Categorization and Coordination*

Vessela Daskalova† and Nicolaas J. Vriend‡

September 14, 2017

Abstract

This paper considers individuals who make predictions in new situations using categories. It presents a framework to examine when decision makers may be better off using fewer rather than more categories, even without exogenous costs of using more. We study three aspects: An individual whose goal is to predict the unobserved value of a new object, decision makers who want to coordinate their predictions with each other, and decision makers who are interested in both. We show that a coordination motive might sometimes provide incentives for coarse categorization.

JEL Classification: C72, D01, D03, D80

Keywords: Categorization, Coordination, Prediction, Bounded rationality

1 Introduction

People often make predictions in new situations with the help of categories.[1] That is, when encountering a new situation, the decision maker assigns it

*We thank colleagues and conference participants at Cambridge, Queen Mary, Stanford, Toulouse School of Economics, Toulouse (*ESEM*), Copenhagen (*Limited Attention, Selective Perception and Bounded Rationality in Decision Making*), Manchester (*RES*), Barcelona (*GSE Workshop on Limited Reasoning and Cognition*), Mannheim (*EEA*), Lisbon (*UECE*), Girona (*SAEe*), East Anglia (*Workshop on Behavioural Game Theory*), and Maastricht (*GAMES*) for comments and discussions. All errors are our own. Vessela Daskalova gratefully acknowledges support through the ANR Labex IAST and Cambridge-INET.

†Corresponding author: Institute for Advanced Study in Toulouse and Toulouse School of Economics, <vessela.daskalova@iast.fr>, <vessela.daskalova@tse-fr.eu>.

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

[1] Abundant research in psychology, cognitive science, and more recently economics suggests so, see e.g. Lakoff (1987), Cohen and Lefebvre (2005), Grimm and Mengel (2012), or Mengel (2012a).

to some category of past experiences, and makes a prediction based on the average experience in that category. Based on the experimental literature in psychology and on previous models in economics, two key features of category-based prediction stand out. First, the average experience in the category is used as a predictor, and second, experiences in other categories do not matter (Murphy and Ross, 1994; Mohlin, 2014). For example, an employer may place a new applicant who is "male and engineer" in a category with others who share these characteristics, and make a prediction about the new applicant's suitability for the job based on the category average.

While people often care about making good predictions, they also often care about *coordinating* their predictions with others. As is well known, the need to coordinate could arise from various factors in different classes of situations - these include strategic complementarity in subsequent actions, reputation concerns, and preference for conformity. Technology adopters may aim to coordinate their predictions on the usefulness of a new technology as their payoffs from adoption increase in the number of other adopters. Counter-terrorism officers may need to coordinate their prediction of the level of danger someone poses in order to later act jointly to disarm the person. Equity researchers may have incentives to coordinate as they might be worried about their reputation if their predictions are too far off from the consensus estimates. Product reviewers, project evaluators or referees on a paper may also have such incentives.[2]

Usually there are many different ways in which an individual can sort her experiences into categories. In particular, she can divide her experiences into many categories (fine categorization) or alternatively, she can split them into just a few categories, each category containing a large number of experiences of different types (coarse categorization). Coarse categorization has been linked to a number of biases in decision-making, including discrimination against minorities, persuasion in advertising, and mispricing in markets.[3] In spite of this, coarse categorization is common. This paper provides a parsimonious framework that helps us develop a formal intuition for why coarse categorization could be beneficial. The use of few rather than many categories could in principle be due to an innate limitation in our brains or to exogenous (e.g. computational) costs of using more categories. Our analysis complements such explanations, as

[2] Some papers discussing these and other situations in which the coordination motive matters include Morris and Shin (2002), Scharfstein and Stein (1990), Meub et al. (2015), Hellwig and Veldkamp (2009) and Swisher (2013).

[3] See Fryer and Jackson (2008), Mullainathan et al. (2008), Barberis and Shleifer (2003) and Steiner and Stewart (2015), respectively, which are discussed in Section 2.

we show when and why coarse categorization can be rational even without such exogenously given costs or limitations.

The starting point is the following problem. Decision makers encounter objects. Each object has a number of observable characteristics plus some *ex ante* unobservable value.[4] We assume that the agent accumulates a number of experiences for which the value is revealed to her. She, then, faces a one-off prediction task. That is, she will encounter a new object, and will need to predict the unobservable value. There is an underlying function relating 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 experiences, which are sorted into categories. That is, she will assign the object to a category on the basis of the observables, and will use her average experience in that category to predict the unobservable value of the newly encountered object.[5]

We consider three variants of this prediction task. To begin with, we examine properties of alternative ways of categorizing for the case of a decision maker who is only interested in predicting the unobserved value of an object, i.e. the situation described above. Second, and this is central to our model, we examine properties of different ways of categorizing for the case of decision makers who only want to coordinate their predictions with each other. We consider a stylized representation of such a coordination problem, abstracting from a variety of specific aspects of situations in which people want to coordinate. Two individuals independently accumulate a number of experiences and their goal is to coordinate predictions on the next object. We represent this situation as a one-shot game in which each player has to choose a categorization to use and payoffs are inversely related to the players' expected prediction error from each other. We characterize the equilibrium properties of this game. The third variant of the prediction task that we consider is the combination of the two benchmark cases just described. That is, decision makers are interested both in correctly predicting the value of an object and in coordinating their prediction with another person. We also analyze the equilibrium properties of this game.

[4]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., and the unobserved value that the agent is trying to predict may be the applicant's suitability for a particular job.

[5]In this model we assume *that* people categorize. On the complementary question as to why decision makers may want to use categories at all to make predictions, see e.g. Peski (2011). He also provides a comparison of a categorizing and a Bayesian decision maker. Our model offers a perspective that is complementary to the Bayesian one.

We use statistics and game theory to derive comparative statics on factors affecting which way of categorizing (in terms of coarseness) will help agents minimize expected prediction errors on the next object. A key insight of the approach is to think of a categorization as a model containing a set of estimators that agents can use to make predictions.The model presented here is thus non-Bayesian as it is closer to classical/frequentist statistics and as such complementary to the Bayesian paradigm. Note in particular that in this model a decision maker does not need to have prior beliefs, and instead it can be seen as one way to model how priors are generated.

We begin by showing that whether fine or coarse categorization is optimal for individual prediction depends on the environment, in particular on the noise level and on the number of experiences the agent has available. If the environment is deterministic, the best the agent can do is to use the finest possible categorization, as it makes unbiased predictions. As noise increases, however, coarser categorizations may make better predictions, as they help to decrease the variance in prediction.

Our key results relate to the case when two players who use categorization as a tool for prediction want to coordinate their predictions. While in a deterministic environment, using the same categorization as the other player is always a best response, and hence all symmetric categorization profiles are equilibria and are equally efficient, this is not the case in a stochastic environment. The more noisy the environment, the more likely it is that only coarser symmetric categorization profiles constitute equilibria and if the noise to sample size ratio is sufficiently high the coarsest one is the unique one. Moreover, efficiency arguments suggest that any coarser symmetric categorization profile is strictly better than any finer categorization profile in any stochastic environment. The intuition is that while using the same categorization helps players decrease the bias in their predictions from each other, coarse categorization helps them decrease their variances in prediction by making predictions less dependent on the idiosyncratic experiences the individuals have accumulated.

In case players care both about individual prediction and about coordination, equilibrium existence and efficiency arguments again point to the optimality of coarse categorization if the weight on coordination is sufficiently high. The sufficient threshold weight is a decreasing function of the noise level and an increasing function of the sample size, and for any given weight on coordination there exists a noise level to sample size ratio that is sufficiently high to render coarse categorization optimal.

On the methodological side, this paper develops a coherent theoretical framework to analyze motives for coarse categorization in both individual decision making and strategic interaction settings. A key contribution is to analyze the interactions of a coordination motive with environmental factors in affecting the optimal way of categorizing. We show that while a coordination motive does not provide a rationale to categorize coarsely in a deterministic environment, it does provide such a rationale in stochastic environments. A tentative implication is that to the extent that coarse categorization can be linked to biases in decisions, such biases could be more likely to arise in stochastic environments in which individuals have incentives to coordinate.

The structure of the paper is as follows. In Section 2 we discuss the paper's relation to the literature. Section 3 describes our model. In Section 4 we present our analysis, followed by Section 5 where we provide a simple illustration of the main results. Section 6 concludes.

2 Relation to the Literature

This paper complements several lines of recent research in economics and is also related to literature in psychology, cognitive science, and computer science. Our approach to analyzing categorizations as containing sets of estimators is broadly inspired by the statistics, econometrics, and machine learning literatures on prediction (Berry and Lindgren, 1996; Varian, 2014; Li and Racine, 2007; Gareth et al., 2013; Bishop, 2007; Murphy, 2012). The literature often distinguishes between parametric and non-parametric models. One can view categorization as a non-parametric model. The finding that coarse categorization could be better for individual prediction in stochastic environments when individuals have limited sample sizes is consistent with Mohlin (2014) and Al-Najjar and Pai (2014). Other papers on categorization in individual prediction include Mullainathan (2002), who compares the accuracy of predictions of an agent who categorizes with the accuracy of predictions of a Bayesian decision maker, and Peski (2011), who 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.[6] As the

[6]There is also a less related literature on categorization that focuses on somewhat different questions than the prediction issue (Manzini and Mariotti, 2012; Mandler et al., 2012; Azrieli and Lehrer, 2007).

model of case-based decision making of Gilboa and Schmeidler (1995), our model of categorization is also a form of decision making on the basis of past experience. In our model, however, as in other categorization models, it is assumed that objects are placed in a category based on their observable attributes and that only objects within a category matter for prediction. This means that we can avoid the need to specify a particular similarity function that compares objects.

Our paper complements the above literature through a number of differences in methodology and focus. A key difference is that whereas these papers focus on individual prediction only, we consider the effect of the coordination motive on the way an individual categorizes. Our contribution to the literature is that we develop a coherent framework to analyze rationales for coarse categorization in both individual decision making and strategic interaction settings. To the best of our knowledge, this is the first paper to look at the question how an individual's optimal prediction model is affected by the strategic motive to coordinate predictions with others.

The paper also belongs to the literature considering the use of categories in strategic interactions. This literature has analyzed players who categorize their opponents in games (Azrieli, 2009, 2010), players who bundle nodes at which other players move into analogy classes (Jehiel, 2005), players who partition their own moves into similarity classes (Jehiel and Samet, 2007), players who learn how to categorize their strategies (Daskalova and Vriend, 2015), and players who categorize games (Mengel, 2012b; Grimm and Mengel, 2012; Heller and Winter, forthcoming).[7] This paper, instead, considers players who make predictions based on their past experiences, and analyzes how a player's choice of a categorization (model) to use depends on the motive to coordinate with others and on the environment.

