Learning frames

Vessela Daskalova and Nicolaas J. Vriend

ISSN 1473-0278

Working Paper No. 929 August 2021

School of Economics and Finance

Queen Mary
University of London

Learning frames*

Vessela Daskalova† and Nicolaas J. Vriend‡

August 12, 2021

Abstract

Players may categorize the strategies available to them. In many games there are different ways to categorize one's strategies (different frames) and which ones players use has implications for the outcomes realized. This paper proposes a model of agents who learn which frames to use through reinforcement. As a case study we fit the model to existing experimental data from coordination games. The analysis shows that the model fits the data well as it matches the key stylized facts. It suggests a trade-off of using coarser versus finer representations of the strategy set when it comes to learning.

Keywords: Variable Frame Theory, Coordination games, Categorization, Reinforcement learning, Focal points, Bounded rationality

JEL Classification: C63, C72, C91, D9

*We thank colleagues at the CEF conferences in London and San Francisco, at the Amsterdam Symposium on Behavioral and Experimental Economics, at the ASSET meeting in Evora, at the ESA conference in Honolulu, at the University of Surrey Constructed Complexities Workshop, at GREQAM (Aix-en-Provence), at IMEBESS Barcelona, at the Barcelona GSE Summer Forum Computational and Experimental Economics Workshop, at the UEA Behavioural Game Theory Workshop, at the University of Cambridge, at the Institute for Advanced Study in Toulouse, at University College Dublin, at LESSAC at the Burgundy School of Business, as well as the editors and two anonymous referees for very helpful comments. All errors are our own. Vessela Daskalova acknowledges funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program), from the IDEX Chair "Information, Deliberation and Collective Choice" at IAST, and from Cambridge-INET.

†School of Economics and Geary Institute, University College Dublin, Belfield, Dublin 4, Ireland and Toulouse School of Economics (IAST), University of Toulouse Capitole, 1 Esplanade de l'Université, 31080 Toulouse, France, <vessela.daskalova@ucd.ie>.

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

1 Introduction

To choose among different actions, a decision maker first needs a mental representation of the available options. In many cases some attribute(s) of the decision situation may prompt the decision maker to categorize the options available to her. Categorization usually implies that the agent treats objects that she has placed in the same category in the same way (Daskalova and Vriend, 2020; Mengel, 2012a; Mohlin, 2014). One important class of situations in which people may categorize is strategic situations. In this paper we consider agents who may group their own strategies into categories. A frame here is a partitioning of the set of strategies available to the agent into categories, with each category containing one or more strategies. Often multiple internal representations of the same situation or multiple frames of it are possible and which one the decision maker chooses has consequences for their behavior and payoffs. Moreover, in many strategic situations which views of the strategy set are best depends on the views of the strategy set chosen by the other players.

How categorizing the strategy set may matter in strategic situations is illustrated through some coordination games, where it may help players solve difficult equilibrium selection issues. To fix ideas consider the following matching objects example. There are five objects and players have to independently choose one of them. If they both choose the same one, each player gets the same positive payoff. If they choose different ones, they get a zero payoff each. If there is nothing to distinguish the five objects, the best the agents can do is put them all in one category and randomize uniform randomly among them. However, imagine that one of the objects is blue, while the other four are green. If a player pays attention to color, she may categorize the blue object separately and may form another category in which she puts all the green objects together. If both players do so and make their subsequent choice by selecting the category blue, they would be able to coordinate better than through pure randomization.[1]

The main goal of this paper is to improve our understanding of factors that may determine which frames of the strategy set players choose and which

[1] This is the simplest possible example of a much more general class of situations where many attributes may be present (Bacharach, 1993; Bacharach and Bernasconi, 1997). In this example, it is a focal point that gives rise to alternative possible representations of the strategy set. Categorizing the strategy set as discussed in this paper, however, need not be focal point based. It could be prompted both by attributes that are and such that are not part of the formal game theoretic description of the situation, e.g. by payoffs or by labels of the actions available.

outcomes emerge as a result.[2] We present a model of agents choosing among alternative frames of a given strategy set. That is, instead of assuming that decision makers have innate cognitive limitations in viewing the strategy set and have an ex ante fixed frame, we assume that ex ante all possible frames of the strategy set are available to each agent. Our model is a dynamic model based on reinforcement learning. Agents learn which frames and categories of strategies to use based on their past experience, and they are more likely to choose those that have performed better in the past. We study how frames that are seen as useful, as well as how the corresponding outcomes, emerge in the process of social interaction. Taking such a dynamic perspective has several advantages. First, using reinforcement learning allows us to keep the model modest in terms of cognitive assumptions. One can thus expect that a relatively broad class of more complex cognitive models may share some of its behavioral properties. Second, we present its application to pure coordination games. A key challenge in these games is the selection of one of a multiplicity of equilibria. Instead of making an a priori assumption of an equilibrium refinement concept, a study of learning dynamics may help to shed some light on the outcomes one may expect in such games. Third, as we will see in our analysis, learning dynamics may present some additional reasons to favor one internal representation over another, reasons that may not be obvious in a static analysis.

As a case study we fit the model to empirical data, examining how it may help us gain insights into human behavior in the coordination games from the laboratory experiment by Bosch-Domènech and Vriend (2013). The analysis shows that the model fits the data from the Bosch-Domènech and Vriend (2013) experiment well, accounting also for differences between experimental treatments. We gain the following insights from analyzing the model under the parameters that best fit the experimental data. First, agents may coordinate even without using the same view of the strategy set. Second, it is not clear that using the finest frame (placing each strategy in a separate category) is always best; there are both advantages and disadvantages to using such a fine representation of the strategy set when we consider learning dynamics.

The factors that determine which internal representation of the strategy set agents learn to use and which outcomes emerge include the number of frames, the number of categories in each frame, the number of actions in each category, as well as the payoffs in the underlying game. Our model highlights that both

[2] As the games we consider in this paper are one-shot, we use the terms actions and strategies interchangeably throughout the paper.

the frame and the category of strategies that agents use matter and it helps us suggest a trade-off of using coarser versus finer representations of the strategy set when it comes to learning.

On the one hand, a coarse frame, i.e. one that splits the strategy set into fewer categories, makes it easier for an agent to find and to learn to use the appropriate category of strategies (as there are fewer options to consider by trial and error and the updating of the perceived strengths of categories will be quicker). This suggests a learning rationale for some coarseness of mental representation, complementary to other rationales discussed in the literature on categorization. On the other hand, a category that contains fewer strategies makes it more likely for agents to choose the right strategy within the category. As we will see, as far as mental representations are concerned this means that some kind of mixture of coarseness and fineness may be advantageous. More specifically this helps to understand why a non-equilibrium focal point may be selected instead of multiple Pareto-superior Nash Equilibria. The focal point alleviates the equilibrium selection problem agents face by prompting them to put it in a separate mental category and to thus distinguish it from other available actions.

To further highlight the importance of frames and categories, we also compare the performance of the model with frames and categories in matching the experimental data with the performance of a basic reinforcement learning model without frames and categories, and show that while the data is well accounted for by the model with frames, it cannot be explained by the model without frames.

The remainder of this paper is organized as follows. Section 2 discusses some related literature. In Section 3 we introduce the model. Section 4 presents the experiment we use as a case study. Section 5 presents the analysis of the model and discusses the findings from the case study. Section 6 concludes.

2 Related literature

This paper sits at the intersection of several different literatures. First, on focal points and equilibrium selection in coordination games; second, on learning in games, in particular on reinforcement learning models; and third, on categorization in games. We now discuss the relation to each of these literatures and to some other relevant papers on the topic of bounded rationality in games.

A common theme between the first strand of the literature and our paper is that the standard normal form representation of the game that is usually used to depict a one-shot strategic situation does not necessarily capture the way a player thinks of the situation and in many cases there may be many alternative representations possible (Bacharach, 1993; Bacharach and Bernasconi, 1997; Bacharach and Stahl, 2000; Casajus, 2000, 2001; Janssen, 2001; Sugden, 1995; Gauthier, 1975; Schelling, 1980).[3] The idea in our model that players may have different frames of the strategy set is most closely inspired by Variable Frame Theory (Bacharach, 1993; Bacharach and Bernasconi, 1997; Bacharach and Stahl, 2000). In our model as in the other papers cited above the focus is on situations where some attributes are distinguished relatively naturally. Given these attributes, there is a collection of possible representations of the situation, i.e. of different frames of the strategy set.

However, there are two main differences between this paper and the Variable Frame Theory (VFT) related literature. In contrast to the literature, we assume that ex ante all these naturally arising frames are available to each agent. This differs from VFT and from Janssen (2001), in which a player has one possible representation of the situation based on the families of attributes she has the cognitive capacity of distinguishing. Thus, our model is not one of limited cognitive capacities in terms of perceiving attributes. We complement the previous literature, by focusing on the question how agents learn to make active use of some frames while leaving others unused.[4] The second main difference is that our model is dynamic. To deal with the issue of selecting among a multiplicity of equilibria, the approach in most of the above literature has been to assume that the payoff dominant equilibrium in expected payoffs will be chosen. This is also known as the Principle of Payoff Dominance, Principle of Coordination in VFT, Principle of Collective Rationality (Sugden, 1995) or Principle of Individual Team Member Rationality (Janssen, 2001).

Instead we follow a dynamic approach based on reinforcement learning with agents being more likely to select frames that have been more successful in the past.[5] Rather than assuming that players coordinate on the Pareto superior

[3]There are also some other less related papers on focal points. Binmore and Samuelson (2006) provide an evolutionary game theory explanation of how focal points may arise and Alberti et al. (2012) present a model of the emergence of salience in recurrent games.

[4]There is some neurophysiological evidence supporting the assumption of agents simultaneously experimenting with alternative representations in an individual decision problem (Wunderlich et al., 2011).

[5]Note that we focus on learning frames of the strategy set that are useful for coordination in a series of one-off interactions. Learning to coordinate in repeated interactions has been considered by Crawford and Haller (1990) and by Goyal and Janssen (1996) among others.

equilibrium in expected payoffs or that they reason as a team, we study whether and in what cases the Pareto superior equilibrium emerges as a result of the interaction. By looking at which outcomes emerge in the process of social interaction we are also able to learn more about the advantages and disadvantages of different frames when it comes specifically to learning.[6]

In this respect our paper is related to a second literature, namely on learning models, and in particular models that are based on some kind of reinforcement learning, examples of which include Bush and Mosteller (1951); Roth and Erev (1995); Börgers and Sarin (1997); Erev and Roth (1998); Camerer et al. (2004).[7]

What we add to the reinforcement learning literature is the idea of learning not actions, but frames and categories of the strategy set. Romero and Rosokha (2019) also consider players who learn to categorize strategies through reinforcement. Their model and applications are complementary to ours in various ways. First, they consider categorization as an adaptive process, rather than as arising from natural attributes of the situation. Second, they focus on a different type of game, the indefinitely repeated Prisoner's Dilemma.

