Information-Constrained Coordination of Economic Behavior[*]

Guy Aridor[†] Rava Azeredo da Silveira[‡] Michael Woodford[§]

January 15, 2024

Abstract

We analyze a coordination game with information-constrained players. The players' actions are based on a noisy compressed representation of the game's payoffs in a particular case, where the compressed representation is a latent state learned by a variational autoencoder (VAE). Our generalized VAE is optimized to trade off the average payoff obtained over a distribution of possible games against a measure of the congruence between the agent's internal model and the statistics of its environment. We apply our model to the coordination game in the experiment of Frydman and Nunnari (2023), and show that it offers an explanation for two salient features of the experimental evidence: both the relatively continuous variation in the players' action probabilities with changes in the game payoffs, and the dependence of the degree of stochasticity of players' choices on the range of game payoffs encountered on different trials. Our approach also provides an account of the way in which play should gradually adjust to a change in the distribution of game payoffs that are encountered, offering an explanation for the history-dependent play documented by Arifovic et al. (2013).

[*] We thank Francesco Grechi for help exploring the numerical properties of VAE models, Herbert Dawid, Evan Friedman, Cary Frydman, Randy Gallistel, Cars Hommes, Tyler Malloy, Chris R. Sims, Jakub Steiner, Mycal Tucker and participants at the Columbia Cognition and Decision Lab Workshop, the Workshop on Information-Theoretic Principles in Cognitive Systems (NeurIPS 2022), the Computational and Experimental Economics Workshop (Simon Fraser University, February 2023), the Summer School on Imperfect Cognition and Economic Behaviour (Hebrew University, July 2023), and the Workshop on Learning and Bounded Rationality (Stony Brook University, July 2023) for helpful discussions, and the Alfred P. Sloan, Jr. Foundation for research support. Silveira also acknowledges the support of CNRS through UMR8023. We are also grateful to Ambreen Chaudhri, Tianhao Liu, and Will Thompson for research assistance, and especially grateful to Cary Frydman and Salvo Nunnari for sharing their data.

[†] Northwestern University, Kellogg School of Management. Email: guy.aridor@kellogg.northwestern.edu.

[‡] ENS Paris and University of Basel. Email: rava@iob.ch.

[§] Columbia University. Email: michael.woodford@columbia.edu.

1 Introduction

An ubiquitous feature of economic decisions observed in the laboratory is that people's decisions appear to be random, as a function of the objective data defining the choice problem. The same experimental subject, confronted again with exactly the same decision, will often not make the same choice as on previous occasions. Thus one must study, not what a given person will *always* choose when presented with options with particular characteristics, but the *probabilities* with which they will make alternative choices. A model of the subject's choice behavior should seek to explain how these probabilities vary with the characteristics of the options presented (Woodford (2020)).

A popular model of imprecise decision making is the rational inattention (RI) model of Sims (2003). According to this theory, decisions are based on an imprecise awareness of the specific situation in which the decision maker acts. The action is optimal (maximizes the decision maker's expected payoff, under some prior distribution over possible situations), subject to its having to be based only on a compressed representation of the current state, rather than an exact description of the state; and the compressed representation itself is also optimal, subject only to a limit on how informative it can be about the true state. Whenever there is a positive cost of information, RI implies that an optimal compressed representation necessarily involves randomization; it thus provides an explanation for the observed stochasticity of choice. RI is also a highly parsimonious theory; if one assumes, with Sims, that the cost of more informative representation should be proportional to the Shannon mutual information between the external state and its internal representation, the model has only a single free parameter (the cost per additional bit of information), and makes correspondingly sharp predictions.

But the theory as formulated by Sims leaves open the question of how the optimal compressed representation, and the associated optimal rule for action choice based on the compression representation, are supposed to be learned for a given environment (defined by a prior distribution over possible decision problems). RI posits that the action associated with any latent state is the one that maximizes the decision maker's expected reward under the posterior distribution over possible external states associated with that latent state; but this raises the further question of how these posteriors are learned.

We propose an alternative model of imprecise decision making with an explicit (and computationally tractable) account both of how compressed internal representations are learned, and of how a distribution over possible external states for each latent state (to be used in action selection) is learned. Our model is based on a popular architecture from the machine learning literature — the "variational autoencoder" (VAE) introduced by Kingma and Welling (2014) — with the criterion used to train the model modified to take account of the usefulness of the internal representations for making better decisions (as in RI). A VAE consists of two statistical models: a *recognition model* that assigns to any external state x a latent state j from some set J of possible representations, and a *generative model* that associates with each latent state j a particular probability distribution of possible external states, which distribution is used (like the posterior distribution in RI) to interpret what the internal representation tells one about the external state. Each of these models is assumed to belong a parametric family of possible statistical models (frequently, a neural network model with some number of tunable connection weights), and the parameters of the models are iteratively adjusted so as to minimize an objective function that is evaluated using the finite sample of values $\{x_i\}$ for the external state in some training dataset (Kingma et al. (2019)).