By deriving rationales for coarse categorization this paper complements papers studying consequences of coarse categorization. Fryer and Jackson (2008) assume an exogenous limit on the number of categories a decision maker has available and 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, in particular under-

[7] Mengel (2012a) takes an evolutionary approach to explain categorization.

and overpricing, observed in financial markets. According to Steiner and Stewart (2015) frequent trading opportunities can lead to price distortions if traders use categories when making predictions. Our paper contributes to understanding why people may categorize coarsely even though coarse categorization can be linked to a number of biases in decision making, and in particular why such biases may be amplified if people want to coordinate predictions. It suggests that in stochastic environments categorizing coarsely rather than finely is not as boundedly rational as it may seem. That is, people may make better predictions by categorizing coarsely rather than finely even if they are not forced to do so by exogenous limitations or by computational costs related to using more categories.

The main message of the paper that the coordination motive may affect the optimal model for an individual and may lead to coarse categorization in stochastic environments is consistent with casual examples of coarse categorization occurring when people try to coordinate. For example, people may use coarse categories and stereotypes when talking to each other, even though they may individually not think in terms of such crude stereotypes. 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). When talking to others, people also typically refer to colors using relatively coarse categories, such as e.g. 'red', 'green', 'blue', even though they may be distinguishing much finer nuances, and even though well-defined, finer color schemes exist (Steels and Belpaeme, 2005; Komarova et al., 2007).

The related literature in psychology, cognitive science, and computer science is abundant. Most related in spirit is the work in cognitive science by Gigerenzer and Brighton (2009), who look at decision making on the basis of heuristics and argue that less information can improve accuracy in prediction. A complementary approach in psychology is to assume that having many categories entails computational costs for the agent and to derive the optimality of coarse categorization under such exogenous costs (Anderson, 1990, 1991). Making decisions on the basis of categories is also related to making decisions on the basis of analogies (Mitchell and Hofstadter, 1996; Hofstadter, 1996), as well as to clustering and classification algorithms in machine learning; see e.g. Bishop (2007) or Murphy (2012). While cognitive scientists and psychologists usually focus on empirical studies and procedural models of how people categorize, as economists we are particularly interested in understanding rationales for coarse categorization.[8] Another key difference is that as economists we are interested in the effect of strategic motives on the way individuals categorize.

3 Model

The basic set-up of the model is the following. An individual will accumulate a number of experiences and will face a one-off prediction task. That is, she will encounter a new object and will have to predict its unobservable value. To do so she will assign the object to a category based on its observed characteristics, and will make a prediction that its unobservable value is equal to the average of the values experienced in that category. The focus here is on the statistical properties of different ways of categorizing her experiences. In particular, we derive comparative statics on factors affecting which ways of categorizing her experiences will help minimize her expected prediction error on the next object.

We consider three cases: An individual interested only in making correct predictions, individuals interested only in coordinating their predictions with each other, and individuals interested both in correct individual prediction and in coordination of predictions. Before analyzing these three cases, we provide some definitions.

3.1 Objects and Categories

An object $o \equiv (x, y)$ is a vector of observed attributes $x \in X$, where X is a finite set of object types, together with an unobserved real-valued feature $y \in \mathbb{R}$ that the individual is trying to predict. Each object's type is determined by its observable attributes, that is by x only. The agent will sample a number of objects before having to make a prediction. We denote the set of all objects the individual will experience before having to make a prediction by O. To facilitate analytical tractability we focus on a symmetric setting in which the agent will sample an equal number n of objects from each object type before making a prediction (with $n > 0$ and n finite).[9] We assume that the values of these experiences are revealed to her before she has to make a prediction.

[8] For some overviews of cognitive science models, see Ashby and Maddox (2005), Zaki et al. (2003), and Smith and Medin (1981). Classical papers on categorization include Rosch (1975, 1978) and Nosofsky (1986).

[9] The analysis can be extended to allow e.g. for objects of each type to be sampled from some probability distribution.

The unobserved feature y that the individual is trying to predict is a noisy function of the observed attributes of the object, i.e. $y = f(x)+\epsilon$. The noise term is i.i.d. normally distributed with mean zero and variance σ^2, i.e. $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Thus, the unobserved value for each object is a random variable $Y \sim \mathcal{N}(f(x), \sigma^2)$. The variance of the noise is homogeneous across agents and across object types.[10] 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. A stochastic environment (i.e. $\sigma^2 > 0$) 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 directly observed when the agent makes a prediction about someone's suitability for a particular job). In the remainder of this paper we focus on stochastic environments ($\sigma^2 > 0$), unless explicitly stated otherwise. We do not make any assumptions on f apart from $f : X \to \mathbb{R}$, where X is the set of vectors of observable attributes and $\mathbb{R}$ is the set of real numbers. The agent does not know the function relating unobserved to observed attributes and therefore uses categories to make predictions. A category C is a subset of the set of objects O.

3.2 Category Beliefs and Categorizations

When encountering a new object, the agent will assign it to a category based on its observable characteristics, and she will predict as y-value of the object the average of the y-values of all her experiences in this category. The category average for category C is $\hat{Y}^C = \frac{1}{|C|}\sum_{o_i \in C} y_i$. Note that a category average is a random variable that depends on the exact realization of the y-values of the objects in the category. It is therefore the average of random variables which are independently drawn from a normal distribution and is also normally distributed. In our framework we think of the category averages as estimators the individual uses to make predictions.

A categorization $P = \{C_1, C_2, ..., C_k\}$ is a decision maker's complete model for making predictions. It is a set of categories that partition the set of objects based on the objects' observable characteristics.[11]

[10] The analysis can be extended to allow for objects of different object types to have different variance of the noise term.

[11] No two categories in the same categorization can cover the same object type, an assumption which is relaxed in the dynamic model in which we also consider hierarchical categorizations;

There are different possible ways to partition a set of experiences into categories. The focus of our analysis is on the coarseness of the different possible categorizations. Coarseness is determined by the number of categories a categorization contains, that is by its cardinality denoted as $|P|$. We say a categorization P is finer than P' if and only if $|P| > |P'|$. P is coarser than P' if and only if $|P| < |P'|$. P and P' are equally coarse if and only if $|P| = |P'|$. The set of all possible categorizations is denoted by $\mathcal{P}$.

To provide an illustration of what categories and categorizations could look like we further assume that the vector of observable attributes is binary of length l, i.e. $x \subset \{0,1\}^l$. Thus, for example, if $l = 2$, then the set of object types is $X = \{11, 10, 01, 00\}$, and the possible categorization types, 15 in this case, are illustrated below.

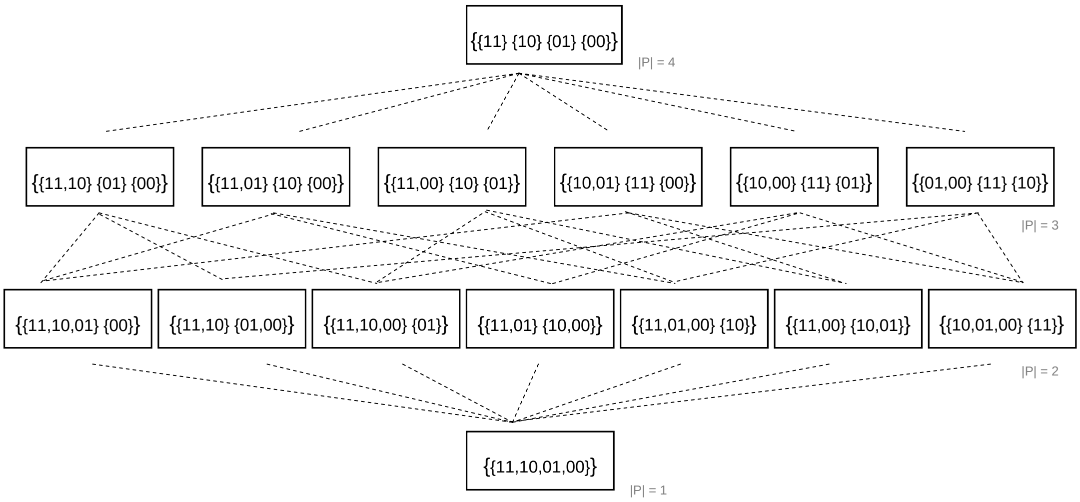

Figure 1: Set of categorizations for the case of four object types

Figure 1 shows all possible categorizations in the case of four object types, distinguishing four levels of coarseness, from the finest possible categorization in which each object type has a separate category at the top ($|P| = 4$), down to the coarsest possible categorization with all object types being assigned to a single category at the bottom ($|P| = 1$). While there is always exactly one finest and one coarsest possible categorization, at intermediate coarseness levels there will be (many) more categorizations, with the exact number depending on the coarseness level and on the number of different attributes describing the objects.

In addition to the concept of categorization coarseness, we define a binary relation "refines" denoted by $\succ$ on the set of possible categorizations $\mathcal{P}$. A

see Daskalova and Vriend (2014).

categorization P refines P', i.e. $P \succ P'$ if for each category $C \in P$, there exists a $C' \in P'$ such that $C \subseteq C'$.[12] Note that while $P \succ P' \Rightarrow |P| > |P'|$, the inverse is not true.

4 Analysis of the Model

The analysis focuses on factors determining which categorizations (in terms of their coarseness) would perform best in different situations. We derive comparative statics on these factors determining optimal coarseness. The characterization of optimal coarseness is not limited to the knowledge attained from observing a particular sample, but takes into account the fact that the experiences in this sample are random.[13] We first analyze the two benchmark cases, i.e., when an individual cares only about predicting the value of the object correctly, and when agents care only about coordinating their predictions about the value of the next object with one another. We then analyze the case when agents care about both. The results and intuition are in the main text, while the proofs are in Appendix A.

4.1 Individual Prediction

The goal of the agent is to make the best possible prediction on the next object she encounters.[14] We assume that the next object is drawn from a discrete uniform distribution of all possible object types.

Definition 1. Expected Prediction Error Individual Prediction

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

[12]For example, in Figure 1 categorization $\{\{11, 10\}\{01\}\{00\}\}$ refines $\{\{11, 10, 01\}\{00\}\}$, but it does not refine $\{\{11, 01, 00\}\{10\}\}$.

[13]In terms of perspective there are some commonalities with the way statisticians study the properties of the sample average as an estimator for the population average. As the statistician is aware of the randomness it does not matter whether he analyzes the problem before or after the realization of the sample. Taking an ex-ante perspective is in line with the approach of classical/frequentist statistics and has the advantage of being less cognitively demanding than the Bayesian approach, as what is optimal can be computed in advance allowing for quicker reactions.

[14]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 would not change. In our complementary dynamic model we consider an agent who experiences a sequence of objects and learns from her experience with each object; see Daskalova and Vriend (2014).