A further principle underlying prediction of which action will be played in the VFT literature is what Bacharach and Bernasconi (1997) call Symmetry Disqualification, and Janssen (2001) refers to as the Principle of Insufficient Reason. Symmetry Disqualification states that if a player sees insufficient reason to distinguish among certain objects, then she will treat them in the same way. Thus, she will randomize uniform randomly among the four green objects in the example from the Introduction. Symmetry Disqualification is also related to the concept of attainable strategies in Crawford and Haller (1990). While this is perhaps not explicitly stated in these papers, we believe that the Principle of Insufficient Reason and Symmetry Disqualification are closely related to the idea of putting objects in mental categories and we phrase our model in terms of categories of strategies.

Hence the third literature this paper belongs to is the literature on categorization, and categorization in strategic situations in particular. Recent game theoretic literature has postulated that players may categorize other players

[6] A further difference with this literature is that in some of these models players with limited cognitive capacities have beliefs about the (weakly) lower cognitive capacities of their opponents (see e.g. Janssen (2001) or Bacharach and Stahl (2000)). In our model agents have no explicit beliefs about which frames the other agents use.

[7] See also Sutton and Barto (2018). This is by no means a comprehensive review of the vast literature on learning, other, less related to our work, learning models include Young (1993) and Arifovic and Ledyard (2011), see also review in Fudenberg and Levine (1998).

(Azrieli, 2009, 2010), they may categorize games (Gibbons et al., 2017; Grimm and Mengel, 2012; Heller and Winter, 2014; LiCalzi and Mühlenbernd, 2019; Mengel, 2012a; Huck et al., 2011; Jehiel, 2005), they may bundle states (Jehiel and Koessler, 2008), nodes other players must move at (Jehiel, 2005; Neilson and Price, 2011), or moves in a game (Jehiel and Samet, 2007) into analogy classes. Daskalova and Vriend (2020) consider players who categorize their own past experiences and are interested in making predictions and coordinating their predictions with one another. LiCalzi and Maagli (2016) consider agents who bargain over a common convex categorization of a given set. Unlike those papers we focus on agents categorizing their own strategies and learning which among alternative possible frames of the strategy set to use. The key assumption here and in most of the above cited categorization literature deriving from previous literature in psychology is that objects within a category are treated in the same way. Papers that show that the use of coarse categories can be optimal in other contexts include Mohlin (2014); Al-Najjar and Pai (2014); Mengel (2012b); Daskalova and Vriend (2020). Here our focus is on the trade-offs that emerge in using coarse and fine frames, and categories containing few versus many strategies, when it comes to learning which view of the strategy set to use.

In terms of papers on bounded rationality in games more generally, Arad and Rubinstein (2012) and Arad and Rubinstein (2019) consider players who think of features of strategies rather than strategies per se. Arad and Rubinstein (2012) is an experimental investigation of the Colonel Blotto game linking it to a multi-dimensional reasoning procedure. Arad and Rubinstein (2019) propose multi-dimensional equilibrium as an alternative to Nash in such situations.

In being about a type of bounded rationality in games, this paper bears some relation to the literature on level-k (Nagel, 1995; Stahl and Wilson, 1994, 1995; Camerer et al., 2004; Costa-Gomes and Crawford, 2006). However, it differs in many aspects from this literature. While in level-k, different levels correspond to differences in strategic sophistication, in our framework different levels correspond to differences in perception of the situation, or more precisely to differences in which attributes agents take into consideration when forming categories. Moreover, a general feature of level-k models is that a higher level is a best response to a lower level, whereas this is not a feature of our model. With the exception of Alaoui and Penta (2015), much of the level-k literature treats levels as given. To the extent that Alaoui and Penta (2015) consider endogenizing levels of strategic reasoning and we consider endogenizing levels of framing of a strategic situation, there is some commonality.

3 Model

This section describes the general set-up of the model. There is a population of n agents who are randomly (re)matched in pairs each period to play the same underlying game Γ. The underlying game Γ is a simultaneous-move one-shot game and thus has a "standard" normal (strategic) form representation. Each agent has an action set $A = \{a_1, a_2, .., a_m\}$ describing her options in the normal form.[8] An agent may categorize the actions available to them. A category C is a subset of the set of actions in the underlying game Γ, i.e. $C \subseteq A$.

Different ways to categorize the action set give rise to different mental representations or different frames. A frame F is a set of categories partitioning the set of actions in the underlying game Γ, that is $F = \{C_1, C_2, ..C_k\}$ such that $\bigcup_i C_i = A$ and $C_i \cap C_j = \emptyset$ for all $C_i, C_j \in F$. The set of frames that are feasible in a given game Γ is denoted by $\Phi = \{F_0, F_1, ..., F_f\}$. We say that a frame F_x is coarser than a frame F_{x+i}, where $x \in \mathbb{N}_0$, $i \in \mathbb{N}$ if and only if F_x partitions the action set A of the underlying game Γ into fewer categories than F_{x+i}.[9] In that case we say that F_{x+i} is finer than F_x.

We focus on cases where categories and frames arise "naturally" from attributes that agents distinguish.[10] This means that the agents in our model cannot form categories by arbitrarily combining actions but can only categorize actions on the basis of perceived attributes. Examples of such attributes are the color of the objects in the matching objects game discussed in the introduction, a focal payoff in the underlying game, or labels attached to actions in the underlying game. Thus, attributes that prompt one to categorize one's strategies might be (e.g. in the case of a focal payoff) or might not be (e.g. in the case of

[8] As we consider cases where both agents have the same options, we leave out indices for players in this description.

[9] Note that in the above definition of frame coarseness we are sidestepping the detail of how many actions any given category contains. Thus, under this definition a frame that contains two categories, each with five actions, is equally coarse as a frame that contains two categories, one with one action and the other with nine actions. This definition has the advantage of simplicy of exposition. We will clarify the importance of the number of actions within a category where this matters specifically for our results.

[10] The related idea of the existence of "natural kinds" has a long history in philosophy, see, for example, Chapter 7 in Book I in John Stuart Mill's *A System of Logic* (1858) or Quine (1969). Aristotle's classification of animals as well as many subsequent taxonomies in the sciences can be seen as examples of classification on the basis of the presence or absence of some attribute.

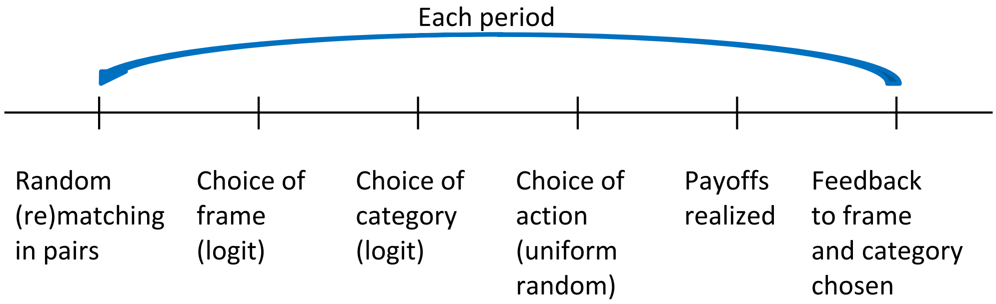


Figure 1: Timeline in the model

labels) part of the formal description of the underlying game in the normal form.

We assume that agents are able to recognize all possible attributes that are distinguished by the modeler. Thus, for example, if an agent needs to choose an object, with the available objects coming in different shapes and colors, we assume that each agent does recognize these shapes and colors.[11]

The question we focus on is which views of the strategy set agents learn to use, and how this may change as they learn in the process of social interaction. Depending on which attributes an agent takes into account, she may look at the strategy set through different frames and use different categories. We think of the different possible frames and categories as competing for the agent's attention, with the attention paid to each frame and category depending on the past success when using them. Some frames and categories may be more successful than others, and this may change over time as all agents may be learning which views of the strategy set to use.

Each period each agent makes a decision and learns from its outcome. The timeline in the model is presented in Figure 1. To undertake an action she needs to make several choices. First, she needs to choose a frame to use. Then she needs to choose a category from the frame she has selected. And finally, she needs to choose an action from the category chosen. Every frame and category have a measure of how good they are, that is their "strength" as determined by their past performance, and frames and categories are chosen on the basis of these perceived strengths. Actions within a category are chosen uniform randomly.

Thus, we need to specify how the perceived strengths of frames and categories

[11]Modeling how the agents have come to recognize attributes as such is beyond the scope of this paper. One possible interpretation is that agents have evolved to recognize certain attributes (Berlin and Kay, 1991). Our model would be complementary to such an evolutionary perspective.

are determined, and how they are taken into account in the choices. We start with the choice of frames. In each period an agent first chooses which frame to use with the logit rule (McFadden, 1974; Cramer, 2003). For an agent i the probability of choosing frame j in period t is equal to

$$p_t(F_{i,j}) = \frac{e^{\beta^F s_t(F_{i,j})}}{\sum_f e^{\beta^F s_t(F_{i,f})}} \,, \tag{1}$$

where $s_t(F_{i,j})$ is the strength of frame j of agent i in period t and β^F is a parameter determining the sensitivity of frame choice to strengths. For $\beta^F = 0$ choice is uniform random, all frames are chosen with equal probability. If $\beta^F > 0$ frames that had a better performance in the past are more likely to be chosen.

After choosing a frame to use, an agent chooses a category from this frame. This choice is made analogically with the logit rule, where $p_t(C_{i,j,c})$ is the probability that agent i in period t chooses category c in frame j,

$$p_t(C_{i,j,c}) = \frac{e^{\beta^C s_t(C_{i,j,c})}}{\sum_k e^{\beta^C s_t(C_{i,j,k})}} \,, \tag{2}$$

$s_t(C_{i,j,c})$ is the strength of category c in frame j of agent i in period t, and β^C is a parameter determining the sensitivity of category choice to strengths.

After choosing a frame and a category, an agent chooses an action within the category chosen uniform randomly.[12]

We now turn to the determination of the perceived strengths of the frames and categories. The initial strengths of frames and categories are captured by the parameters s_0^F and s_0^C, respectively. Thus, each agent starts out with all possible frames and values all of them equally in period 0.[13]

The strength of the frame j that agent i uses in period t, $s_t^F(F_{i,j})$, is updated according to the following rule:

$$s_t^F(F_{i,j}) = s_{t-1}^F(F_{i,j})(1 - \alpha^F) + \pi_t \alpha^F \tag{3}$$

where $s_{t-1}^F(F_{i,j})$ is the strength of this frame in the previous period, α^F is the

[12]Treating all objects within a category equally is a basic feature of many categorization models (Mohlin, 2014; Mengel, 2012b). It is also in line with the Principle of Insufficient Reason.

[13]In principle one could start with any random valuation of the different frames, but it would be difficult to conceptually justify why one would start with one valuation rather than another. We adopt here the Principle of Insufficient Reason, which dictates that absent any reason to treat objects differently one should treat them in the same way.

weight the agent places on the latest interaction, and π_t is the payoff from the interaction in period t. The higher the α^F (with $0 \leq \alpha^F \leq 1$), the greater the weight the agent places on more recent experiences.

The updating of the strength of the category used is analogical. That is, the strength of category c in frame j of agent i in period t, $s_t^C(C_{i,j,c})$, is updated according to:

$$s_t^C(C_{i,j,c}) = s_{t-1}^C(C_{i,j,c})(1 - \alpha^C) + \pi_t \alpha^C \tag{4}$$

where α^C is the weight placed on the most recent experience. Note that some categories may exist in more than one frame of the agent. In that case the strength of a category is updated with the payoff it generated whenever it is used in each frame that it is an element of.

