L(i)) + Var(Y) + Bias^2(P^L(i)) \leq Var(P^{L^+}) + Var(Y) + Bias^2(P^{L^+})\\ &\Leftrightarrow Var(P^L(i)) - Var(P^{L^+}) \leq Bias^2(P^{L^+}) - Bias^2(P^L(i))\end{aligned} \tag{7}$$

The representation of the comparison of EPE^{IP} of two categorizations in terms of differences in variances on the LHS and differences in squared biases on the RHS is useful for proving some of our later results. These conditions show that for a categorization $P^L(i)$ to be optimal for IP, the difference between its variance and the variance of any other categorization the agent could use has to be smaller or equal to the difference in squared biases between the other categorization and $P^L(i)$.

Proposition 2. *Existence of Optimal Categorization for Individual Prediction*
There always exists an optimal categorization for individual prediction.

The proof of the above result is trivial as the set of possible categorizations is finite.

A key question we are interested in is how the optimal way of categorizing depends on the environment the agent is in - whether it is deterministic or stochastic. Proposition 3 describes how the coarseness of the categorization(s) that minimize(s) the EPE^{IP}, depends on the exogenously given noise level and on the sample size of each object type. These are two of the key parameters describing the environment, i.e. the amount of stochasticity in the environment and the amount of information from past experiences that the agent has.

Proposition 3. *Coarseness of the Optimal Categorization(s) for Individual Prediction The coarseness of the optimal categorization(s) for IP is increasing in the noise level σ^2 and decreasing in the sample size n.*

Proposition 3 establishes that the higher the noise in the environment and the smaller the sample size of past experiences, the coarser the optimal categorization(s). The sketch of the proof is as follows. We examine the effect of the noise on the fulfillment of the optimality conditions for a categorization given in Definition 4. A change in the noise will only affect the differences in variances (i.e. the LHS of the three conditions). In Proposition 1 Part 1 we showed that as noise increases the difference in variance between a finer and a coarser categorization increases. Therefore as noise increases coarser categorizations become relatively more attractive compared to finer categorizations than before. We use this to show that an increase in the noise will never lead to a finer or an equally coarse categorization becoming optimal if it was not optimal before, but it can lead to a coarser categorization becoming optimal.

The proof for an increase in sample size is analogical. The increase in sample size also affects only the LHS of the conditions in Definition 4. More precisely, an increase in sample size means that the difference in variance between a finer and a coarser categorization becomes smaller. We use this insight from Proposition 1 Part 1 to show that as n increases only a finer categorization can become optimal if it was not optimal before. To sum up, the intuition is that noisy environments in which the sample size is limited make coarse categorizations attractive as coarse categorizations decrease the variance in prediction. The smaller the sample size, the greater the benefit of categorizing coarsely. Tables 2 and 3 in Appendix A illustrate the proposition.

4.2 Coordination

In this section we consider the case when players care only about coordinating their predictions with each other (see Definition 2). The setup is as follows. We assume that there are two players. Each of them has independently accumulated n experiences of each object type in the past. That is, in general each of them has sampled different objects. The two individuals face a one-off prediction problem in which they both observe the same object. They see the object's observable characteristics and have to make a prediction about the object's unobserved value. For simplicity, we assume that this next object is drawn from the discrete uniform distribution of object types. Each individual makes a prediction about the object's unobserved value equal to the average value in the category of past experiences she puts the object in. The goal of the two players is simply to coordinate their prediction on the next object, i.e. to minimize their expected prediction error from each other. In this setting the individual is not interested in the true unobserved value, but only in the other person's prediction. As before there are many alternative ways in which each of the two players can categorize her past experiences. Which way each of them chooses matters because it will determine their expected prediction error from each other. We represent the above situation formally as a one-shot game in which each player independently chooses a categorization. We then analyze the equilibrium properties of this game.

We begin by introducing Lemma 2 which extends our approach of bias-variance decomposition of expected prediction error to the case of coordination. Our main result in this section is Proposition 4 in which we analyze the equilibrium properties of the coordination game. In Proposition 5 we establish sufficient conditions for a connection between optimality for individual prediction and existence of Nash equilibria in the coordination game.

Lemma 2 shows that the $EPE^C(P_1, P_2)$ can be decomposed into a bias and a variance component.

Lemma 2. *Bias-Variance Decomposition of* $EPE^C(P_1, P_2)$

$$EPE^C(P_1,P_2) = \sum_{C_k \in P_1} \sum_{C_l \in P_2} \sum_{x_j \in (C_k^T, C_l^T)} p_j E[(\hat{Y}^{C_k} - \hat{Y}^{C_l})^2]$$

$$\begin{aligned} &= \sum_{C_k \in P_1} \sum_{x_j \in C_k^T} p_j Var(\hat{Y}^{C_k}) + \sum_{C_l \in P_2} \sum_{x_j \in C_l^T} p_j Var(\hat{Y}^{C_l}) \\ &\quad + \sum_{C_k \in P_1} \sum_{C_l \in P_2} \sum_{x_j \in (C_k^T, C_l^T)} p_j (E[\hat{Y}^{C_k}] - E[\hat{Y}^{C_l}])^2 \end{aligned} \tag{8}$$

$$= Var(P_1) + Var(P_2) + Bias^2(P_1, P_2)$$

That is, the $EPE^C(P_1, P_2)$ is equal to the sum of the expected variance of the category beliefs of the categorization P_1 that Player 1 uses, the expected variance of the category beliefs of the categorization P_2 that Player 2 uses, and the expected bias of the category beliefs of the two players from each other. It is increasing in each of these terms. Note that whenever both players use the exact same categorization, they will have the same expected value for each category belief and the bias of their categorizations from each other will be zero.

We want to understand the determinants of optimal categorizations when players want to coordinate their predictions.

Definition 5. Coordination Game

Our coordination game is a two-player game in which each player independently chooses a categorization from the set of available categorizations $\mathcal{P}$.[24] Note that all possible categorizations are available to each player, so that $\mathcal{P}$ is the same for both players. The preference relations of the players are represented by the $EPE^C(P_1, P_2)$, which is a function $f : \mathcal{P} \times \mathcal{P} \to \mathcal{R}$ mapping from the set of possible categorization profiles to the set of real numbers.

Next, focusing on pure strategy NE, we consider the relevant equilibrium conditions for the coordination game. Let $(P_1^L(i), P_2^M(j))$ denote a categorization profile such that $P_1^L(i)$ is the categorization that Player 1 uses and $P_2^M(j)$ is the categorization that Player 2 uses. The subscripts 1,2 denote the players. The two categorizations could be the same or different. The level of coarseness M could be a finer level of coarseness than L, it could be the same one, or it could be a coarser one.

Definition 6. NE Conditions of the Coordination Game

For $(P_1^L(i), P_2^M(j))$ to be a Nash Equilibrium (NE) the following conditions have to hold. Given the categorization of the opponent: i) no player should have an incentive to deviate to a finer categorization than the one she is currently using; ii) no player should have an incentive to deviate to another categorization at the same level of coarseness as the one she is currently using; and iii) no player should have an incentive to deviate to a coarser categorization than the one she is currently using.

Using the bias-variance decomposition from Lemma 2 this leads to the following conditions from the perspective of Player 1. The conditions for Player 2 are analogical. For a complete derivation and all conditions for both players, see Appendix B.

For all $P_1^{L^-} \prec P_1^L(i)$ it has to hold that:

$$\begin{aligned} &EPE(P_1^L(i), P_2^M(j)) \leq EPE(P_1^{L^-}, P_2^M(j)) \\ &\Leftrightarrow Var(P_1^L(i)) - Var(P_1^{L^-}) \leq Bias^2(P_1^{L^-}, P_2^M(j)) - Bias^2(P_1^L(i), P_2^M(j)) \end{aligned} \tag{9}$$

For all $P_1^L(k) \sim P_1^L(i)$ it has to hold that:

$$\begin{aligned} &EPE(P_1^L(i), P_2^M(j)) \leq EPE(P_1^L(k), P_2^M(j)) \\ &\Leftrightarrow Var(P_1^L(i)) - Var(P_1^L(k)) \leq Bias^2(P_1^L(k), P_2^M(j)) - Bias^2(P_1^L(i), P_2^M(j)) \end{aligned} \tag{10}$$

[24]Whether our game is a coordination game in the usual sense of a game with multiple equilibria is something to be analyzed in this section.

For all $P_1^{L+} \succ P_1^L(i)$ it has to hold that:

$$\begin{aligned} & EPE(P_1^L(i), P_2^M(j)) \leq EPE(P_1^{L+}, P_2^M(j)) \\ & \Leftrightarrow Var(P_1^L(i)) - Var(P_1^{L+}) \leq Bias^2(P_1^{L+}, P_2^M(j)) - Bias^2(P_1^L(i), P_2^M(j)) \end{aligned} \tag{11}$$

A categorization profile is a NE in the coordination game if for each player the variance of her categorization minus the variance of any categorization she could deviate to is smaller or equal to the difference in squared bias between the two players' categorizations, if she were to deviate minus the respective squared bias if she does not deviate. Note that this representation of $EPE^C(P_1, P_2)$ has its LHS (i.e. the difference in variance before and after a deviation) in common with our representation of $EPE^{IP}(P)$. This will be useful in our proofs connecting optimality in IP with equilibrium existence in the coordination game.

In Proposition 4 we analyze some equilibrium properties of the coordination game. This is one of our main Propositions. In Part 1 we rule out the possibility that a large class of asymmetric categorization profiles constitute NE. In Part 2 we show which symmetric equilibrium always exists in the coordination game and what the existence of other symmetric equilibria depends on. And in Part 3 we Pareto-rank all symmetric categorization profiles.

Proposition 4. *Equilibrium Properties of the Coordination Game*

Part 1. Ruling Out Existence of Asymmetric Equilibria

Categorization profiles such that the coarseness of the categorization of Player 1 is different from the coarseness of the categorization of Player 2 cannot be a NE for any $\sigma^2 > 0$. Categorization profiles such that the two players use different categorizations at the same level of coarseness cannot be a NE if the two categorizations have different expected values of the estimators for least one object type.

Part 2. Existence of Symmetric Equilibria

The number of symmetric categorization profiles that constitute a NE is a decreasing function of σ^2. Both players using the coarsest possible categorization profile is always a NE and may be the only one if σ^2 is sufficiently high.

Part 3. Pareto-ranking of Symmetric Categorization Profiles

The efficiency of symmetric categorization profiles increases with coarseness.

We now sketch briefly the proofs for Part 1. We first rule out the existence of asymmetric equilibria in which players use categorizations at different levels of coarseness. To rule out the existence of such equilibria it is sufficient to show that for all categorization profiles such that players use categorizations at different levels of coarseness, there exists at least one profitable deviation for at least one player. We show that the player who uses a finer categorization than the opponent always has an incentive to deviate to using the exact same categorization as the opponent (same coarseness and same categorization at that level). The intuition is that this is profitable because by moving to a coarser categorization this player decreases her variance in prediction. Moreover, using the same categorization as the opponent means that they will have no bias in their predictions from each other. Thus, both the bias and the variance components of $EPE^C(P_1, P_2)$ decrease if the player who uses a finer categorization switches to using the same categorization as the opponent.

The argument to rule out the existence of asymmetric equilibria in which players use different categorizations at the same level of coarseness is similar. Again one needs to show that there exists at least one profitable deviation. It is guaranteed that whenever the two categorizations have different expected values of the estimators for least one object type it is profitable for a player to deviate to using the same categorization as the opponent, as this will reduce their bias from each other to zero.

We now turn to Part 2. We consider all symmetric categorization profiles, i.e. all categorization profiles such that both players use the exact same categorization. We have shown in Part 1 that for any positive noise level no player ever has an incentive to use a categorization that is finer than the one the opponent is using. Likewise no player has an incentive to deviate to another equally coarse categorization different from the one the opponent is using, as this would create a bias in their predictions from each other without changing either person's variance in prediction. What remains to be considered is the question whether a player has an incentive to deviate to a coarser categorization than the one the opponent is using and if so, in what cases. In Appendix B we show that this depends on the noise level. We know from Proposition 1 Part 1 that keeping all else equal, the higher the noise level, the greater the decrease in variance from moving to a coarser categorization (note that the bias is not affected by a change in the noise level). Thus, even if a categorization profile constitutes a NE at some noise level, it might not be a NE at a higher noise level, as a deviation to a coarser categorization may become profitable. The only exception is the coarsest possible categorization because at the coarsest possible level there is no coarser categorization to deviate to. Thus, both using the coarsest possible categorization is always a NE. If noise is sufficiently high, it may be the only NE.

The proof of Part 3 proceeds in the following way. For all symmetric profiles, since players are using the exact same categorizations their bias of prediction from each other is always $Bias^2(P_1^L(i), P_2^L(i)) = 0$. Thus, the $EPE^C(P_1, P_2)$ of these categorization profiles depends only on the variances of the category beliefs: $EPE^C(P_1, P_2) = Var(P_1^L(i)) + Var(P_2^L(i))$. We know from Proposition 1 that the coarser a categorization, the smaller its variance. This means that for any positive noise level a coarser symmetric categorization profile always has a smaller $EPE^C(P_1, P_2)$ than a finer symmetric categorization profile. Symmetric categorization profiles are therefore Pareto-ranked with coarser profiles being more efficient than finer profiles for any positive noise level. In the case of $\sigma^2 = 0$, we have $Var(P) = 0$ and hence all symmetric categorization profiles are equally good. Note also that as all categorizations at the same level of coarseness have an equal variance (Proposition 1 Part 1), if there are multiple symmetric categorization profiles at the same level of coarseness, then they will be equally efficient.

The result that the categorization profile in which both players use the coarsest possible categorization is always a NE and that it is the Pareto-superior NE for any positive noise level is perhaps somewhat counterintuitive, as the prediction error with respect to the true object value may be large if both players use the coarsest possible categorization. However, if players only want to coordinate their predictions, what matters for them is their prediction error with respect to each other. By both using the same categorization they minimize the bias component of $EPE^C(P_1^L, P_2^L)$, and by both using the coarsest categorization they minimize its variance component. Note that whether there is a unique NE in the coordination game depends on the noise level and the sample size.

We now analyze the connection between optimality for individual prediction and NE in the coordination game. In Proposition 5 we identify sufficient conditions under which we can connect optimality for individual prediction with NE existence in the coordination game. In particular, we identify a set of sufficient conditions under which all NE such that players use categorizations at a finer level of coarseness than the individually optimal one are ruled out and in NE players will always categorize equally coarsely or more coarsely than the optimal level for individual prediction.

Proposition 5. *Connection Optimality IP and NE in Coordination Game*

Part 1. NE if Both Use Individually Optimal Categorization

If $P^L(i)$ is optimal in IP and $Bias^2(P^{L+}(l)) - Bias^2(P^L(i)) \leq Bias^2(P^{L+}(l), P^L(i))$ for all $P^{L+}(l) \succ P^L(i)$ such that $P^{L^+}(l) \notin R(P^L(i))$ then $(P^L(i), P^L(i))$ is a NE in the coordination game.

Part 2. No Equilibria in Categorizations Finer than Individually Optimal[25]

If $P^L(i)$ is the finest optimal categorization in IP, then there exists no symmetric equilibrium $(P^{L^-}(j), P^{L^-}(j))$ at any level finer than $P^L(i)$ involving any categorization that is connected by a path to $P^L(i)$, i.e. for any $P^{L^-}(j) \in R(P^L(i))$. Additionally, if $Bias^2(P^L(i)) - Bias^2(P^{L^-}(k)) \geq Bias^2(P^{L^-}(k), P^L(i))$ for all $P^{L^-}(k) \prec P^L(i)$ such that $P^{L^-}(k) \notin R(P^L(i))$ then there exists no symmetric equilibrium $(P^{L^-}(k), P^{L^-}(k))$ at any level finer than $P^L(i)$ involving any categorization that is not connected by a path to $P^L(i)$.

Part 3. Symmetric Equilibria at Levels Coarser than the Individually Optimal

If $P^L(i)$ is optimal in IP and the condition $Var(P^{L^+}(j)) - Var(P^{L^{++}}) \leq Bias^2(P^{L^{++}}, P^{L^+}(j))$ holds for each $P^{L^{++}} \succ P^{L^+}(j)$, then $(P^{L^+}(j), P^{L^+}(j))$ is a NE in the coordination game.

Proposition 5 Part 1 provides a sufficient (though not necessary) condition for the optimality of a categorization for individual prediction to imply that a categorization

[25]Note that by ruling out existence of equilibria in symmetric categorizations at a level of coarseness below that of the finest individually optimal categorization, we are ruling out the existence of any equilibrium that involves a level of coarseness finer than the finest individually optimal, as the existence of asymmetric equilibria was ruled out in Proposition 4 Part 1.

profile such that both players use this same individually optimal categorization is a NE in the coordination game. The sketch of the proof is the following. For $(P^L(i), P^L(i))$ to be a NE we need that no player has an incentive to deviate to a finer, to another equally coarse or to a coarser categorization. Proposition 4 Part 1 implies that no player ever has an incentive to deviate to a finer categorization than the one his opponent uses or to another categorization that is equally coarse but different than the one the other player uses. So to show that $(P^L(i), P^L(i))$ is a NE we need to show that no player has a profitable deviation to a coarser categorization. There are two cases that we need to consider: a deviation to a coarser categorization that is connected by a path to $P^L(i)$ and a deviation to a coarser categorization that is not connected by a path to $P^L(i)$.

In Appendix B we use Proposition 1 to show that if $P^L(i)$ is optimal for individual prediction, no player has an incentive to deviate to any coarser categorization that is connected by a path to $P^L(i)$. We cannot show that in general a player would not have a profitable deviation to a categorization that is not connected by a path to $P^L(i)$. We establish a sufficient condition under which individual optimality of a categorization implies that no player has a profitable deviation to another categorization that is not connected by a path to $P^L(i)$. This sufficient condition for no profitable deviation to a coarser categorization not connected by a path to $P^L(i)$ is that the difference in bias for individual prediction of any coarser categorization that is not connected by a path and the bias of the categorization that is optimal for individual prediction is smaller than or equal to the bias of the two players' predictions from each other in the coordination game if they use the above categorizations, respectively. Note that this is a sufficient (but not necessary) condition for individual optimality to imply that both players using the individually optimal categorization is a NE in the coordination game.