In the classic VAE formulation, the training objective seeks to achieve as great as possible a congruence between the joint distribution over latent states and external states predicted by the generative model and the joint distribution that results from using the recognition model to classify the external states occurring in the training database. Because both models must be selected from particular parametric families, perfect congruence will generally not be possible. (Perfect congruence is instead assumed in RI, insofar as the interpretation of the latent states is assumed to be based on exact Bayesian inference.) It would not make sense, however, to demand perfect congruence when the training dataset is itself only a finite sample from the prior distribution, and thus only an approximation to the true distribution; contenting oneself with only finding the best fit within a parametric family of models allows one to avoid over-fitting to the particular sample values in the training dataset. The fact that we consider only a finitely-parameterized class of possible models also makes it possible for all of the parameters to be learned from a finite training dataset.

In the applications that motivated this proposal, the goal is simply to be able to reconstruct the external state as accurately as possible on the basis of the compressed internal representation. One might, however, use the interpretation of the latent states provided by the generative model to select other kinds of responses, the payoff to which depends on the degree of suitability of the action a that is selected for the external state x. To deal with such cases, we add to the usual VAE training objective a second term, the average reward obtained from actions chosen on the basis of the compressed representation, with a weight that reflects the relative importance assigned to the two considerations in adapting one's internal model to one's environment.[1] In general, we find that there is a non-trivial trade-off between the two subgoals represented by the two terms in our training objective: the compressed representations that will allow the greatest degree of congruence between the two models are not generally exactly the same as those that would allow the most accurate state-contingent action selection.

This tradeoff then implies that our cognitive model, even when fully trained on the basis of an arbitrarily large training dataset, will often imply imprecise (and more specifically, random) deci-

[1] Other recent proposals that generalize the VAE architecture to include the rewards achieved from action selection in the objective used for training include Malloy et al. (2022) and Tucker et al. (2022).

sion making. Even when a precise classification of states, of the kind needed for perfectly accurate state-contingent action selection, is allowed by the parametric family of recognition models that is considered, an imprecise classification (implying correspondingly imprecise action selection) may be selected by the training objective because it allows greater congruence between the recognition model and the generative model. The preference for congruence between the two models can thus serve as a substitute for the information cost function in RI theory.

Our alternative model has advantages over RI theory as an account of how imprecise internal representations are endogenously adapted to the statistics of a particular environment, that begin with its greater computational tractability and the fact that it comes with a model of how the internal representations are learned. But there are other advantages as well, that are well illustrated by an application of the model to observed behavior in an experimental coordination game. The coordination game that we consider is a two-player, one-shot simultaneous-move game studied experimentally by Frydman and Nunnari (2023); it represents a stylized version of the kind of strategic interaction that arises in the case of a bank run or a speculative attack on a currency peg. The payoff matrix in the game depends on a state variable x that varies from trial to trial in the experiment (an independent draw on each trial from some prior distribution), but is revealed to the players before they choose their actions.

Players' action choices are observed to be random in the case of a given value of x, but with their action probabilities varying systematically with x: the probability of a "run" occurring (or of an "attack" on the currency peg) is steadily increasing as "economic fundamentals" deteriorate. RI theory can explain the randomness of behavior in what would seem to be a game with full information (and that does not vary with x in the way that a mixed-strategy Nash equilibrium would). However, application of RI theory to the game, as in Yang (2015), implies that action probabilities should jump discontinuously at one or more particular values of x,[2] while in experimental data the probabilities appear to change only gradually with progress changes in x. And RI theory implies that varying the dispersion of the values of x in the prior distribution should not change the equilibrium choice probability associated with a given value of x, while Frydman and Nunnari (2023) show that drawing values of x from a more dispersed distribution makes choice probabilities change more gradually with increases in x.

Our model instead predicts that the choice probability should be a continuous function of x, and that it should be a more gradually decreasing function in the case of the more dispersed prior, in conformity with the experimental evidence.[3] We believe that these are additional reasons for

[2] See section A of the Appendix for discussion of why this is the case.

[3] Frydman and Nunnari (2023) also present a model of endogenous variation in the precision of a noisy internal representation of the state that is consistent with both of these aspects of their data, based on a model of efficient coding. Our interpretation of the experimental data is broadly consistent with theirs, but we optimize over a more flexibly specified family of possible internal representations. In particular, their formulation requires the precision of

interest in the particular way that we propose to model information-constrained decision making.

2 Modeling Imprecise Action Selection

In this section we provide the details of our computational model of imprecise action selection. We first provide an overview of our approach in the context of an individual decision problem in which a decision maker (DM) must select an action a from some finite set A of possible actions, and the payoff from the action chosen depends on a state x, which can take any of a continuum of values, and to which the DM may be imperfectly attentive. We then extend the model to deal with the case of strategic interaction between two players who choose their actions simultaneously, as in the coordination game treated in the next section.

2.1 A VAE Model of Individual Decision Making

We consider a decision problem in which an external state x is drawn independently on each occasion (each experimental trial) from a continuous frequency distribution $\pi(x)$. On a given occasion, the current state x is encoded as one of a finite set $\mathcal{J}$ of possible latent states; the bound on the number of possible latent states is taken as a constraint. The VAE structure (Kingma and Welling (2014), Kingma et al. (2019)) consists of a *recognition model*, or encoding model, that specifies the conditional probabilities $p(j\,|x)$ of assigning any state x in the support of the prior to a given latent state j, and a *generative model*, or decoding model, that is used to interpret the meaning of a given latent state in terms of a distribution of possible external states that may have given rise to it.