To sum up, there are in principle the following free parameters in this model: the initial strengths of frames s_0^F and of categories s_0^C, the learning rates for frames α^F and for categories α^C, and β^F for choice of frames and β^C for choice of categories. In addition, in principle the parameter values could also vary between agents and over time. For the sake of parsimony, in this paper we present a version of the model where we use the same parameter values at both the frame and the category level, i.e. $s_0^F = s_0^C = s_0$, $\alpha^F = \alpha^C = \alpha$, and $\beta^F = \beta^C = \beta$. These parameter values are the same for each agent and they do not vary over time. Thus, the model has three free parameter values.[14]

Our assumption here is thus that all agents start out with all possible views of the strategy set and they weigh them in the same way. This specification also implies that they use the same learning mechanism, which does not differ between treatments and does not adapt over time. Thus, *ex ante* agents are identical. However, what is interesting is that even so *ex post* they might learn to use different views of the strategy set.

The agents do not receive any feedback on the frames other agents use, on the categories, or on the strategies chosen by them. Neither do they receive feedback on the other agent's payoff. They receive feedback only on their own payoff and use this to update the strength of the frames and categories they have chosen. In this sense our model is complementary to some other learning models (such as e.g. fictitious play, or experience weighted attraction learning).

[14] The results for a version of the model where the parameters for frames and categories differ, that is with six free parameters (s_0^F, s_0^C, α^F, α^C, β^F, β^C), are available from the authors upon request.

There are various reasons to restrict feedback in this model to an agent's own payoff only. First, it is difficult to envisage how one could justify an assumption of observing another agent's internal representation of the strategy set. Second, giving only own payoff as feedback is consistent with the experiments that we use as a case study. Third, it seems interesting to investigate whether a model as simple as ours can capture the stylized facts observed in an experiment with human subjects.

Related to the limited feedback, the agents do not form any explicit beliefs about the frames and categories that other agents use. Thus, our approach is complementary to a Bayesian approach where they would have beliefs over others' types, with a type consisting of a particular frame/category combination. Here allowing for explicit beliefs would complicate the model unnecessarily as each agent would need to form beliefs over both the frames and the categories that other agents could use, taking also into account that a category may exist under multiple frames. However, the fact that we use a reinforcement learning model should not be seen as some kind of fundamentalist anti-strategic reasoning approach. Instead of viewing reinforcement learning in complete contrast to players' forming beliefs about other players' actions, one could think of reinforcement learning itself as a form of (adaptive) implicit belief learning. Note that as the perceived strengths of frames and categories that they use change over time as a result of their interaction with others, agents in our model are learning and updating their valuations of how good the different options are.

4 Case study: the underlying game and experiment

In this section we introduce the underlying game, its static analysis, as well as the experimental data from the experiment by Bosch-Domènech and Vriend (2013), which we use for the case study of our model.

4.1 The underlying game

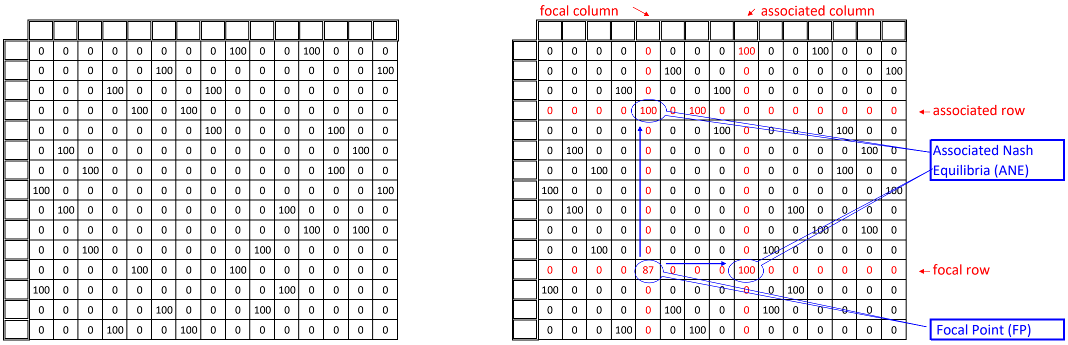
We introduce the non-equilibrium focal point game from the experiment by Bosch-Domènech and Vriend (2013) (henceforth BDV). The main question this experiment addresses is whether a non-equilibrium focal point can help players coordinate better. At the beginning of each experimental session players are randomly assigned to be either a row player or a column player. Each player receives the payoff matrix of the game. Players are randomly and anonymously (re)matched in pairs each period. A player's task is to independently choose one of 15 possible rows (columns) of the payoff matrix. The payoff matrix from the control treatment is presented in Figure 2a, where the payoff in each cell represents the monetary reward that is given to each player.[15] There are overall 30 payoff-equivalent Nash equilibria in pure strategies (NE). These NE are scattered in the payoff matrix in such a way that there are always two NE per row (column) in an attempt to avoid any equilibrium being more conspicuous than the others (e.g. there are no NE in the corners or the center of the payoff matrix).

There are three other treatments. In each of them a focal point is introduced by reducing the payoff of one of the NE as in the example in Figure 2b. The three treatments involve a payoff of 46, 87, and 99, respectively. Figure 2b shows the 87 treatment (the payoff matrices for the other two treatments are analogical, but for the magnitude of the focal point). The idea is that after reducing the NE payoff, the corresponding action profile is no longer a NE. Given that one player plays the focal action, the other player has an incentive to deviate to an Associated Nash Equilibrium (ANE), i.e. the NE in the same row (column) as the focal payoff. Note, however, that since there are two ANE (one in the same row and one in the same column as the focal point), if both players were to deviate from the focal action in the hope of realizing an ANE, they would miscoordinate. Thus, in the focal point treatments instead of 30 NE, there are 29 NE and one non-equilibrium focal point.

We now discuss the set of feasible frames of the strategy set in the non-equilibrium focal point game, starting with the frames that arise in the treatments with a focal point. They are illustrated in Figure 3. A player can put all possible strategies in one category, the category "any" and then choose uniform randomly among them. We call this frame F_0.

However, in the focal point treatments a player may observe that there is a payoff in one of the cells of the payoff matrix that differs from all others. Thus, the player can split the category "any" into the category "focal" and the category "any with two NE". This view of the strategy set corresponds to F_1 in Figure 3. In the first category, there is only one action - i.e. the focal action, which

[15] A number of variations of this matrix were used in BDV, consisting basically of rotating or mirroring the payoff matrix, which implies that subjects in different sessions may face slightly differently configured payoff matrices.



a) Control - without focal point b) 87 treatment - with focal point

Figure 2: Game from Bosch-Domènech and Vriend (2013)

the player plays with probability 1 if she chooses this category. In the category "any with two NE" there are 14 actions corresponding to the remaining 14 rows (columns).

A player may further distinguish the associated row (column) (see Figure 2b). Thus, under F_2 a player partitions the category "any with 2 NE" into the two categories "associated" and "any other with 2 NE". Note that the category "associated" can be distinguished only after distinguishing the category "focal". The next level is F_3. We assume that under F_3 a player splits the category "any other with 2 NE" into the 13 rows (columns) that it comprises.[16]

Next, we turn to the control treatment (the treatment without a focal point). There are two possible frames of the strategy set. Just as in the focal point treatments, a player can use frame F_0, putting all possible strategies in one category ("any") and then choosing uniform randomly among them. Alternatively, a player could put each strategy in a separate category and have fifteen separate categories, each containing one strategy corresponding to the choice of the respective row (column). We call this frame F_1.[17]

[16]In principle there can be a number of frames between F_2 and the frame that we described as F_3. After distinguishing the ANE, a player may realize that there is another NE in the same row (column) as the ANE, and thus she may have a frame that distinguishes the corresponding actions from the remaining rows (columns) with two NE. Following such steps, eventually all rows (columns) could be distinguished. Note, however, that while the associated action has a clearly distinct feature in that a deviation from the focal to the associated action may lead to a strict payoff gain, there is no such payoff gain from further deviations, as all NE are equivalent. In other words, the distinction between these other rows (columns) does not appear that prominent. Therefore, we assume that under F_3 a player splits the category "any other with 2 NE" directly into the 13 rows (columns) it comprises.

[17]Note that this frame essentially corresponds to frame F_3 in the treatments with a focal point.

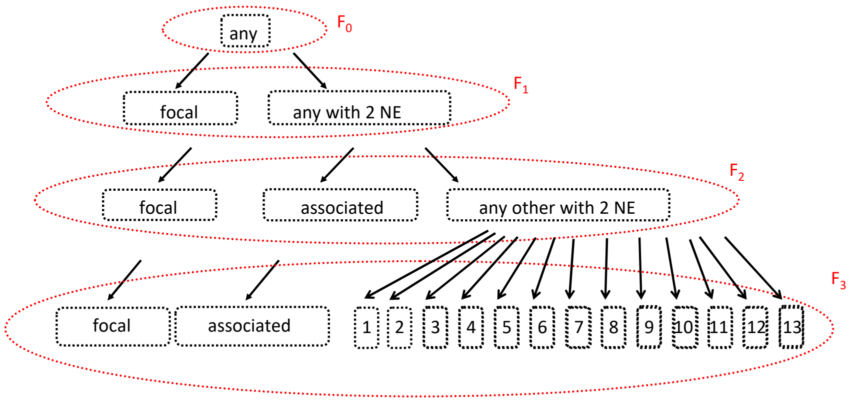


Figure 3: Frame levels in focal point treatments in the BDV game

4.2 Static analysis

What are the expected payoff consequences when the players apply the frames just discussed? We now discuss our interpretation of the predictions that VFT would make for the game we consider. By doing so we derive predictions from a static model in order to then compare them with the findings of our dynamic model.

For *given* frame profiles of the players, the underlying normal form of the game can be transformed into a payoff matrix of expected payoffs from alternative category profiles (as in VFT). Figure 4 shows the consequences in terms of the expected payoffs when both players use the same frame. If both players use F_0, they put all 15 strategies in the same category "any". Each player randomizes among the 15 strategies and her expected payoff is equal to $\frac{1}{15} \times \frac{1}{15} \times (100 \times 29 + 87) = 13.28$.[18] This is shown in the first table of Figure 4. If both players use F_1, each of them has two categories of strategies: "focal" and "any with 2 NE". As we can see, if both players use F_1 there are two equilibria: ("focal", "focal") and ("any with 2 NE", "any with 2 NE"). The equilibrium in which both use the focal category is Pareto superior to the one in which both choose the category "any with 2 NE". Variable Frame Theory, using the Principle of Payoff Dominance to select among alternative equilibria, would predict that ("focal", "focal") will be played if both players make a decision under frame F_1.

Next, we consider the case of both players using F_2. Each player has three categories of strategies and there are three equilibria, two asymmetric and one

[18] We are considering here the expected payoffs in the 87 focal point treatment. The expected payoffs in the other treatments can be calculated in the same way.

symmetric: ("focal", "associated"), ("associated", "focal"), ("any other with 2 NE", "any other with 2 NE"). The two equilibria in which one player plays the focal category and the other the associated category are the Pareto dominant equilibria in the case when both players use F_2 and hence the VFT prediction. However, VFT would not predict which of the two equilibria would be played.