Proposition 5 Part 2 guarantees that there is no symmetric NE in the coordination game in which players use a categorization that is finer than the finest individually optimal categorization $P^L(i)$ and connected by a path to it. It also gives a sufficient condition under which, if $P^L(i)$ is the finest individually optimal categorization, there are also no NE in symmetric categorizations that are finer than the finest individually optimal categorization and not connected by a path to it. We present a brief sketch of the proof. We need to show that if $P^L(i)$ is the finest optimal categorization for individual prediction, then there exists a profitable deviation for at least one player from any symmetric categorization profile at a finer level of coarseness that is connected by a path to $P^L(i)$. For that purpose we consider the deviation to the finest individually optimal categorization and show that this is always profitable. This proves the first part of the claim. We then derive sufficient conditions for individual optimality of $P^L(i)$ to imply the ruling out of existence of any equilibria in finer symmetric categorizations that are not connected by a path to $P^L(i)$.

Proposition 5 Part 3 gives a sufficient condition to for the existence of NE in symmetric categorizations at levels of coarseness above the individually optimal one. To show the existence of a symmetric equilibrium we need to show that no player has an incentive to deviate to a finer, to another equally coarse or to a coarser categorization. The first two conditions are guaranteed to hold, as we have shown in Proposition 4 Part 1. For the last condition to hold it is necessary and sufficient that the difference in variance between the categorization $P^{L^+}(j)$ and any categorization coarser than it is smaller or equal to the bias of the two players from using the respective categorizations.

Thus, Proposition 5 gives us conditions that are sufficient to ensure that if a categorization is optimal for IP, then both players using this categorization is a NE; both players using any finer symmetric categorization profile cannot be a NE; and there exist NE in symmetric categorizations at levels of coarseness above the individually optimal one.

4.3 Individual Prediction and Coordination

In the previous two sections we analyzed the properties of optimal categorizations in case an individual is interested only in making a correct prediction about the true unobserved value of the object and the properties of different categorization profiles in the game in which players are only interested in coordinating their predictions. In many cases economic agents will care both about making correct predictions and about coordinating their predictions with each other. In this section we represent this situation formally with what we call an *IP&C* game and we discuss the properties of different categorization profiles in this game. In Lemma 3 we extend the bias-variance decomposition to the $EPE^{IP\&C}$. Below we present it from the perspective of Player 1.

Lemma 3. *General Bias-Variance Decomposition* $EPE_1^{IP\&C}(P_1, P_2)$

$$\begin{aligned}EPE_1^{IP\&C}(P_1, P_2) = w\Big[\sum_{C_k\in P_1}\sum_{x_j\in C_k^T} p_j(Var(\hat{Y}^{C_k}) + Var(Y_j) + (E[\hat{Y}^{C_k}] - \mu_j)^2)\Big]\\ + (1-w)\Big[\sum_{C_k\in P_1}\sum_{C_l\in P_2}\sum_{x_j\in(C_k^T,C_l^T)} p_j(Var(\hat{Y}^{C_k}) + Var(\hat{Y}^{C_l}) + (E[\hat{Y}^{C_k}] - E[\hat{Y}^{C_l}])^2)\Big]\end{aligned}$$

$$\begin{aligned}= wVar(P_1) + wVar(Y) + wBias^2(P_1)\\ + (1-w)Var(P_1) + (1-w)Var(P_2) + (1-w)Bias^2(P_1, P_2)\end{aligned} \tag{12}$$

The $EPE_2^{IP\&C}(P_1, P_2)$ of Player 2 can be written in an analogical way. The $EPE^{IP\&C}$ is the convex combination of the EPE^{IP} and the EPE^{C} with a weight w on individual prediction and $(1-w)$ on coordination, respectively. It is increasing in the variance of the categorization the player uses, in the variance of the underlying object types in the population, in the bias of the categorization that the player uses, in the variance of the other player's categorization, and in the biases of their predictions from each other. We now formalize the situation described above and we call it an *IP&C* game.

Definition 7. *IP&C* game

An *IP&C* game is a version of the coordination game such that the preference relations of the two players are represented by $EPE_1^{IP\&C}$ and $EPE_2^{IP\&C}$, respectively, with each player i placing a weight w_i on individual prediction and a weight $(1-w_i)$ on coordination.

We now describe the NE conditions in the $IP\&C$ game. For simplicity we assume in the analysis below that $w_1 = w_2 = w$, i.e. the two players have identical preferences.

Definition 8. NE conditions of the $IP\&C$ game

For $(P_1^L(i), P_2^M(j))$ to be a Nash Equilibrium (NE) in the $IP\&C$ game the following conditions have to hold. Given the categorization of the opponent: i) no player should have an incentive to deviate to a finer categorization than the one she is currently using; ii) no player should have an incentive to deviate to another categorization at the same level of coarseness as the one she is currently using; and iii) no player should have an incentive to deviate to a coarser categorization than the one she is currently using. Using the bias-variance decomposition from Lemma 3 to write the above conditions, we get the following from the perspective of Player 1. The conditions for Player 2 are analogical. For a complete derivation and all conditions for both players, see Appendix B.

i) For all $P_1^{L-} \prec P_1^L(i)$ it should hold that:

$$\begin{aligned} &EPE_1^{IP\&C}(P_1^L(i), P_2^M(j)) \leq EPE_1^{IP\&C}(P_1^{L-}, P_2^M(j)) \\ &\Leftrightarrow Var(P_1^L(i)) - Var(P_1^{L^-}) \\ &\quad \leq w\Big[Bias^2(P_1^{L^-}) - Bias^2(P_1^L(i))\Big] \\ &\quad + (1-w)\Big[Bias^2(P_1^{L^-}, P_2^M(j)) - Bias^2(P_1^L(i), P_2^M(j))\Big] \end{aligned} \tag{13}$$

ii) For all $P_1^L(k) \sim P_1^L(i)$ it should hold that:

$$\begin{aligned} &EPE_1^{IP\&C}(P_1^L(i), P_2^M(j)) \leq EPE_1^{IP\&C}(P_1^L(k), P_2^M(j)) \\ &\Leftrightarrow Var(P_1^L(i)) - Var(P_1^L(k)) \\ &\quad \leq w\Big[Bias^2(P_1^L(k)) - Bias^2(P_1^L(i))\Big] \\ &\quad + (1-w)\Big[Bias^2(P_1^L(k)), P_2^M(j)) - Bias^2(P_1^L(i), P_2^M(j))\Big] \end{aligned} \tag{14}$$

iii) For all $P_1^{L^+} \succ P_1^L(i)$ it should hold that:

$$\begin{aligned} &EPE_1^{IP\&C}(P_1^L(i), P_2^M(j)) \leq EPE_1^{IP\&C}(P_1^{L^+}, P_2^M(j)) \\ &\Leftrightarrow Var(P_1^L(i)) - Var(P_1^{L^+}) \\ &\quad \leq w\Big[Bias^2(P_1^{L^+}) - Bias^2(P_1^L(i))\Big] \\ &\quad + (1-w)\Big[Bias^2(P_1^{L^+}, P_2^M(j)) - Bias^2(P_1^L(i), P_2^M(j))\Big] \end{aligned} \tag{15}$$

For a categorization profile to be a NE in the $IP\&C$ game, the change in variance from deviating to any other categorization has to be smaller or equal to the weighted sum of the difference between the new and the old squared bias in individual prediction and the difference between the new and the old squared bias in the coordination game.

In Proposition 6 we characterize some equilibrium properties of the $IP\&C$ game. Parts 1 and 2 deal with the connection between a categorization's optimality in IP with NE existence in the $IP\&C$ game. In Part 3 we present a condition under which we can

Pareto rank all symmetric categorization profiles in the $IP\&C$ game.

Proposition 6. *Equilibrium Properties of the $IP\&C$ game*

Part 1. Ruling Out Existence of Some Asymmetric Equilibria

If $P^L(i)$ is the finest optimal categorization for IP, then any categorization profile $(P^{L^-}, P^L(i))$ can never be a NE in the $IP\&C$. Furthermore, if $P^L(i)$ is the finest optimal categorization for IP and $Bias^2(P^{L^-}(j), P^{L^{--}}(k)) \leq Bias^2(P^L(i), P^{L^{--}}(k))$, any categorization profile $(P^{L^-}(j), P^{L^{--}}(k))$ such that $P^{L^-}(j) \in R(P^L(i))$ and $P^{L^{--}}(k) \in R(P^L(i))$ and $P^{L^{--}}(k) \in R(P^{L^-}(k))$ can never be a NE in the $IP\&C$ game.

Part 2. Symmetric Equilibria

If $P^L(i)$ is optimal for IP and $(P^L(i), P^L(i))$ is a NE in C, then $(P^L(i), P^L(i))$ is a NE in $IP\&C$. Furthermore, if $P^L(i)$ is the finest optimal categorization for IP, then any $(P^{L^-}(k), P^{L^-}(k))$ such that $P^{L^-}(k) \prec P^L(i)$ and $P^{L^-}(k) \in R(P^L(i))$ is not a NE in $IP\&C$.

Part 3. Pareto-ranking of Symmetric Profiles

Any coarser symmetric categorization profile $(P^{L^+}(j), P^{L^+}(j))$ is Pareto-superior to any finer symmetric categorization profile $(P^L(i), P^L(i))$ in the $IP\&C$ game if and only if the following condition holds for any $P^{L^+}(j) \succ P^L(i)$:

$$w \leq \frac{2\Big[Var(P^L(i)) - Var(P^{L^+}(j))\Big]}{Bias^2(P^{L^+}(j)) - Bias^2(P^L(i)) + Var(P^L(i)) - Var(P^{L^+}(j))} \tag{16}$$

Proposition 6 Part 1 says that if a categorization $P^L(i)$ is the finest optimal categorization for individual prediction, then there exists no asymmetric equilibrium in the $IP\&C$ game such that one player uses the categorization that is optimal for individual prediction and the other player uses a finer categorization. Moreover, it provides a sufficient (but not necessary) condition to ensure that there exists no asymmetric equilibrium in the $IP\&C$ game such that the two players use categorizations at different levels of coarseness both finer than the finest individually optimal one if those categorizations are linked by a path to each other as well as to the finest individually optimal categorization. The sketch of the proof is the following. To rule out the existence of an equilibrium we need to show that there always exists a profitable deviation from the respective categorization profile for at least one player. In the case of any categorization profile such that one player uses the individually optimal categorization and the other player uses a finer one, we show that it is always profitable for the player using the finer categorization to deviate to the individually optimal one. Deviating to the individually optimal categorization is profitable both because the player then makes a smaller error in individual prediction and because by using the same categorization as the other player there will be no bias in their predictions with respect to each other. To show that there exists no NE such that players use categorizations at different levels of coarseness below the finest individually optimal one, we show that it is always profitable for the player who uses the finer categorization of the two to deviate to the individually optimal categorization.

The first statement in Part 2 says that if a categorization is optimal for individual prediction and if both players using this categorization is also a NE in the coordination case, then it will also be a NE in the $IP\&C$ game. The proof is straightforward as optimality in individual prediction and NE existence in the coordination game imply that there is no rationale to deviate from this categorization profile in the $IP\&C$ game, neither from an individual prediction perspective nor from a coordination point of view. The next statement is that if a categorization is the finest optimal for individual prediction, then both players using a finer symmetric categorization profile that is connected by a path to the individually optimal categorization is not a NE in the $IP\&C$ case. This follows because from the individual perspective it is profitable to deviate to the individually optimal categorization, and from a coordination perspective we established in Proposition 5 Part 2 that there is no finer symmetric NE in categorizations connected by a path to the individually optimal categorization.

Part 3 of Proposition 6 gives us necessary and sufficient conditions under which any coarser symmetric categorization profile is Pareto-superior to any finer symmetric categorization profile in the $IP\&C$ case. As w is the weight placed on individual predictions, this result means that symmetric categorization profiles can be Pareto ranked on their coarseness, with coarser profiles being the more efficient, also in the $IP\&C$ case as long as individuals assign sufficient weight to being coordinated with each other.

4.4 Summary of the Static Model

In our static model we considered the following one-off prediction problem. An individual encounters an object and has to make a prediction of its unobserved value. The individual makes this prediction by placing the object in a category based on its observable characteristics and predicts that its unobserved value will be equal to the average unobserved value of all past experiences she has in this category. In our static framework, we assumed that the agent has already accumulated a number of past experiences, for simplicity, we take this to be n observations of each object type. She has many alternative ways of organizing her past experiences into categories. We considered what determines which categorization of her past experiences will help her make the best prediction on the next object encountered. We analyzed three variants of this problem.

First, we considered the case when the individual is only interested in predicting the true values of the next object. We find that the level of coarseness of the optimal categorization(s) for IP depends on the noise level in the environment and on the sample size of past experiences the agent has available. In deterministic environments the finest categorization is the best as its category beliefs are unbiased and there is no variance in prediction, so they are also as consistent as the beliefs of any other categorization. In a stochastic environment in which the agent has a limited amount of past experiences, however, the higher the noise and the smaller the sample size, the coarser the optimal categorization. In such situations coarse categorizations might perform better than fine categorizations as by lumping together a larger number of objects per category they decrease the variance in prediction. Thus, although their category beliefs are biased estimators, they may have a lower EPE^{IP} as their beliefs are more consistent estimators than the category beliefs of finer categorizations. The above result suggests one rationale for why economic agents may use coarse categorizations even if there are no costs of categorizing finely and no bounds on computational power. That is, they may categorize coarsely as coarse categorization leads to better predictions in stochastic environments.

Second, we considered a one-off coordination problem in which two agents want to coordinate their predictions on the next object encountered. We find that if players have an incentive to coordinate, using categorizations at different levels of coarseness is not optimal. There can be only NE in which players use categorizations at the same level of coarseness. But a given symmetric categorization profile may or may not be a NE depending on the noise level. If the environment is deterministic, all symmetric categorization profiles are NE and are equally efficient. The higher the noise in the environment, the greater the incentive to deviate to a coarser categorization. Thus, as noise increases, finer symmetric categorization profiles gradually stop being NE. If noise is sufficiently high, the unique NE is both categorizing at the coarsest possible level. Interestingly, both using the coarsest possible categorization is a NE for any noise level. Moreover, coarser symmetric categorization profiles are Pareto-superior to finer symmetric categorization profiles for any positive noise level. The intuition is that, on the one hand, by using a symmetric categorization profile, players have no bias from one another. On the other hand, by using coarse categorizations, they minimize their variance in prediction. These two results suggest pressure to use coarse categorization in environments in which players want to coordinate their predictions. When analyzing the connection between optimality of a categorization for individual prediction and NE in the coordination game, we establish sufficient conditions under which if a given categorization is optimal for IP, there are no NE in categorizations finer than the finest individually optimal one but only NE at the same level of coarseness or coarser. These results together suggest that the attempt to coordinate predictions with others may be a further rationale for coarse categorization, additional and complementary to those so far discussed in the literature.

Third, we considered the case when players want both to predict the true object value correctly and to coordinate their predictions with each other. Naturally, the analysis of this case is much more complex. We analyzed some equilibrium properties and showed that we can rule out some classes of profiles involving categorizations below the individually optimal level of coarseness.

5 Dynamic Model

There are many alternative approaches to analyzing categorization. Our static model presented and analyzed in sections 3 and 4 provides one possible perspective. In it we made some assumptions in order to facilitate analytical tractability. In this section we present a complementary dynamic framework based on a number of alternative assumptions. The basic underlying problem that we consider is the same. An agent observes some characteristics of an object and has to predict its unobserved value. She uses her past experiences stored in categories to make predictions. Our focus is again on analyzing the properties of different alternative ways of categorizing and the tension between fine and coarse categorization. And as before we consider three variants of this problem: individual prediction, coordination, and individual prediction and coordination.

The main conceptual differences between the static and the dynamic model are the following. The first one relates to the nature of the prediction situation we consider. In the static model we assumed that the agent has already accumulated a number of past experiences and faces the task of making a one-off prediction. In the dynamic model, we consider a dynamic situation instead. The agent faces a stream of objects, one in each period. She has to make a prediction of the object's unobserved value each period as she encounters it. Second, in the static model we compared the expected prediction error of all possible categorizations without considering how an agent learns to categorize. In the dynamic model we consider the question whether an agent can learn to adapt by learning the most appropriate categorization for a given environment. The agent starts out with a random set of possible categorizations and searches for the best categorizations to use. This makes the dynamic framework also suitable to analyze adaptation in non-stationary environments. Third, a further assumption that we relax here is that we allow not only for disjoint partitionings of the object set but also for hierarchical and incomplete categorizations. Hierarchical categorization abound in the real world. For example, someone could distinguish the category 'any workers' as a possible input factor, and at the same time the subordinate category 'British workers' and the category 'other workers'. An incomplete categorization is one that lacks a category for some type of objects, e.g. the categorization containing only categories 'European food' and 'Asian food'.

5.1 Basic setup

We now present the basic setup of the dynamic model that is used in all three cases of individual prediction, coordination, and $IP\&C$.[26] An agent starts out with a pool containing a limited number of different categorizations. This number is much smaller than the number of all possible categorizations. At the outset the agent starts with a random set of categorizations. Each period she encounters an object. She chooses a categorization from the pool to use and a category within this categorization. She makes a prediction of the unobserved values of the object equal to the category belief for the respective category. Afterwards the true object values are revealed and the agent uses them to calculate her payoff, to update her beliefs for the category and categorization

[26]Our dynamic model is non-Bayesian. For discussions and comparisons of the predictions of categorical and Bayesian decision makers, we refer to Peski (2011) and Mullainathan (2002).

she has used, and to update her valuation of their performance.