The recognition model is specified by conditional probabilities $p_\phi(j|x)$, defined for each possible external state x. Here ϕ is a finite-dimensional vector of parameters, indicating which member of some parametric family of possible encoding rules is used; the recognition model is optimized only within this family. The encoding rule, combined with the environmental distribution $\pi(x)$, implies a joint distribution for external states and their labels given by

$$p_\phi(j, x) \;=\; \pi(x) \cdot p_\phi(j\,|x).$$

The generative model also specifies a joint distribution for the labels and states, parameterized by a finite-dimensional vector of parameters θ. It consists of learned frequencies of occurrence $\{q_\theta(j)\}$

encoding to be the same for all values of x, so that the kind of internal representation required for perfectly accurate choice is precluded as infeasible. We instead allow encoding rules that can make arbitrarily sharp distinctions between values of x above and below a particular threshold, and hence that can approximate arbitrarily well the distinction between high-x and low-x states that is required for perfectly accurate choice.

for each of the latent categories, and a learned distribution of external states $\tilde{p}_\theta(x\,|j)$ for each of the categories. The implied generative model for the joint distribution of external states and their labels is then given by

$$\tilde{p}_\theta(j,x) \;=\; q_\theta(j) \cdot \tilde{p}_\theta(x\,|j).$$

The process by which a system of compressed internal representation is learned involves adjusting the parameters ϕ and θ of the two models so as to achieve as close as possible a degree of congruence between these two different joint distributions of labels and states. Given exposure to a body of experience of values for the external state $\{x_i\}$, the recognition model assigns labels to them (generally in a probabilistic way), resulting in a database of labeled observations $\{x_i, j_i\}$ that can be used to "train" the generative model. Here the goal is to find parameter values θ for which the generative model provides a good description of the database of labeled observations. At the same time, sampling values from the joint distribution of labels and states specified by the generative model can produce a database of simulated observations that can be used to "train" the recognition model. In this case the goal is to find parameter values ϕ so that any observation x will be mapped to a frequency distribution of labels j similar to the frequency with which states near x are assigned the label j in data sampled from the generative model. Thus the parameters (ϕ, θ) can be jointly learned through an iterative updating process (discussed in section 7), in which the parameters θ are adjusted so as to better fit the joint distribution generated on the basis of external experience and the current parameter values ϕ; the parameters ϕ are then adjusted so as to better fit the joint distribution generated on the basis of the current parameter values θ; and so on until both parameter vectors converge.

In typical applications of VAE modeling, the model once trained is used to classify new observations, assigning to any observation x a compressed representation j; this then allows reconstruction of the external state (generation of an estimated value $\hat{x}$) on the basis of the compressed representation, by sampling from the distribution $\tilde{p}_\theta(x\,|j)$ provided by the generative model. We can however use the same architecture to model a decision maker that must choose an action a from a set A that need not correspond to different possible values of the external state. We do this by adding to the VAE structure a decision rule that specifies an action $a(j) \in A$ for each of the latent states j.[4] The decision rule is learned on the basis of simulations of the generative model; that is, for each latent state j, the system learns an action $a(j)$ that should be desirable if the external state is drawn from the distribution $\tilde{p}_\theta(x\,|j)$. Thus in our case, as in more standard applications of

[4] More generally, one might suppose that the decision rule specifies a probability distribution over A for each latent state j. But in the kind of problems with which we are concerned here, there will be no advantage for a DM in randomizing their response conditional on a given latent state, and we simplify both our notation and our discussion of learning by supposing that the algorithm only considers deterministic decision rules. The randomness of experimental subjects' observed responses conditional on the external state x will then have to be attributed entirely to randomness of the classification of states by the recognition model, as in RI analyses.

VAEs, the latent state is "decoded" using the generative model, treating the distribution $\tilde{p}_\theta(x\,|j)$ as a sort of "posterior belief" when the external state is internally represented by the latent state x.

2.2 Alternative Training Objectives

As just explained, a key idea is that the parameters ϕ and θ should both be adjusted to achieve as great as possible a degree of congruence between the joint distributions of labels and states implied by the recognition model and generative model respectively. A natural approach would be to follow Kingma and Welling (2014); Kingma et al. (2019) and suppose that ϕ and θ are jointly optimized so as to minimize the Kullback-Leibler divergence of the joint distribution implied by the encoder relative to that implied by the decoder, $D_{KL}(p_\phi(j,x)||\tilde{p}_\theta(j,x))$.