Figure 4 finally shows the case when both players use F_3. There are 29 equilibria just as in the original normal form: the 2 ANE plus the other 27 NE. All of these equilibria are equally efficient and entail a payoff of 100. VFT would suggest that any of them is a possible solution.

The asymmetric cases, when the players use different frames, are represented in Figure 5. The equilibrium analysis is analogical to that for the symmetric cases, and the VFT predictions are indicated in the figure.

The static analysis above is for *given* frame combinations. In our learning model, however, the players actually choose which frame to use. Before turning to the dynamic analysis (fitting our model to the experimental data), as a referee suggested, we can extend the static analysis to identify the 'equilibrium' profiles of frames. This provides a benchmark for the experimental data and for our learning model. To do this, Figure 6 effectively summarizes the predictions of Variable Frame Theory for different frame profiles, - i.e. selecting for all possible frame profiles the payoffs resulting from the Variable Frame Equilibria.

$F_0 \times F_0$	any
any	13.28

$F_1 \times F_1$	focal	any with 2 NE
focal	87	7.14
any with 2 NE	7.14	13.78

$F_2 \times F_2$	focal	associated	any other with 2 NE
focal	87	100	0
associated	100	0	7.69
any other with 2 NE	0	7.69	14.79

$F_3 \times F_3$	focal	associated	column 1	column 2	column 3	column 4	column 6	column 7	column 8	column 10	column 11	column 12	column 13	column 14	column 15
focal	87	100	0	0	0	0	0	0	0	0	0	0	0	0	0
associated	100	0	0	0	0	0	0	100	0	0	0	0	0	0	0
row 1	0	100	0	0	0	0	0	0	0	0	0	100	0	0	0
row 2	0	0	0	0	0	0	100	0	0	0	0	0	0	0	100
row 3	0	0	0	0	0	100	0	0	100	0	0	0	0	0	0
row 5	0	0	0	0	0	0	0	0	100	0	0	0	100	0	0
row 6	0	0	0	100	0	0	0	0	0	0	0	0	0	100	0
row 7	0	0	0	0	100	0	0	0	0	0	0	0	100	0	0
row 8	0	0	100	0	0	0	0	0	0	0	0	0	0	0	100
row 9	0	0	0	100	0	0	0	0	0	0	100	0	0	0	0
row 10	0	0	0	0	0	0	0	0	0	0	0	100	0	100	0
row 11	0	0	0	0	100	0	0	0	0	100	0	0	0	0	0
row 13	0	0	100	0	0	0	0	0	0	0	100	0	0	0	0
row 14	0	0	0	0	0	0	100	0	0	100	0	0	0	0	0
row 15	0	0	0	0	0	100	0	100	0	0	0	0	0	0	0

Notes: VFT prediction (shaded cells)

Figure 4: Expected payoffs under symmetric frame profiles BDV game (87 treatment)

$F_0 \times F_1$	focal	any with 2 NE
any	12.47	13.33

$F_0 \times F_2$	focal	associated	any other with 2 NE
any	12.47	13.33	13.33

$F_0 \times F_3$	focal	associated	column 1	column 2	column 3	column 4	column 6	column 7	column 8	column 10	column 11	column 12	column 13	column 14	column 15
any	12.47	13.33	13.33	13.33	13.33	13.33	13.33	13.33	13.33	13.33	13.33	13.33	13.33	13.33	13.33

$F_1 \times F_2$	focal	associated	any other with 2 NE
focal	87	100	0
any with 2 NE	7.14	7.14	14.29

$F_1 \times F_3$	focal	associated	column 1	column 2	column 3	column 4	column 6	column 7	column 8	column 10	column 11	column 12	column 13	column 14	column 15
focal	87	100	0	0	0	0	0	0	0	0	0	0	0	0	0
any with 2 NE	7.14	7.14	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29	14.29

$F_2 \times F_3$	focal	associated	column 1	column 2	column 3	column 4	column 6	column 7	column 8	column 10	column 11	column 12	column 13	column 14	column 15
focal	87	100	0	0	0	0	0	0	0	0	0	0	0	0	0
associated	100	0	0	0	0	0	0	100	0	0	0	0	0	0	0
any other with 2 NE	0	7.69	15.38	15.38	15.38	15.38	15.38	7.69	15.38	15.38	15.38	15.38	15.38	15.38	15.38

Notes: VFT prediction (shaded cells)

Figure 5: Expected payoffs under asymmetric frame profiles BDV game (87 treatment)

	F_0	F_1	F_2	F_3
F_0	13.28	13.33	13.33	13.33
F_1	13.33	87	<u>100</u>	<u>100</u>
F_2	13.33	<u>100</u>	<u>100</u>	<u>100</u>
F_3	13.33	<u>100</u>	<u>100</u>	<u>100</u>

Figure 6: Expected payoffs under Pareto-dominant equilibrium for each frame profile BDV game (87 treatment)

As we can see in Figure 6, where the outcomes of frame equilibrium profiles are underlined, frame F_0 is strictly dominated by the other frames, and thus, an individual player always has an incentive to go at least one frame level higher than F_0. Also from F_1xF_1, an individual player has an incentive to unilaterally go one frame level higher, as moving to F_2 allows a player to get the ANE payoff of 100 instead of the focal payoff of 87. While any profiles involving F_0 as well as the symmetric profile with F_1 are ruled out by such a frame equilibrium analysis, otherwise any combination of F_1, F_2 and F_3 may constitute a frame equilibrium.[19]

We can make the following observations. First, the asymmetric frame equilibria that we find in this static analysis suggest that even in equilibrium individual players may have different views of the world. Second, almost all frame equilibria imply that the players go for one of the ANE,[20] which means that both associated and focal actions are key. Third, given the multiplicity and nature of these frame equilibria, the solution to the underlying coordination problem is still far from trivial, and it may be that some kind of miscoordination (with outcomes that are neither Pareto efficient nor a NE) is not unlikely.

This is where the analysis of our learning model in the next few subsections comes in. It is complementary to the static equilibrium analysis presented here, focusing in particular on the dynamics. Studying these dynamics may be interesting for two reasons. First, as the trajectories observed in the experimental data presented in the next section show, shining some light on the behavior of the players in the intermediate run may be interesting. Second, a study of these dynamics may provide some insights as to which of these profiles we should consider as more plausible in the longer run. This applies to the equilibrium

[19] Note that in addition to these eight payoff-equivalent frame equilibrium profiles, there will be a number of mixed frame equilibria.

[20] The only exceptions are the following additional frame equilibria. In the symmetric frame equilibrium F_3xF_3 all the NE of the underlying game may be relevant, and in the asymmetric frame equilibrium F_2xF_3 we also find the equivalent (associated action, column 7) equilibrium.

selection of categories for a given combination of frames, as well as the selection among the frame equilibrium profiles mentioned above.

Focusing on these dynamics will also allow us to highlight the importance of some aspects of the learning dynamics and the differences between the frames in this respect. More specifically, while frame F_3 may be seen as weakly-dominant in Figure 6, looking at Figures 4 and 5 illustrates that there may be some hidden cost. That is, under the finer frame F_3 there are many more possible categories to choose from, and hence learning which category to use may take longer. This is a key part of the learning trade-off we suggest in the paper.

4.3 Experimental data

We now summarize some key findings of the BDV experiment. In the experiment there were 6-8 sessions for each treatment, with 18 participants in each session. They were randomly divided into an equal number of row and column players at the beginning of the session. They kept their roles throughout the session. The game was played for fifty rounds with random rematching in each period. The only feedback that the participants received after each round was the payoff from the interaction. They did not receive further feedback on the other player's action.

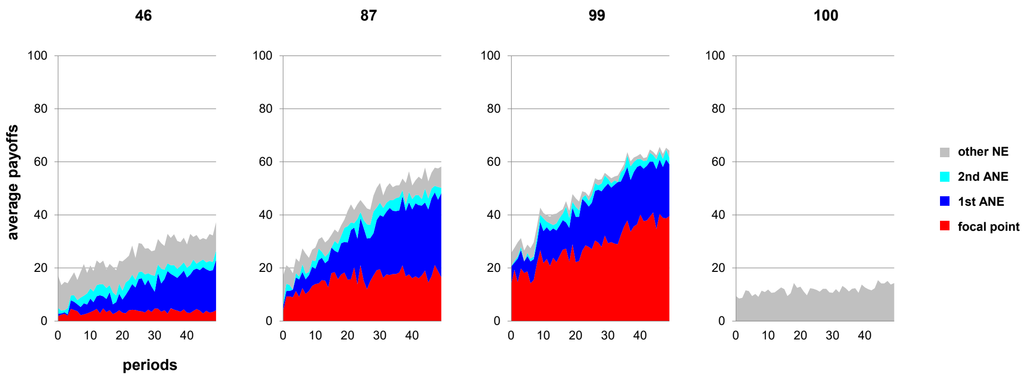


Figure 7: BDV experiment: average expected payoffs per player

Figure 7 shows the experimental data on average expected payoffs per player per type of outcome for each period for each treatment.[21] Conditional on coordinating, we can distinguish four types of outcomes in the BDV game:

[21]The authors report expected rather than actual payoffs in order to clear the data of pure random matching effects. The average expected payoffs are based on all pairs that could have been formed in a session rather than just those that were actually formed.

coordination on the focal point, coordination on the 1st Associated Nash Equilibrium (ANE), coordination on the 2nd ANE, and coordination on any other NE.[22]

We observe the following stylized facts in Figure 7. First, the average expected payoffs increase significantly over time in all three treatments with a focal point, indicating that participants learn to coordinate better over time, whereas in the control treatment there is only some almost negligible increase. Second, as a result, in the focal point treatments participants learn to achieve higher average expected payoffs than in the control treatment. Thus, the existence of a focal point helps participants to learn to coordinate in the experiment. Third, the higher the focal payoff, the higher the average expected payoffs players achieve towards the end of the fifty periods. That is, the average expected payoffs towards the end of the fifty periods are higher in the 99 than in the 87 treatment, and higher in the 87 than in the 46 treatment, with the difference between the 87 and the 46 treatments being more pronounced than the difference between the 99 and 87 treatments. Fourth, the increase in average expected payoffs in the focal point treatments is driven either by increased coordination on the focal point or by increased coordination on the ANE or by both, but not by increased coordination on the other NE. Fifth, the average expected payoffs realized through the ANE increase over time in all focal point treatments. Sixth, the higher the focal payoff, the higher the frequency of coordination on the focal point relative to coordination on the ANE at the end of the fifty periods. Seventh, the higher the focal payoff, the less frequent the coordination on the other NE.

Following BDV, we focus on the average expected payoffs as they capture various important aspects of coordination success, including the frequency of coordination as well as the different outcomes players coordinate on.[23]

[22]Note that while *ex ante* we expect that each of the two ANE has an equal probability of being realized, in a *given* experimental session one ANE is usually by chance realized more often than another. Once one of the ANE tends to prevail more in a given session, this difference tends to get reinforced as row-deviations and column-deviations from the focal to the associated action cannot co-exist successfully. Thus, *ex post* for each session we would expect a substantial difference between the frequencies of these row-deviations and column-deviations. Which of the ANE prevails will vary from session to session. The label "1st ANE" is given to the ANE that is realized more often in a given session. Hence, which of the two ANE is labeled "1st ANE" may differ from session to session.