To analyze the properties of the categorizations that agents learn if they are only interested in coordinating their predictions with each other, we additionally assume that there are a number of agents and that two of them are randomly matched each period. The two matched agents encounter the same object and each of them independently chooses some categorization and category to use as described above, and makes her prediction about the unobserved value of the object. After predictions are made both the true object values and the prediction of the other are revealed to the agent. Agents use the true object value to update their beliefs. As they only care about coordinating their predictions with each other, they use their prediction error from the other agent's prediction to update their valuation of how good the category and categorization used are for coordination.

The set up for the case when they care both about predicting the true object value and about coordinating their prediction with each other is analogical. The only difference is that each agent uses a weighted average of the prediction error from the true object value and the prediction error from the other agent to update her valuation of how good the category and categorization used are.

5.2 Using Categories to Predict

The agent encounters an object each period. The object's observable characteristics are described by a vector $x \in \{0,1\}^l$ and the object's unobservable characteristics by a vector $y \in \mathbb{R}^m$. The agent makes a prediction about the object's unobservable characteristics by putting the object in a category based on its observable characteristics. A category is represented as a rule consisting of a filter string and a belief attached to it. The category filter can consist of 1, 0, and #, where # means that the attribute in this position is not taken into account.[27] As an example, in the case of two object attributes, the set of possible filters is given by $\mathcal{F} = \{11, 10, 01, 00, 1\#, 0\#, \#1, \#0, \#\#\}$.[28] The filter 11 can only be used on objects of type 11. The filter 1# is equivalent to the category type $\{11, 10\}$ from our static model and can be applied both to objects type 11 and 10. The filter ## is applicable to all four object types: $\{11, 10, 01, 00\}$. Thus, the more # signs in the category filter, the more object types will be matched by this category and therefore the coarser it is.[29]

A categorization is a finite set of rules. For example, the categorization $\{0\#, 01, 1\#\}$

[27] The use of filter strings is convenient for the computational model. For a similar use of filters, see Arthur et al. (1997).

[28] The number of possible categories here is different than in the static model. This is due to the well-known linear separability issue discussed in Minsky and Papert (1969) and Rumelhart and McClelland (1986). That is, it is impossible to have a filter that places two objects that differ on all possible object attributes together without including all other objects. E.g. the objects 01 and 10 cannot be put in one separate category without including 11 and 00. It is easy to solve this by adding an additional dimension that is a nonlinear combination of the present features. We simply abstract from this issue as solving it would unnecessarily complicate the model without adding new insights.

[29] The number of category filters is given by $|\mathcal{F}| = 3^l$.

has three categories. In contrast to the static model, a categorization here is not necessarily a disjoint partitioning of the object set. For example, in the categorization $\{0\#, 01, 1\#\}$, both $0\#$ and 01 can be used to make predictions about objects type of 01. The agent can also have incomplete categorizations, i.e. such that do not completely cover the set of object types. For example, the categorization $\{1\#, 00\}$ does not contain a rule for objects type 01. We assume that a categorization cannot contain the same rule filter twice.[30]

At any given time the agent has a pool containing a number of categorizations. The agent keeps track of the performance of each categorization using a measure called strength. Each period she chooses a categorization from the pool using the standard discrete choice framework known as the logit rule (McFadden, 1973). The probability of choosing a categorization i in period t is:

$$p_t(i) = \frac{e^{\beta s_t(i)}}{\sum_j e^{\beta s_t(j)}} , \tag{17}$$

where $s_t(i)$ ($s_t(j)$) is the strength of categorization i (j) in period t and β is a parameter determining the sensitivity of choice to categorization strength. For $\beta = 0$ choice is uniform random, i.e. independent of strength. If $\beta > 0$ stronger categorizations have a higher probability of being selected. If β is sufficiently large, the strongest categorization is selected with probability 1.

If there is only one category filter in the selected categorization that is applicable to the object's observed attributes, the respective category is used. If more than one category filter fits the object's attributes, the agent selects among them with the logit rule and is more likely to choose the one that has made better predictions in the past (positive β).

The agent makes a prediction of the unobserved values of the object equal to the belief of the selected category. If no category filter in the categorization fits the object's attributes, the agent makes a prediction of the unobserved value of the object using the belief of the categorization. If the object has been put in a category that did not contain any observations yet, the agent makes a random prediction (drawn from the distribution of objects for this object type in the environment).

5.3 Learning

After the prediction is made the true unobserved values of the object are revealed. She uses this true value to update the beliefs of the categorization and category she has

[30] Allowing for non-disjoint partitionings of the object set means that there are even more categorizations or models of the world that the agent could use to make predictions than in the case considered in our static model. If we restrict attention to those categorizations whose number of categories is smaller or equal to the number of different object types in the population, the number of possible categorizations including those that are not a disjoint partitioning of the object set is: $|\mathcal{P}| = \sum_{i=1}^{2^l} \binom{3^l}{i}$. For example, with $l = 2$ attributes and $|O^T| = 2^l = 4$, the number of possible categorizations is: $|\mathcal{P}| = \sum_{i=1}^{4} \binom{9}{i} = \binom{9}{1} + \binom{9}{2} + \binom{9}{3} + \binom{9}{4}$.

used. The belief of a category (or categorization, respectively) in period t is a weighted average of the y values of the objects experienced with this category (or categorization, respectively) in the past, putting more weight on recent experiences. It is equal to:

$$\hat{y}_t = \frac{\hat{y}_{t-1}(1-d) + y_t}{n_{t-1}(1-d) + z_t} , \tag{18}$$

where $\hat{y}_{t-1}$ is the category (categorization) belief from the previous period; n_{t-1} is the number of objects that have been put in this category (categorization) in the past; y_t is the y-value of the object in period t if it is put in this category (categorization); and $z_t \in \{0, 1\}$ is an indicator variable showing whether the category (categorization) was used in period t. We discount past observations with a factor d, $0 < d < 1$. The discount factor d is crucial for adaptation in changing environments, as it allows the agent to forget old irrelevant experiences. Note that effectively the discount factor operates as a proxy for sample size. The higher the value of d the smaller the weight of previous experiences and thus the smaller the sample size that the agent uses. A $d = 0$ would be equivalent to no discounting.

We discount the observations in each category and categorization each period. This means that if a category is not used, its category belief stays the same ($y_t = 0$ and $z_t = 0$), but the information in it is discounted. Note that if a category is more general it will be used more often. The more often a category is used, the greater the sample size the category belief is based on and thus the smaller its variance in prediction (as in the static model).

The prediction error of an agent in period t is equal to the mean squared error between the predicted values and the true unobserved values of the object, i.e.:

$$PE_t = \sum_i (\hat{y}_{ti} - y_{ti})^2 , \tag{19}$$

where $\hat{y}_{ti}$ is the predicted value of the i-th unobserved attribute of the object encountered in period t and y_{ti} is the respective true value. The agent receives a payoff in each period that is equal to

$$\pi_t = c - PE_t , \tag{20}$$

where c is a constant. If the prediction error is 0, the category and categorization get the highest possible payoff c. As prediction error increases, payoff decreases.[31] The agent uses the payoff to update her valuation of how good the category/categorization used is:

$$s_t(i) = s_{t-1}(i)(1-\alpha) + \pi_t \alpha , \tag{21}$$

where $s_t(i)$ is the strength of category/ categorization i in period t, $s_{t-1}(i)$ is the strength of the category/ categorization i in period $t-1$, π_t is the payoff in period t, and $0 \leq \alpha \leq 1$ is a learning rate parameter. The higher the α, the greater the weight placed on how

[31]The constant is chosen so that the payoff is non-negative.

useful the category and categorization were for the most recent experience relative to previous experiences.[32]

Besides the updating of beliefs and the reinforcement of the categorizations and categories currently considered by the agent, she also uses her experience to update the set of categorizations that she considers. The agent learns from experience which categorizations are useful in a given environment. That is, at any given time the agent has a pool containing a limited number of categorizations that she uses. At the outset the categorizations in this pool are random. She tries them out for a number of periods, updating her valuation of how good they are. Every e periods she throws the weakest categorization (the one with the lowest strength) out of the pool and creates a new one to experiment with. To create a new categorization she chooses two of the existing categorizations from the pool with the logit rule (positive β). She randomly picks categories from these selected categorizations to incorporate them in the new categorization. Each bit in a new category may be mutated with some small probability p_m. The new category receives the strength of the category it is derived from. If the sum of the number of categories in the two original categorizations is odd, then the new categorization consists of the average number of categories in the two old categorizations, rounded up or down with equal probability. If the sum of the number of categories in the two original categorizations is even, then the new categorization consists of the average number of categories in the two old categorizations, rounded up or down by one with some small probability p_c. The strength of the new categorization is equal to the average of the strengths of the two categorizations it is derived from.[33]

This search mechanism allows the agent to adapt her view of the world to the environment. The idea is that as the agent gathers experiences she reevaluates her way of looking at the world. Periodically she looks at the experience that she has gathered and she abandons views of the world that performed bad in the past. She forms new views of the world by combining useful information that she has obtained and by experimenting. For example, if categorizations in the pool consisting of few categories have been more successful in the past, the agent is more likely to choose them as a basis for forming a new categorization. This new categorization is then also more likely to consist of few categories.

6 Analysis of the Dynamic Model

In this section we present a numerical analysis of the dynamic model. The analysis focuses on the dynamics of some properties of the categorizations that an agent learns

[32] Placing a higher weight on recent experiences is useful for the agent when she knows that the environment she is in is changing. It helps her adapt her view of the world more quickly.

[33] The above mechanism is a form of a genetic algorithm, closely related to rule-based systems; see e.g. Holland et al. (1989). A main difference to commonly used rule-based systems however is that while in rule-based systems the genetic step is performed with respect to rules, in our algorithm it is performed with respect to categorizations, i.e. sets of rules.

over time. We investigate how the categorizations that the agent learns are determined by the situation she is in. To facilitate tractability we consider the case of two object attributes and four object types, but the model is applicable to higher numbers of object attributes. There are a number of general parameters that we keep fixed throughout the analysis. They are shown in Table 1. At any given time the agent has eight categorizations in the pool to choose from. The evolution step is performed every $e = 800$ periods. All categorizations and categories are initialized with a strength of 0.5 and we rescale strengths between 0 and 1 each period. The β for the categorization choice is 5 both for the choice of categorization to use each period and for the choice of categorizations in the genetic step. The β for category choice is 2. The learning rate α both for updating the strength of categories and categorizations is 0.001. All results reported are based on an average of 1000 runs. We use a moving average window of 1000 periods. In the analysis we report we use the function $y_i = x_i + \epsilon$.

Table 1: Parameters Dynamic Model

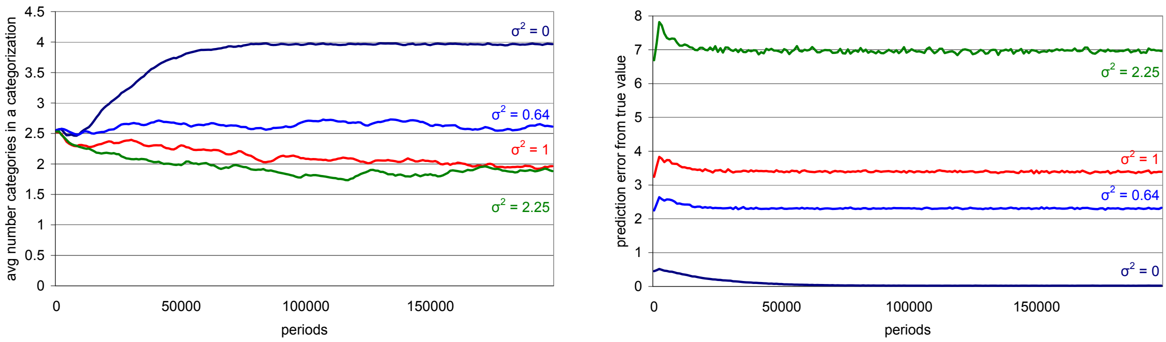
Parameter	Value
Number of categorizations in the pool	8
Number of periods between evolution steps	800
Initial categorization strength	0.5
Initial category strength	0.5
Rescaling of strength $[0, 1]$	yes
β categorization choice	5.0
β category choice	2.0
α	0.001
p_m	0.333
p_c	0.066
Number of runs	1000
Number of periods moving average window	1000
σ^2	0.00 - 2.25 (see details in text)
Discounting	0.05 - 0.50 (see details in text)
Function $y = f(x) + \epsilon$	$y_i = x_i + \epsilon$

We begin by analyzing the case when the agent cares only about individual prediction, followed by the case when she cares only about coordination. Finally, we present the case when agents care about both.

6.1 Individual Prediction

We first consider the case of an individual who is only interested in predicting the true object values correctly. We analyze the effect of the stochasticity of the environment and of the discounting parameter on the coarseness of the categorizations that the agent learns.

Figure 2a represents the coarseness of the categorizations that the agent learns over time for various levels of noise in the environment and for a fixed discounting factor. The indicator of coarseness on the y-axis is the average number of categories in the categorizations in the agent's pool. The higher the average number of categories in a categorization, the finer it is. Thus, a low average number of categories in a categorization in the agent's pool indicates coarse categorizing. We focus on the long horizon as we are interested in the properties of the categorizations that the agent learns in alternative environments in the long term. Figure 2b illustrates the corresponding dynamics of the prediction error that the agent makes with respect to the true values of the object. The parameters used in these figures are the same as in Table 1. In Figure 2 we vary the level of σ^2 keeping the discount factor fixed at 0.3.



a. Coarseness

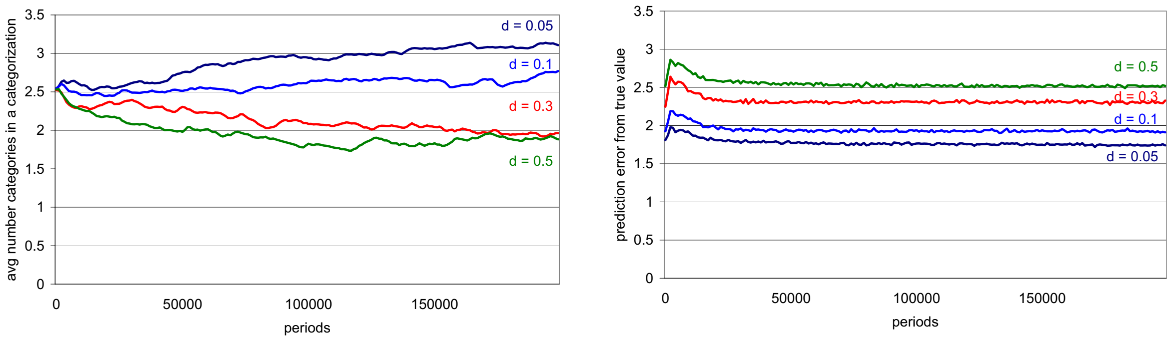
b. Prediction Error

Figure 2: Effect of noise on coarseness

A noise level of $\sigma^2 = 0$ corresponds to a deterministic environment. We can see in Figure 2a that in such an environment the categorizations that the agent learns consist on average of the maximum possible number of categories. As the noise level increases to $\sigma^2 = 0.64$, the categorizations that the agent learns become coarser with an average number of categories around 2.6. For even higher noise levels of $\sigma^2 = 1$ and $\sigma^2 = 2.25$, the categorizations that the agent learns at the end of a large number of periods converge towards an average number of categories just above and below 2, respectively. In Figure 2b we can see that in a deterministic environment the agent learns to make no prediction error by using categorizations with the highest possible number of categories. As the noise level increases, it is not possible for the agent not to make any prediction errors, but the prediction error she makes decreases as she learns categorizations which are more suitable for the respective type of environment.

Numerical Result 1. *For a given (positive) level of discounting, the categorizations that the agent learns are coarser the higher the noise level.*

Figure 3a represents the effect of the discounting factor on the coarseness of the categorizations that the agent learns over time for a given noise level. The discounting is a proxy for sample size. Thus, we see that the higher the discounting, i.e. the smaller the sample size of past experiences the agent has available, the coarser the categorizations that the agent learns. Under a discount factor of $d = 0.05$ the agent learns categorizations that consist on average of more than three categories. Under a discount factor $d = 0.5$ she learns categorizations that consist on average of less than two categories. In Figure 3b we see the corresponding prediction error that the agent makes. The prediction error is higher the higher the discounting, but in each case it decreases over time as the agent learns the categorizations that are suitable. The parameters used are again the ones shown in Table 1. The noise is fixed at $\sigma^2 = 0.64$.



a. Coarseness

b. Prediction Error

Figure 3: Effect of discounting on coarseness

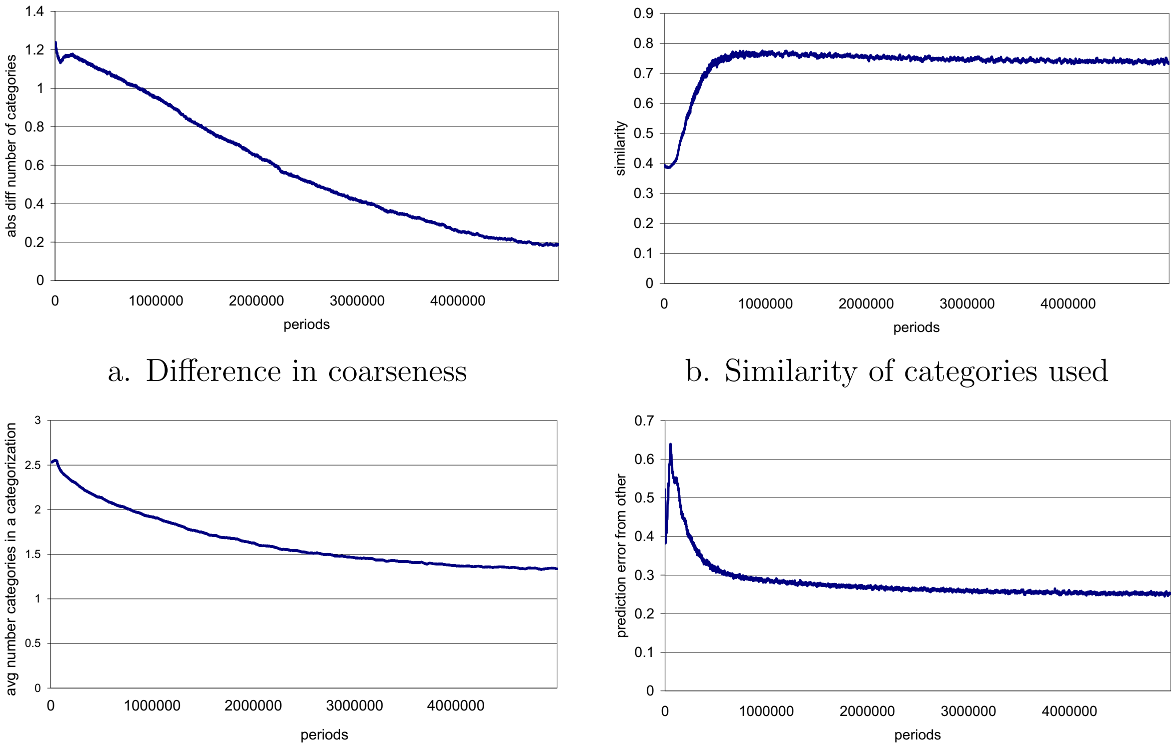
Numerical Result 2. *For a given (positive) level of noise, the categorizations that the agent learns are coarser the higher the discounting.*

In the static model we developed earlier we showed that the higher the noise level and the smaller the sample size the agent has available, the coarser the optimal categorization(s) (Proposition 3). Our Numerical Results 1 and 2 are in line with this finding. We have shown that the agent of our dynamic model learns to categorize finely in a deterministic environment and to categorize more and more coarsely as the noise level increases and as her discounting of past observations increases.