The indicator that 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 Definition 1 shows, the expected prediction error on an object type x_j is the expected mean squared error between the category average of the category the object is assigned to ($\hat{Y}^{C_k}$) and the object's unobserved value (Y_j).[15] Note that $\hat{Y}^{C_k}$ and Y_j are capitalized as they denote random variables. The expected prediction error of a categorization P combines the expected prediction error on all object types, with $p = \frac{1}{|X|}$.

Lemma 1. *Bias-Variance Decomposition of EPE^{IP}*

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

Lemma 2 decomposes the expected prediction error of a categorization in the individual prediction case into a bias and a variance component.[16] It shows that the $EPE^{IP}[P]$ increases if any of the following components increases: the expected variance of the category averages 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 averages of the categorization $Bias^2[P]$. For expositional convenience from now on we write bias and variance instead of expected bias and expected variance component of the expected prediction error.[17]

Lemma 2. *Comparative Statics of Bias and Variance Components of EPE^{IP}*

Part 1. For any categorization P, $Var[P] + Var[Y] = \frac{p\sigma^2|P|}{n} + \sigma^2$ is strictly increasing in σ^2 and strictly decreasing in n. It is strictly increasing in the

[15]The mean squared error is a standard way of measuring how good an estimator is; see e.g. Berry and Lindgren (1996). Minimizing mean squared error is equivalent to assuming maximization of a utility function that is simply the negative of the mean squared error.

[16]This is an analogy to bias and variance as properties of an estimator, but here we apply the concepts to the phenomenon of categorization, which in our framework contains a set of estimators, i.e. a set of category averages.

[17]The term variance can have different meanings in other contexts. Throughout this paper, we use it exclusively to denote the variance component of an EPE.

number of categories $|P|$.

Part 2. For any $P \succ P'$: $Bias^2[P] \leq Bias^2[P']$.

Lemma 2 Part 1 shows that the higher the noise in the environment and/or the smaller the sample size (in any stochastic environment), the higher the variance of a categorization. The reason is that each category belief is a normally distributed random variable with a variance equal to the variance in the environment divided by the number of objects in the respective category. A key insight is that in stochastic environments coarser categorizations have lower variance than finer categorizations. The intuition is that coarser categorizations partition a given set of objects into fewer categories and thus there are on average more objects per category in a coarser than in a finer categorization.[18]

We now look at Part 2. The bias of each category average with respect to a given object type is the expected mean squared error between the expected value of the category average (the estimator $E[\hat{Y}^{C_k}]$) and the mean of this object type (μ_j for object type x_j). Note that the category average of a category in which there is only one object type will be an unbiased estimator of the mean of this object type, as the expected value of the category average will be equal to the mean for this object type. However, the category average of a category that contains object types with different means will be generally a biased estimator for a particular object type as its expected value will be equal to the average of the means and thus the estimator will be making a mistake towards the means of at least some object types.[19] In Appendix A we show that any category formed through merging of two categories will therefore have a weakly greater bias than the sum of the biases of the categories that were merged to form it. As a result, any categorization P that refines P' will have a weakly smaller bias.[20]

Lemma 2 illustrates the trade-off of using fine versus coarse categorizations. The benefit of coarser categorizations, on the one hand, is that they have lower

[18]Note that for some given categorization, each category, regardless of how many different object types it combines, makes exactly the same contribution to the overall variance component of the EPE of that categorization. To illustrate with an example from Figure 1, the category $\{11, 10\}$ and the category $\{01\}$ have the same contribution to the variance component of the EPE of the categorization $\{\{11, 10\}\ \{01\}\ \{00\}\}$. On the one hand, a category containing more object types has a smaller variance due to the larger number of objects in it. On the other hand, the next object drawn is more likely to be from this category than from a category containing only one object type. Under our symmetry assumptions, the two effects cancel out exactly. As a result, the variance of all categorizations of equal coarseness is the same.

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

[20]The biases of categorizations that are not related by the "refines" relation are not directly comparable without making additional assumptions.

variance in noisy environments. The benefit of refinements of categorizations, on the other hand, is that they decrease the bias component of the EPE^{IP}.

We next characterize what an optimal categorization for individual prediction means. A categorization P^* is optimal for individual prediction if and only if its EPE^{IP} is smaller than or equal to the EPE^{IP} of any categorization P from the set of available categorizations $\mathcal{P}$. Using the bias-variance decomposition from Lemma 2, this is equivalent to:

$$\begin{aligned} & EPE^{IP}[P^*] \leq EPE^{IP}[P] && \forall P \\ \Leftrightarrow\ & Var[P^*] - Var[P] \leq Bias^2[P] - Bias^2[P^*] && \forall P \end{aligned} \tag{3}$$

The latter representation shows that for a categorization P^* to be optimal for IP, the difference between its variance and the variance of any other categorization available to the agent has to be smaller or equal to the difference in squared biases between the other categorization and P^*. A key question we are interested in is how the optimal way of categorizing (in terms of coarseness) depends on the environment the agent is in - whether it is deterministic or stochastic. Proposition 1 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.

Proposition 1. *Coarseness of the Optimal Categorization(s) for IP*

The coarseness of the optimal categorization(s) for IP is weakly increasing in σ^2 and weakly decreasing in n.

As a starting point consider a deterministic environment. This is an environment without noise and the variance component of any categorization is zero. Hence the finest possible categorization is optimal as it has zero bias. As the noise increases, however, the variance component becomes more important. The difference in variance between a finer and a coarser categorization increases and coarser categorizations become more attractive compared to finer ones than before as the decrease in variance that they offer outweighs the increase in bias.[21] The same is true for a decrease in sample size (for any positive noise level). As a result, if the noise level to sample size ratio is sufficiently small, an individual can make better predictions using coarser rather than finer categorizations.

[21] This follows from Lemma 2 (see proof in Appendix A).

4.2 Coordination

We now consider the case when players care only about coordinating their predictions. The setup is as follows. There are two players. Each of them will independently accumulate n experiences of each object type.[22] The two individuals, then, will face a one-off prediction problem. They will both observe the same object and they independently will have to make a prediction about the object's unobserved value. We assume that this next object will be drawn from the discrete uniform distribution of all possible object types. As before, agents use categorizations to make predictions. As there are many alternative ways for each of the two players to categorize her experiences, which categorization each of them chooses matters because it will determine their expected prediction error from each other. Here 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 of it.[23]

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

[22]Note that the assumption that the players sample the experiences they base their prediction on independently does not exclude the possibility that they observe the same object types. The essential differences in the samples between the players concern only the idiosyncratic y-values experienced. As we can think of these differences as being due to differences in the subjective perceptions of the objects, our analysis allows for the players to observe exactly the same observable characteristics, as long as the subjective noise term is independently drawn for every instance that an object is perceived by a player. For example, two people observing the performance of the same worker may have different subjective perceptions of his performance. We can think of these independent samples as a benchmark case. An alternative benchmark would be if the sampled and perceived histories of the two individuals are identical. That is, they experience not only the same object types, but perceive in every instance identical y-values as well. In Appendix A we derive a general expression for bias-variance decomposition of the EPE^C, covering these two benchmark cases as well as intermediate cases in which the players' perceived values are (imperfectly) correlated. In our analysis below we focus on the benchmark case of independent samples, leaving intermediate cases in which the players' perceived values are (imperfectly) correlated for further research.

[23]One could of course also use the same framework to analyze the case in which players care not about minimizing but about maximizing their expected prediction error from coordination.

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

$$EPE^C[P_1, P_2] = \sum_{C_k \in P_1} \sum_{C_l \in P_2} \sum_{x_j \in (C_k, C_l)} pE[(\hat{Y}^{C_k} - \hat{Y}^{C_l})^2] \tag{4}$$

We want to understand the determinants of optimal categorization coarseness when players want to coordinate their predictions. We model this situation as a two-player one-shot game in which a player's strategy is a categorization from the set of available categorizations $\mathcal{P}$. Players choose independently from each other. All 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 $h : \mathcal{P} \times \mathcal{P} \to \mathbb{R}_+$ mapping from the set of possible categorization profiles to the set of non-negative real numbers.

Lemma 3. *Bias-Variance Decomposition of EPE^C*

$$\begin{aligned} EPE^C[P_1, P_2] &= \sum_{C_k \in P_1} \sum_{C_l \in P_2} \sum_{x_j \in (C_k, C_l)} pE[(\hat{Y}^{C_k} - \hat{Y}^{C_l})^2] \\ &= \sum_{C_k \in P_1} \sum_{x_j \in C_k} Var[\hat{Y}^{C_k}] + \sum_{C_l \in P_2} \sum_{x_j \in C_l} pVar[\hat{Y}^{C_l}] \\ &\quad + \sum_{C_k \in P_1} \sum_{C_l \in P_2} \sum_{x_j \in (C_k, C_l)} p(E[\hat{Y}^{C_k}] - E[\hat{Y}^{C_l}])^2 \\ &= Var[P_1] + Var[P_2] + Bias^2[P_1, P_2] \end{aligned} \tag{5}$$

Lemma 3 decomposes the expected prediction error in the coordination case into a bias and a variance component. The $EPE^C[P_1, P_2]$ is equal to the sum of the expected variance of the categorization P_1 that Player 1 uses, the expected variance of the categorization P_2 that Player 2 uses, and the expected bias of the two players' categorizations from each other. It is increasing in each of these terms. The variance components of the $EPE^C[P_1, P_2]$ are analogical to the variance component of a player's categorization in individual prediction. The bias component is, however, different from the one in the individual prediction case. Here we define the bias of the two players' predictions on a given object type as the mean squared error between the expected value of the category average that one player is using and the expected value of the category average that the other player is using.[24]

We now consider the pure strategy Nash equilibria of the coordination game, where the players' strategies are the categorizations they use to make predictions. An equilibrium is a categorization profile, one categorization for each player, such that the expected prediction error for each player is minimized given the categorization of the other player, which means that no player has an incentive to deviate to a different categorization.[25]

Let $[P_1^*, P_2^*]$ denote a categorization profile such that P_1^* is the categorization that Player 1 uses and P_2^* is the categorization that Player 2 uses. The two categorizations could be the same or different, and in case different they could be of the same or different level of coarseness. Using the bias-variance decomposition from Lemma 3, and denoting any categorization from the set of available categorizations with $P \in \mathcal{P}$, we can write the Nash equilibrium condition from the perspective of Player 1 as:[26]

$$
\begin{aligned}
EPE^C[P_1^*, P_2^*] &\leq EPE^C[P_1, P_2^*] && \forall P_1 \\
\Leftrightarrow Var[P_1^*] - Var[P_1] &\leq Bias^2[P_1, P_2^*] - Bias^2[P_1^*, P_2^*] && \forall P_1
\end{aligned}
\tag{6}
$$

A categorization profile is a NE in the coordination game if and only if for each player the variance of the categorization she is using 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.