[23]We can think of the average expected payoffs as a *weighted* frequency of coordination success. Each successful coordination counts as one and is weighted by the payoff from the outcome that the players have coordinated on, while miscoordination counts as zero. Thus, coordination on a Nash equilibrium is weighted by a payoff of 100, and coordination on a focal

Note that in principle a difference in average expected payoffs between two treatments could be due either to differences in individual behavior (the actions chosen) or simply to the difference in the magnitude of the focal payoff as such. In the Appendix we disentangle these two effects, computing precisely which effect is responsible for which proportion of the difference in average expected payoffs. As Table 1 in the Appendix shows, the increases in average expected payoffs when the focal payoff increases are predominantly due to differences in behavior. Depending on the specific focal payoff change considered, 72% to 82% of the increase in average expected payoffs is due to differences in the behavior, and only 18% to 28% is the direct result of the increase of the magnitude of the focal payoff as such.

To add some further perspective to this, when it comes to measuring the degree of coordination success, instead of using the average expected payoffs (i.e. the *weighted* frequencies of coordination), we can also consider the *unweighted* frequencies of coordination, simply counting each coordination on a Nash equilibrium or on a focal point as a success. Figure 13 in the Appendix shows the equivalent of Figure 7, focusing instead on the *unweighted* frequencies of coordination by type of outcome, and confirming the following key stylized facts. First, as we increase the focal payoff in the various treatments ("46", "87", and "99"), the unweighted frequencies of coordination increase as well. Second, the relative importance of the focal action increases with the magnitude of the focal payoffs, as the unweighted frequency of coordination involving the focal action (leading to either the focal point or an ANE) increases, relative to all others.

Thus, this confirms that the increases in average expected payoffs (when comparing the 46 to the 87 and 99 treatments) are not an artefact simply due to the changes of the magnitude of the focal payoffs as such, and do correspond to changes in individual behavior. Therefore, in our analysis we will focus mostly on the average expected payoffs as a measure of coordination success.

5 Case study: model analysis

We now fit the model presented in Section 3 to the data from the BDV experiment presented in Figure 7 and we analyse the model under the parameters that best fit the experimental data in order to learn more about which views of the strategy

point is weighted by the magnitude of the focal payoff.

set are consistent with behaviour observed in the experiment. At the end of this section we also consider whether a simpler reinforcement learning model (without frames and categories) could account for the data.

5.1 Fitting the model to the experimental data

As we explained above, for each of the parameters considered we assume the same value across all agents, all periods and all decisions (choice of frames and choice of categories), as well as all treatments (including the control treatment). Hence, the model has three free parameters: the initial strength s_0, the temperature parameter in the logit rule β, and the learning rate α. To look for the model parameter values that best fit the average behavior observed in the experimental data as summarized in Figure 7, we conducted an $11 \times 11 \times 11 = 1,331$ grid search of the parameter space.[24]

For each treatment, we compute the four average expected payoffs per agent for each of the fifty periods, distinguishing the expected payoffs per agent realized through the focal point, through each of the two ANE and through any of the other NE, for the experimental data as well as for the data generated by the model.[25] We then compute the mean squared errors between the experimental data and the model data, summing the mean squared errors over the four types of expected payoffs, over all periods and over all treatments, and we look for the parameter values that minimize this sum.

We found that the mean squared error between the experimental data and the model data is minimized by the following set of parameter values: $s_0 = 0.1$, $\beta = 5$, and $\alpha = 0.7$, giving a total Mean Squared Error of 13,353 or 3.71 per agent per period.[26]

The goal of the model is to match aggregate behavior in the experiment. What we have is not a model of a representative agent, and also not of an average agent. An essential aspect of this model is that we do not reduce behavior to a single (representative or average) agent, not even a single row and a single column

[24] Having all payoffs normalized to [0.0, 1.0], for s_0 we considered all values in the range $[0.0, 1.0]$ in increments of 0.1, for β all values in the range $[0, 50]$ in increments of 5, and for α all values in the range $[0.0, 1.0]$ in increments of 0.1.

[25] The experimental data for each treatment are averaged over all (6-8) sessions in that treatment. The model data for each treatment are averaged over 10,000 runs of the model. In each experimental session as well as in each model run, there are nine row and nine column players.

[26] To check the robustness of this optimization result, we did the grid search 10 times, each time for 10,000 runs, and this is the optimal parameter set for each of these 10 times.

agent. Instead, it is based on a number of individual agents (corresponding to the number in the experiment). One thing to note is that although ex ante all individuals are identical in our model, some heterogeneity in the individual behavior emerges. This is due not just directly to the stochastic elements in the choices but also to the random matching and the different experiences. And this heterogeneity matters, as it is the distribution of the behaviors of the agents that matters, in the experiment as well as in the model. A model with one representative or average row agent and one such column agent would not be able to match the stylized facts of the experiment. We would also like to emphasize that modelling a group of individuals rather than a single (representative or average) agent is important in the following sense. While a given individual may change their behavior, at the same time their environment is changing as well, as the other agents may be changing their behavior, thus changing what this given individual needs to learn. Our model captures this changing of individual behavior and changing of the environment at the same time.

5.2 Analysis of the model fitted to the experimental data

We now analyze the behavior of the model under the parameters that best fit the experimental data. Figure 8 shows the average payoffs realized through the various types of outcomes in the four treatments. As we can see, the fitted model matches each of the stylized facts of the experimental data listed above.

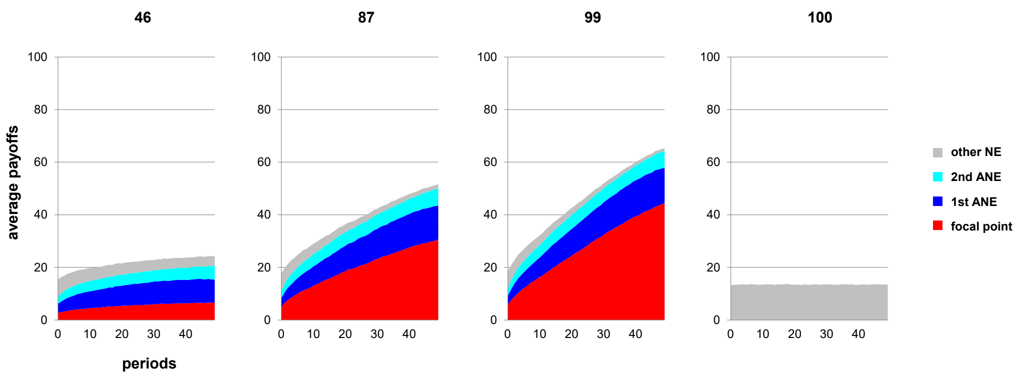


Figure 8: Learning model: Average payoffs per agent

Table 1 in the Appendix makes precise how much of the differences in average payoffs between treatments as captured by the learning model is due to differences in the magnitude of the focal payoff and how much is due to differences in the behaviour of the agents. The decomposition shows that just as for the

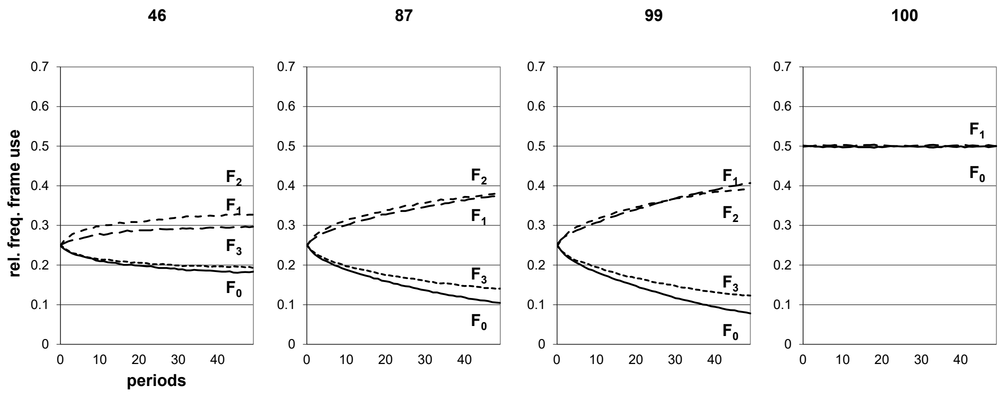


Figure 9: Learning model: relative frequency frame use

BDV experiment it is also true for the model that the differences in individual behavior are most important, accounting for 64% to 74% of the differences in expected payoffs, with only 26% to 36% the direct result of differences of the magnitude of the focal payoff as such. It also seems interesting to note that for the experimental data as well as for the learning model it is the case that the larger the change of the magnitude of the focal payoffs as such, the larger the relative importance of the change in behavior becomes.

In the Appendix we also consider the *unweighted* frequency of coordination for the learning model. Figure 13 in the Appendix confirms that also for the model the behavior of the agents is different in the various treatments and that the differences in payoffs are not simply a direct result of a change in the magnitude of the focal payoff as such.

Before considering the dynamics of the choices of frames and categories in more detail, let us provide a short interpretation of the specific parameter values found. The optimal initial strength of 0.1 implies that there will be some exploration of the strategy set as far as the frame and category choices are concerned. But this exploration may be somewhat limited as the strengths of some choices may increase above 0.1. This is also related to the other two parameters. The relatively high $\alpha = 0.7$ suggests that agents react relatively fast with respect to their most recent experience as to which frames and categories to use, and that they quickly forget earlier experiences. The $\beta = 5$ means that frames and categories that have performed better in the past have a higher chance of being selected.

In the experiment we observe the actions of the players, the extent of coordination, and the specific outcomes that they coordinate on. What is not observed in the experiment are the views of the strategy set that players use to make their choices. We can use the model to analyze the dynamics of use of frames and categories under the parameters that best fit the experimental data to gain some insights into the question which views of the strategy set would be compatible with the behavior observed in the experiment.

The analysis of the reinforcement learning model with frames suggests several insights into the behavior of the players in the experiment. By modeling frames of the strategy set and the learning of which one to use, we make explicit how the players may be viewing the various options presented to them with the focal point game, which attributes they may be taking into account, as well as how all this may be changing over time. In the model these dynamics themselves are governed by the three parameters that we used to fit the experimental data.

In order to gain some insights into which views of the game are most consistent with the behavior observed in the experiment, we present the dynamics of the relative frequency of frame use. Figure 9 illustrates the relative frequency of frame use over time for all four treatments (averaged over 10,000 runs). We see that in all focal point treatments agents increase their use of F_2 and F_1 and decrease their use of F_0 and F_3 over time. In the 46 and 87 focal point treatments F_2 is used slightly more often than F_1 at the end of the fifty periods, whereas in the 99 treatment F_1 is used slightly more often than F_2. The two frames in the control treatment are used equally often. Under F_0 in the control treatment an agent puts all actions in one category and chooses uniform randomly among them. Under F_1 in the control treatment each action is put in a separate category and these categories are then reinforced through the realized payoffs. As we can see, averaged over multiple runs of the model, there is no difference in payoffs between these two frames. That is, for the given parameters found, in the control treatment frame F_1 turns out to be just as good as the random choice under F_0.