6.2 Coordination

Next, we consider the case when agents are only interested in coordinating their predictions with each other. We would like to understand better whether if they are only interested in coordinating, they will learn ways of categorizing that are similar to each other. We consider two alternative indicators of similarity. The first one is a measure of the similarity of the categorizations that the agents learn, i.e. of the categorizations in their pools. The second one measures similarity of the categories that they use.

The analysis was conducted with a total number of 20 agents in the pool, two of whom are randomly matched each period. We used a $\sigma^2 = 0.25$ and a discount factor of 0.05. In Figure 4a, we present our first measure, the absolute difference in the number of categories in the categorizations in the pools of the agents that are matched in a given period. Figure 4a shows that over time the absolute difference in the number of categories



a. Difference in coarseness

b. Similarity of categories used

c. Coarseness

d) Prediction error from other

Figure 4: Similarity of categorizations in case of coordination

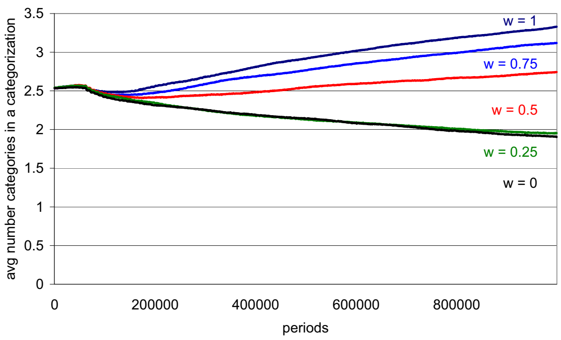
in the categorizations in the pools of the players declines substantially and reaches a value of 0.2. This indicates that over time players learn categorizations of more or less similar level of coarseness. This is interesting as it is in line with Proposition 4 Part 1 in which we showed that in the static setup if players are only interested in coordinating there are no NE in which the two players use categorizations of different levels of coarseness.

Figure 4b presents our second indicator, similarity of the categories used by the players. The similarity of two categories is calculated as the ratio of the number of objects that can be categorized by both categories to the number of objects that can be categorized by either category. Similarity ranges between 0 and 1. In case one player has a category for the object encountered and the other does not, similarity is 0. In case both players do not have a category for the object encountered, similarity is 1. We see that over time the similarity measure almost doubles reaching a level of 0.8 out of 1. It does not reach the maximum value of 1, which is related to the fact that under these parameters players learn to use relatively coarse categorizations and there can be different coarse categorizations that make similar predictions (see Figure 4c). Figure 4c shows the average number of categories in the categorizations that players use. Note that in a pool containing a large number of agents, agents need to use coarse categorizations in order to coordinate, as each of them has had different experiences in the past. There are many coarse categorizations with which agents can minimize their prediction error from one another. Thus, they do not necessarily need to use the same categorization. Figure 4d shows the dynamics of prediction error from the other's prediction. We see that prediction error from the other decreases over time as the agents learn to use more similar categories.[34]

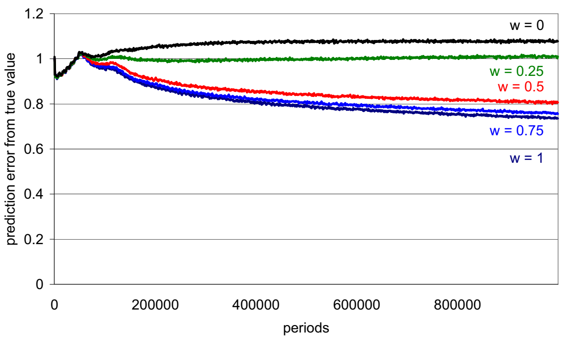
Numerical Result 3. *If agents aim to coordinate predictions they learn more similar ways of categorizing.*

6.3 Individual Prediction and Coordination

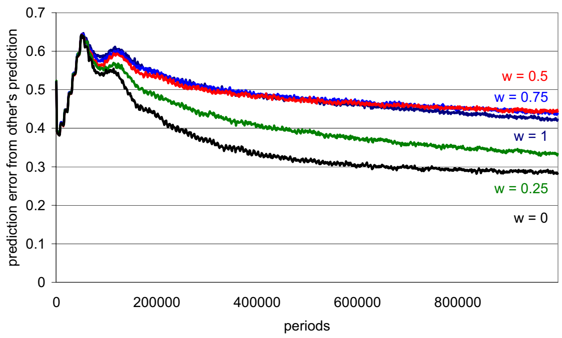
Finally, we consider the properties of the categorizations the players learn if they are interested both in predicting the true object value correctly and in coordinating their predictions with each other. More precisely we are interested in how the attempt to coordinate with another person affects the way an individual categorizes.



a. Coarseness



b. Prediction error from true value



c. Prediction error from other

Figure 5: Effect of higher weight on coordination

The analysis was conducted with a pool of 20 agents, two of whom were matched each period. We used a $\sigma^2 = 0.25$ and a discount factor $d = 0.05$.

[34] In Figure 4d there is an initial increase in prediction error from other's prediction. This is related to the fact that there are a large number of agents in the pool, two of whom are matched each period. As each agent in the pool has different past experiences, it takes time for agents to find the optimal level of coarseness.

Figure 5a shows the average number of categories in a categorization in the agent's pool over time for alternative weights on individual prediction and coordination. A weight of $w = 1$ means that an agent is only interested in IP. A weight of $w = 0.75$ indicates that this is the respective emphasis on individual prediction with the remaining $1 - w = 0.25$ weight placed on coordination. We see that the higher the weight agents place on coordination, the coarser the categorizations that they learn. Figures 5b and 5c show the prediction errors from the true value and from the other player's prediction, respectively. We observe that the higher the weight on IP, the lower the error with respect to the true object value in the long term. There is a non-monotonicity of the dependence of prediction error from the other on the weight w.

Numerical Result 4. *The higher the weight on coordination, the coarser the categorizations that players learn.*

6.4 Summary of the Results of the Dynamic Model

In sections 5 and 6 we presented a dynamic framework for analyzing categorization. We considered the following problem. There is an agent who observes some characteristics of an object each period and has to predict its unobserved value. The agent searches for the best categorizations or models of the world to use in a given environment. She starts out with a random set of categorizations and learns through trial and error which categorizations are useful in a given environment by searching the space of possible categorizations. We find that if the agent is only interested in predicting the true object value correctly, the higher the noise in the environment and the higher the discounting of past observations, the coarser the categorizations that she learns. If players are only interested in coordinating their predictions with each other, they learn similar ways of categorizing. These results are in line with the intuition developed in the static model.

The dynamic model allowed us to analyze how the categorizations that the agent learns depend on how much the agent cares about individual prediction and about coordination, respectively. We observe that the attempt to coordinate with others changes the way an individual categorizes. We find that the more an agent is interested in coordinating with others, the coarser the categorizations that she learns. This is consistent with the intuition developed in the static model regarding higher efficiency of categorization profiles at a higher level of coarseness. Agents can coordinate with each other better if they use coarser categorizations because by lumping more objects together in a single category each of them is better able to decrease her variance in prediction. The lower the variance in prediction of each agent, the less the players' predictions are going to differ from one another and hence the better the coordination.

7 Concluding Remarks

This paper investigates the usage of categories to make decisions. We focused on the basic properties of different ways of categorizing - addressing in particular the question which factors influence whether it is more useful to categorize finely or to categorize coarsely. We construct two complementary models to analyze the above described problem - a static and a dynamic model. We consider three different variants of the problem - individual prediction, coordination, and the convex combination of the two.

Our findings on individual prediction are in line with some recent literature in economics, which shows that coarse categorization may be optimal in stochastic environments in which agents have a limited number of past observations. Thus, avoiding overfitting in prediction might be one rationale for agents to categorize coarsely. The key contribution of this paper consists of considering properties of categorization if players want to coordinate predictions with each other, or if they care both about predicting the true value correctly and about coordinating their predictions with each other. Our results show that incentives to coordinate may be a further rationale for coarse categorization, additional and complementary to those hitherto discussed in the literature.

Given that the use of categorization seems to be widespread, that the literature has shown that coarse categorization may be linked to biased outcomes ranging from discrimination against a minority group to biases in financial markets, and that the world abounds with situations requiring some coordination, this finding seems important.

References

Nabil I. Al-Najjar and Mallesh M. Pai. Coarse decision making and overfitting. *Journal of Economic Theory*, 150(0):467 – 486, 2014.

John R. Anderson. *The Adaptive Character of Thought*. Psychology Press, 1990.

John R. Anderson. The adaptive nature of human categorization. *Psychological Review*, 98(3):409, 1991.

W. Brian Arthur, John H. Holland, Blake LeBaron, Richard Palmer, and Paul Tayler. Asset pricing under endogenous expectations in an artificial stock market. In W.B. Arthur, S.N. Durlauf, and D.A. Lane, editors, *The Economy as an Evolving Complex System II*, pages 15–44. Reading MA: Addison-Wesley, 1997.

Gregory F. Ashby and Todd W. Maddox. Human category learning. *Annual Review of Psychology*, 56:149–178, 2005.

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

Yaron Azrieli. Categorization and correlation in a random-matching game. *Journal of Mathematical Economics*, 46(3):303–310, 2010.

Yaron Azrieli and Ehud Lehrer. Categorization generated by extended prototypes: An axiomatic approach. *Journal of Mathematical Psychology*, 51(1):14–28, 2007.

Nicholas Barberis and Andrei Shleifer. Style investing. *Journal of Financial Economics*, 68(2):161–199, 2003.

Donald A. Berry and Bernard W. Lindgren. *Statistics: Theory and Methods*. Duxbury Press Belmont, CA, 1996.

Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2007.

Kilian Bizer, Lukas Meub, Till Proeger, and Markus Spiwoks. Strategic coordination in forecasting: An experimental study. Discussion paper, Center for European Governance and Economic Development Research, 2014.

Jacques Cremer, Luis Garicano, and Andrea Prat. Language and the theory of the firm. *Quarterly Journal of Economics*, 122(1):373–407, 2007.

Roland Fryer and Matthew O. Jackson. A categorical model of cognition and biased decision-making. *BE Journal of Theoretical Economics*, 8(1):1–42, 2008.

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

Itzhak Gilboa and Larry Samuelson. Preferring simplicity. Mimeo, 2009.

Itzhak Gilboa and David Schmeidler. Case-based decision theory. *Quarterly Journal of Economics*, 110(3):605–639, 1995.

Veronika Grimm and Friederike Mengel. An experiment on learning in a multiple games environment. *Journal of Economic Theory*, 147(6):2220–2259, 2012.

Yuval Heller and Eyal Winter. Rule rationality. Mimeo, 2014.

Douglas R. Hofstadter. *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. Basic Books, 1996.

John H. Holland, Keith J. Holyoak, Richard E. Nisbett, and Paul R. Thagard. *Induction: Processes of Inference, Learning, and Discovery*. MIT Press, 1989.

Philippe Jehiel. Analogy-based expectation equilibrium. *Journal of Economic Theory*, 123(2):81–104, 2005.

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

Natalia L. Komarova, Kimberly A. Jameson, and Louis Narens. Evolutionary models of color categorization based on discrimination. *Journal of Mathematical Psychology*, 51 (6):359–382, 2007.

Barton L. Lipman. Why is language vague? Mimeo, 2009.

Michael Mandler, Paola Manzini, and Marco Mariotti. A million answers to twenty questions: Choosing by checklist. *Journal of Economic Theory*, 147(1):71–92, 2012.

Paola Manzini and Marco Mariotti. Categorize then choose: Boundedly rational choice and welfare. *Journal of the European Economic Association*, 10(5):1141–1165, 2012.

Friederike Mengel. On the evolution of coarse categories. *Journal of Theoretical Biology*, 307:117–124, 2012a.

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

Marvin Minsky and Seymour Papert. *Perceptron: an Introduction to Computational Geometry*. MIT Press, 1969.

Melanie Mitchell and Douglas Hofstadter. Perspectives on copycat: Comparisons with recent work. In Douglas Hofstadter, editor, *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*, pages 275–299. New York: Basic Books, 1996.

Erik Mohlin. Optimal categorization. *Journal of Economic Theory*, 2014.

Stephen Morris and Hyung Song Shin. Social value of public information. *American Economic Review*, 92(5):1521–1534, 2002.

Sendhil Mullainathan. Thinking through categories. Working paper, NBER, 2002.

Sendhil Mullainathan, Joshua Schwartzstein, and Andrei Shleifer. Coarse thinking and persuasion. *Quarterly Journal of Economics*, 123(2):577–619, 2008.

Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

Robert M. Nosofsky. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115(1):39, 1986.

Marcin Peski. Prior symmetry, similarity-based reasoning, and endogenous categorization. *Journal of Economic Theory*, 146(1):111–140, 2011.

Eleanor Rosch. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3):192, 1975.

Eleanor Rosch. Principles of categorization. In Eric Margolis and Stephen Laurence, editors, *Cognition and Categorization*, pages 27–48. Hillsdale, NJ: Erlbaum, 1978.

Gian-Carlo Rota. The number of partitions of a set. *American Mathematical Monthly*, 71(5):498–504, 1964.

David E Rumelhart and James L McClelland. *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1. Foundations*. MIT Press, 1986.

David S. Scharfstein and Jeremy C. Stein. Herd behavior and investment. *American Economic Review*, 80:3:465–479, 1990.

Edward E. Smith and Douglas L. Medin. *Categories and Concepts*. Harvard University Press Cambridge, MA, 1981.

Luc Steels, Tony Belpaeme, et al. Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and brain sciences*, 28(4):469–488, 2005.

Scott N. Swisher. Information aggregation and product reviews in the entertainment industry. Mimeo, 2013.

Safa R. Zaki, Robert M. Nosofsky, Roger D. Stanton, and Andrew L. Cohen. Prototype and exemplar accounts of category learning and attentional allocation: a reassessment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29:1160–1173, 2003.

A Appendix: Examples

This appendix provides a numerical illustration of a number of aspects of Proposition 1 and Proposition 3. Assuming a specific function relating unobservable to observed attributes, a given sample size n, and a given noise level σ^2, Table 2 shows the $Var(P)$, $Var(Y)$, $Bias^2(P)$, and $EPE(P)$, calculated according to the formula in Lemma 1, for different categorizations with a range of coarseness levels.[35] Table 3 provides the same information, but for a higher noise level. Column 2 shows that a coarser categorization has a smaller variance than a finer categorization. Column 4 illustrates that the bias of coarser categorizations connected by a path to a finer categorization is always greater or equal to the bias of the finer categorizations. Comparing column 5 in Tables 2 and 3 shows that an increase in σ means that only a categorization that is coarser than the original can become optimal (if it was not optimal before).

Table 2: Example $l = 2$, $y = 2x_1 + x_2 + \epsilon$ with $\sigma^2 = 0.5$, $n = 2$

(1) Categorization	(2) $Var(P)$	(3) $Var(Y)$	(4) $Bias^2(P)$	(5) $EPE(P)$
$\{\{11\}\{10\}\{01\}\{00\}\}$	0.2500	0.5000	0.0000	0.7500
$\{\{11,10\}\{01\}\{00\}\}$	0.1875	0.5000	0.1250	0.8125
$\{\{11,10,01\}\{00\}\}$	0.1250	0.5000	0.5000	1.1250
$\{\{11,10\}\{01,00\}\}$	0.1250	0.5000	0.2500	0.8750
$\{\{11,10,01,00\}\}$	0.0625	0.5000	1.2500	1.8125

Table 3: Example $l = 2$, $y = 2x_1 + x_2 + \epsilon$ with $\sigma^2 = 4.0$, $n = 2$

(1) Categorization	(2) $Var(P)$	(3) $Var(Y)$	(4) $Bias^2(P)$	(5) $EPE(P)$
$\{\{11\}\{10\}\{01\}\{00\}\}$	2.000	4.000	0.000	6.000
$\{\{11,10\}\{01\}\{00\}\}$	1.500	4.000	0.125	5.625
$\{\{11,10,01\}\{00\}\}$	1.000	4.000	0.500	5.500
$\{\{11,10\}\{01,00\}\}$	1.000	4.000	0.250	5.250
$\{\{11,10,01,00\}\}$	0.500	4.000	1.250	5.750

[35] Only a subset of the possible categorizations is included for reasons of clarity.