[24] Thus, we extend the definition of bias of an estimator with respect to a 'true' population mean to an analogical concept in a strategic interaction setting, defining the bias with respect to another person's prediction rather than with respect to a 'true' population mean.

[25] There are at least two interpretations of how such an equilibrium can arise in this game. The first possibility is that if players happened to choose a categorization profile that constitutes an equilibrium by chance or if someone would advise them of one, no player would have an incentive to deviate from their equilibrium strategy (Holt and Roth, 2004). A second interpretation is that even in the case players initially lack information on how to play equilibrium, they would learn to play equilibrium over time. Although such learning over time might suggest that players are collecting a higher number of observations and hence finer categorizations are becoming optimal, this need not be the case. In particular, if players face a non-stationary noisy environment (i.e. changing $f(x)$) or if players have memory bounds and forget some of their past experiences in a noisy environment, they would not necessarily be able to accumulate a large enough number of observations to push them towards the finest possible categorization profile. While it is beyond the scope of this paper to model non-stationary environments or forgetting, we think of the limited number of observations n introduced in this model as a proxy for such situations.

[26] Analogous conditions can be written down for Player 2.

Our main results in this section are Propositions 2 and 3, in which we analyze the equilibrium properties (existence and efficiency) of the coordination game, focusing in particular on some comparative statics, i.e. the effect of changes in the environmental parameters on the equilibrium outcomes. Before doing so, in Lemma 4 we give some results on the types of outcomes that can be ruled out in this game.

Lemma 4. *Ruling Out Some Outcomes in the Coordination Game*

Part 1. There are no NE involving two categorizations at different levels of coarseness for any $\sigma^2 > 0$. A profile of two different categorizations at the same level of coarseness cannot be a NE if the two categorizations make different expected predictions on at least one object type.

Part 2. If P^* is the finest optimal categorization in IP, then there exists no symmetric equilibrium $[P, P]$ with any $P \succ P^*$.

Lemma 4 Part 1 rules out the existence of asymmetric equilibria with players using categorizations at different levels of coarseness. Such equilibria do not exist because the player who uses the finer categorization always has an incentive to deviate to the same categorization as the opponent. By doing so she decreases both components of the EPE^C. The bias component is reduced to zero if they both use the same categorization. The variance component decreases if a player switches to a coarser categorization. That is, using a categorization that is finer than the one the other player is using is suboptimal in any stochastic environment. Part 1 also gives a sufficient condition to rule out the existence of any asymmetric equilibria involving two categorizations at the same level of coarseness. Next, in Part 2 we show that there does not exist any symmetric equilibrium in which the two players use any categorization that refines the finest categorization that is optimal for individual prediction, as a player always has an incentive to deviate to the finest optimal categorization for individual prediction. The latter statement is a first argument supporting the point made in this paper that players might have incentives to categorize (in this case: at least weakly) more coarsely if they want to coordinate predictions than if they care about individual prediction.

Proposition 2. *Existence of Equilibria in the Coordination Game*

The number of symmetric NE is weakly decreasing in σ^2 and weakly increasing in n. Both players using the coarsest possible categorization is always a NE and may be the unique one if $\frac{\sigma^2}{n}$ is sufficiently high.

As a starting point consider the situation without any noise in the environment. Any symmetric categorization profile is an equilibrium, as there is no variance and the bias of the two players' categorizations from each other is zero when they use the same categorization and hence no player has an incentive to deviate. Proposition 2 shows that as noise levels increase sufficiently a player will have an incentive to deviate to a coarser categorization and finer symmetric categorization profiles will stop being equilibria. While in a deterministic environment using the same categorization as the other player is always a best response, this is not necessarily true in a stochastic environment. The intuition is that as noise increases variance becomes more important and by deviating to a coarser categorization a player can decrease the variance in her prediction. The game is symmetric and therefore this holds for either player. The effect for a decrease in sample size in case of a stochastic environment is analogical to that of an increase in noise. Lemma 4 shows that only profiles such that both players use the same categorization can be mutual best responses in stochastic environments when players' objectives are simply to coordinate their predictions. While in a deterministic environment, all profiles such that the two players use the same categorization are mutual best responses, an insight from Proposition 2 is that not all profiles such that players use the same categorization are mutual best responses in stochastic environments. The reason is that in a stochastic environment a player may have an incentive to unilaterally deviate to a coarser categorization as this helps the players to coordinate better. The coarsest possible categorization is always an equilibrium, as there is no coarser categorization to deviate to. It will be the unique equilibrium if the noise level to sample size ratio is sufficiently high. This is an equilibrium existence argument in favor of coarse categorization if players want to coordinate predictions in stochastic environments.

Proposition 3. *Efficiency of Equilibria in the Coordination Game*

The efficiency of symmetric NE is strictly increasing in categorization coarseness.

We have shown in Lemma 4 that for any positive noise level there are no asymmetric equilibria. For symmetric categorization profiles, the bias component is zero and the variance component of any coarser categorization profile is smaller than that of any finer categorization profile for any $\sigma^2 > 0$. While in a deterministic environment all symmetric categorization profiles are equally efficient, in stochastic environments any coarser symmetric categorization profile is more efficient than any finer symmetric categorization profile. This also means that any coarser symmetric equilibrium is more efficient than any finer symmetric equilibrium. Proposition 3 is an equilibrium efficiency argument for coarse categorization when players want to coordinate in stochastic environments.

The result that the categorization profile in which both players use the coarsest possible categorization is always an equilibrium and that it is the Pareto-superior one in any stochastic environment is perhaps somewhat counterintuitive, as the prediction error with respect to the 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 $EPE^C[P_1, P_2]$, and by both using the coarsest categorization they minimize its variance component. The intuition is that when categorizing coarsely the two individuals are placing a larger number of experiences in each category and by doing so they are decreasing the dependence of each person's prediction on the randomness of her past experience.

In this model players use a categorization and a category average within that categorization to make a prediction. If we did not impose this structure, note that as players care only about coordinating predictions, they could in principle solve or circumvent the coordination problem by predicting some 'fixed' number, independently from the observed characteristics of the object type drawn. From a coordination point of view, however, the question is which 'fixed' number they should use. Salience might matter. But different numbers may be salient depending on the specific circumstances in which players need to coordinate. The analysis here abstracts from such circumstantial factors, and it is in that sense complementary to any such salience theory of coordination. One obvious candidate for a salient number that is 'fixed', i.e. independent both from the specific object type drawn and also from the circumstances, may be the number corresponding to the expected y-value, taking into account the range of possible object types and their corresponding means. However, the players do not know this value. Still, they could form a prediction. In fact, if players use categorization to form their predictions, and if they use the coarsest possible categorization, lumping all objects into one category, then the expected value of their prediction will be equal to this expected y-value. As the analysis points out, the coarsest categorization profile is always an equilibrium and it constitutes the Pareto-superior one in any stochastic environment. Thus, one can think of our analysis as a formal justification of this idea to solve the coordination problem by using the expected y-value as salient prediction.

4.3 Individual Prediction and Coordination

In the previous two sections we analyzed factors determining optimal categorization coarseness for an individual who is interested only in making a correct prediction about the unobserved value of an object, and for agents who are only interested in coordinating their predictions. In some 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 as an $IP\&C$ game, and we analyze equilibrium properties of this game. An $IP\&C$ game is a version of the coordination game such that the preference relations of player i are determined by $EPE_i^{IP\&C}$.

Definition 3. Expected Prediction Error Individual Prediction & Coordination

$$EPE_i^{IP\&C}[P_1, P_2] = wEPE_i^{IP}[P_i] + (1-w)EPE^C[P_1, P_2] \quad (7)$$

The $EPE_i^{IP\&C}[P_1, P_2]$ of player i is the convex combination of the expected prediction errors from Definition 1 and from Definition 2. Let w denote the weight that the individual places on making correct predictions about the object value, and $(1-w)$ denote the weight she places on coordinating her predictions with the other $(0 \leq w < 1)$. We assume that both players place the same weight w on individual prediction.[27] Note that if the two players use different categorizations, then in the $IP\&C$ case their expected prediction errors will normally be different from each other as they would be making different mistakes with respect to the object value, even though they would be making the same error with respect to each other.

In Lemma 5 we derive a bias-variance decomposition of the $EPE_i^{IP\&C}$. The $EPE_i^{IP\&C}$ 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 bias of their predictions from each other.

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

Lemma 5. *Bias-Variance Decomposition of $EPE^{IP\&C}$*

$$\begin{aligned} EPE_i^{IP\&C}[P_1, P_2] &= wEPE_i^{IP} + (1-w)EPE^C \\ &= w(Var[P_i] + Var[Y] + Bias^2[P_i]) \\ &\quad + (1-w)(Var[P_1] + Var[P_2] + Bias^2[P_1, P_2]) \end{aligned} \tag{8}$$

We now derive the NE conditions in the $IP\&C$ game. For $[P_1^*, P_2^*]$ to be a NE in the $IP\&C$ game, no player should have an incentive to deviate to any other (finer, equally coarse or coarser) categorization given the categorization of the opponent. Using the bias-variance decomposition from Lemma 5 to write this condition, and denoting with P_1 any categorization in the set of possible categorizations, we get the following from the perspective of Player 1.

$$\begin{aligned} & EPE_1^{IP\&C}[P_1^*, P_2^*] \leq EPE_1^{IP\&C}[P_1, P_2^*] \qquad \forall P_1 \\ & \Leftrightarrow Var[P_1^*] - Var[P_1] \\ & \quad \leq w\Big[Bias^2[P_1] - Bias^2[P_1^*]\Big] \\ & \quad + (1-w)\Big[Bias^2[P_1, P_2^*] - Bias^2[P_1^*, P_2^*]\Big] \qquad \forall P_1 \end{aligned} \tag{9}$$

For a categorization profile to be an equilibrium in the $IP\&C$ game, the change in variance from deviating to any other categorization has to be weakly smaller than 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.

Our main results in this section are Propositions 4 and 5, in which we characterize some equilibrium properties of the $IP\&C$ game. We now first give results on the type of outcomes that cannot be equilibria in this game.

Lemma 6. *Ruling Out Some Outcomes in the $IP\&C$ game*

Part 1. There are no NE involving two categorizations at different levels of coarseness for any $\sigma^2 > 0$ if w is sufficiently low. The sufficient threshold is increasing in $\frac{\sigma^2}{n}$. A profile of two different categorizations at the same level of coarseness cannot be a NE if the two categorizations make different expected predictions on at least one object type.

Part 2. If P^* is the finest optimal categorization in IP, there exists no symmetric equilibrium $[P, P]$ with any $P \succ P^*$ in $IP\&C$.