We observe that in all three focal point treatments, agents tend to learn the frames of intermediate coarseness levels rather than the most or the least detailed ones. Moreover, the higher the focal point, the steeper the learning curves over the fifty periods. The analysis suggests that while players learn to coordinate more and more successfully, they do not necessarily achieve this by learning finer and finer frames.[27]

[27]Note that the frames are based on attributes of the game that agents perceive, and that a finer frame means that agents pay attention to more attributes. These levels of coarseness should not be confused with the levels of reasoning of the level-k literature, where each higher level of reasoning is a best response to players' using the previous level of reasoning. In our model a finer frame is not necessarily a best response to coarser frames.

Agents' views of the strategy set remain quite mixed in all focal point treatments, even after fifty periods. The use of the frames F_1 and F_2 is higher than the use of the frames F_0 and F_3 in all focal point treatments. Under frame F_2 agents distinguish the focal as well as the associated action from each other and from all the other possible actions, but besides or instead of this view, often the agents learn to use frame F_1 in which they distinguish only the focal action from all other actions as the most relevant one. This mixing between different frames may happen at the individual as well as at the population level. Thus, what we found is that when two agents meet they may actually look in a different way at the strategy set (although the options each of them is facing in the underlying game are the same).

The next question we consider is, conditional on a given frame being used, which categories within this frame do agents choose? Figure 10 represents the relative frequency of category use over time under the different frames (F_1 to F_3) for each of the three focal point treatments. The control treatment is omitted as agents in the control treatment under F_1 use all 15 categories corresponding to the different rows (columns) more or less equally often, with a relative frequency of about 0.07 each (at the end of the fifty periods).

We now analyze the category use within each frame in the different focal point treatments. In all treatments the frame F_0 contains only the category "any". Thus, whenever F_0 is used, the category "any" is used. Therefore no graph of category use under F_0 is presented. Figure 10 shows that in all three focal point treatments whenever agents use F_1, they use predominantly the category "focal". The use of the category "any 2 NE" under F_1 decreases substantially in all three treatments. Conditional on choosing F_2, agents use predominantly the categories "focal" and "associated" in all three treatments. In all three treatments, under F_3 agents also use the categories "focal" and "associated" more often than the other categories. In terms of between treatment comparisons, the higher the magnitude of the focal point, the higher the use of the category focal under F_1 at the end of the 50 periods. Moreover, both under F_2 and F_3, the higher the focal point, the higher the relative use of the category "focal" compared to the category "associated" at the end of the 50 periods.

Note that overall the difference in relative frequency of use of those categories whose use increases compared to those categories whose use decreases in each treatment is more pronounced under F_1 than under F_2 than under F_3. That is, we observe that the finer the frame of the strategy set the agent has (e.g. compare

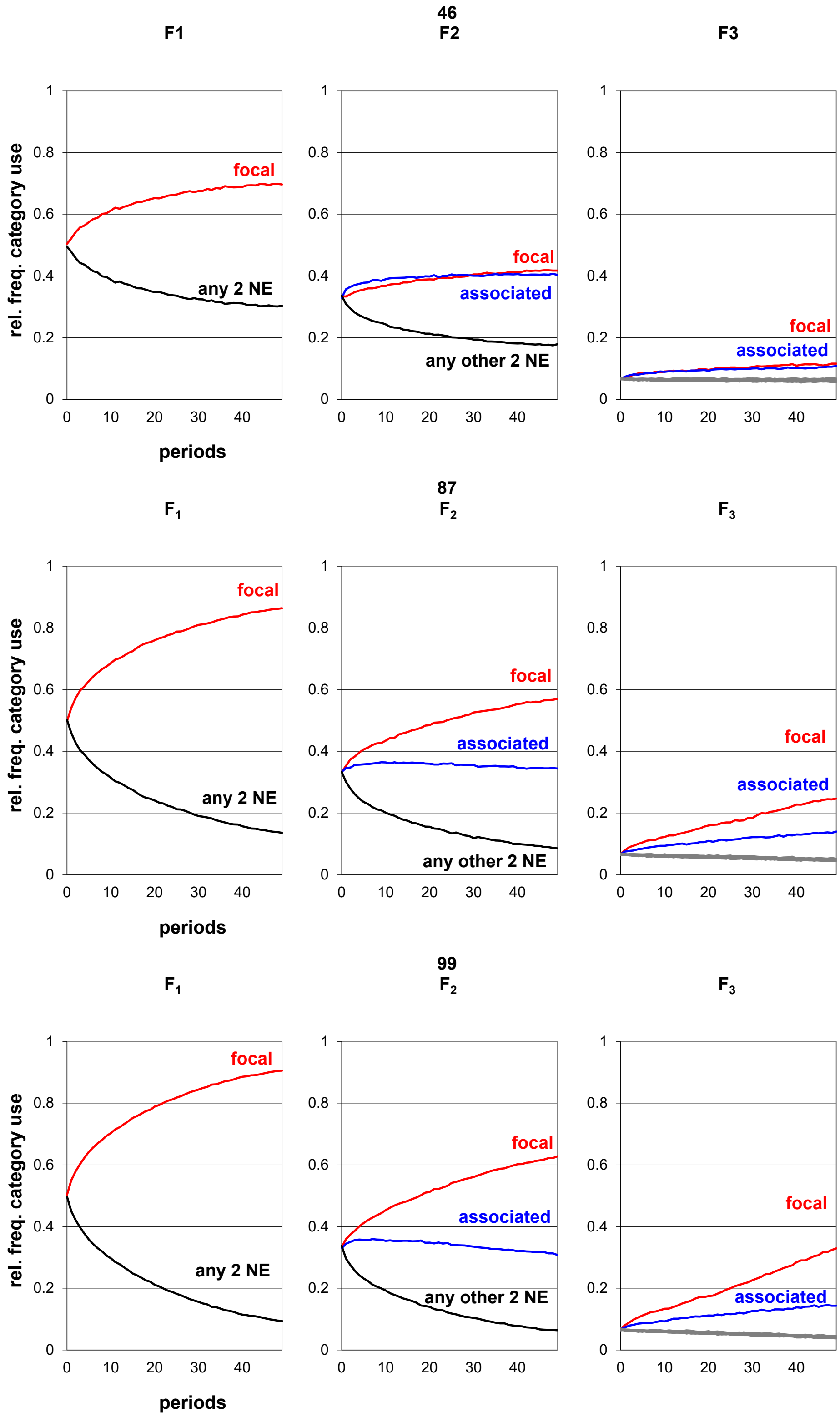


Figure 10: Learning model: relative frequency category use

F_3 to F_2 to F_1 to F_0), the longer it takes the agent to learn which category within this frame to use.[28] This suggests that when it comes to learning, there is some tension between using a more and a less detailed representation of the strategy set.

On the one hand, a coarser frame (with fewer categories) has the advantage that the agents can learn relatively quickly the value of each category within such a frame, and they can learn quickly on which category to coordinate. On the other hand, a coarser frame (with on average more options per category) may have the disadvantage that while coordinating on some category may be facilitated, there will be a good chance of miscoordination *within* such a category. Thus, from a learning perspective, the best views of the strategy set are characterized by a mixture of coarseness and fineness: few categories, of which some contain few actions (those on which to base the successful coordination), while the other categories contain many actions. It is this trade-off that gives frame F_1 a clear learning advantage, as it has only two categories, "focal" and "any with 2 NE". The former contains only the focal action, making coordination on it easy once this category is chosen, while the latter includes all other actions.

Thus, the analysis of the dynamics of our learning model suggest a rationale for some coarse framing of the strategy set, a rationale that may not be apparent in a static analysis.[29] It also helps us understand why even if a number of frame profiles involving F_3 give an expected payoff of 100 and are the Pareto dominant equilibrium selected according to Variable Frame Theory (see Figure 6), the use of F_3 decreases over time.

The dynamic analysis of the reinforcement learning model with frames shows that agents may not always learn to coordinate on a (Pareto-superior) NE and that agents may learn instead to coordinate on a non-equilibrium focal point (in line with the experimental findings). The model provides an explanation for this, as there is a trade-off between using a finer frame of the game that contains the Pareto-dominant equilibria and using a coarser frame of the game that helps agents to learn to coordinate faster. Agents may thus choose representations of the game (frame F_1) under which they cannot coordinate on the Pareto-superior equilibrium in expected payoffs. To what extent agents coordinate on the Pareto-superior equilibrium in expected payoffs and to what extent they coordinate on

[28]Note that this is true even though if a category is used by the agent it is reinforced under all frames under which it exists.

[29]Note that this rationale for using coarser frames does not depend on some inability to perceive a more detailed representation of the action set.

the focal point depends on the difference between the Pareto-superior NE payoff and the payoffs under the alternatives. Hence the difference in outcomes observed in the different treatments of the experiment can be understood as resulting from different incentives to learn to use a finer frame. The higher the focal payoff (which can be achieved also under F_1), the less of an incentive there is to learn to use the finer frame F_2.

5.3 Basic reinforcement learning without frames

The analysis above shows that the reinforcement learning model with frames, in which the agents learn which frame of the strategy set to use, fits the experimental data well in the sense that it is consistent with the stylized facts observed in the experiment. To provide some further perspective, in this section we will compare it with a closely related reinforcement learning model - without frames. That is, in this model the agents use reinforcement learning to respond to feedback about past performance as in our model, but they do not use frames, i.e. they do not categorize the action set. This allows us to assess how important the categorization of the action set is. Would a simpler reinforcement learning model without frames account for the experimental data equally well? Note that a significantly better performance of the model with frames would suggest that the idea of players categorizing the strategy set and learning which frame to use may be relevant as an explanation for the behavior observed in the experiment.

The key difference of this more basic reinforcement learning model with the model proposed in Section 3 is that here there are no frames and categories. That is, the agents have only one possible internal representation of the game, corresponding to the standard normal form of the underlying game, i.e. the payoff matrix in Figure 2. Otherwise the model corresponds to our model. Thus, each period two agents are randomly matched to play the game, each agent chooses simply an action, i.e. one of the fifteen possible rows (columns) using the logit rule, payoffs are realized, and the strength of each agent's action is updated with the payoff from the interaction.

In order to compare the two models we carry out a grid search for the parameters of the model without frames that best fit the experimental data. As before, the parameters are the initial strength s_0 (here of each action), the β (for choice between the fifteen actions), and the α (to reinforce the strength of each action). And just as before, we minimize the average mean squared error, taking into account the four types of payoffs realized in the experiment (see Figure 7) and those generated by the model. Again the grid search is run 10 times, each time taking the average over 10,000 runs of the model in order to clear out effects due to randomness.

For 6 out of the 10 times, the optimal α found is equal to 0, and for the remaining 4 times, the optimal β found is equal to 0. Note that these findings for the 10 runs are perfectly consistent as these parameter values imply persistent random behavior without any learning. An $\alpha = 0$ means that the strengths of the various choices are not updated on the basis of experience and remain equal to the initial strengths. Since these initial strengths were the same for all choices, the agents will continue to choose their actions uniform randomly. A $\beta = 0$ means that the agents do not respond to any differences in strength between the various choices and that they make their choices uniform randomly.