B Appendix: Proofs

Lemma 1. *Bias-Variance Decomposition of $EPE^{IP}(P)$*

Proof.

$$
\begin{aligned}
&EPE^{IP}(P) = \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j E[(\hat{Y}^{C_k} - Y_j)^2] \\
&= \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j E[(\hat{Y}^{C_k})^2 - 2\hat{Y}^{C_k} Y_j + (Y_j)^2] \\
&= \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j \Big[E[(\hat{Y}^{C_k})^2] - 2E[\hat{Y}^{C_k} Y_j] + E[(Y_j)^2]\Big] \\
&= \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j \Big[E[(\hat{Y}^{C_k})^2] - 2E[\hat{Y}^{C_k}]E[Y_j] + E[(Y_j)^2]\Big] \\
&= \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j \Big[Var(\hat{Y}^{C_k}) + (E[\hat{Y}^{C_k}])^2 - 2E[\hat{Y}^{C_k}]E[Y_j] + Var(Y_j) + (E[Y_j])^2\Big] \\
&= \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j \Big[Var(\hat{Y}^{C_k}) + Var(Y_j) + (E[\hat{Y}^{C_k}] - E[Y_j])^2\Big] \\
&= \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j Var(\hat{Y}^{C_k}) + \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j Var(Y_j) + \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j (E[\hat{Y}^{C_k}] - E[Y_j])^2 \\
&= \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j Var(\hat{Y}^{C_k}) + \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j Var(Y_j) + \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j (E[\hat{Y}^{C_k}] - \mu_j)^2 \\
&= Var(P) + Var(Y) + Bias^2(P)
\end{aligned} \tag{22}
$$

□

In the proof above Line 3 is equivalent to Line 4 since for two independent random variables X and Y, $E[XY] = E[X]E[Y]$. To see that $\hat{Y}^{C_k}$ and Y_j are independent, consider the following. $\hat{Y}^{C_k}$ is the category belief for category C_k. It is a random variable because its realization depends on the random draws of the y-values of the objects that have been sampled in this category. $\hat{Y}^{C_k}$ is therefore based on random draws made in the past. Y_j is a random variable because it depends on the y-value that will be drawn for the next object of type j. Therefore $P(Y_j = y_j) = P(Y_j = y_j | \hat{Y}^{C_K} = \hat{y}^{C_K})$ for any pair $(\hat{y}^{C_K}, y_j)$.

Line 4 is equivalent to Line 5 since for any random variable X, $Var(X) = E[X^2] - (E[X])^2$ and therefore $E[X^2] = Var(X) + (E[X])^2$ (see e.g. Berry and Lindgren, 1996, p. 92 and the derivation below). This expression is applied to $E[(\hat{Y}^{C_k})^2]$ and to $E[(Y_j)^2]$

to move from line 4 to line 5 above.

$$\begin{aligned}
&Var(X) = E[(X-\mu)^2] \\
&= E[X^2 - 2X\mu + \mu^2] \\
&= E[X^2 - 2XE[X] + \mu^2] \\
&= E[X^2] - 2E[X]E[X] + (E[X])^2 \\
&= E[X^2] - 2(E[X])^2 + (E[X])^2 \\
&= E[X^2] - (E[X])^2
\end{aligned} \tag{23}$$

Proposition 1. *Comparative Statics Bias and Variance Components $EPE^{IP}(P)$*

Part 1. Variances

Proof. From the bias-variance decomposition in Lemma 1, we know that the variance component of the $EPE^{IP}(P)$ of a categorization is equal to:

$$Var(P) + Var(Y) = \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j Var(\hat{Y}^{C_k}) + \sum_{C_k \in P} \sum_{x_j \in C_k^T} p_j Var(Y_j) \ , \tag{24}$$

where p_j denotes the probability that the next object encountered is of type j. We assumed that the next object will be drawn with equal probability from the different object types, $p_j = p$ for all j and therefore $p = \frac{1}{|O^T|}$. Hence, we can also write the above as:

$$Var(P) + Var(Y) = p \sum_{C_k \in P} \sum_{x_j \in C_k^T} Var(\hat{Y}^{C_k}) + p \sum_{C_k \in P} \sum_{x_j \in C_k^T} Var(Y_j) \tag{25}$$

We now look at these two terms separately:
First term: $Var(P) = p \sum_{C_k \in P} \sum_{x_j \in C_k^T} Var(\hat{Y}^{C_k})$

We know that as the category belief is a random variable equal to the average of a number of random variables, independently drawn from the normal distribution, the variance of any category belief is equal to σ^2 divided by the number of objects in the category. The number of objects in the category is equal to the number of different object types in this category, which we denote by t_k for category k times the sample size n of each object type (as we assumed that an agent has sampled n objects of each type. Then the above becomes:

$$p \sum_{C_k \in P} \sum_{x_j \in C_k^T} Var(\hat{Y}^{C_k}) = p \sum_{C_k \in P} \sum_{x_j \in C_k^T} \frac{\sigma^2}{t_k n} = p \sum_{C_k \in P} \frac{t_k \sigma^2}{t_k n} = \frac{pk\sigma^2}{n} \tag{26}$$

Thus, under the assumption of equal variance for each object type in the population, equal probability of observing an object from each type, and of equal number of observations from each type in the agent's past experience, the variance of category belief for each category in a given categorization is the same.

Second term: $Var(Y) = p \sum_{C_k \in P} \sum_{x_j \in C_k^T} Var(Y_j)$

We assumed that $Var(Y_j)$ is the same for all j and that it is equal to σ^2. Then the above becomes:

$$p \sum_{C_k \in P} \sum_{x_j \in C_k^T} Var(Y_j) = p \sum_{C_k \in P} \sum_{x_j \in C_k^T} \sigma^2 = \frac{1}{|O^T|}|O^T|\sigma^2 = \sigma^2 \tag{27}$$

Connecting our results on the two terms, the variance component of the $EPE(P)$ is equal to:

$$Var(P) + Var(Y) = \frac{pk\sigma^2}{n} + \sigma^2 \tag{28}$$

This expression shows that the variance component of the $EPE(P)$ of a categorization is increasing in the number of categories k, increasing in the noise level σ^2, and decreasing in n.

We now compare the variance of a finer categorization P^L with the variance of a coarser categorization P^{L+}. Assume that the finer categorization has k categories. The coarser categorization has $k - m$ categories (with $m > 0$).

$$Var(P^L) - Var(P^{L+}) = \frac{pk\sigma^2}{n} + \sigma^2 - \left(\frac{p(k-m)\sigma^2}{n} + \sigma^2\right) = \frac{pm\sigma^2}{n} \tag{29}$$

That is, a finer categorization has a greater variance than a coarser categorization (for any $\sigma^2 > 0$). Moreover, the difference in variance between a finer and a coarser categorization is equal to $\frac{p\sigma^2}{n}$ times the difference in number of categories between the two (m). We see from the above expression that the difference in variance between a finer and a coarser categorization is increasing in their difference in coarseness m and in σ^2, and is decreasing in n. □

Part 2. Biases

Proof. Part 2 says that the $Bias^2$ of any (coarser) *categorization* $P^{L+}(j)$ that can be formed by merging two or more categories of a (finer) *categorization* $P^L(i)$ is always greater than or equal to the $Bias^2$ of the finer *categorization* $P^L(i)$. To prove this, we show that the $Bias^2$ of the category belief of any coarser *category* containing $t + v$ object types is always greater than or equal to the sum of the biases of the category beliefs of any two finer *categories* with t and with v object types, respectively, through the merging of which it can be formed. This implies that the $Bias^2$ of any coarser *categorization* that was formed by merging two *categories* of a finer *categorization* $P^L(i)$ is always greater than or equal to the $Bias^2$ of the finer *categorization.* Therefore for categorizations that are connected by a path, the $Bias^2$ of any coarser categorization $P^{L^+}(j)$ is greater than or equal to the $Bias^2$ of any finer categorization $P^L(i)$.

The $Bias^2$ of the category belief of a category with t object types is equal to: $Bias^2(\hat{y}_t) = \sum_{i=1}^{t} p_i(E[\hat{y}_t] - \mu_i)^2$, where $E[\hat{y}_t] = \frac{\sum_{i=1}^{t} \mu_i}{t}$ is the expected value of category belief and μ_i is the mean of object type i in the population. We assumed that $p_i = \frac{1}{|O_T|} = p$ for all

i. Therefore $\sum_{i=1}^{t} p_i(E[\hat{y}_t] - \mu_i)^2 = p\sum_{i=1}^{t}(E[\hat{y}_t] - \mu_i)^2$. Note that $E[\hat{y}_t] = \frac{\sum_{i=1}^{t}\mu_i}{t}$ since we assumed that the agent has sampled n objects of each object type. Thus, the $Bias^2$ can be rewritten in the following way:

$$\begin{aligned}
Bias^2(\hat{y}_t) &= p\sum_{i=1}^{t}(E[\hat{y}_t] - \mu_i)^2 \\
&= p\sum_{i=1}^{t}\Big[(E[\hat{y}_t])^2 - 2E[\hat{y}_t]\mu_i + (\mu_i)^2\Big] \\
&= p\Big[\sum_{i=1}^{t}(E[\hat{y}_t])^2 - 2\sum_{i=1}^{t}E[\hat{y}_t]\mu_i + \sum_{i=1}^{t}\mu_i^2\Big] \\
&= p\Big[\sum_{i=1}^{t}(E[\hat{y}_t])^2 - 2E[\hat{y}_t]\sum_{i=1}^{t}\mu_i + \sum_{i=1}^{t}\mu_i^2\Big] \\
&= p\Big[\sum_{i=1}^{t}(E[\hat{y}_t])^2 - 2E[\hat{y}_t]E[\hat{y}_t]t + \sum_{i=1}^{t}\mu_i^2\Big] \\
&= p\Big[t(E[\hat{y}_t])^2 - 2t(E[\hat{y}_t])^2 + \sum_{i=1}^{t}\mu_i^2\Big] \\
&= p\Big[\sum_{i=1}^{t}\mu_i^2 - t(E[\hat{y}_t])^2\Big] \\
&= p\Big[\sum_{i=1}^{t}\mu_i^2 - \frac{t(\sum_{i=1}^{t}\mu_i)^2}{t^2}\Big] \\
\Leftrightarrow Bias^2(\hat{y}_t) &= p\Big[\sum_{i=1}^{t}\mu_i^2 - \frac{(\sum_{i=1}^{t}\mu_i)^2}{t}\Big]
\end{aligned} \tag{30}$$

In line 5 we use that $E[\hat{y}_t]t = \sum_{i=1}^{t}\mu_i$, which follows directly from the definition of expected value of category belief.

For a category that has v object types the corresponding expression for the squared bias of category belief is equal to:

$$Bias^2(\hat{y}_v) = p\Big[\sum_{j=1}^{v}\mu_j^2 - \frac{(\sum_{j=1}^{v}\mu_j)^2}{v}\Big] \tag{31}$$

And for a category that has $t + v$ object types the corresponding expression for the squared bias of category belief is equal to:

$$Bias^2(\hat{y}_{t+v}) = p\Big[\sum_{k=1}^{t+v}\mu_k^2 - \frac{(\sum_{k=1}^{t+v}\mu_k)^2}{t+v}\Big] \tag{32}$$

We start by considering a finer categorization that has two categories, one with t and one with v object types, respectively. This categorization may also contain other categories but we abstract from them. We now show that the $Bias^2$ of a coarser categorization that was formed by merging the category with t and the category with v object types into one category with $t+v$ object types (keeping all other categories from the finer categorization unchanged) is greater than the $Bias^2$ of the finer categorization. We do this by showing that the $Bias^2$ of the coarser categorization is greater than or equal to the sum of the $Biases^2$ of the two categories from the finer categorization, which were merged to create it. That is, we show that $Bias^2(\hat{y}_{t+v}) \geq Bias^2(\hat{y}_t) + Bias^2(\hat{y}_v)$.

$$
\begin{aligned}
& Bias^2(\hat{y}_{t+v}) \geq Bias^2(\hat{y}_t) + Bias^2(\hat{y}_v) \\
& \Leftrightarrow \sum_{k=1}^{t+v} \mu_k^2 - \frac{(\sum_{k=1}^{t+v} \mu_k)^2}{t+v} \geq \sum_{i=1}^{t} \mu_i^2 - \frac{(\sum_{i=1}^{t} \mu_i)^2}{t} + \sum_{j=t+1}^{t+v} \mu_j^2 - \frac{(\sum_{j=t+1}^{t+v} \mu_j)^2}{v} \\
& \Leftrightarrow \sum_{k=1}^{t} \mu_k^2 + \sum_{k=t+1}^{t+v} \mu_k^2 - \frac{(\sum_{k=1}^{t} \mu_k + \sum_{k=t+1}^{t+v} \mu_k)^2}{t+v} \\
& \quad \geq \sum_{i=1}^{t} \mu_i^2 - \frac{(\sum_{i=1}^{t} \mu_i)^2}{t} + \sum_{j=t+1}^{t+v} \mu_j^2 - \frac{(\sum_{j=t+1}^{t+v} \mu_j)^2}{v} \\
& \Leftrightarrow -\frac{(\sum_{k=1}^{t} \mu_k)^2}{t+v} - \frac{2(\sum_{k=1}^{t} \mu_k)(\sum_{k=t+1}^{t+v} \mu_k)}{t+v} - \frac{(\sum_{k=t+1}^{t+v} \mu_k)^2}{t+v} \\
& \quad \geq -\frac{(\sum_{i=1}^{t} \mu_i)^2}{t} - \frac{(\sum_{j=t+1}^{t+v} \mu_j)^2}{v} \\
& \Leftrightarrow -\frac{(\sum_{k=1}^{t} \mu_k)^2}{t} + \frac{v(\sum_{k=1}^{t} \mu_k)^2}{t(t+v)} - \frac{2(\sum_{k=1}^{t} \mu_k)(\sum_{k=t+1}^{t+v} \mu_k)}{t+v} - \frac{(\sum_{k=t+1}^{t+v} \mu_k)^2}{t+v} \\
& \quad \geq -\frac{(\sum_{i=1}^{t} \mu_i)^2}{t} - \frac{(\sum_{j=t+1}^{t+v} \mu_j)^2}{v} \\
& \Leftrightarrow \frac{v(\sum_{k=1}^{t} \mu_k)^2}{t(t+v)} - \frac{2(\sum_{k=1}^{t} \mu_k)(\sum_{k=t+1}^{t+v} \mu_k)}{t+v} - \frac{(\sum_{k=t+1}^{t+v} \mu_k)^2}{t+v} \geq -\frac{(\sum_{j=t+1}^{t+v} \mu_j)^2}{v} \\
& \Leftrightarrow \frac{v(\sum_{k=1}^{t} \mu_k)^2}{t(t+v)} - \frac{2(\sum_{k=1}^{t} \mu_k)(\sum_{k=t+1}^{t+v} \mu_k)}{t+v} - \frac{v(\sum_{k=t+1}^{t+v} \mu_k)^2}{v(t+v)} + \frac{(t+v)(\sum_{k=t+1}^{t+v} \mu_j)^2}{v(t+v)} \geq 0 \\
& \Leftrightarrow \frac{v(\sum_{k=1}^{t} \mu_k)^2}{t(t+v)} - \frac{2(\sum_{k=1}^{t} \mu_k)(\sum_{k=t+1}^{t+v} \mu_k)}{t+v} - \frac{v(\sum_{k=t+1}^{t+v} \mu_k)^2}{v(t+v)} \\
& \quad + \frac{v(\sum_{k=t+1}^{t+v} \mu_k)^2}{v(t+v)} + \frac{t(\sum_{j=t+1}^{t+v} \mu_j)^2}{v(t+v)} \geq 0 \\
& \Leftrightarrow \frac{v(\sum_{k=1}^{t} \mu_k)^2}{t(t+v)} - \frac{2(\sum_{k=1}^{t} \mu_k)(\sum_{k=t+1}^{t+v} \mu_k)}{t+v} + \frac{t(\sum_{j=t+1}^{t+v} \mu_j)^2}{v(t+v)} \geq 0
\end{aligned} \tag{33}
$$

Let $\sum_{k=1}^{t} \mu_k = A$ and $\sum_{j=t+1}^{t+v} \mu_j = B$.

$$
\begin{aligned}
&\frac{vA^2}{t(t+v)} - \frac{2AB}{t+v} + \frac{tB^2}{v(t+v)} \geq 0 \\
&\Leftrightarrow \frac{v^2A^2 - 2tvAB + t^2B^2}{tv(t+v)} \geq 0 \\
&\Leftrightarrow \frac{(vA - tB)^2}{tv(t+v)} \geq 0 \\
&\Leftrightarrow \frac{tv\Big[(1/t)A - (1/v)B\Big]^2}{t+v} \geq 0 \\
&\Leftrightarrow \frac{tv\Big[(1/t)\sum_{k=1}^{t} \mu_k - (1/v)\sum_{j=1}^{v} \mu_j\Big]^2}{t+v} \geq 0 \\
&\Leftrightarrow \frac{tv(\hat{y}_t - \hat{y}_v)^2}{t+v} \geq 0
\end{aligned}
\tag{34}
$$

As $t > 0$ and $v > 0$ and the remaining term is squared, this will always be true. Thus, $Bias^2(\hat{y}_{t+v}) \geq Bias^2(\hat{y}_t) + Bias^2(\hat{y}_v)$. This implies that for any two categorizations that are connected by a path (i.e. such that the coarser categorization was formed only through merging of two or more categories of the finer categorization), the $Bias^2$ of the coarser categorization $P^{L^+}(j)$ is greater or equal to the $Bias^2$ of the finer categorization $P^L(i)$. □

Proposition 2. *Existence of Optimal Categorization for Individual Prediction*

Proof. The proof of existence is trivial. The function $EPE^{IP}(P)$ that maps from the set of possible categorizations to the set of real numbers is guaranteed to attain a minimum on the set of all possible categorizations $\mathcal{P}$, as the set of all possible categorizations is finite (see Sundaram, 2008, p. 90). The set of possible categorizations $\mathcal{P}$ is finite because the number of different object types is finite. The cardinality of the set of possible categorizations is equal to the number of possible categorizations that form a disjoint partitioning of the object set. This number is given by the Bell number B_d with $d = |O^T|$. □

Proposition 3. *Coarseness of the Optimal Categorization(s) for Individual Prediction*

Proof. Denote any optimal categorization by $P^L(i)$, any finer than the optimal by P^{L-}, any equally coarse by $P^L(j)$, and any coarser by P^{L+}. Note that there may be more than one optimal categorization and that the optimal categorizations may be at different coarseness levels. In Definition 4 we established that the following conditions have to hold for any optimal categorization: i) the EPE^{IP} of any optimal categorization is smaller or equal to the EPE^{IP} of any categorization finer than it; ii) the EPE^{IP} of any optimal categorization is smaller or equal to the EPE^{IP} of any categorization that is equally coarse; iii) the EPE^{IP} of any optimal categorization is smaller or equal to the EPE^{IP} of any categorization that is coarser than it.