Lemma 6 Part 1 describes sufficient conditions under which asymmetric equilibria can be ruled out. Given two categorizations at different levels of coarseness, the player with the finer categorization has an incentive to deviate to the same categorization as the other player if the weight placed on coordination $(1-w)$ is high enough. As we show in Appendix A, what constitutes a sufficiently high weight on coordination depends on the parameters of the environment. The noisier the environment and the smaller the sample size (in any noisy environment), the less weight on coordination is sufficient to rule out asymmetric equilibria.

Part 2 states that any symmetric profile consisting of both players using a categorization that refines the finest individually optimal categorization is not an equilibrium in the $IP\&C$ game. The reason is that a player can unilaterally profitably deviate to the finest individually optimal categorization. This will decrease her EPE^{IP} as well as make it easier to coordinate with the other player.

Proposition 4. *Existence of Equilibria in the $IP\&C$ game*

The number of symmetric NE is weakly decreasing in σ^2 and weakly increasing in n. Both players using the coarsest possible categorization is always a NE, regardless of σ^2 and n as long as $w \leq 1/2$. In addition, for any $w > 1/2$, the coarsest possible categorization profile is a NE if $\frac{\sigma^2}{n}$ is sufficiently high.

Proposition 4 shows that also in the $IP\&C$ game finer symmetric categorization profiles will stop being equilibria as the noise level increases or the sample size decreases, as a player will eventually have an incentive to unilaterally deviate to a coarser categorization to decrease the variance in her prediction. To check that the coarsest possible categorization is an equilibrium, we only need to examine unilateral deviations to a finer categorization. We show that a sufficient though not necessary condition for any such deviation not to be profitable in any stochastic environment, regardless of the noise level and sample size, is that the players assign at least as much weight to being coordinated as to predicting the value of the object correctly. Even in case the players assign more weight to the latter, the coarsest categorization profile is an equilibrium if the noise level to sample size ratio is high enough. The intuition is that if the players assign more weight to IP finer categorizations might have an advantage as they could reduce the bias in individual prediction. However, the higher the noise level and the smaller the sample size, the lower the importance of such a decrease in bias compared to the decrease in variance that coarser categorizations offer.

Proposition 5. *Efficiency of Equilibria in the $IP\&C$ game*

Any $[\widetilde{P}, \widetilde{P}]$ *such that* $|\widetilde{P}| < |P|$ *is Pareto-superior to any* $[P, P]$ *if and only if one of the following holds:*

(i) If $Bias^2[\widetilde{P}] > Bias^2[P]$ *and* w *is sufficiently low. The sufficient threshold is strictly decreasing in* σ^2 *and strictly increasing in* n*.*

(ii) If $Bias^2[\widetilde{P}] \leq Bias^2[P]$ *regardless of* w*,* $\sigma^2 > 0$*, and* n*.*

Proposition 5 provides necessary and sufficient conditions under which any coarser symmetric categorization profile is Pareto-superior to any finer symmetric categorization profile in the $IP\&C$ game.

It first states that when the bias of any coarser categorization is greater than the bias of any finer categorization, we can Pareto-rank any coarser symmetric categorization profile as Pareto-superior to any finer symmetric categorization profile if the weight on coordination $(1 - w)$ is sufficiently high. The sufficient threshold for this Pareto-ranking depends on the parameters in the environment. The higher the noise in the environment and the smaller the sample size, the lower the weight on coordination that is sufficient. Any given weight on coordination (i.e. any $0 < 1-w \leq 1$) can induce a Pareto-ranking where any coarser symmetric categorization profile is more efficient than any finer symmetric categorization profile under a high enough noise level σ^2 to sample size n ratio.

Note that (i) is relevant when comparing any categorizations such that the finer ones refines the coarser one, as Lemma 2 states that for such categorizations the condition on the biases specified for case (i) will always be satisfied. If the finer categorization does not refine the coarser, this condition on the biases may or may not hold. If it does hold, case (i) will still be relevant. If this condition on the biases does not hold, the second part of Proposition 5 is the relevant one.

Case (ii) says that if the bias of any coarser categorization is weakly smaller than the bias of any finer categorization, then Pareto-ranking (with coarser symmetric categorization profiles being better) is guaranteed regardless of w, $\sigma^2 > 0$ and n. This is true because in this case finer categorizations do not have any advantage in terms of lower bias in individual prediction.

5 An Illustration

5.1 Individual Prediction

We now use one of our motivating examples from the introduction to give an illustration of the results in the simplest possible setting.[28] Consider a British counter-terrorism intelligence agent whose goal is to predict the level of threat a person poses. In her training the agent has studied a number of profiles. A profile contains information about some observable attributes, e.g. hair colour, skin colour, height, gender, whether the person carries a suitcase, etc. For each profile, the agent also knows the level of threat eventually experienced. When encountering the next person, the agent will have to predict the threat level that this person poses.

Note that this example is simply in order to illustrate the comparative statics derived in the previous section. Thus, the presentation here is not from the point of view of the agent who makes predictions, but from our point of view as outside observers and analysts who know the environmental parameters and analyze the situation. The agent does not have the knowledge to calculate the errors for the particular situation we are assuming, but she can by herself derive comparative statics on how the coarseness of the optimal categorization(s) is affected by changes in the environmental parameters, as derivation of comparative statics does not rely on knowledge of particular parameters.

While we stress that the results in our model hold for any finite number of observable characteristics, in this example, for the sake of presentational convenience, we assume that people have only one observable attribute - hair colour - blond or dark, which carry numerical values of 0 and 1 respectively. Assume that the value of the threat level is equal to the value of the observed attribute plus some normally distributed noise term (with mean zero and variance σ^2), with higher values indicating higher threat levels.[29]

The agent organizes her experiences (the profiles she has studied) in categories. In the case of one observable attribute, the agent has two possible ways of categorizing people. Either put all dark-haired people in one category and all blond people in another, i.e. use the finest possible categorization. Or put both dark-haired and blond people in the same category, i.e. use the coarsest possible categorization. Assume that the agent has sampled two people of each

[28]The results could be illustrated using alternative examples.

[29]Thus, dark-haired people are expected to pose more of a threat.

type during her training, that is two dark-haired and two blond people.

First, consider the benchmark case of a deterministic environment, i.e. $\sigma^2 = 0$. Here observed attributes can perfectly predict threat levels. The variance component of the expected prediction error is zero. Thus, the best the agent can do is to put dark-haired people in one category and blond people in another, as then the bias component of expected prediction error is also zero. If she puts both dark-haired and blond people in the same category, she will be making a biased prediction, as the expected value of her predictor (0.5) differs from the expected level of threat of dark-haired (1) as well as of blond people (0). Thus, her expected prediction error will be positive. See Table 1.

	Categorization	$\vert P\vert$	$Var[P]$	$Var[Y]$	$Bias^2[P]$	$EPE^{IP}[P]$
$\sigma^2 = 0$	Fine $\{\{0\}\{1\}\}$	2	0	0	0	0
	Coarse $\{\{0,1\}\}$	1	0	0	0.25	0.25
$\sigma^2 = 2$	Fine $\{\{0\}\{1\}\}$	2	1	2	0	3
	Coarse $\{\{0,1\}\}$	1	0.5	2	0.25	2.75

Table 1: Expected Prediction Errors in Example $l = 1$, $y = x + \epsilon$, $n = 2$

Next, we look at the more realistic case of a stochastic environment ($\sigma^2 = 2$), where not every individual who looks the same poses the same threat. In this case there is a trade-off of using fine versus coarse categorizations. On the one hand, fine categorizations help decrease the bias component of expected prediction error. Thus, as before, the bias component is 0 if the agent puts all dark-haired people in one category and all blond people in another (see Table 1). On the other hand, coarse categorizations, although biased, help decrease the variance component $Var[P]$ by having more experiences in a given category. Thus, if the noise level to sample size ratio is sufficiently high, a coarse categorization may have a lower expected prediction error (in this example 2.75) than a finer categorization (3.0) as is illustrated in Table 1. We have illustrated that as the environment becomes noisy coarse categorizations can become better for prediction than finer categorizations (Proposition 1).

5.2 Coordination

We now illustrate the results for the coordination case by extending the example above. Assume that there are two counter-terrorism intelligence agents - a British agent and a Russian agent. There is a terrorism threat, and their goal is to predict the level of threat posed by the next person they encounter. Depending on the predicted threat posed by this person, the British and Russian agencies will take some action to neutralize the person or not. Due to various strategic complementarities they want to take the same action.

The agents will encounter the same person and we assume this person is equally likely to be a dark-haired person or a blond person. The two agents will have no possibility to communicate with each other prior to making their prediction. Each agent has been trained by her own agency and thus each person's past experience is randomly and independently drawn from the database of observations. We assume that in her training each agent has seen two cases of dark-haired and two cases of blond people. This situation can be represented as a one-shot simultaneous move game where each agent chooses which categorization to use from the set of possible categorizations, and their goal is simply to minimize their expected prediction error, where the prediction error is the difference in prediction between the British and the Russian agent. The question is which categorization profiles are NE, and which are the Pareto superior ones.

Figure 2a below illustrates the case of a deterministic environment. In a deterministic environment ($\sigma^2 = 0$), all categorization profiles such that both players use the same categorization are Nash equilibria. They are also all equally efficient. Thus, in this case it does not matter whether each of the two agents puts dark-haired and blond people in separate categories or if each of the two agents uses only one category for dark-haired and blond people. As long as they are using the same model, their expected prediction error from each other will be zero. If one of them uses a coarse and the other a fine categorization, their expected prediction error from each other will be positive due to the bias component. Next, consider a stochastic environment. If the environment is stochastic, then not necessarily all profiles where both players use the same categorization constitute equilibria (Proposition 2). In the example in Figure 2b, both players using the finest possible categorization and placing dark-haired people in one category and blond people in another is no longer an equilibrium. The reason is that because of the noise, individuals who look the same do not

	Fine {{1}{0}}	Coarse {{1,0}}
Fine {{1}{0}}	0 , 0	0.25 , 0.25
Coarse {{1,0}}	0.25 , 0.25	0 , 0

a) $\sigma^2 = 0$

	Fine {{1}{0}}	Coarse {{1,0}}
Fine {{1}{0}}	2 , 2	1.75 , 1.75
Coarse {{1,0}}	1.75 , 1.75	1 , 1

b) $\sigma^2 = 2$

Figure 2: EPE^C in example $l = 1$, $y = x + \epsilon$, $n = 2$

necessarily pose the same threat, and at this noise level an agent has an incentive to unilaterally deviate to a coarser categorization as coarser categorizations decrease the variance. Furthermore, for any positive noise level coarser symmetric categorization profiles will be more efficient than finer symmetric categorization profiles (Proposition 3). In the example in Figure 2b, both players placing both dark-haired and blond people in one category Pareto-dominates both agents placing dark-haired people in one category and blond people in another (lower EPE^C).