Note that as a mathematical identity we can write

$$D_{KL}(p_\phi||\tilde{p}_\theta) \;=\; D_{KL}(q_\phi||q_\theta) + \sum_j q_\phi(j) D_{KL}(\pi_\phi(\cdot|j)\,||\,\tilde{p}_\theta(\cdot|j)), \qquad (1)$$

where $q_\phi(j)$ is the frequency with which the label j occurs in the joint distribution p_ϕ,

$$q_\phi(j) \;\equiv\; \int p_\phi(j,x)\,dx, \qquad (2)$$

and $\pi_\phi(x\,|j)$ is the posterior probability that the state is x when the label is j, again as implied by the joint distribution p_ϕ. Thus if the parametric family of possible generative models $\tilde{p}_\theta$ is flexible enough to include this possibility, $D_{KL}(p_\phi||\tilde{p}_\theta)$ is minimized by choosing $q_\theta = q_\phi$ and $\tilde{p}_\theta(\cdot|j) = \pi_\phi(\cdot|j)$ for each j. Hence even when the family of possible generative models does not allow precisely this solution, choosing θ so as to minimize $D_{KL}(p_\phi||\tilde{p}_\theta)$ can be regarded as a variational approximation to the interpretation of the latent states that would be given by exact Bayesian inference, using correct knowledge of both the distribution from which the external states are drawn and the probabilistic classifications produced by the recognition model ϕ.

Moreover, if we use a finite sample from p_ϕ as the "empirical distribution" p^{emp} to which we wish to make $\tilde{p}_\theta$ congruent, then choosing θ to minimize $D_{KL}(p^{emp}||\tilde{p}_\theta)$ is equivalent to choosing the parameters θ to maximize the likelihood (under the generative model) of the sample. Thus choosing θ to minimize this objective amounts to maximum likelihood estimation of the parameters of the generative model.

Similarly, suppose that we fix the generative model θ, and for any recognition model ϕ, define the joint distribution

$$\hat{p}_\phi(j,x) \;=\; \pi_\theta(x)\cdot p_\phi(j\,|x),$$

where $\pi_\theta(x)$ is the marginal distribution for x implied by the generative model. Then it is also a mathematical identity that

$$D_{KL}(\hat{p}_\phi || \tilde{p}_\theta) \;=\; \mathrm{E}_\theta[D_{KL}(p_\phi(\cdot|x) \,||\, \tilde{p}_\theta(\cdot|x))],$$

where $\tilde{p}_\theta(j|x)$ is the conditional probability of the label being j when the state is x, according to the generative model, and $\mathrm{E}_\theta[\cdot]$ denotes an expectation over values of x drawn from the distribution π_θ. It follows that if the state x is drawn from a distribution that is correctly described by the generative model, and the family of recognition models considered is flexible enough to allow this, $D_{KL}(p_\phi || \tilde{p}_\theta)$ will be minimized by choosing the encoding rule to be $p_\phi(j|x) = p_\theta(j|x)$ for each state x. Thus the proposed learning process would lead to an encoding rule that classifies a state x on the basis of Bayesian inference about the latent state j that has given rise to it, under an assumption that the generative model correctly describes the process through which states x arise (so that there can be thought to be a "true" latent state to infer from the observation of x), as in Helmholtz's theory of perceptual judgments. Even when the family of possible encoding models does not allow precisely this solution, choosing ϕ so as to minimize $D_{KL}(p_\phi || \tilde{p}_\theta)$ can be viewed as a variational approximation to a model of perceptual classification of this kind.

But as noted by Bowman et al. (2016); Chen et al. (2017), training the VAE in this way ensures that the generative model will provide a reasonable approximation to the environmental distribution $\pi(x)$, but does not necessarily lead to a meaningful latent representation. Indeed, in the example that we present below, training our model to minimize $D_{KL}(p_\phi || \tilde{p}_\theta)$ leads to latent states that are completely uninformative about the external state x. A cognitive model of this kind is clearly not useful as a basis for action choice. Much of the recent VAE machine learning literature follows Alemi et al. (2018), who propose extending the objective function used in Kingma and Welling (2014); Kingma et al. (2019) to explicitly incentivize the model to learn a more meaningful representation. Their "β-VAE" approach allows for more "disentangled" representations by providing an additional bonus for classification schemes in which the different categories are more informative about the underlying stimuli (the objective proposed in "infomax" theories of efficient coding).

For applications of the kind that we consider here, however, it is more natural to provide a bonus not simply for encoding rules that differentiate between different external states, regardless of whether the particular external states that they distinguish are ones that the DM needs to tell apart in order to make good decisions, but instead to provide a bonus for encoding rules that support more efficient action selection. Suppose that the DM suffers a loss $\mathcal{L}(a; x)$ from choosing action a when the external state is x.[5] Then for any recognition model ϕ and generative model θ,

[5] Here we specify the DM's problem in terms of minimization of expected loss rather than maximization of expected reward, to conform to the typical specification of the training objective in the VAE literature as a criterion to be minimized.

we can compute

$$L \;\equiv\; \mathrm{E}[\mathcal{L}(a(j);\, x)], \tag{3}$$

the expected loss when decisions are made using this VAE and an arbitrary decision rule $a(j)$. Here $\mathrm{E}[\cdot]$ refers to an expectation over the values of (j, x) when x is drawn from the environment and j is assigned by the recognition model.