This is equivalent to (see Definition 4):

$$\begin{aligned}
&\text{i) } Var(P^L(i)) - Var(P^{L^-}) \leq Bias^2(P^{L^-}) - Bias^2(P^L(i)) \\
&\quad \text{for all } P^{L^-} \prec P^L(i) \\
&\text{ii) } Var(P^L(i)) - Var(P^L(j)) \leq Bias^2(P^L(j)) - Bias^2(P^L(i)) \\
&\quad \text{for all } P^L(j) \sim P^L(i) \\
&\text{iii) } Var(P^L(i)) - Var(P^{L^+}) \leq Bias^2(P^{L^+}) - Bias^2(P^L(i)) \\
&\quad \text{for all } P^{L^+} \succ P^L(i)
\end{aligned} \tag{35}$$

We want to derive how the coarseness of the optimal categorization(s) changes with changes in σ^2 and in n. Note that changes in σ^2 and in n do not affect the bias terms. This is because the squared bias term of a categorization's EPE^{IP} is equal to the expected bias of its category beliefs. The squared bias of a category belief depends on the differences between the expected value of category belief and the true population mean for each object type in this category. Since neither the expected value of category belief nor the true population mean is affected by changes in σ^2 and n, the bias term of a categorization will not be affected by such changes. Therefore, in the inequalities above only the LHS will change. To examine the changes on the LHS, we will use our result from Proposition 1 Part 1 where we established how the difference in variance between a finer and a coarser categorization is affected by changes in σ^2 and in n.

We first show how an increase in σ^2 will affect the coarseness of the optimal categorization(s) by examining the effect of an increase in σ^2 on the fulfillment of the above conditions for optimality.

i) In Proposition 1 Part 1 we showed that the difference in the variance between a finer and a coarser categorization is positive and increases with an increase in σ^2. That is, the difference between a coarser and a finer categorization is negative and increasingly negative with increases in σ^2. If before $Var(P^L(i)) - Var(P^{L-}) \leq Bias^2(P^{L-}) - Bias^2(P^L(i))$, as σ^2 increases the LHS becomes even more negative (and the RHS does not change). Thus, if before using a categorization $P^L(i)$ was equally good or better than using a finer categorization P^{L-}, this will be even more so as σ^2 increases. Therefore, no categorization finer than $P^L(i)$ can become optimal after an increase in σ^2 if it was not optimal before the increase.

ii) The change in σ^2 has no effect on $Var(P^L(i)) - Var(P^L(j)) \leq Bias^2(P^L(j)) - Bias^2(P^L(i))$ because $Var(P^L(i)) - Var(P^L(j)) = 0$, as the variances of all categorizations at the same level of coarseness are equal. Thus, this equation will continue to hold after the change in σ^2. No categorization that is equally coarse as $P^L(i)$ can become optimal after an increase in σ^2 if it was not optimal before the increase.

iii) If before $Var(P^L(i)) - Var(P^{L+}) \leq Bias^2(P^{L+}) - Bias^2(P^L(i))$, as σ^2 increases the LHS increases while the RHS stays the same, so that eventually the LHS has to become greater than the RHS, thus making use of a coarser categorization profitable.

We have thus shown that as σ^2 increases, it cannot become optimal to use a categorization that was not optimal before and that is finer than or equally coarse as the initially optimal one(s). Only coarser categorizations can become optimal.

Note that the above result implies that if before the change in σ^2 there were several optimal categorizations at different levels of coarseness, no categorization that was not optimal before and that is finer than or equally coarse as the *coarsest* of those that were initially optimal can become optimal after the increase in σ^2.

We now turn to the effect of changes in the sample size n. Let again $P^L(i)$ be the optimal categorization(s) before the increase in n with the same conditions for optimality as above.

i) If before the increase in n, $EPE^{IP}(P^L(i)) \leq EPE^{IP}(P^{L-})$, now this may no longer be fulfilled. More precisely, we know that for $P^L(i)$ to be optimal before the increase in sample size it has to hold that:

$Var(P^L(i)) - Var(P^{L-}) \leq Bias^2(P^{L-}) - Bias^2(P^L(i))$.

From Proposition 1 we know that the LHS is negative and equal to: $Var(P^L(i)) - Var(P^{L-}) = -\frac{pm\sigma^2}{n}$, where m is the difference in the number of categories in the finer and in the coarser categorization.

We can see that $\lim_{n\to\infty}(Var(P^L(i)) - Var(P^{L-})) = 0$

In other words, as the sample size increases in the limit the LHS will tend to zero while the RHS stays unchanged. Thus, as n increases, the LHS may no longer be smaller than the RHS and it may become better to use a finer categorization than the one that was optimal before the increase in n.

ii) We also know (again from Proposition 1 Part 1) that as n increases the difference between the EPE^{IP} of the optimal categorization and of others of equal coarseness will not be affected as $Var(P^L(i)) - Var(P^L(j)) = 0$. Thus, one will not have an incentive to switch to another categorization of equal coarseness because of the increase in sample size.

iii) The condition $Var(P^L(i)) - Var(P^{L+}) \leq Bias^2(P^{L+}) - Bias^2(P^L(i))$, will continue to hold after the increase in n since before the increase $Var(P^L(i)) - Var(P^{L+})) \geq 0$ and as n increases $\lim_{n\to\infty}(Var(P^L(i)) - Var(P^{L+})) = 0$. Thus, one will never have an incentive to move to a coarser or another equally coarse categorization if n increases. We have thus shown that as n increases only a finer categorization than the initially optimal one(s) can become profitable.

Note that the above result implies that if before the change in n there were several optimal categorizations at different levels of coarseness, no categorization that was not optimal before and that is coarser than or equally coarse as the *finest* of those that were initially optimal can become optimal after the increase in n. □

Lemma 2. *Bias-Variance Decomposition of* $EPE^C(P_1, P_2)$

Proof.

$$\begin{aligned}
EPE^C(P_1,P_2) &= \sum_{C_k\in P_1}\sum_{C_l\in P_2}\sum_{x_j\in(C_k^T,C_l^T)} p_j E[(\hat{Y}^{C_k}-\hat{Y}^{C_l})^2] \\
&= \sum_{C_k\in P_1}\sum_{C_l\in P_2}\sum_{x_j\in(C_k^T,C_l^T)} p_j E[(\hat{Y}^{C_k})^2 - 2\hat{Y}^{C_k}\hat{Y}^{C_l} + (\hat{Y}^{C_l})^2] \\
&= \sum_{C_k\in P_1}\sum_{C_l\in P_2}\sum_{x_j\in(C_k^T,C_l^T)} p_j \Big[E[(\hat{Y}^{C_k})^2] - 2E[\hat{Y}^{C_k}\hat{Y}^{C_l}] + E[(\hat{Y}^{C_l})^2]\Big] \\
&= \sum_{C_k\in P_1}\sum_{C_l\in P_2}\sum_{x_j\in(C_k^T,C_l^T)} p_j \Big[E[(\hat{Y}^{C_k})^2] - 2E[\hat{Y}^{C_k}]E[\hat{Y}^{C_l}] + E[(\hat{Y}^{C_l})^2]\Big] \\
&= \sum_{C_k\in P_1}\sum_{C_l\in P_2}\sum_{x_j\in(C_k^T,C_l^T)} p_j \Big[Var(\hat{Y}^{C_k}) + (E[\hat{Y}^{C_k}])^2 - 2E[\hat{Y}^{C_k}]E[\hat{Y}^{C_l}] \\
&\quad + Var(\hat{Y}^{C_l}) + (E[\hat{Y}^{C_l}])^2\Big] \\
&= \sum_{C_k\in P_1}\sum_{C_l\in P_2}\sum_{x_j\in(C_k^T,C_l^T)} p_j \Big[Var(\hat{Y}^{C_k}) + Var(\hat{Y}^{C_l}) + (E[\hat{Y}^{C_k}] - E[\hat{Y}^{C_l}])^2\Big] \\
&= \sum_{C_k\in P_1}\sum_{x_j\in C_k^T} p_j Var(\hat{Y}^{C_k}) + \sum_{C_l\in P_2}\sum_{x_j\in C_l^T} p_j Var(\hat{Y}^{C_l}) \\
&\quad + \sum_{C_k\in P_1}\sum_{C_l\in P_2}\sum_{x_j\in(C_k^T,C_l^T)} p_j (E[\hat{Y}^{C_k}] - E[\hat{Y}^{C_l}])^2 \\
&= Var(P_1) + Var(P_2) + Bias^2(P_1,P_2)
\end{aligned} \tag{36}$$

□

Note that line 3 is equivalent to line 4 since for any two independent variables $E[XY] = E[X]E[Y]$. To see that $\hat{Y}^{C_k}$ and $\hat{Y}^{C_l}$ are independent note that $\hat{Y}^{C_k}$, the category belief of category C_K of Player 1, is a random variable and its realization depends on the y-values of the objects that Player 1 has sampled in this category in the past. The same applies to $\hat{Y}^{C_l}$, the category belief for category C_l of Player 2. These two variables are independent from each other as the y-values of the objects that Player 1 and Player 2 encounter are randomly and independently drawn. Thus, it holds that $P(\hat{Y}^{C_k} = \hat{y}^{C_k}) = P(\hat{Y}^{C_K} = \hat{y}^{C_k}|\hat{Y}^{C_l} = \hat{y}^{C_l})$ for any pair $(\hat{y}^{C_k}, \hat{y}^{C_l})$.

To get from line 4 to line 5 we apply the expression $E[X^2] = Var(X) + (E[X])^2$ for $E[(\hat{Y}^{C_k})^2]$ and $E[(\hat{Y}^{C_l})^2]$. For a derivation of this expression, see Lemma 1.

Definition 6. NE Conditions of Coordination Game

Let $(P_1^L(i), P_2^M(j))$ denote a categorization profile such that $P_1^L(i)$ is the categorization that Player 1 uses and $P_2^M(j)$ is the categorization that Player 2 uses. The two categorizations could be the same or different. For $(P_1^L(i), P_2^M(j))$ to be a Nash Equilibrium (NE) in the coordination game the following conditions have to hold. Given the categorization of the opponent:

i) No player should have an incentive to deviate to a finer categorization than the one she is currently using

- Player 1. For all $P_1^{L-} \prec P_1^L(i)$ it has to hold that:

$$\begin{aligned}
&EPE(P_1^L(i), P_2^M(j)) \leq EPE(P_1^{L^-}, P_2^M(j)) \\
&\Leftrightarrow Var(P_1^L(i)) + Var(P_2^M(j)) + Bias^2(P_1^L(i), P_2^M(j)) \\
&\quad \leq Var(P_1^{L^-}) + Var(P_2^M(j)) + Bias^2(P_1^{L^-}, P_2^M(j)) \\
&\Leftrightarrow Var(P_1^L(i)) - Var(P_1^{L^-}) \leq Bias^2(P_1^{L^-}, P_2^M(j)) - Bias^2(P_1^L(i), P_2^M(j))
\end{aligned} \tag{37}$$

- Player 2. For all $P_2^{M-} \prec P_2^M(j)$ it has to hold that:

$$\begin{aligned}
&EPE(P_1^L(i), P_2^M(j)) \leq EPE(P_1^L(i), P_2^{M^-}) \\
&\Leftrightarrow Var(P_1^L(i)) + Var(P_2^M(j)) + Bias^2(P_1^L(i), P_2^M(j)) \\
&\quad \leq Var(P_1^L(i)) + Var(P_2^{M^-}) + Bias^2(P_1^L(i), P_2^{M^-}) \\
&\Leftrightarrow Var(P_2^M(j)) - Var(P_2^{M^-}) \leq Bias^2(P_1^L(i), P_2^{M^-}) - Bias^2(P_1^L(i), P_2^M(j))
\end{aligned} \tag{38}$$

ii) No player should have an incentive to deviate to another categorization at the same level of coarseness as the one she is currently using

- Player 1. For all $P_1^L(k) \sim P_1^L(i)$ it has to hold that:

$$\begin{aligned}
&EPE(P_1^L(i), P_2^M(j)) \leq EPE(P_1^L(k), P_2^M(j)) \\
&\Leftrightarrow Var(P_1^L(i)) + Var(P_2^M(j)) + Bias^2(P_1^L(i), P_2^M(j)) \\
&\quad \leq Var(P_1^L(k)) + Var(P_2^M(j)) + Bias^2(P_1^L(k), P_2^M(j)) \\
&\Leftrightarrow Var(P_1^L(i)) - Var(P_1^L(k)) \leq Bias^2(P_1^L(k), P_2^M(j)) - Bias^2(P_1^L(i), P_2^M(j))
\end{aligned} \tag{39}$$

- Player 2. For all $P_2^M(l) \sim P_2^M(j)$ it has to hold that:

$$\begin{aligned}
&EPE(P_1^L(i), P_2^M(j)) \leq EPE(P_1^L(i), P_2^M(l)) \\
&\Leftrightarrow Var(P_1^L(i)) + Var(P_2^M(j)) + Bias^2(P_1^L(i), P_2^M(j)) \\
&\quad \leq Var(P_1^L(i)) + Var(P_2^M(l)) + Bias^2(P_1^L(i), P_2^M(l)) \\
&\Leftrightarrow Var(P_2^M(j)) - Var(P_2^M(l)) \leq Bias^2(P_1^L(i), P_2^M(l)) - Bias^2(P_1^L(i), P_2^M(j))
\end{aligned} \tag{40}$$

iii) No player should have an incentive to deviate to a coarser categorization than the one she is currently using

- Player 1. For all $P_1^{L+} \succ P_1^L(i)$ it has to hold that:

$$\begin{aligned}
&EPE(P_1^L(i), P_2^M(j)) \leq EPE(P_1^{L^+}, P_2^M(j)) \\
&\Leftrightarrow Var(P_1^L(i)) + Var(P_2^M(j)) + Bias^2(P_1^L(i), P_2^M(j)) \\
&\quad \leq Var(P_1^{L^+}) + Var(P_2^M(j)) + Bias^2(P_1^{L^+}, P_2^M(j)) \\
&\Leftrightarrow Var(P_1^L(i)) - Var(P_1^{L^+}) \leq Bias^2(P_1^{L^+}, P_2^M(j)) - Bias^2(P_1^L(i), P_2^M(j))
\end{aligned} \tag{41}$$

- Player 2. For all $P_2^{M+} \succ P_2^M(j)$ it has to hold that:

$$\begin{aligned}
&EPE(P_1^L(i), P_2^M(j)) \leq EPE(P_1^L(i), P_2^{M+}) \\
&\Leftrightarrow Var(P_1^L(i)) + Var(P_2^M(j)) + Bias^2(P_1^L(i), P_2^M(j)) \\
&\quad \leq Var(P_1^L(i)) + Var(P_2^{M+}) + Bias^2(P_1^L(i), P_2^{M+}) \\
&\Leftrightarrow Var(P_2^M(j)) - Var(P_2^{M+}) \leq Bias^2(P_1^L(i), P_2^{M+}) - Bias^2(P_1^L(i), P_2^M(j))
\end{aligned} \tag{42}$$

Proposition 4. *Equilibrium Properties of the Coordination Game*

Part 1. Ruling Out Existence of Asymmetric Equilibria

Proof. We first prove that categorization profiles such that the two players use categorizations at different levels of coarseness cannot be a NE. Consider any categorization profile $(P_1^L(i), P_2^{L^+}(j))$ such that Player 2 uses a coarser categorization than Player 1. We now show that Player 1 always has an incentive to deviate to the exact same categorization that Player 2 uses. That is, we show that for $P_1^{L^+}(j) \succ P_1^L(i)$:

$$\begin{aligned}
&EPE(P_1^L(i), P_2^{L^+}(j)) > EPE(P_1^{L^+}(j), P_2^{L^+}(j)) \\
&\Leftrightarrow Var(P_1^L(i)) - Var(P_1^{L^+}(j)) > Bias^2(P_1^{L^+}(j), P_2^{L^+}(j)) - Bias^2(P_1^L(i), P_2^{L^+}(j))
\end{aligned} \tag{43}$$

Note that the LHS is the difference in variance between a finer and a coarser categorization. We know from Proposition 1 Part 1 that the difference in variance between a finer and a coarser categorization will be greater than zero for any $\sigma^2 > 0$. Therefore, if $\sigma^2 > 0$, the LHS is always positive. The RHS is either negative or zero, as the first term on the RHS is always zero. Hence the above inequality always holds. We have shown that a categorization profile in which two players use categorizations at different levels of coarseness cannot be a NE for any $\sigma^2 > 0$.