Lemma 4 showed that there can be no equilibria in the coordination game that involving categorizations that refine the finest individually optimal categorization. We saw previously that if $\sigma^2 = 2$, the finest optimal categorization for individual prediction is the categorization $\{\{1, 0\}\}$. The example above shows that if $\sigma^2 = 2$, the only equilibrium is such that both players use this same categorization, and there are no equilibria involving finer categorizations. Finally, we note that the results extend to any number of coarseness levels.[30]

5.3 Individual Prediction & Coordination

We now consider the case when the two counter-terrorism intelligence agents care both about individual prediction and about coordination. All parameter values are the same as before. The weight on individual prediction is $w = \frac{1}{2}$. In a deterministic environment, see Figure 3a, both symmetric categorization profiles are Nash equilibria. The finer symmetric categorization profile is Pareto-superior

[30] Above we use the case of one attribute and two possible categorizations simply to make the illustration easier, as with two attributes there are already fifteen possible categorizations and a 15×15 matrix.

	Fine {{1}{0}}	Coarse {{1,0}}
Fine {{1}{0}}	0 , 0	0.125 , 0.25
Coarse {{1,0}}	0.25 , 0.125	0.125 , 0.125

a) $\sigma^2 = 0$

	Fine {{1}{0}}	Coarse {{1,0}}
Fine {{1}{0}}	2.5 , 2.5	2.375 , 2.25
Coarse {{1,0}}	2.25 , 2.375	1.875 , 1.875

b) $\sigma^2 = 2$

Figure 3: $EPE^{IP\&C}$ in example $l = 1$, $y = x + \epsilon$, $n = 2$, $w = \frac{1}{2}$

as when using the finest possible categorization, the two agents have zero bias in individual prediction, whereas if they use the coarsest categorization they have a positive bias.

Next, consider a stochastic environment. As we see in Figure 3b, both players using the finest possible categorization is no longer a Nash equilibrium, as a player has an incentive to unilaterally deviate to a coarser categorization to decrease the variance. Both players using the coarsest possible categorization is the unique equilibrium in this situation. Note also that as weight on coordination is sufficiently high there is no equilibrium such that players use categorizations at different levels of coarseness (see Lemma 6).

6 Concluding Remarks

Motivated by the observations that people often use categories to make predictions and that situations in which they have incentives to coordinate their predictions abound, this paper aims to improve our understanding of when and why decision makers may be better off categorizing coarsely rather than finely.

We develop a framework that looks at categorizations as containing sets of estimators on the basis of which agents make predictions. We apply this framework to derive comparative statics on factors affecting optimal coarseness in settings in which individuals care about predicting correctly and in settings in which they want to coordinate their predictions with each other, as well as in the convex combination of these two types of situations.

The paper gives an insight into how the optimal model for an individual is affected by the environment and by strategic considerations. The key contribution, relying on both equilibrium existence and efficiency arguments, is to show that incentives to coordinate can provide a rationale for coarse categorization in stochastic environments. The theory thus enhances our understanding of why people may categorize coarsely. As coarse categorization can be linked to various biases in decision making ranging from mispricing in markets to discrimination against minorities, our analysis suggests a possible role for the coordination motive in understanding such biases.

References

Nabil I. Al-Najjar and Mallesh M. Pai. Coarse Decision Making and Overfitting. *Journal of Economic Theory*, 150:467 – 486, 2014.

John R. Anderson. *The Adaptive Character of Thought*. Lawrence Erlbaum, 1990.

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

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.

Henri Cohen and Claire Lefebvre, editors. *Handbook of Categorization in Cognitive Science*. Elsevier, 2005.

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

Vessela Daskalova and Nicolaas J. Vriend. Categorization and Coordination. Working Paper No. 719, School of Economics and Finance, Queen Mary University of London, and Working Paper 2014/11, Cambridge-INET Institute, 2014.

Vessela Daskalova and Nicolaas J. Vriend. Learning Categorizations of Strategy Spaces. Unpublished Manuscript, 2015.

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.

James Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*, volume 6. Springer, 2013.

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 David Schmeidler. Case-based Decision Theory. *The 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. *International Economic Review*, forthcoming.

Christian Hellwig and Laura Veldkamp. Knowing What Others Know: Coordination Motives in Information Acquisition. *The Review of Economic Studies*, 76(1):223–251, 2009.

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

Charles A. Holt and Alvin E. Roth. The Nash Equilibrium: A Perspective. *Proceedings of the National Academy of Sciences*, 101(12):3999–4002, 2004.

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.

George Lakoff. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press, 1987.

Qi Li and Jeffrey Scott Racine. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, 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. Learning across Games. *Games and Economic Behavior*, 74 (2):601–619, 2012a.

Friederike Mengel. On the Evolution of Coarse Categories. *Journal of Theoretical Biology*, 307:117–124, 2012b.

Lukas Meub, Till Proeger, Kilian Bizer, and Markus Spiwoks. Strategic Coordination in Forecasting: An Experimental Study. *Finance Research Letters*, 13:155–162, 2015.

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*, 152:356–381, 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.

Gregory L Murphy and Brian H Ross. Predictions from Uncertain Categorizations. *Cognitive psychology*, 27(2):148–193, 1994.

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.

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 and Tony Belpaeme. Coordinating Perceptually Grounded Categories through Language: A Case Study for Colour. *Behavioral and Brain Sciences*, 28(4):469–488, 2005.

Jakub Steiner and Colin Stewart. Price Distortions under Coarse Reasoning with Frequent Trade. *Journal of Economic Theory*, 159:574–595, 2015.

Scott N. Swisher. Information Aggregation and Product Reviews in the Entertainment Industry. Mimeo, 2013.

Hal R. Varian. Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, 28(2):3–28, 2014.

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: Proofs

Lemma 2. Bias-Variance Decomposition of EPE^{IP}

Proof.

$$
\begin{aligned}
&EPE^{IP}[P] = \sum_{C_k \in P} \sum_{x_j \in C_k} pE[(\hat{Y}^{C_k} - Y_j)^2] \\
&= \sum_{C_k \in P} \sum_{x_j \in C_k} p\Big[E[(\hat{Y}^{C_k})^2] - 2E[\hat{Y}^{C_k} Y_j] + E[(Y_j)^2]\Big] \\
&= \sum_{C_k \in P} \sum_{x_j \in C_k} p\Big[E[(\hat{Y}^{C_k})^2] - 2E[\hat{Y}^{C_k}]E[Y_j] + E[(Y_j)^2]\Big] \\
&= \sum_{C_k \in P} \sum_{x_j \in C_k} p\Big[Var[\hat{Y}^{C_k}] + (E[\hat{Y}^{C_k}])^2 - 2E[\hat{Y}^{C_k}]E[Y_j] + Var[Y_j] + (E[Y_j])^2\Big] \\
&= \sum_{C_k \in P} \sum_{x_j \in C_k} p\Big[Var[\hat{Y}^{C_k}] + Var[Y_j] + (E[\hat{Y}^{C_k}] - E[Y_j])^2\Big] \\
&= \sum_{C_k \in P} \sum_{x_j \in C_k} pVar[\hat{Y}^{C_k}] + \sum_{C_k \in P} \sum_{x_j \in C_k} pVar[Y_j] + \sum_{C_k \in P} \sum_{x_j \in C_k} p(E[\hat{Y}^{C_k}] - E[Y_j])^2 \\
&= \sum_{C_k \in P} \sum_{x_j \in C_k} pVar[\hat{Y}^{C_k}] + \sum_{C_k \in P} \sum_{x_j \in C_k} pVar[Y_j] + \sum_{C_k \in P} \sum_{x_j \in C_k} p(E[\hat{Y}^{C_k}] - \mu_j)^2 \\
&= Var[P] + Var[Y] + Bias^2[P]
\end{aligned}
\tag{10}
$$

□

In the proof above, Line 2 is equivalent to Line 3 since for two independent random variables X_1 and X_2, $E[X_1 X_2] = E[X_1]E[X_2]$. To see that $\hat{Y}^{C_k}$ and Y_j are independent, consider the following. $\hat{Y}^{C_k}$ is the category average 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 are sampled in this category. Y_j is a random variable because it depends on the y-value that will be drawn for the next object of type x_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 3 is equivalent to Line 4 since for any random variable X, $E[X^2] = Var[X] + (E[X])^2$. We use this expression for $E[(\hat{Y}^{C_k})^2]$ and for $E[(Y_j)^2]$.

Lemma 2. Comparative Statics of Bias and Variance Components of EPE^{IP}

Part 1.

Proof. We now look at the two terms of the variance component from Lemma 2 separately. We know that as a category average is a random variable equal to the average of the realizations of random variables, independently drawn from the normal distribution, the variance of any category average 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 first term becomes:

$$p \sum_{C_k \in P} \sum_{x_j \in C_k} Var[\hat{Y}^{C_k}] = p \sum_{C_k \in P} \sum_{x_j \in C_k} \frac{\sigma^2}{t_k n} = p \sum_{C_k \in P} \frac{t_k \sigma^2}{t_k n} = \frac{p\sigma^2 |P|}{n}$$

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 an equal number of observations from each type in the agent's experience, the variance of category average for each category in a given categorization is the same.

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

$$p \sum_{C_k \in P} \sum_{x_j \in C_k} Var[Y_j] = p \sum_{C_k \in P} \sum_{x_j \in C_k} \sigma^2 = \frac{1}{|X|} |X| \sigma^2 = \sigma^2$$

Connecting our results on the two terms, the variance component of the $EPE^{IP}[P]$ is equal to:

$$Var[P] + Var[Y] = \frac{p\sigma^2 |P|}{n} + \sigma^2$$

This expression shows that the variance component of the $EPE^{IP}[P]$ of a categorization is strictly increasing in the noise level σ^2, and strictly decreasing in n. It is also strictly increasing in the number of categories $|P|$ (for any $\sigma^2 > 0$).

We now compare the variance of a finer categorization P with the variance of a coarser categorization P'.

$$Var[P] - Var[P'] = \frac{p\sigma^2 |P|}{n} + \sigma^2 - \left(\frac{p\sigma^2 |P'|}{n} + \sigma^2 \right) = \frac{p\sigma^2 (|P| - |P'|)}{n}$$

That is, any finer categorization has a greater variance than any 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 ($|P| - |P'|$). The difference in variance between a finer and a coarser categorization is strictly increasing in their difference in coarseness ($|P| - |P'|$) and in σ^2, and is strictly decreasing in n. All categorizations at the same level of coarseness have the same variance. □

Part 2.