Our proposed model of information-constrained choice assumes that the parameters of the VAE are adjusted so as to solve the problem

$$\min_{\phi,\theta} D_{KL}(p_\phi || \tilde{p}_\theta) \;+\; \beta L, \tag{4}$$

where L is the expected loss measure defined in (3) and $\beta > 0$ indicates the relative weight placed on the two desiderata in the training objective. A key goal of our analysis is to characterize the trade-off between the two alternative objectives reflected in (4), and hence the way in which predicted behavior varies depending on the size of β. Here we have defined the problem of optimally training the VAE, for a given decision rule $a(j)$, but we can also optimize the decision rule so as to achieve the lowest possible value of the training objective (4).

While our primary interest in decision processes that minimize (4) for one value of β or another, it may also be of interest to consider a more general form of training objective. Alemi et al. (2018) note that the traditional VAE training objective can be expressed as

$$D_{KL}(p_\phi || \tilde{p}_\theta) \;=\; -H \,+\, D \,+\, R,$$

where

$$H \;\equiv\; -\mathrm{E}[\log \pi(x)]$$

is the entropy of the prior distribution for x,

$$D \;\equiv\; -\mathrm{E}\Big[\sum_j p_\phi(j|x) \log \tilde{p}_\theta(x|j)\Big]$$

is a cross-entropy measure of the average distortion involved in decoding the latent states using the generative model θ, and

$$R \;\equiv\; \mathrm{E}[D_{KL}(p_\phi(\cdot|x) || q_\theta)] \tag{5}$$

is the rate at which information must be transmitted over a channel that takes x as an input and yields a random classification j as its output, when the operation of the channel is optimized for a frequency distribution q_θ of occurrence of the different output signals. Since H is independent of the VAE parameters, the traditional VAE objective is equivalent to minimizing the value of $D + R$.

Alemi et al. (2018) propose that, rather than requiring that equal weight be put on the minimization of the terms D and R, as in the objective proposed by Kingma and Welling (2014), one can train a VAE by minimizing a generalized objective $D + \beta R$, where the positive weight β need not equal 1. (This is their "β-VAE" model.) Regardless of the value of $\beta > 0$, the probabilities q_θ that minimize this objective will be given by $q_\theta = q_\phi$, in which case R is equal to the Shannon mutual information between the external state x and the internal representation j. A value $\beta < 1$ thus corresponds to assigning a bonus to internal representations that are more informative, for any given value of $D_{KL}(p_\phi || \tilde{p}_\theta)$.

We can allow for this concern as well, by letting the parameters of the VAE be adjusted so as to solve a more general problem of the form

$$\min_{\phi,\theta} D + \beta_1 R + \beta_2 L, \tag{6}$$

where β_1, β_2 are both positive coefficients. This family of models nests the "β-VAE" of Alemi et al. (2018) as a limiting case (the one in which $\beta_2 = 0$). If we instead write (6) in the form

$$\min_{\phi,\theta} D_{KL}(p_\phi || \tilde{p}_\theta) + (\beta_1 - 1)R + \beta_2 L, \tag{7}$$

we see that our basic model is another special case of it (the one in which $\beta_1 = 1$). An assumption that $\beta_1 < 1$ implies a preference (other things being equal) for more informative (more "disentangled") internal representations, while an assumption that $\beta_1 > 1$ would instead imply an aversion (other things being equal) to more informative (more "complex") internal representations. While the β-VAE literature argues for specifications with $\beta_1 < 1$, RI models assume that $\beta_1 > 1$.[6]

RI theory assumes that $\beta_1 > 1$ because (under further assumptions of those models) it is only in this case that the theory implies that decision making should be at all imprecise. We instead show that our model can predict stochastic action selection even with $\beta_1 \leq 1$ (see section 6.1 below), if the generative model must be selected from a restricted family. The β-VAE literature argues for assuming that $\beta_1 < 1$ because (in some cases) $\beta_1 \geq 1$ implies that the optimal structure will involve a completely uninformative latent state. This is indeed true in our model (see section 5.3 below), if we assume that $\beta_2 = 0$, as in the model of Alemi et al. (2018); but we show that the optimal latent state in our model is informative, even when $\beta_1 \geq 1$, when we assume that $\beta_2 > 0$. Thus while the most general formulation of our model allows us to consider values of β_1 either larger or smaller than 1, assuming a value different from 1 is not important for any of our qualitative conclusions.

[6] See Appendix section A for further discussion.

2.3 Extension to a Setting with Strategic Interaction

Suppose now that multiple information-constrained DMs each choose actions simultaneously, and that each one's reward will depend on the actions chosen by all. To simplify notation, we consider the case of a two-person game (as in the application below). Suppose that a player who chooses action a will suffer a loss $\mathcal{L}(a, a'; x)$ if the other player chooses action a' and the external state is x. We can again suppose that a player's action must be based on their internal representation j of the external state. This internal representation is generated by a recognition model, and decoded using a generative model, just as in the individual decision problem discussed above.