We now show that that if $Bias^2(P_1^L(i), P_2^L(j)) > 0$, i.e. if the categorizations that the two players use have different expected values of the estimators for at least one object type, then there exists no asymmetric equilibrium $(P_1^L(i), P_2^L(j))$ such that the two players use different categorizations at the same level of coarseness. We do this by showing that it is profitable to deviate to using the exact same categorization at the same level of coarseness rather than using different categorizations. That is for any $P_1^L(i) \sim P_1^L(j)$:

$$\begin{aligned}
&EPE(P_1^L(i), P_2^L(j)) > EPE(P_1^L(j), P_2^L(j)) \text{ if } Bias^2(P_1^L(i), P_2^L(j)) > 0 \\
&\Leftrightarrow Var(P_1^L(i)) - Var(P_1^L(j)) > -Bias^2(P_1^L(i), P_2^L(j)))
\end{aligned} \tag{44}$$

We know from Proposition 1 that the variance of any two categorizations at the same level of coarseness is the same and therefore the LHS $= 0$. If $Bias^2(P_1^L(i), P_2^L(j)) > 0$, then $LHS > RHS$ and we have shown that any categorization profile such that the two players use different categorizations at the same level of coarseness is not a NE. □

Part 2. Existence of Symmetric Equilibria

Proof. To show that a symmetric categorization profile $(P^L(i), P^L(i))$ is a NE we need to show that:

$$\begin{array}{l} \text{i) } EPE(P^L(i), P^L(i)) \leq EPE(P^{L^-}, P^L(i)) \text{ for all } P^{L^-} \prec P^L(i) \\ \text{ii) } EPE(P^L(i), P^L(i)) \leq EPE(P^L(j), P^L(i)) \text{ for all } P^L(j) \sim P^L(i) \\ \text{iii) } EPE(P^L(i), P^L(i)) \leq EPE(P^{L^+}, P^L(i)) \text{ for all } P^{L^+} \succ P^L(i) \end{array} \tag{45}$$

We have omitted the indices for the players, as the categorization profile considered is symmetric. Note that it is guaranteed that the first two conditions always hold as we have shown in Part 1 that no player ever has an incentive to use a categorization that is finer than the one the opponent uses and that no player ever has an incentive to deviate to using a different categorization at the same level of coarseness as the opponent. We therefore focus our attention on the third NE condition, which can be rewritten as:

$$Var(P^L(i)) - Var(P^{L^+}) \leq Bias^2(P^{L^+}, P^L(i)) \tag{46}$$

Observe that the squared bias component of the two players' EPE from each other will always be finite and that it is not affected by changes in σ^2. From Proposition 1 we know that the difference in variance between a finer and a coarser categorization is positive and increasing in σ^2. Thus, the LHS of the above equation is positive and increasing in σ^2. For a given $Bias^2(P^{L^+}, P^L(i))$, as σ^2 increases, the LHS increases and eventually for each pair of categorizations $(P^L(i), P^L(i))$ there will be a point at which the $LHS > RHS$, i.e. a player will have an incentive to deviate to a coarser categorization and $(P^L(i), P^L(i))$ will no longer be a NE. Thus, the number of symmetric NE is a decreasing function of σ^2. At the coarsest level there is no coarser categorization to deviate to and thus both players using the coarsest possible categorization is always a NE. If σ^2 is sufficiently high, it will be the only one. □

Part 3. Pareto-ranking of Symmetric Categorization Profiles

Proof. We now show that we can Pareto-rank all symmetric profiles in the coordination game, i.e. all outcomes such that both players use the exact same categorizations. Let us denote a symmetric categorization profile by $(P_1^L(i), P_2^L(i))$. For all symmetric profiles, since players are using the exact same categorizations their bias from each other is always $Bias^2(P_1^L(i), P_2^L(i)) = 0$. Thus, the EPE^C of these categorization profiles depends only on the variance. It is equal to $EPE^C(P_1^L(i), P_2^L(i)) = Var(P_1^L(i)) + Var(P_2^L(i))$. We know from Proposition 1 that the coarser the categorization, the smaller its variance. This means that for any positive noise level the EPE^C of a coarser symmetric categorization profile will always have a smaller variance component than the EPE^C of a finer symmetric categorization profile. Symmetric categorization profiles are therefore Pareto-ranked with coarser outcomes being more efficient than finer outcomes for any positive noise level. In the case of $\sigma^2 = 0$, $Var(P_1^L(i)) = Var(P_2^L(i)) = 0$, and therefore all symmetric categorization profiles are equally efficient.

Note also that as all categorizations at the same level of coarseness have an equal variance, all symmetric categorization profiles at the same level of coarseness will be equally efficient at any noise level in the coordination game (e.g. $EPE^C(P^L(i), P^L(i)) = EPE^C(P^L(j), P^L(j))$ for all i, j at a given level of coarseness). □

Proposition 5. *Connection Optimality IP and NE in Coordination Game*

Part 1. NE if Both Use Individually Optimal Categorization

Proof. The statement of the proposition gives a sufficient (though not necessary) condition for the optimality of $P^L(i)$ for IP to imply that the respective symmetric categorization profile in which both players use the individually optimal categorization is a NE in the coordination game. Note that to show that $(P^L(i), P^L(i))$ is a NE in the coordination game we need to show that the following conditions hold.

i) No player has a profitable deviation to a finer categorization. This condition is always fulfilled as we have shown in Proposition 4 Part 1 that it cannot be profitable for a player to use a categorization finer than the one the opponent uses.

ii) There is no profitable deviation to a different categorization of equal coarseness. This condition is also always fulfilled as the variance of all categorizations of equal coarseness is the same and thus from the perspective of the variance there is no benefit of moving from using the same categorization to using a different categorization than the opponent at the same coarseness level. Also, deviating to using a different categorization than the opponent, can only lead to an increase in the bias component of the EPE^C.

iii) There is no profitable deviation to a coarser categorization. This last condition can be decomposed in two parts.

iiia) The first one is that if $P^L(i)$ is optimal for IP, there is no profitable deviation to a coarser categorization $P^{L^+}(k) \in R(P^L(i))$ that is connected by a path to $P^L(i)$, i.e. a categorization that can be formed by merging categories that exist in $P^L(i)$.

iiib) The second part is that if $P^L(i)$ is optimal for IP, there is no profitable deviation to a coarser categorization $P^{L^+}(k) \notin R(P^L(i))$, i.e. a coarser categorization that is not connected by a path to $P^L(i)$.

We first consider iiia). We want to show that in the coordination game, if both players choose $P^L(i)$, then no player has an incentive to deviate to a coarser categorization $P^{L+}(k)$ that is connected by a path to $P^L(i)$. That is, for all $P^{L^+}(k) \succ P^L(i)$ such that $P^{L^+}(k) \in R(P^L(i))$:

$$\begin{aligned} &EPE^{IP}(P^L(i), P^L(i)) \leq EPE^{IP}(P^{L^+}(k), P^L(i)) \\ &\Leftrightarrow Var(P^L(i)) - Var(P^{L^+}(k)) \leq Bias^2(P^{L^+}(k), P^L(i)) \end{aligned} \tag{47}$$

We know that as $P^L(i)$ is optimal for IP the following holds:

$$\begin{aligned} &EPE^{IP}(P^L(i)) \leq EPE^{IP}(P^{L^+}(k)) \text{ for all } P^{L^+}(k) \succ P^L(i) \\ &\Leftrightarrow Var(P^L(i)) - Var(P^{L^+}(k)) \leq Bias^2(P^{L^+}(k)) - Bias^2(P^L(i)) \end{aligned} \tag{48}$$

Note that the LHS of the two inequalities is the same. We now compare the RHS.

We know from Proposition 1 Part 2 that for all $P^{L+1}(k) \in R(P^L(i))$:

$$Bias^2(P^{L+1}(k)) - Bias^2(P^L(i)) = \frac{tv(\hat{y}_t - \hat{y}_v)^2}{t+v} \tag{49}$$

Now consider $Bias^2(P^{L+1}(k), P^L(i))$. Assume that one player uses $P^L(i)$, which has two categories, one with t different object types and the other one with v different object types, with category beliefs $\hat{y}_t = \frac{\sum_{i=1}^{t}\mu_i}{t}$ and $\hat{y}_v = \frac{\sum_{j=1}^{v}\mu_j}{v}$ respectively. The other player uses a coarser categorization $P^{L+1}(k)$ which has a category which merges the category with t and the category with v object types and her category belief is $\hat{y}_{t+v} = \frac{\sum_{k=1}^{t+v}\mu_k}{t+v}$. Note that under the assumption of an equal number of objects of each type $\hat{y}_{t+v} = \frac{t\hat{y}_t + v\hat{y}_v}{t+v}$. All their other categories are equal. The $Bias^2$ of the two players' predictions from each other is:

$$\begin{aligned}
&Bias^2(P^{L+1}(k), P^L(i)) = \sum_{i=1}^{t}(\hat{y}_t - \hat{y}_{t+v})^2 + \sum_{j=1}^{v}(\hat{y}_v - \hat{y}_{t+v})^2 \\
&= t(\hat{y}_t - \hat{y}_{t+v})^2 + v(\hat{y}_v - \hat{y}_{t+v})^2 \\
&= t\left(\hat{y}_t - \frac{t\hat{y}_t + v\hat{y}_v}{t+v}\right)^2 + v\left(\hat{y}_v - \frac{t\hat{y}_t + v\hat{y}_v}{t+v}\right)^2 \\
&= t\left(\hat{y}_t^2 - 2\hat{y}_t\frac{t\hat{y}_t + v\hat{y}_v}{t+v} + \frac{(t\hat{y}_t + v\hat{y}_v)^2}{(t+v)^2}\right) + v\left(\hat{y}_v^2 - 2\hat{y}_v\frac{t\hat{y}_t + v\hat{y}_v}{t+v} + \frac{(t\hat{y}_t + v\hat{y}_v)^2}{(t+v)^2}\right) \\
&= \frac{tv(\hat{y}_t - \hat{y}_v)^2}{t+v}
\end{aligned} \tag{50}$$

Thus, we have shown that $Bias^2(P^{L+1}(k), P^L(i)) = Bias^2(P^{L+1}(k)) - Bias^2(P^L(i))$. This holds for any two categorizations with consecutive coarseness levels. Analogically, $Bias^2(P^{L+c}, P^L(i)) = Bias^2(P^{L+c}) - Bias^2(P^L(i))$ holds for all $P^{L+c} \succ P^L(i)$ such that $P^{L+c} \in R(P^L(i))$, with $c \in Z^+$. Hence if $P^L(i)$ is optimal for IP there is no profitable deviation to a coarser categorization $P^{L+c}(k)$ that is connected by a path to $P^L(i)$.

We now consider iiib), i.e. deviations to a coarser categorization $P^{L+d}(l) \notin R(P^L(i))$ that is not connected by a path to $P^L(i)$, with $d \in Z^+$.

iiib) The following condition guarantees that no player has an incentive to deviate from the symmetric categorization profile in which both use the individually optimal categorization to a coarser categorization on a different path.

$Var(P^L(i)) - Var(P^{L+d}(l)) \leq Bias^2(P^L(i), P^{L+d}(l))$

A sufficient (but not necessary) condition for the above inequality to hold is that:

$Bias^2(P^{L+d}(l)) - Bias^2(P^L(i)) \leq Bias^2(P^{L+d}(l), P^L(i))$

Without making further assumptions on the biases of different categorizations, this condition does not necessarily hold for all coarser categorizations that are not connected by a path to $P^L(i)$. But when it does hold it is sufficient to ensure that the categorization profile such that both players use the individually optimal categorization is a NE in the coordination game. □

Part 2. No Symmetric Equilibria in Categorizations Finer than Individually Optimal

Proof. If $P^L(i)$ is the finest optimal categorization for IP this means that all categorizations finer than $P^L(i)$ have a greater EPE in IP. Thus:

$$\begin{aligned} &EPE(P^{L^-}(j)) > EPE(P^L(i)) \text{ for all } P^{L^-}(j) \prec P^L(i) \\ &\Leftrightarrow Var(P^{L^-}(j)) - Var(P^L(i)) > Bias^2(P^L(i)) - Bias^2(P^{L^-}(j)) \end{aligned} \tag{51}$$

We know from Proposition 5 Part 1 that for any $P^{L^-}(j) \in R(P^L(i))$:
$Bias^2(P^L(i)) - Bias^2(P^{L^-}(j)) = Bias^2(P^L(i), P^{L^-}(j))$
Thus, the above is equivalent to:
$Var(P^{L^-}(j)) - Var(P^L(i)) > Bias^2(P^L(i), P^{L^-}(j))$
Now consider that for any $(P^{L^-}(j), P^{L^-}(j))$ to be a NE it is necessary that no player has an incentive to deviate to a coarser categorization. That is:

$$\begin{aligned} &EPE(P^{L^-}(j), P^{L^-}(j)) \leq EPE(P^L(i), P^{L^-}(j)) \\ &\Leftrightarrow Var(P^{L^-}(j)) - Var(P^L(i)) \leq Bias^2(P^L(i), P^{L^-}(j)) \end{aligned} \tag{52}$$

But this contradicts that $P^L(i)$ is the finest optimal categorization for IP. Therefore we have shown that if $P^L(i)$ is the finest optimal categorization for IP, then there exists no symmetric NE $(P^{L^-}(j), P^{L^-}(j))$ at a level finer than the finest individually optimal categorization such that $P^{L^-}(j)$ is connected by a path to $P^L(i)$.

We now give a sufficient condition under which there exists no symmetric equilibrium $(P^{L^-}(k), P^{L^-}(k))$ such that $P^{L^-}(k) \notin R(P^L(i))$. We need to show that:

$$\begin{aligned} &EPE(P^{L^-}(k), P^{L^-}(k)) > EPE(P^L(i), P^{L^-}(k)) \\ &\Leftrightarrow Var(P^{L^-}(k)) - Var(P^L(i)) > Bias^2(P^L(i), P^{L^-}(k)) \end{aligned} \tag{53}$$

We know from the fact that $P^L(i)$ is the finest optimal categorization in IP that:

$$\begin{aligned} &EPE(P^{L^-}(k)) > EPE(P^L(i)) \\ &\Leftrightarrow Var(P^{L^-}(k)) - Var(P^L(i)) > Bias^2(P^L(i)) - Bias^2(P^{L^-}(k)) \end{aligned} \tag{54}$$

Whenever the condition
$Bias^2(P^L(i)) - Bias^2(P^{L^-}(k)) \geq Bias^2(P^{L^-}(k), P^L(i))$
holds, it must be true that
$Var(P^{L^-}(k)) - Var(P^L(i)) > Bias^2(P^L(i), P^{L^-}(k))$.
Thus, a player then always has an incentive to deviate to the individually optimal categorization. □

Part 3. Symmetric equilibria at Levels Coarser than the Individually Optimal

Proof. For $(P^{L^+}(j), P^{L^+}(j))$ to be a NE in the coordination game it is necessary that no player has an incentive to deviate to any finer, to any other equally coarse or to any coarser categorization. We showed in Proposition 4 Part 1 that no player ever has an incentive to deviate from a symmetric categorization profile to using a finer or another equally coarse categorization. If the condition $Var(P^{L^+}(j)) - Var(P^{L^{++}}) \leq Bias^2(P^{L^{++}}, P^{L^+}(j))$ holds for each $P^{L^{++}} \succ P^{L^+}(j)$ then no player has an incentive to deviate to a coarser categorization. □

Lemma 3. *General Bias-Variance Decomposition* $EPE^{IP\&C}(P_1, P_2)$

Proof. This follows directly from combining Lemma 1 and Lemma 2. □

Definition 8. NE conditions of the $IP\&C$ game

For $(P^L(i), P^M(j))$ to be a Nash Equilibrium (NE) in the $IP\&C$ game the following conditions have to hold. Given the categorization of the opponent: i) no player should have an incentive to deviate to a finer categorization than the one she is currently using; ii) no player should have an incentive to deviate to another categorization at the same level of coarseness as the one she is currently using; and iii) no player should have an incentive to deviate to a coarser categorization than the one she is currently using.