Proof. To prove Part 2, we show that the $Bias^2$ of the category average of any *category* containing $t + v$ object types is always greater than or equal to the sum of the biases of the category averages of any two *categories* with t and with v object types, respectively, through the merging of which it can be formed. This implies that the $Bias^2[P']$ is weakly greater than $Bias^2[P]$, where P is any categorization that refines P'.

The $Bias^2$ of the category average 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 average and μ_i is the mean of object type i in the population. Note that $E[\hat{Y}_t] = \frac{\sum_{i=1}^{t}\mu_i}{t}$ since we assumed that the agent has sampled an equal number of objects of each object type. Thus, after some algebraic transformations the $Bias^2$ can be rewritten as:

$$Bias^2[\hat{Y}_t] = p\Big[\sum_{i=1}^{t} \mu_i^2 - \frac{(\sum_{i=1}^{t}\mu_i)^2}{t}\Big] \tag{11}$$

Analogically, for a category that has v or $t+v$ object types the corresponding expression for the squared bias of the category average can be written by simply replacing t in equation (11) by v or $t+v$, respectively. We now show that:

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

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

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

As $t > 0$ and $v > 0$ and the remaining term is squared, the above inequality always holds. $\square$

Proposition 1. Coarseness of the Optimal Categorization(s) for Individual Prediction

Proof. For a categorization P^* to be optimal for individual prediction, its EPE^{IP} has to be smaller than or equal to the EPE^{IP} of any other categorization from the set of possible categorizations, i.e. any $P \in \mathcal{P}$ (see inequality (3)). We now 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 averages. The squared bias of a category average depends on the differences between the expected value of category average and the true population mean for each object type in this category. Since neither of the two is affected by changes in σ^2 and n, the bias term of a categorization will not be affected by such changes. Therefore, in inequality (3) only the LHS will change.

We show how an increase in σ^2 will affect the coarseness of the optimal categorization(s).

i) First, consider a potential deviation to a finer categorization. In Lemma 2 Part 1 we showed that the difference in the variance between a finer and a coarser categorization is positive and increasing in σ^2. If before inequality (3) was fulfilled, as σ^2 increases the LHS becomes even more negative (and the RHS does not change). Thus, if before using a categorization P^* was equally good or better than using a finer categorization, this will be even more so as σ^2 increases. Therefore, no categorization finer than P^* can become optimal after an increase in σ^2 if it was not optimal before the increase.

ii) Next, consider a potential deviation to a categorization of equal coarseness. The change in σ^2 has no effect on the fulfillment of inequality (3), as the variances of all categorizations at the same level of coarseness are equal. No categorization that is equally coarse as P^* can become optimal after an increase in σ^2 if it was not optimal before the increase.

iii) Finally, consider a potential deviation to a coarser categorization. If before inequality (3) held, as σ^2 increases the LHS increases while the RHS stays the same. 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. Note that from Lemma 2 Part 1 we know that the difference in variance between a finer and coarser categorization is decreasing in the sample size n. We can use the same arguments as above to show 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 3. Bias-Variance Decomposition of EPE^C

Proof.

$$
\begin{aligned}
EPE^C[P_1, P_2] &= \sum_{C_k \in P_1} \sum_{C_l \in P_2} \sum_{x_j \in (C_k, C_l)} pE[(\hat{Y}^{C_k} - \hat{Y}^{C_l})^2] \\
&= \sum_{C_k \in P_1} \sum_{C_l \in P_2} \sum_{x_j \in (C_k, C_l)} p\Big[E[(\hat{Y}^{C_k})^2] - 2E[\hat{Y}^{C_k}\hat{Y}^{C_l}] + E[(\hat{Y}^{C_l})^2]\Big] \\
&= \sum_{C_k \in P_1} \sum_{C_l \in P_2} \sum_{x_j \in (C_k, C_l)} p\Big[Var[\hat{Y}^{C_k}] + (E[\hat{Y}^{C_k}])^2 - 2Cov[\hat{Y}^{C_k}, \hat{Y}^{C_l}] \\
&\quad - 2E[\hat{Y}^{C_k}]E[\hat{Y}^{C_l}] + Var[\hat{Y}^{C_l}] + (E[\hat{Y}^{C_l}])^2\Big] \\
&= \sum_{C_k \in P_1} \sum_{C_l \in P_2} \sum_{x_j \in (C_k, C_l)} p\Big[Var[\hat{Y}^{C_k}] + Var[\hat{Y}^{C_l}] - 2Cov[\hat{Y}^{C_k}, \hat{Y}^{C_l}] + (E[\hat{Y}^{C_k}] - E[\hat{Y}^{C_l}])^2\Big] \\
&= \sum_{C_k \in P_1} \sum_{x_j \in C_k} pVar[\hat{Y}^{C_k}] + \sum_{C_l \in P_2} \sum_{x_j \in C_l} pVar[\hat{Y}^{C_l}] - 2 \sum_{C_k \in P_1} \sum_{C_l \in P_2} \sum_{x_j \in (C_k^T, C_l^T)} pCov[\hat{Y}^{C_k}, \hat{Y}^{C_l}] \\
&\quad + \sum_{C_k \in P_1} \sum_{C_l \in P_2} \sum_{x_j \in (C_k, C_l)} p(E[\hat{Y}^{C_k}] - E[\hat{Y}^{C_l}])^2 \\
&= Var[P_1] + Var[P_2] - 2Cov[P_1, P_2] + Bias^2[P_1, P_2]
\end{aligned}
\tag{12}
$$

□

Line 2 is equivalent to Lines 3-4 since we know that for any two random variables $E[XY] = Cov[X, Y] + E[X]E[Y]$ and for any random variable $E[X^2] = Var[X] + (E[X])^2$. We apply the first expression to $E[\hat{Y}^{C_k}\hat{Y}^{C_l}]$ and the second expression to $E[(\hat{Y}^{C_k})^2]$ and to $E[(\hat{Y}^{C_l})^2]$.

If the two players have sampled their experiences independently from each other, then $Cov[P_1, P_2] = 0$ and

$$
EPE^C[P_1, P_2] = Var[P_1] + Var[P_2] + Bias^2[P_1, P_2] \tag{13}
$$

By the definition of correlation we have that:

$$
Corr[\hat{Y}^{C_k}, \hat{Y}^{C_l}] = \frac{Cov[\hat{Y}^{C_k}, \hat{Y}^{C_l}]}{\sqrt{Var[\hat{Y}^{C_k}]Var[\hat{Y}^{C_l}]}}
$$

Note that if the two players sample identical experiences, $Corr[\hat{Y}^{C_k}, \hat{Y}^{C_l}] = 1$. Then $Cov[\hat{Y}^{C_k}, \hat{Y}^{C_l}] = \sqrt{Var[\hat{Y}^{C_k}]Var[\hat{Y}^{C_l}]}$. If the two players are using exactly the same categorization, then for each object type and for each category $\hat{Y}^{C_k} = \hat{Y}^{C_l}$ and the above is equivalent to: $Cov[\hat{Y}^{C_k}, \hat{Y}^{C_k}] = \sqrt{Var[\hat{Y}^{C_k}]Var[\hat{Y}^{C_k}]}$.

This means that for any symmetric categorization profile

$$\begin{aligned}&EPE^C[P,P]\\&= Var[P] + Var[P] - 2Cov[P,P] + Bias^2[P,P] = 0\end{aligned}$$

In this benchmark case of exactly identical samples, regardless of the noise level, all symmetric categorization profiles are NE in the coordination game and are equally efficient. In the proofs below we focus on the case of independently sampled experiences (based on equation (13) above).

Lemma 4. Ruling Out Some Outcomes in the Coordination Game

Part 1.

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 such that Player 2 uses a coarser categorization than Player 1. Player 1 always has an incentive to deviate to the exact same categorization that Player 2 uses as it would strictly reduce her EPE^C. This is beneficial both from a variance and a bias perspective. Moving to a coarser categorization reduces the variance (see Lemma 2 Part 1). Moving to the same categorization as the other player eliminates the bias.

Analogically, there exists no asymmetric equilibrium $[P_1, P_2]$ such that the two players use different categorizations at the same level of coarseness as long as $Bias^2[P_1, P_2] > 0$, i.e. as long as the categorizations that the two players use make different expected predictions on at least one object type. It would be then profitable for either player to deviate to using the exact same categorization at the same level of coarseness as the other player rather than using different categorizations at that level, as this would decrease bias to zero while variance stays constant. □

Part 2.

Proof. To prove that any $[P, P]$ with $P \succ P^*$ cannot be a NE, we show that Player 1 has an incentive to deviate to P^*, i.e. to the finest optimal categorization in IP. That is, we show:

$$EPE[P_1, P_2] > EPE[P_1^*, P_2]$$

Applying the result derived in inequality (6), this is equivalent to:

$$Var[P_1] - Var[P_1^*] > Bias^2[P_1^*, P_2] \tag{14}$$

Assume that one player uses a categorization P', which has, besides some other categories, two categories of which one with t different object types and the other one with v different object types, with category averages $\hat{Y}_t = \frac{\sum_{i=1}^{t} \mu_i}{t}$ and $\hat{Y}_v = \frac{\sum_{i=1}^{v} \mu_i}{v}$ respectively. The other player uses a coarser categorization P'', which has a category that merges the category with t and the category with v object types, and her category average is $\hat{Y}_{t+v} = \frac{\sum_{i=1}^{t+v} \mu_i}{t+v}$. Note that under the assumption of an equal number of objects of each type $\hat{Y}_{t+v} = \frac{t\hat{Y}_t + v\hat{Y}_v}{t+v}$. As all their other categories are equal, the $Bias^2$ of the two players' predictions from each other is:

$$\begin{aligned}
Bias^2[P', P''] &= \sum_{i=1}^{t} (\hat{Y}_t - \hat{Y}_{t+v})^2 + \sum_{i=1}^{v} (\hat{Y}_v - \hat{Y}_{t+v})^2 \\
&= t \left(\hat{Y}_t - \frac{t\hat{Y}_t + v\hat{Y}_v}{t+v} \right)^2 + v \left(\hat{Y}_v - \frac{t\hat{Y}_t + v\hat{Y}_v}{t+v} \right)^2 \\
&= \frac{tv(\hat{Y}_t - \hat{Y}_v)^2}{t+v}
\end{aligned}$$

We know from Lemma 2 Part 2 that:

$$Bias^2[P'] - Bias^2[P''] = \frac{tv(\hat{Y}_t - \hat{Y}_v)^2}{t+v}$$

Thus, we have shown that $Bias^2[P', P''] = Bias^2[P'] - Bias^2[P'']$. This holds for any two categorizations with consecutive coarseness levels where the second refines the first. Analogically, $Bias^2[P, P^*] = Bias^2[P^*] - Bias^2[P]$ holds for all $P \succ P^*$.