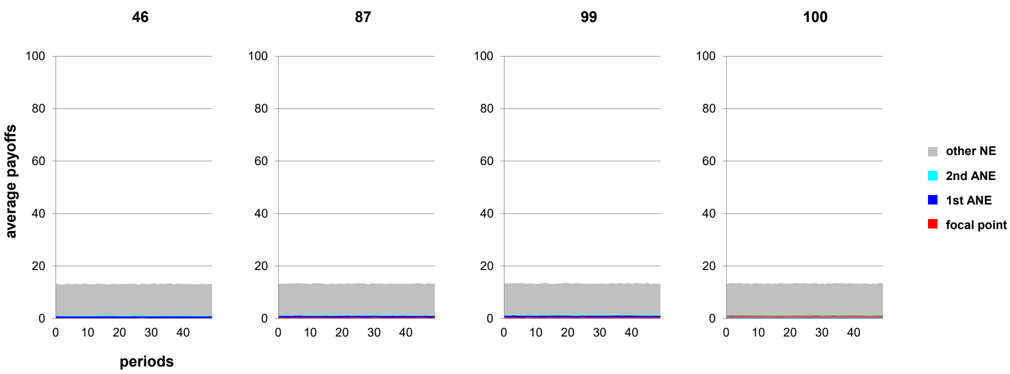


Figure 11: Learning model without frames: average payoffs per agent

This is illustrated in Figure 11, showing the average payoffs per agent per type of outcome for each period under the optimal parameter values for the model without frames, i.e. $s_0 = 0.8$, $\alpha = 0.0$, $\beta = 5$. As we can see, in each treatment, including the control treatment, the overall payoffs stay constant over the fifty periods at the level expected for random choices, and the payoffs realized through the focal point or the ANE stay constant at the level expected for random behavior as well, which means that almost all the payoffs are realized through the other NE.

In fact, none of the stylized facts observed in the experiment (as illustrated in Figure 7) is captured by this more basic reinforcement learning model in which the agents do not categorize the strategy set, whereas our model (see Figure 8) captured each of these stylized facts. Thus, the basic reinforcement learning model without frames cannot explain the experimental data any better than uniform random choice of the available actions.

This does *not* mean that agents in a basic reinforcement learning model without frames as such cannot learn to coordinate successfully in a game like the focal point game. This is illustrated in Figure 12. Instead of fitting the model to the BDV experiment distinguishing the four types of coordination outcomes as in Figure 7, we here take into account only the overall payoffs realized in each period. The optimal parameter values found are $s_0 = 0.0$, $\alpha = 0.1$ and $\beta = 50.0$. As we can see, the overall payoffs per agent increase over time in all treatments. Note, however, that although agents in this reinforcement learning model successfully learn to coordinate, none of the stylized facts observed in the experiment are explained by this model without frames. For example, the overall payoff differences between the treatments that we observed in the experiment cannot be explained by this reinforcement learning model, as here the payoff differences between the treatments become negligible, with even the control ("100") treatment showing the same increase in total payoffs as the focal point treatments. Moreover, with this learning model without frames the focal point and the ANE do not play any significant role, as almost all the payoffs are realized through the other NE.

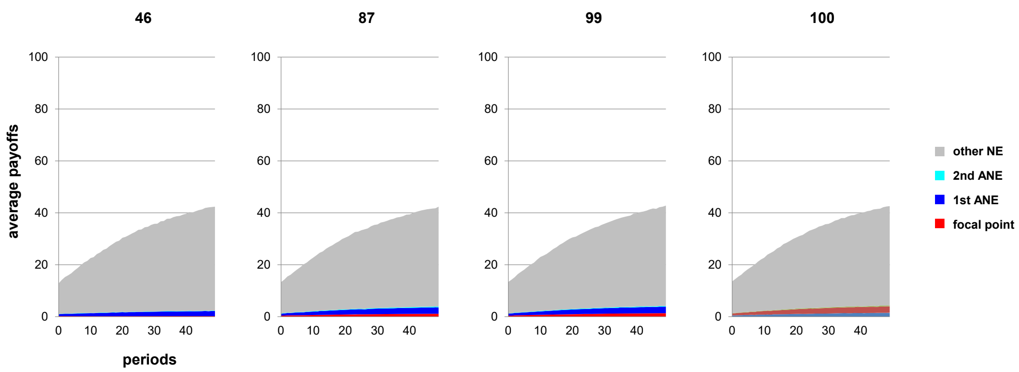


Figure 12: Learning model without frames: average payoffs per agent (fitting overall payoffs only)

6 Concluding remarks

Starting with the idea that often there may be many alternative ways to view a given strategic situation, this paper presents a model of agents who categorize their own actions. The agents learn which frames and categories of actions to use in the process of social interaction. The model is modest in terms of cognitive assumptions. It postulates that agents are more likely to choose frames and categories that have performed better in the past. As a case study we applied the model to improve our understanding of the dynamics of behavior in the non-equilibrium focal point game from the experiment by Bosch-Domènech and Vriend (2013).

The model suggests why different players may learn to use different views of the strategy set even in the long term, and why we may get outcomes that are neither Pareto efficient nor a Nash equilibrium. Even if agents have all possible frames available to them, they do not necessarily learn to use the finest view of the strategy set. If agents need to learn to coordinate, there may be some tension between fineness and coarseness of the frames they want to use. On the one hand, learning to coordinate is easier if there are fewer categories, which favors coarse frames. On the other hand, miscoordination within a category is more likely the more actions there are within a category and this may favor finer frames.

We conjecture that understanding how players learn which categorization of the strategy set to use may be useful to understand the dynamics of more games.

References

Nabil I. Al-Najjar and Mallesh M. Pai. Coarse decision making and overfitting. *Journal of Economic Theory*, 150:467–486, 2014. doi: https://doi.org/10.1016/j.jet.2013.12.003.

Larbi Alaoui and Antonio Penta. Endogenous depth of reasoning. *The Review of Economic Studies*, 83(4):1297–1333, 2015. doi: https://doi.org/10.1093/restud/rdv052.

Federica Alberti, Robert Sugden, and Kei Tsutsui. Salience as an emergent property. *Journal of Economic Behavior & Organization*, 82(2-3):379–394, 2012. doi: https://doi.org/10.1016/j.jebo.2011.10.016.

Ayala Arad and Ariel Rubinstein. Multi-dimensional iterative reasoning in action: the case of the colonel blotto game. *Journal of Economic Behavior & Organization*, 84(2):571–585, 2012. doi: https://doi.org/10.1016/j.jebo.2012.09.004.

Ayala Arad and Ariel Rubinstein. Multi-dimensional reasoning in games: framework, equilibrium and applications. *American Economic Journal: Microeconomics*, 11(3):285–318, 2019. doi: https://doi.org/10.1257/mic.20170322.

Jasmina Arifovic and John Ledyard. A behavioral model for mechanism design: individual evolutionary learning. *Journal of Economic Behavior & Organization*, 78(3):374–395, 2011. doi: https://doi.org/10.1016/j.jebo.2011.01.021.

Yaron Azrieli. Categorizing others in a large game. *Games and Economic Behavior*, 67(2):351–362, 2009. doi: https://doi.org/10.1016/j.geb.2009.01.004.

Yaron Azrieli. Categorization and correlation in a random-matching game. *Journal of Mathematical Economics*, 46(3):303–310, 2010. doi: https://doi.org/10.1016/j.jmateco.2009.11.015.

Michael Bacharach. Variable universe games. In Piero Tani Ken Binmore, Alan Kirman, editor, *Frontiers of Game Theory*, pages 255–275. Cambridge, Massachusetts and London, England: MIT Press, 1993.

Michael Bacharach and Michele Bernasconi. The variable frame theory of focal points: an experimental study. *Games and Economic Behavior*, 19(1):1–45, 1997. doi: https://doi.org/10.1006/game.1997.0546.

Michael Bacharach and Dale O. Stahl. Variable-frame level-n theory. *Games and Economic Behavior*, 32(2):220–246, 2000. doi: https://doi.org/10.1006/game.2000.0796.

Brent Berlin and Paul Kay. *Basic Color Terms: Their Universality and Evolution*. Berkeley and Los Angeles, California: University of California Press, 1991.

Ken Binmore and Larry Samuelson. The evolution of focal points. *Games and Economic Behavior*, 55(1):21–42, 2006. doi: https://doi.org/10.1016/j.geb.2005.03.005.

Tilman Börgers and Rajiv Sarin. Learning through reinforcement and replicator dynamics. *Journal of Economic Theory*, 77(1):1–14, 1997. doi: https://doi.org/10.1006/jeth.1997.2319.

Antoni Bosch-Domènech and Nicolaas J. Vriend. On the role of non-equilibrium focal points as coordination devices. *Journal of Economic Behavior & Organization*, 94:52–67, 2013. doi: https://doi.org/10.1016/j.jebo.2013.07.014.

Robert R. Bush and Frederick Mosteller. A mathematical model for simple learning. *Psychological Review*, 58(5):313–323, 1951. doi: https://psycnet.apa.org/doi/10.1037/h0054388.

Colin F. Camerer, Teck-Hua Ho, and Juin-Kuan Chong. A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3):861–898, 2004. doi: https://doi.org/10.1162/0033553041502225.

André Casajus. Focal points in framed strategic forms. *Games and Economic Behavior*, 32(2):263–291, 2000. doi: https://doi.org/10.1006/game.1999.0771.

André Casajus. *Focal Points in Framed Games: Breaking the Symmetry*, volume 499. Berlin Heidelberg: Springer-Verlag, 2001.

Miguel A. Costa-Gomes and Vincent P. Crawford. Cognition and behavior in two-person guessing games: an experimental study. *American Economic Review*, 96(5):1737–1768, 2006. doi: https://doi.org/10.1257/aer.96.5.1737.

Jan Salomon Cramer. *Logit Models from Economics and Other Fields*. Cambridge, England: Cambridge University Press, 2003.

Vincent P. Crawford and Hans Haller. Learning how to cooperate: optimal play in repeated coordination games. *Econometrica*, 58(3):571–595, 1990. doi: https://doi.org/10.2307/2938191.

Vessela Daskalova and Nicolaas J. Vriend. Categorization and coordination. *European Economic Review*, 129:103519, 2020. doi: https://doi.org/10.1016/j.euroecorev.2020.103519.

Ido Erev and Alvin E. Roth. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, 88(4):848–881, 1998.

Drew Fudenberg and David K. Levine. *The Theory of Learning in Games*, volume 2. Cambridge, Massachusetts and London, England: MIT Press, 1998.

David Gauthier. Coordination. *Dialogue: Canadian Philosophical Review/Revue canadienne de philosophie*, 14(2):195–221, 1975. doi: https://doi.org/10.1017/S0012217300043365.

Robert S. Gibbons, Marco LiCalzi, and Massimo Warglien. What Situation is This? Coarse Cognition and Behavior over a Space of Games. Working Paper 2017/09, Department of Management, Università Ca' Foscari Venezia, 2017.

Sanjeev Goyal and Maarten C.W. Janssen. Can we rationally learn to coordinate? *Theory and Decision*, 40:29–49, 1996. doi: https://doi.org/10.1007/bf00133159.