Using the bias-variance decomposition from Lemma 3 to write the above conditions, we can rewrite the conditions in the following way.

i) No player should have an incentive to deviate to a finer categorization than the one she is currently using

- Player 1: For all $P_1^{L-} \prec P_1^L(i)$ it should hold that:

$$
\begin{aligned}
& EPE_1^{IP\&C}(P_1^L(i), P_2^M(j)) \leq EPE_1^{IP\&C}(P_1^{L-}, P_2^M(j)) \\
\Leftrightarrow\ & w\Big[Bias^2(P_1^L(i)) + Var(P_1^L(i))\Big] \\
& \quad + (1-w)\Big[Var(P_1^L(i)) + Var(P_2^M(j)) + Bias^2(P_1^L(i), P_2^M(j))\Big] \\
& \quad \leq w\Big[Bias^2(P_1^{L^-}) + Var(P_1^{L^-})\Big] \\
& \quad + (1-w)\Big[Var(P_1^{L^-}) + Var(P_2^M(j)) + Bias^2(P_1^{L^-}, P_2^M(j))\Big] \\
\Leftrightarrow\ & wBias^2(P_1^L(i)) + wVar(P_1^L(i)) + (1-w)Var(P_1^L(i)) \\
& \quad + (1-w)Var(P_2^M(j)) + (1-w)Bias^2(P_1^L(i), P_2^M(j)) \\
& \quad \leq wBias^2(P_1^{L^-}) + wVar(P_1^{L^-}) + (1-w)Var(P_1^{L^-}) \\
& \quad + (1-w)Var(P_2^M(j)) + (1-w)Bias^2(P_1^{L^-}, P_2^M(j)) \\
\Leftrightarrow\ & Var(P_1^L(i)) - Var(P_1^{L^-}) \\
& \quad \leq w\Big[Bias^2(P_1^{L^-}) - Bias^2(P_1^L(i))\Big] \\
& \quad + (1-w)\Big[Bias^2(P_1^{L^-}, P_2^M(j)) - Bias^2(P_1^L(i), P_2^M(j))\Big]
\end{aligned} \tag{55}
$$

- Player 2: For all $P_2^{M^-} \prec P_2^M(j)$ it should hold that:

$$\begin{aligned}
&EPE_2^{IP\&C}(P_1^L(i), P_2^M(j)) \leq EPE_2^{IP\&C}(P_1^L(i), P_2^{M^-}) \\
&\Leftrightarrow Var(P_2^M(j)) - Var(P_2^{M^-}) \\
&\quad \leq w\Big[Bias^2(P_2^{M^-}) - Bias^2(P_2^M(j))\Big] \\
&\quad + (1-w)\Big[Bias^2(P_1^L(i), P_2^{M^-}) - Bias^2(P_1^L(i), P_2^M(j))\Big]
\end{aligned} \tag{56}$$

ii) No player should have an incentive to deviate to another categorization at the same level of coarseness as the one she is currently using

- Player 1: For all $P_1^L(k) \sim P_1^L(i)$ it should hold that:

$$\begin{aligned}
&EPE_1^{IP\&C}(P_1^L(i), P_2^M(j)) \leq EPE_1^{IP\&C}(P_1^L(k), P_2^M(j)) \\
&\Leftrightarrow Var(P_1^L(i)) - Var(P_1^L(k)) \\
&\quad \leq w\Big[Bias^2(P_1^L(k)) - Bias^2(P_1^L(i))\Big] \\
&\quad + (1-w)\Big[Bias^2(P_1^L(k)), P_2^M(j)) - Bias^2(P_1^L(i), P_2^M(j))\Big]
\end{aligned} \tag{57}$$

- Player 2: For all $P_2^M(l) \sim P_2^M(j)$ it should hold that:

$$\begin{aligned}
&EPE_2^{IP\&C}(P_1^L(i), P_2^M(j)) \leq EPE_2^{IP\&C}(P_1^L(i), P_2^M(l)) \\
&\Leftrightarrow Var(P_2^M(j)) - Var(P_2^M(l)) \\
&\quad \leq w\Big[Bias^2(P_2^M(l)) - Bias^2(P_2^M(j))\Big] \\
&\quad + (1-w)\Big[Bias^2(P_1^L(i), P_2^M(l)) - Bias^2(P_1^L(i), P_2^M(j))\Big]
\end{aligned} \tag{58}$$

iii) No player should have an incentive to deviate to a coarser categorization than the one she is currently using

- Player 1: For all $P_1^{L^+} \succ P_1^L(i)$ it should hold that:

$$\begin{aligned}
&EPE_1^{IP\&C}(P_1^L(i), P_2^M(j)) \leq EPE_1^{IP\&C}(P_1^{L^+}, P_2^M(j)) \\
&\Leftrightarrow Var(P_1^L(i)) - Var(P_1^{L^+}) \\
&\quad \leq w\Big[Bias^2(P_1^{L^+}) - Bias^2(P_1^L(i))\Big] \\
&\quad + (1-w)\Big[Bias^2(P_1^{L^+}, P_2^M(j)) - Bias^2(P_1^L(i), P_2^M(j))\Big]
\end{aligned} \tag{59}$$

- Player 2: For all $P_2^{M^+} \succ P_2^M(j)$ it should hold that

$$\begin{aligned}
&EPE_2^{IP\&C}(P_1^L(i), P_2^M(j)) \leq EPE_2^{IP\&C}(P_1^L(i), P_2^{M^+}) \\
&\Leftrightarrow Var(P_2^M(j)) - Var(P_2^{M^+}) \\
&\quad \leq w\Big[Bias^2(P_2^{M^+}) - Bias^2(P_2^M(j))\Big] \\
&\quad + (1-w)\Big[Bias^2(P_1^L(i), P_2^{M^+}) - Bias^2(P_1^L(i), P_2^M(j))\Big]
\end{aligned} \tag{60}$$

Proposition 6. *Equilibrium Properties of the IP&C game*

Part 1. Ruling Out Existence of Some Asymmetric Equilibria

Proof. To show that $(P_1^{L^-}, P_2^L(i))$ can never be an equilibrium in $IP\&C$, we need to show that there always exists a profitable deviation for at least one player. Consider the deviation in which the player who uses the finer categorization switches to using the same individually optimal categorization as the opponent. We now show that this is always profitable. That is, we show that for all $P^{L^-} \prec P^L(i)$:

$$\begin{aligned}
&EPE_1^{IP\&C}(P_1^{L-}, P_2^L(i)) > EPE_1^{IP\&C}(P_1^L(i), P_2^L(i)) \\
&\Leftrightarrow w\Big[Var(P_1^{L^-}) + Bias^2(P_1^{L-})\Big] \\
&\qquad + (1-w)\Big[Var(P_1^{L-}) + Var(P_2^L(i)) + Bias^2(P_1^{L-}, P_2^L(i))\Big] \\
&\qquad > w\Big[Var(P_1^L(i)) + Bias^2(P_1^L(i))\Big] \\
&\qquad + (1-w)\Big[Var(P_1^L(i)) + Var(P_2^L(i)) + Bias^2(P_1^L(i), P_2^L(i))\Big] \\
&\Leftrightarrow wVar(P_1^{L^-}) + wBias^2(P_1^{L^-}) \\
&\qquad + (1-w)Var(P_1^{L^-}) + (1-w)Var(P_2^L(i)) + (1-w)Bias^2(P_1^{L^-}, P_2^L(i)) \\
&\qquad > wVar(P_1^L(i)) + wBias^2(P_1^L(i)) \\
&\qquad + (1-w)Var(P_1^L(i)) + (1-w)Var(P_2^L(i)) + (1-w)Bias^2(P_1^L(i), P_2^L(i)) \\
&\Leftrightarrow Var(P_1^{L^-}) - Var(P_1^L(i)) \\
&\qquad > w\Big[Bias^2(P_1^L(i)) - Bias^2(P_1^{L^-})\Big] - (1-w)Bias^2(P_1^{L^-}, P_2^L(i))
\end{aligned} \tag{61}$$

As $P^L(i)$ is the finest optimal for IP we know that $Var(P^{L^-}) - Var(P^L(i)) > Bias^2(P^L(i)) - Bias^2(P^{L^-})$. This implies that the LHS of the last inequality, $Var(P_1^{L^-}) - Var(P_1^L(i))$, is greater than $w\Big[Bias^2(P^L(i)) - Bias^2(P^{L^-})\Big]$ for any $0 \le w \le 1$. Note that as the other term on the RHS, $(1-w)Bias^2(P_1^{L^-}, P_2^L(i))$, is nonnegative, the LHS is greater than the RHS. We have shown that there exists no asymmetric equilibrium in the $IP\&C$ game such that one player uses the individually optimal categorization and the other player uses a finer one.

The second statement in Part 1 is that if $P^L(i)$ is the finest optimal categorization for IP, then any categorization profile $(P_1^{L^-}(j), P_2^{L^{--}}(k))$ such that players use categorizations at different levels of coarseness, both below the finest individually optimal and connected by a path to each other as well as to the finest individually optimal categorization, can never be a NE in the $IP\&C$ game. To show this, we need to show that there always exists a profitable deviation from $(P_1^{L^-}(j), P_2^{L^{--}}(k))$ for at least one player. Consider the deviation such that the first player switches to the individually optimal categorization $P^L(i)$. In the proposition we have given a sufficient condition under which this is profitable, i.e. under which we show that $EPE_1^{IP\&C}(P_1^{L^-}(j), P_2^{L^{--}}(k)) >$

$EPE_1^{IP\&C}(P_1^L(i), P_2^{L^{--}}(k))$.

$$
\begin{aligned}
& w\Big[Bias^2(P_1^{L^-}(j)) + Var(P_1^{L^-}(j))\Big] \\
& \quad + (1-w)\Big[Var(P_1^{L^-}(j)) + Var(P_2^{L^{--}}(k)) + Bias^2(P_1^{L^-}(j), P_2^{L^{--}}(k))\Big] \\
& \quad > w\Big[Bias^2(P_1^L(i)) + Var(P_1^L(i))\Big] \\
& \quad + (1-w)\Big[Var(P_1^L(i) + Var(P_2^{L^{--}}(k)) + Bias^2(P_1^L(i), P_2^{L^{--}}(k))\Big] \\
& \Leftrightarrow wBias^2(P_1^{L^-}(j)) + wVar(P_1^{L^-}(j)) + (1-w)Var(P_1^{L^-}(j)) \\
& \quad + (1-w)Var(P_2^{L^{--}}(k)) + (1-w)Bias^2(P_1^{L^-}(j), P_2^{L^{--}}(k)) \\
& \quad > wBias^2(P_1^L(i)) + wVar(P_1^L(i)) + (1-w)Var(P_1^L(i)) \\
& \quad + (1-w)Var(P_2^{L^{--}}(k)) + (1-w)Bias^2(P_1^L(i), P_2^{L^{--}}(k)) \\
& \Leftrightarrow Var(P_1^{L^-}(j)) - Var(P_1^L(i)) \\
& \quad > w\Big[Bias^2(P_1^L(i)) - Bias^2(P_1^{L^-}(j))\Big] \\
& \quad - (1-w)\Big[Bias^2(P^{L^-}(j), P^{L^{--}}(k)) - Bias^2(P_1^L(i), P_2^{L^{--}}(k))\Big]
\end{aligned} \tag{62}
$$

Since $P^L(i)$ is the finest optimal for IP we know that:
$Var(P^{L^-}(j)) - Var(P^L(i)) > Bias^2(P^L(i)) - Bias^2(P^{L^-}(j))$.

Therefore the LHS is greater than $w\Big[Bias^2(P_1^L(i)) - Bias^2(P_1^{L^-}(j))\Big]$ for any $0 \leq w \leq 1$. Since we assumed that $Bias^2(P_1^{L^-}(j), P_2^{L^{--}}(k)) \leq Bias^2(P_1^L(i), P_2^{L^{--}}(k))$, the LHS is greater than the RHS. We have thus given a sufficient condition to ensure that there exists no equilibrium in the $IP\&C$ game such that the two players use categorizations at different levels of coarseness, both finer than the individually optimal one and both connected by a path to the individually optimal one. □

Part 2. Symmetric Equilibria

Proof. We need to show that if $P^L(i)$ is optimal for IP and $(P^L(i), P^L(i))$ is a NE in the coordination game, then there exists no profitable deviation from $(P^L(i), P^L(i))$ to any finer, any equally coarse or any coarser categorization in $IP\&C$.

i) We first show that there is no profitable deviation to a finer categorization. According to Definition 8:

$$
\begin{aligned}
& EPE_1^{IP\&C}(P_1^L(i), P_2^L(i)) \leq EPE_1^{IP\&C}(P_1^{L-}, P_2^L(i)) \\
& \Leftrightarrow w\Big[Var(P_1^L(i)) + Bias^2(P_1^L(i))\Big] + (1-w)\Big[Var(P_1^L(i)) + Var(P_2^L(i))\Big] \\
& \quad + Bias^2(P_1^L(i), P_2^L(i)) \\
& \quad \leq w\Big[Var(P_1^{L^-}) + Bias^2(P_1^{L^-})\Big] \\
& \quad + (1-w)\Big[Var(P_1^{L^-}) + Var(P_2^L(i)) + Bias^2(P_1^{L^-}, P_2^L(i))\Big] \\
& \Leftrightarrow Var(P_1^L(i)) - Var(P_1^{L^-}) \\
& \quad \leq w\Big[Bias^2(P_1^{L^-}) - Bias^2(P_1^L(i))\Big] + (1-w)Bias^2(P_1^{L^-}, P_2^L(i))
\end{aligned} \tag{63}
$$

Note that the LHS $\leq Bias^2(P_1^{L^-}) - Bias^2(P_1^L(i))$ follows directly from the optimality of $P^L(i)$ for IP. LHS $\leq Bias^2(P_1^{L^-}, P_2^L(i))$ follows directly from $(P^L(i), P^L(i))$ being a NE in the coordination game. Therefore, the LHS is always smaller than the RHS and no player has a profitable deviation to a finer categorization.

ii) There is no profitable deviation to another equally coarse categorization.

$EPE_1^{IP\&C}(P_1^L(i), P_2^L(i)) \leq EPE_1^{IP\&C}(P_1^L(j), P_2^L(i))$

The same reasoning applies as above.

iii) There is no profitable deviation to a coarser categorization.

$EPE_1^{IP\&C}(P_1^L(i), P_2^L(i)) \leq EPE_1^{IP\&C}(P_1^{L+}, P_2^L(i))$

The same reasoning applies as above.

We now consider the second statement from Part 2. We need to show that if $P^L(i)$ is the finest optimal for IP, then $(P^{L^-}(k), P^{L^-}(k))$ is not a NE in the $IP\&C$ game for any $P^{L^-}(k) \in R(P^L(i))$. We show this by showing that a player has a profitable deviation to $P^L(i)$, i.e. $EPE_1^{IP\&C}(P_1^{L^-}(k), P_2^{L^-}(k)) > EPE_1^{IP\&C}(P_1^L(i), P_2^{L^-}(k))$.

$$
\begin{aligned}
& w\Big[Var(P_1^{L^-}(k)) + Bias^2(P_1^{L^-}(k))\Big] \\
&\quad + (1-w)\Big[Var(P_1^{L^-}(k)) + Var(P_2^{L^-}(k)) + Bias^2(P_1^{L^-}(k), P_2^{L^-}(k))\Big] \\
&\quad > w\Big[Var(P_1^L(i)) + Bias^2(P_1^L(i))\Big] \\
&\quad + (1-w)\Big[Var(P_1^L(i)) + Var(P_2^{L^-}(k)) + Bias^2(P_1^L(i), P_2^{L^-}(k))\Big] \\
& \Leftrightarrow wVar(P_1^{L^-}(k)) + wBias^2(P_1^{L^-}(k)) \\
&\quad + (1-w)Var(P_1^{L^-}(k)) + (1-w)Var(P_2^{L^-}(k)) \\
&\quad > wVar(P_1^L(i)) + wBias^2(P_1^L(i)) + (1-w)Var(P_1^L(i)) \\
&\quad + (1-w)Var(P_2^{L^-}(k)) + (1-w)Bias^2(P_1^L(i), P_2^{L^-}(k)) \\
& \Leftrightarrow Var(P_1^{L^-}(k)) - Var(P_1^L(i)) \\
&\quad > w\Big[Bias^2(P_1^L(i)) - Bias^2(P_1^{L^-}(k))\Big] + (1-w)Bias^2(P_1^L(i), P_2^{L^-}(k))
\end{aligned} \tag{64}
$$

We know from Proposition 5 Part 1 that for categorizations that are connected by a path $Bias^2(P^L(i)) - Bias^2(P^{L^-}(k)) = Bias^2(P^{L^-}(k), P^L(i))$.

Thus, the last two lines become:

$$
Var(P_1^{L^-}(k)) - Var(P_1^L(i)) > Bias^2(P_1^L(i)) - Bias^2(P_1^{L^-}(k)) \tag{65}
$$

We know that if $P^L(i)$ is the finest optimal for IP, then the LHS is indeed greater than the RHS. We have thus shown that there exists no equilibrium in symmetric categorizations that are finer and that are on the same path as the individually optimal one in the $IP\&C$ game. □

Part 3. Pareto-ranking of Symmetric Profiles

Proof. We now derive a necessary and sufficient condition for any coarser categorization profile to be Pareto-superior to any finer categorization profile in the $IP\&C$ game. That is, we are interested in when it is true for all $P^{L^+}(j) \succ P^L(i)$ that:

$$\begin{aligned}
&EPE_1(P_1^{L^+}(j), P_2^{L^+}(j)) \leq EPE_1(P_1^L(i), P_2^L(i)) \\
&\Leftrightarrow w\Big[Var(P_1^{L^+}(j)) + Bias^2(P_1^{L^+}(j))\Big] \\
&\quad + (1-w)\Big[Var(P_1^{L^+}(j)) + Var(P_2^{L^+}(j)) + Bias^2(P_1^{L^+}(j), P_2^{L^+}(j))\Big] \\
&\quad \leq w\Big[Var(P_1^L(i)) + Bias^2(P_1^L(i))\Big] \\
&\quad + (1-w)\Big[Var(P_1^L(i)) + Var(P_2^L(i)) + Bias^2(P_1^L(i), P_2^L(i))\Big] \\
&\Leftrightarrow wVar(P_1^{L^+}(j)) + wBias^2(P_1^{L^+}(j)) + Var(P_1^{L^+}(j)) \\
&\quad + Var(P_2^{L^+}(j)) - wVar(P_1^{L^+}(j)) - wVar(P_2^{L^+}(j)) \\
&\quad \leq wVar(P_1^L(i)) + wBias^2(P_1^L(i)) + Var(P_1^L(i)) \\
&\quad + Var(P_2^L(i)) - wVar(P_1^L(i)) - wVar(P_2^L(i)) \\
&\Leftrightarrow wBias^2(P_1^{L^+}(j)) - wVar(P_2^{L^+}(j)) - wBias^2(P_1^L(i)) + wVar(P_2^L(i)) \\
&\quad \leq 2\Big[Var(P^L(i)) - Var(P^{L^+}(j))\Big] \\
&\Leftrightarrow w\Big[Bias^2(P^{L^+}(j)) - Bias^2(P^L(i)) + Var(P^L(i)) - Var(P^{L^+}(j))\Big] \\
&\quad \leq 2\Big[Var(P^L(i)) - Var(P^{L^+}(j))\Big] \\
&\Leftrightarrow w \leq \frac{2\Big[Var(P^L(i)) - Var(P^{L^+}(j))\Big]}{Bias^2(P^{L^+}(j)) - Bias^2(P^L(i)) + Var(P^L(i)) - Var(P^{L^+}(j))}
\end{aligned} \tag{66}$$

□