Thus, inequality (14) is equivalent to:

$Var[P] - Var[P^*] > Bias^2[P^*] - Bias^2[P]$

This can be re-written as $EPE[P] > EPE[P^*]$. And since we assumed P^* is the finest optimal categorization for IP this is by definition true, because all categorizations finer than P have a greater EPE^{IP}. □

Proposition 2. Existence of Equilibria in the Coordination Game

Proof. To show that a symmetric categorization profile $[P^*, P^*]$ is a NE we need to show that the condition specified in inequality (6) is fulfilled. We omit the indices for the players, as the categorization profile considered is symmetric. We have shown in Lemma 4 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 a potential deviation to a coarser categorization $|P| < |P^*|$. Applying the result derived in inequality (6), this can be rewritten as:

$$Var[P^*] - Var[P] \leq Bias^2[P, P^*]$$

Observe that the squared bias component of the two players' EPE^C from each other will always be finite and that it is not affected by changes in σ^2 or n. From Lemma 2 we know that the difference in variance between a finer and a coarser categorization is positive, strictly increasing in σ^2, and strictly decreasing in n (for any $\sigma^2 > 0$). Thus, for a given $Bias^2[P, P^*]$, as σ^2 increases on n decreases, the LHS increases and eventually for any pair of categorizations $[P^*, P^*]$ there will be a point at which the LHS > RHS for some $|P| < P^*$, i.e. a player will have an incentive to deviate to some coarser categorization and $[P^*, P^*]$ will no longer be a NE. Thus, the number of symmetric NE decreases as σ^2 increases or n decreases. 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 large and n sufficiently small it will be the only one. □

Proposition 3. Efficiency of Equilibria in the Coordination Game

Proof. We 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 $[P, P]$. For all symmetric profiles, since players are using the exact same categorizations their bias from each other is always $Bias^2[P, P] = 0$. Thus, the EPE^C of these categorization profiles depends only on the variance. It is equal to $EPE^C[P,] = Var[P] + Var[P]$. We know from Lemma 2 that the coarser the categorization, the smaller its variance. This means that in any stochastic environment ($\sigma^2 > 0$) 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 $\sigma^2 = 0$, $Var[P] = 0$ for all categorizations, 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 given noise level in the coordination game. □

Lemma 5. Bias-Variance Decomposition of $EPE^{IP\&C}$

Proof. This follows directly from combining Lemma 2 and Lemma 3. □

Lemma 6. Ruling Out Some Outcomes in the $IP\&C$ Game

Part 1.

Proof. Consider any categorization profile with two categorizations at different levels of coarseness $[P_1, \widetilde{P}_2]$ where $|\widetilde{P}| < |P|$. We derive a sufficient condition to rule out such categorization profiles from being Nash equilibria in the $IP\&C$

game. Under this condition Player 1 has an incentive to deviate to the same coarser categorization that Player 2 uses. That is, from $EPE^{IP\&C}[P_1, \widetilde{P}_2] > EPE^{IP\&C}[\widetilde{P}_1, \widetilde{P}_2]$.

Using inequality (9), for any $|\widetilde{P}_1| < |P_1|$, player 1 will have an incentive to deviate if:

$$
\begin{aligned}
&Var[P_1] - Var[\widetilde{P}_1] > \\
&\quad w\Big[Bias^2[\widetilde{P}_1] - Bias^2[P_1]\Big] - (1-w)Bias^2[P_1, \widetilde{P}_2]
\end{aligned}
$$

$$
\begin{aligned}
&w\Big[Bias^2[\widetilde{P}_1] - Bias^2[P_1] + Bias^2[P_1, \widetilde{P}_2]\Big] \\
&\quad < Var[P_1] - Var[\widetilde{P}_1] + Bias^2[P_1, \widetilde{P}_2]
\end{aligned}
$$

If the LHS is negative or equal to zero, then the last inequality will always be satisfied (for any $0 \leq w \leq 1$), as the RHS is always positive. Hence the remaining condition to be checked is the case that the LHS is positive. In this case we can rewrite the inequality as:

$$
w < \frac{Var[P_1] - Var[\widetilde{P}_1] + Bias^2[P_1, \widetilde{P}_2]}{Bias^2[\widetilde{P}_1] - Bias^2[P_1] + Bias^2[P_1, \widetilde{P}_2]}
$$

Note that the only terms affected by σ^2 and n are the variances in the numerator of the RHS of this inequality, with the difference between these two increasing if σ^2 increases and/or n decreases. This means that the maximum weight on individual prediction that can be allowed for this condition to hold increases as σ^2 goes up and/or n goes down. For any given w (with $0 \leq w \leq 1$) there will be some $\frac{\sigma^2}{n}$ that is sufficiently high to satisfy the threshold condition.

Finally, to rule out asymmetric NE, we need to consider two players using a different categorization at the same coarseness level. If they make a different prediction in expectation for at least one type of object, then one of the two players has an incentive to deviate to the categorization used by the other player. This is true for any $0 \leq w < 1$ as the variance is not affected, but one of the two categorizations will have a smaller bias component of the EPE^{IP} and it is always good to move to using the same categorization from a coordination point of view. □

Part 2.

Proof. We need to show that if P^* is the finest optimal for IP, then $[P, P]$ is not a NE in the $IP\&C$ game for any $P \succ P^*$. We show this by showing that Player 1 has a profitable deviation to P^*, i.e. $EPE_1^{IP\&C}[P_1, P_2] > EPE_1^{IP\&C}[P_1^*, P_2]$.

Applying again the result derived in inequality (9), we get:

$$\begin{aligned}\Leftrightarrow\ & Var[P_1] - Var[P_1^*] \\ & > w\Big[Bias^2[P_1^*] - Bias^2[P_1]\Big] + (1-w)Bias^2[P_1^*, P_1]\end{aligned}$$

We know from Lemma 4 Part 2 that $Bias^2[P_1^*, P_2] = Bias^2[P_1^*] - Bias^2[P_1]$ for any $P \succ P^*$. Thus, the inequality becomes:

$$Var[P_1] - Var[P_1^*] > Bias^2[P_1^*] - Bias^2[P_1]$$

We know that this holds as P^* is the finest optimal categorization for IP. □

Proposition 4. Existence of Equilibria in the $IP\&C$ game

Proof. Assume a symmetric categorization profile is a NE in the $IP\&C$ game for a given σ^2 and n. Consider an increase in σ^2. Can this lead to a player having an incentive to unilaterally deviate to a finer categorization? The RHS (last two lines) of inequality (9) is not affected by a change in σ^2. The LHS is the difference in variances between a coarser and a finer categorization. This difference is smaller than or equal to zero in any stochastic environment, and increasingly negative with any increases in σ^2. Thus, if the inequality (9) held before the change in σ^2, it will continue to hold after an increase in σ^2. Can a player have an incentive to unilaterally deviate to another equally coarse categorization after an increase in σ^2? Lemma 2 showed that categorizations of equal coarseness have the same variances regardless of the noise level. Thus, both the LHS and the RHS are not affected by the change in noise level and a player who had no incentive to unilaterally deviate to another categorization at the same level of coarseness before the increase, will still not have an incentive to do so after the increase. Can a player have an incentive to unilaterally deviate to a coarser categorization? In this case the LHS is the difference in variances between a finer and a coarser categorization. It is positive and increasing in σ^2. The RHS is not affected. Thus, if σ^2 is sufficiently high, the LHS will become greater than the RHS and it will be profitable for a player to unilaterally deviate to a coarser categorization. This means that as σ^2 increases, finer symmetric categorization profiles that were NE before the noise increase may stop being NE, thus reducing the number of symmetric NE. The proof for a decrease of the sample size n is analogous.

Next, we derive a sufficient condition for the coarsest possible categorization profile, denoted here by $[P_1^*, P_2^*]$, to be a NE in the $IP\&C$ game regardless of the noise level and the sample size. Since there is a unique coarsest categorization, there is no other categorization at the same level of coarseness or coarser to deviate to. We only need to examine unilateral deviations to a finer categorization. Let $P_1 \in \mathcal{P}$ denote any categorization finer than the coarsest. For $[P_1^*, P_2^*]$ to be a NE, we need inequality (9) to hold. In this case, the LHS is the difference in variance between a coarser and a finer categorization which is always ≤ 0. A sufficient though not necessary condition for inequality (9) to hold is thus that the RHS is ≥ 0. This is guaranteed if $(1-w)[Bias^2[P_1, P_2^*] \geq w\Big[Bias^2[P_1^*] - Bias^2[P_1]\Big]$ for any $P_1 \in \mathcal{P}$ finer than the coarsest. Note that as long as $w \leq \frac{1}{2}$ this condition always holds for any $P \succ P^*$ and hence for all possible categorizations as all of them refine P^*. For $w > \frac{1}{2}$, the LHS would be smaller than the RHS for sufficiently high $\frac{\sigma^2}{n}$. □

Proposition 5. Efficiency of Equilibria in the $IP\&C$ Game

Proof. We derive when $EPE^{IP\&C}[\widetilde{P}_1, \widetilde{P}_2] \leq EPE^{IP\&C}[P, P]$ for any $|\widetilde{P}| < |P|$. Using equation (5) and a number of algebraic transformations we get the following condition:

$$\begin{aligned} & w\Big[Bias^2[\widetilde{P}] - Bias^2[P] + Var[P] - Var[\widetilde{P}]\Big] \\ & \quad \leq 2\Big[Var[P] - Var[\widetilde{P}]\Big] \end{aligned} \tag{15}$$

We know that in any stochastic environment $Var[P] - Var[\widetilde{P}] > 0$.

(i). If $Bias^2[\widetilde{P}] - Bias^2[P] > 0$ then we can write inequality (15) as:

$$w \leq \frac{2\Big[Var[P] - Var[\widetilde{P}]\Big]}{Bias^2[\widetilde{P}] - Bias^2[P] + Var[P] - Var[\widetilde{P}]}$$

We know from Lemma 2 Part 1 that $Var[P] - Var[\widetilde{P}]$ is strictly increasing in σ^2 and strictly decreasing in n (for any $\sigma^2 > 0$). Hence as σ^2 increases or n decreases, the RHS of the above inequality increases. This means that the noisier the environment and the smaller the sample size, the lower the threshold weight on coordination that is sufficient to Pareto-rank any coarser symmetric Nash equilibrium as more efficient than any finer symmetric Nash equilibrium. What is more, for any $0 \leq w < 1$, there will be some noise σ^2 to sample size n ratio that is high enough to allow for such a Pareto-ranking of the Nash equilibria in the $IP\&C$ game.

(ii). If $Bias^2[\widetilde{P}] - Bias^2[P] \leq 0$, then it is straightforward to show that inequality (15) holds for any $0 \leq w \leq 1$, any $\sigma^2 > 0$, and any $n \geq 1$. □