We again consider the problem of optimally training the players' VAEs, taking as given their respective decision rules $a(j)$ and $a'(j')$. The congruence measure $D_{KL}(p_\phi || \tilde{p}_\theta)$ in each player's training objective again depends only on the player's own VAE parameters (ϕ, θ), and the prior distribution $\pi(x)$ for the external state. However, the loss measure L for a given player now depends not only on their own recognition model (and their decision rule), but also on the other player's recognition model (and that player's decision rule). The definition (3) must now be generalized to state that

$$L \;\equiv\; \mathrm{E}[\mathcal{L}(a(j), a'(j');\, x)], \tag{8}$$

where $\mathrm{E}[\cdot]$ now refers to an expectation over the values of (j, j', x) when x is drawn from the prior, j is assigned by the player's own recognition model ϕ, and j' is assigned by the opponent's recognition model ϕ'. Each player's VAE is trained to minimize the objective (4) as above, but now using the definition (8) for the expected loss term.

Under this proposal, each player has a VAE of their own, the parameters of which are trained based on data (observations of the external state) generated by the environment, but also based on the observed actions of the other player — which depend on the other player's recognition model. Thus the two VAEs are coupled systems, each producing data that are used to train the other one. The state of belief to which the process converges will thus represent a sort of information-constrained Nash equilibrium.

As in the case of the individual decision problem, we can generalize the model by assuming that each player's VAE is adjusted to solve the problem (7) rather than (4). In this case, there are two parameters β_1, β_2 to specify the relative weights placed on the alternative stabilization objectives.

3 A Coordination Game

Here we apply our method to a coordination game studied experimentally by Frydman and Nunnari (2023). Two players each choose whether to "stay" or "leave." By leaving, either player can guarantee for themselves an amount that is independent of what the other does; if instead they stay, they obtain a payoff that is larger if the other player also chooses to stay. This captures the essential strategic logic of situations like bank runs or a speculative attack on a currency peg.[7] Games of this kind have been extensively studied, both theoretically (e.g., Morris and Shin (2003), Yang (2015)) and in laboratory experiments (e.g., Heinemann et al. (2004), Heinemann et al. (2009), Arifovic et al. (2013), Arifovic and Jiang (2019)). We begin by reviewing the structure of the game, and then describe how our approach can be applied to it.

3.1 The Game and its Symmetries

The payoffs in the game considered by Frydman and Nunnari (2023) are of the form shown in Table 1.[8] Here the payoffs a and b that are obtained by a player that chooses to stay satisfy $b > a$, and are the same on every trial; the parameter θ (the value of the outside option) instead varies randomly from trial to trial. Because a and b are always the same, a cognitive system that is optimized for the prior distribution over possible games that can be faced in the experiment will be optimized for these particular values (while it must instead allow for random variation in θ); and here we simplify our discussion of adaptation by supposing that the values of a and b are known precisely, rather than having to be learned.[9]

Table 1: *Coordination Game*

	Leave	Stay
Leave	$\theta,\ \theta$	$\theta,\ a$
Stay	$a,\ \theta$	$b,\ b$

Assuming that the DM's objective is to maximize their expected payoff, the strategic considerations in a game of this kind are unchanged if we re-scale payoffs using any monotonically increasing affine transformation.[10] We thus use normalized payoffs from here on in our analysis, so that our numerical results are equally applicable to any game of this kind, regardless of the

[7] Models of these phenomena generally assume a large number of players who make simultaneous decisions, perhaps even a continuum, with the payoff of an individual player depending both on their own action and on the aggregate action of the mass of other players. However, the essential issues raised by such models can be analyzed using a two-player game, as for example in (Yang (2015)).

[8] In their case, the parameters are equal to $a = 47$ and $b = 63$.

[9] We could suppose that the DM must learn a generative model of the joint distribution of these parameters along with θ and the other player's action frequencies; but since in the training sample it should be observed that the numerical values of a and b are always the same, the DM should learn a generative model in which these parameters are only able to take those specific values. We simplify our notation by not having to specify beliefs about these parameters as part of the generative model.

[10] To state this more precisely: action probabilities π^{row} for the row player and π^{col} for the column player represent a Nash equilibrium of the original game if and only if the pair (π^{row}, π^{col}) are a Nash equilibrium for the game with the transformed payoffs.

numerical values of a and b in any particular application.[11] Specifically, we choose a scale for the payoffs in terms of which a is represented by the value -2 and b is represented by the value $+2$. We let the value of θ in terms of these rescaled units be denoted $2x$. Thus the random real number x is proportional to the amount by which θ is greater than the midpoint between a and b on a given trial.

We further observe that the strategic analysis of such a game is unchanged if we add some amount to a player's payoffs that may depend on what their opponent does, but that (for any choice by the opponent) is the same regardless of what they themselves do.[12] Thus we obtain an equivalent game if we add $-x + 1$ to a player's payoff in the event that their opponent leaves, and an amount $-x - 1$ to their payoff in the event that their opponent stays. With this modification, the payoff matrix becomes the one shown in Table 2. We can without loss of generality suppose that the payoff matrix is this latter one. To any equilibrium of the game with the payoff matrix shown in Table 1, there must be a corresponding equilibrium of the game with the payoff matrix shown in Table 2; thus it suffices to study the equilibria of the latter game.