Veronika Grimm and Friederike Mengel. An experiment on learning in a multiple games environment. *Journal of Economic Theory*, 147(6):2220–2259, 2012. doi: https://doi.org/10.1016/j.jet.2012.05.011.

Yuval Heller and Eyal Winter. Rule rationality. *International Economic Review*, 57(3):997–1026, 2014. doi: https://doi.org/10.1111/iere.12183.

Steffen Huck, Philippe Jehiel, and Tom Rutter. Feedback spillover and analogy-based expectations: a multi-game experiment. *Games and Economic Behavior*, 71(2):351–365, 2011. doi: https://doi.org/10.1016/j.geb.2010.06.007.

Maarten C.W. Janssen. Rationalizing focal points. *Theory and Decision*, 50: 119–148, 2001. doi: https://doi.org/10.1023/A:1010349014718.

Philippe Jehiel. Analogy-based expectation equilibrium. *Journal of Economic Theory*, 123(2):81–104, 2005. doi: https://doi.org/10.1016/j.jet.2003.12.003.

Philippe Jehiel and Frédéric Koessler. Revisiting games of incomplete information with analogy-based expectations. *Games and Economic Behavior*, 62(2):533–557, 2008. doi: https://doi.org/10.1016/j.geb.2007.06.006.

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

Marco LiCalzi and Nadia Maagli. Bargaining over a common categorisation. *Synthese*, 193(3):705–723, 2016. doi: https://doi.org/10.1007/s11229-015-0790-5.

Marco LiCalzi and Roland Mühlenbernd. Categorization and cooperation across games. *Games*, 10(1):1–21, 2019. doi: https://doi.org/10.3390/g10010005.

Daniel L. McFadden. Conditional logit analysis of qualitative choice behavior. In Paul Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. New York: Academic Press, 1974.

Friederike Mengel. Learning across games. *Games and Economic Behavior*, 74 (2):601–619, 2012a. doi: https://doi.org/10.1016/j.geb.2011.08.020.

Friederike Mengel. On the evolution of coarse categories. *Journal of Theoretical Biology*, 307:117–124, 2012b. doi: https://doi.org/10.1016/j.jtbi.2012.05.016.

John Stuart Mill. *A System of Logic, Ratiocinative and Inductive*, volume 2. New York: Harper & Brothers, 1858.

Erik Mohlin. Optimal categorization. *Journal of Economic Theory*, 152:356–381, 2014. doi: https://doi.org/10.1016/j.jet.2014.03.007.

Rosemarie Nagel. Unraveling in guessing games: an experimental study. *American Economic Review*, 85(5):1313–1326, 1995.

William S. Neilson and Michael K. Price. Focal Strategies and Behavioral Backward Induction in Centipede Games. Manuscript, University of Tennessee, Knoxville, 2011.

Willard Van Orman Quine. Natural kinds. In Nicholas Rescher, editor, *Essays in Honor of Carl G. Hempel*, pages 5–23. Springer, 1969.

Julian Romero and Yaroslav Rosokha. A Model of Adaptive Reinforcement Learning. Working Paper March 11, 2019, SSRN, 2019.

Alvin E. Roth and Ido Erev. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8(1):164–212, 1995. doi: https://doi.org/10.1016/S0899-8256(05)80020-X.

Thomas C Schelling. *The Strategy of Conflict*. Cambridge, Massachusetts and London, England: Harvard University Press, 1980.

Dale O. Stahl and Paul W. Wilson. Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization*, 25(3):309–327, 1994. doi: https://doi.org/10.1016/0167-2681(94)90103-1.

Dale O. Stahl and Paul W. Wilson. On players' models of other players: theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254, 1995. doi: https://doi.org/10.1006/game.1995.1031.

Robert Sugden. A theory of focal points. *Economic Journal*, 105:533–550, 1995. doi: https://doi.org/10.2307/2235016.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. Cambridge, Massachusetts and London, England: MIT Press, 2018.

Klaus Wunderlich, Ulrik R. Beierholm, Peter Bossaerts, and John P. O'Doherty. The human prefrontal cortex mediates integration of potential causes behind observed outcomes. *Journal of neurophysiology*, 106(3):1558–1569, 2011. doi: https://doi.org/10.1152/jn.01051.2010.

H. Peyton Young. The evolution of conventions. *Econometrica*, 61(1):57–84, 1993. doi: https://doi.org/10.2307/2951778.

A Appendix: coordination success across treatments

In the main text we use the average payoffs by type of outcome in the different treatments as the leading indicator of coordination success. In this appendix we decompose the differences in average payoffs between the 46 and the 87 treatments, as well as between the 46 and the 99 treatments and between the 87 and the 99 treatments, by disentangling two effects on these average payoffs when considering different treatments. On the one hand, there are the differences in the magnitude of the focal payoff as such, and on the other hand there are the differences in behavior (actions chosen).

The decomposition works as follows. Take, e.g., the 46 and 87 treatments, compute for each the expected overall payoffs over all 50 periods, and take the difference.[30] To find what proportion of this difference is due simply to the change in the focal payoff from 46 to 87, we *fix* the behavior (i.e. frequency of choices of all actions) in the 46 treatment and then simply re-compute the

[30]This analysis is robust to considering different periods instead.

	BDV experiment		**Learning model**	
comparison of treatments	due to magnitude of focal payoff	due to difference in behavior	due to magnitude of focal payoff	due to difference in behavior
46 and 87	20%	80%	29%	71%
87 and 99	28%	72%	36%	64%
46 and 99	18%	82%	26%	74%

Table 1: Decomposition of differences in average payoffs

expected payoffs using a payoff of 87 instead of 46 for successful coordination on the focal point. Any remaining difference between the expected payoffs in the 46 and 87 treatments is due to differences in behavior. The analysis to decompose the difference in payoffs between the 46 and the 99 treatments, and between the 87 and the 99 treatments is analogous.

Table 1 presents this decomposition of the differences in the average payoffs between the treatments in the experiment, and similarly for the learning model. Starting with the experiment, as the table shows, the increases in average payoffs when the focal payoff is increased are predominantly due to differences in behavior. Depending on the exact focal payoff increase considered, 72% to 82% of the increase in average payoffs is due to differences in the behavior, and only 18% to 28% is the direct result of the increase of the focal payoffs as such. This is matched by the learning model as well. As the decomposition of average payoff differences for the model in Table 1 shows, the differences in individual behavior are again most important, accounting for 64% to 74% of the increase in expected payoffs, with only 26% to 36% the direct result of the increases of the magnitude of the focal payoff as such. Note that the pattern of the relative importance of differences in the magnitude of the focal payoffs as such and of differences in behavior is the same for the experimental data and for the learning model, with the largest increase of the focal payoff (from 46 to 99) making the smallest contribution when it comes to explaining the average expected payoffs (18% and 26% for BDV experiment and learning model respectively).

Next, while in the main text we focus on the average expected payoffs to measure the *weighted* relative frequencies of coordination, in this appendix we show in Figure 13 the *unweighted* frequencies of coordination by type of outcome and by treatment for the BDV experiment and for the learning model, as discussed in Subsections 4.3 and 5.2 respectively.

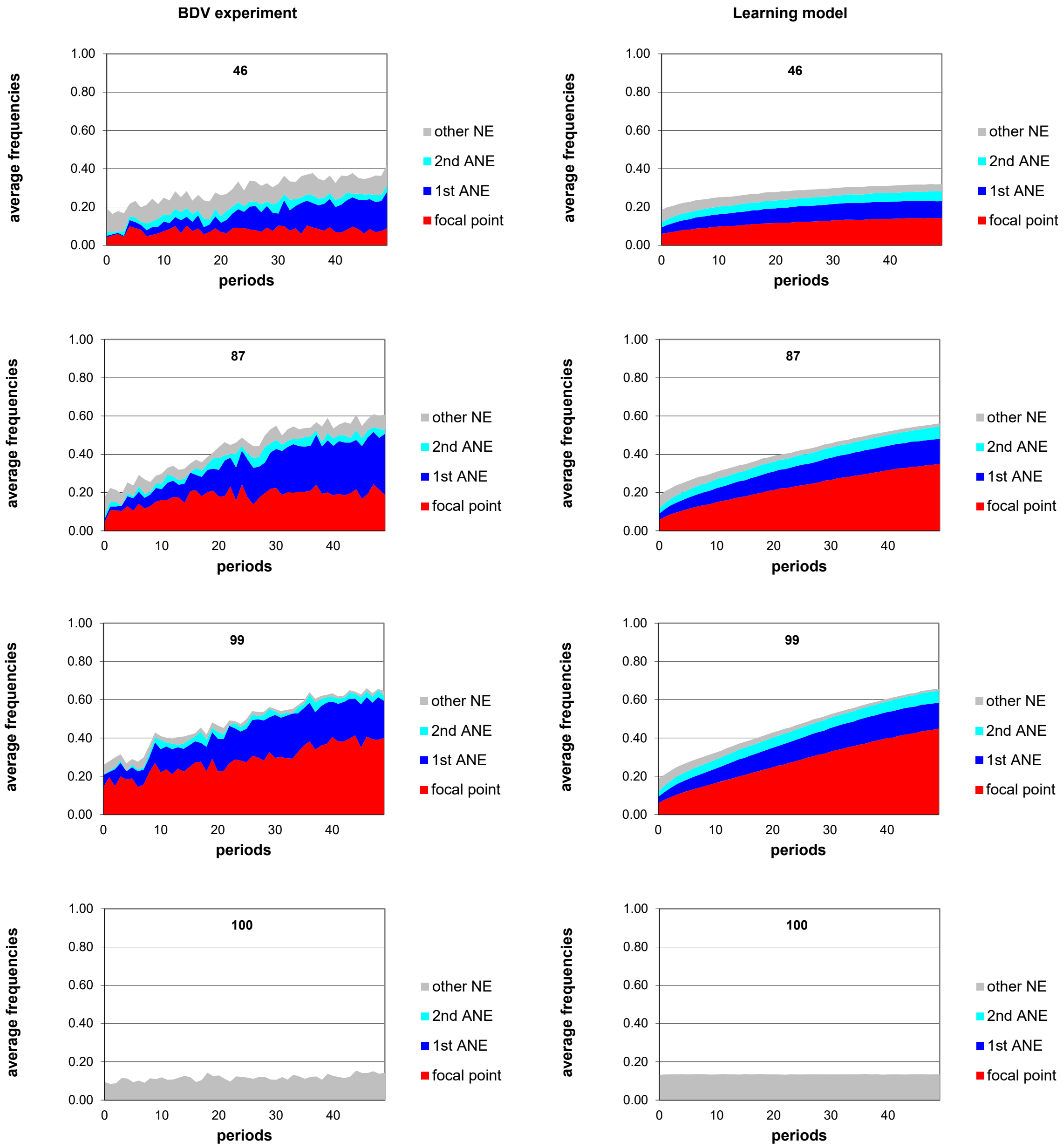


Figure 13: BDV experiment (left) and Learning model (right): Unweighted frequencies of coordination by type of outcome and by treatment

School of Economics and Finance

This working paper has been produced by the School of Economics and Finance at Queen Mary University of London

School of Economics and Finance Queen Mary University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 7356
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/research/workingpapers/