Table 2: *The Game with Transformed Payoffs*

	Leave	Stay
Leave	$x + 1,\ x + 1$	$x - 1,\ -x - 1$
Stay	$-x - 1,\ x - 1$	$1 - x,\ 1 - x$

The advantage of using the transformed payoff matrix shown in Table 2 is that it makes clear the symmetries of this kind of game. First, it is already evident from Table 1 that the payoff matrix is invariant under a transformation that reverses the labels of the two players. But it is evident in Table 2 that there is a second symmetry as well: the payoff matrix is invariant under a transformation that (i) reverses the labels of the two possible actions, for both players, and (ii) reverses the sign of x. Because of these symmetries, if there is an equilibrium in which the row player chooses to stay with probability $p(x)$ when the state is x, and in which the column player chooses to stay with probability $q(x)$, then we know that there must also be three additional equilibria: not just $(p(x),\ q(x))$, but also $(q(x),\ p(x))$, and in addition $(1 - p(-x),\ 1 - q(-x))$ and $(1 - q(-x),\ 1 - p(-x))$. As we discuss below, the set of equilibria of the game between boundedly-rational players modeled by VAEs is invariant under these same transformations, under some (relatively weak) assumptions about the classes of parametric models that are considered by the VAEs.

[11] Note however that the numerical value of the weight β_2 associated with a particular solution will change when the scale of the values in the payoff matrix is different.

[12] Again, a change in the payoff matrix of this kind results in no change in the set of pairs (π^{row}, π^{col}) that constitute Nash equilibria. It similarly has no consequences for our analysis of boundedly-rational equilibria, since the term L in our proposed training objective (the only term that involves the game payoffs) depends only the differences ΔU between the payoffs received by a given player under different possible actions, holding fixed the action of the other player.

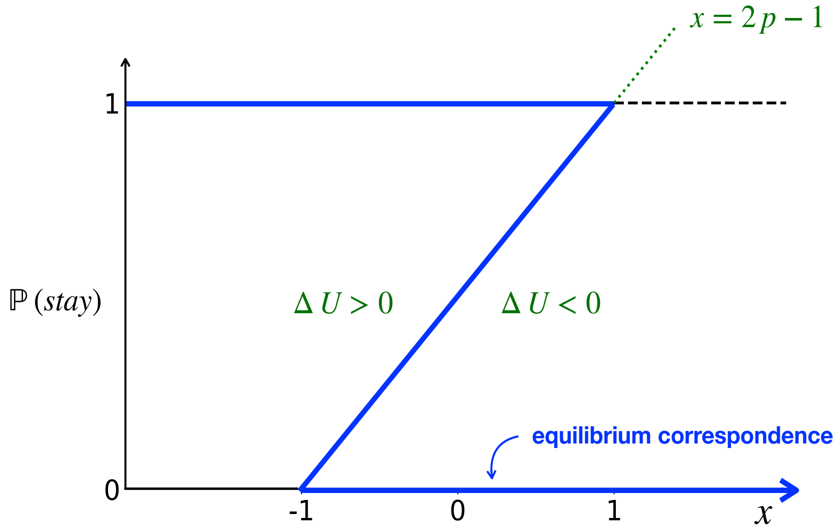


Figure 1: The equilibrium correspondence in the case of common knowledge of the external state.

3.2 The Imprecision of Observed Behavior

We first consider what equilibrium behavior should be like in the case that players are able to respond precisely to the value of x on a given trial (and to the way that their opponent plays in that state). To simplify the discussion, we consider only equilibria that are invariant under an exchange of the labels of the two players (one of the two symmetries discussed above). Thus we assume that $q(x) = p(x)$, and ask what the function $p(x)$ must be like in an equilibrium.

Given a value of x (that we assume for now to be common knowledge between the two players), if the other player chooses to stay with probability p, then a player's expected payoff from staying exceeds their expected payoff from leaving by the amount

$$\Delta U \;=\; 2[(2p-1) \;-\; x]. \tag{9}$$

It is thus strictly preferable for the player to stay in the case of any pair (x, p) above and to the left of the diagonal dotted line in Figure 1, and strictly preferable for the player to leave in the case of any pair below and to the right of the line. We can then use the figure to graph the *equilibrium correspondence:* the set of pairs (x, p) with the property that if the external state is x and the other player stays with probability p, it is a best response for oneself to stay with probability p as well. This correspondence consists of the set $\mathcal{E}$ of points identified by the thick bars in Figure 1 (a Z-shaped graph).

A function $p(x)$ specifying the state-contingent (and probabilistic) behavior of each of the players then represents a symmetric equilibrium of the game specified by the payoff matrix in

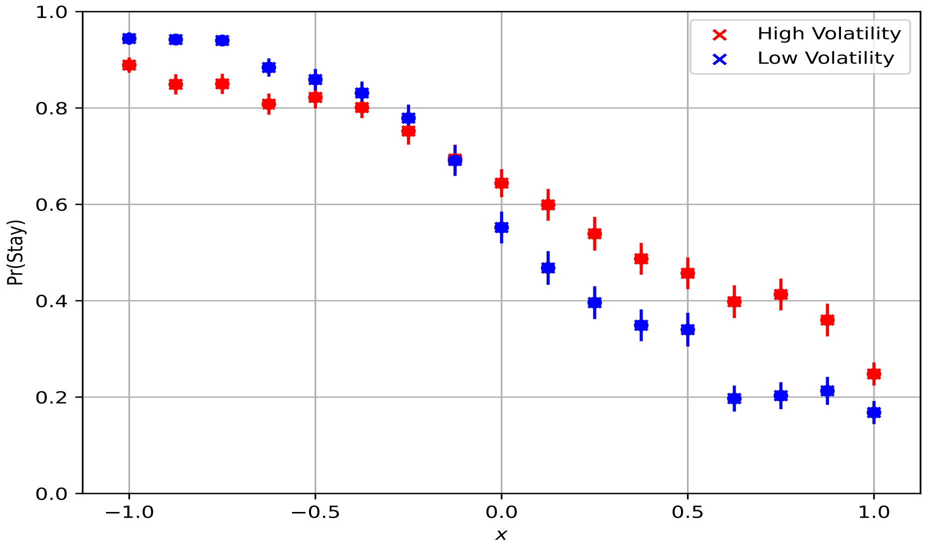


Figure 2: State-contingent behavior in the experiment of Frydman and Nunnari (2023).

Table 2 if and only if $p(x)$ is a single-valued function defined for all x on the real line, the graph of which belongs entirely to the equilibrium correspondence $\mathcal{E}$ shown in Figure 1. It is evident from the figure that we cannot have such a function, unless it involves a discontinuous jump at at least one value of x. If there is only one jump, it must be a downward jump at some value $\bar{x}$, with $p(x) = 1$ for all $x < \bar{x}$, and $p(x) = 0$ for all $x > \bar{x}$. The unique solution of this kind that also satisfies the other symmetry discussed above (i.e., such that $p(-x) = 1 - p(x)$ for all x) is the one in which $\bar{x} = 0$.

This kind of discontinuous behavior obviously requires extreme precision in the players' recognition of the value of the state x, and it is not what we see in experimental implementations of such games. As an illustration, Figure 2 plots data from the experiment of Frydman and Nunnari (2023), using the normalized state space notation introduced here. The vertical axis plots the frequency with which players choose to stay, as a function of that trial's value of x on the horizontal axis. The two curves are data from two alternative treatments, that differ only in the value of ω, the standard deviation of the prior distribution. (In their "low volatility" treatment, $\omega = \omega_{low} \equiv \sqrt{5}/2$, while in the "high volatility" treatment, $\omega = \omega_{high} \equiv 5$ in our normalized units.[13]) In neither treatment do we observe that the probability of staying drops discontinuously from $p(x) = 1$ to $p(x) = 0$ at some critical value $\bar{x}$; instead, in both cases the probability of staying falls gradually over the range from $x = -1$ to $x = 1$. Neither graph can be a selection from the equilibrium correspondence shown in Figure 1; similar gradual declines are observed in other experiments such as those of Heinemann et al. (2004).

[13] The two alternative priors are shown graphically in top left panel of Figure 12 in the Appendix.

4 The Coordination Game with VAE Players

We now consider the implications of applying the VAE model from section 2 to this coordination game. In order to apply the model, we must choose specific parametric families of recognition and generative models. In this section, we confine ourselves to the characterization of equilibria of the game between VAEs, without asking how such coordination of the two VAEs occurs. Later (in section 7), we discuss the way in which the players' VAE parameters can be learned from a training dataset generated by their own play.

4.1 Symmetries and Equilibria of the Game Between VAEs

In section 3.1 above, we have discussed the symmetries of our coordination game when it is assumed that the value of x is observed precisely by both players, so that their action probabilities can be conditioned upon it. In this case it is possible to define what it means for the strategy for the row player defined by a function $p(x)$ to be a best response to the strategy for the column player defined by function $q(x)$ by considering play in each individual state x separately. Hence the set of equilibria is independent of any assumption about the prior from which the values of x are drawn, and no symmetry assumption on the prior was needed for us to conclude that the set of equilibria should be invariant under two different symmetry transformations. If we assume instead that actions must be conditioned upon a coarse latent state rather than the exact value of x, slightly stronger assumptions are needed in order to demonstrate the invariance of the set of equilibria under these transformations; but corresponding symmetries continue to exist under some relatively simple assumptions.

The game between VAEs possesses the first symmetry (invariance under reversal of the labels of the players) if we suppose that the classes of possible specifications of two players' VAEs are the same (as we shall assume below): if parameters (ϕ, θ) represent a possible VAE for the row player, they also represent a possible VAE for the column player, and vice versa.

The game between VAEs will also possess the second symmetry under some additional assumptions. First, we must assume that x is drawn from a symmetric prior, i.e., a prior density with the property that $\pi(x) = \pi(-x)$ for all x. In our main analysis (everywhere except in section 7.2 below), we assume a prior of the form $N(0,\, \omega^2)$, for some value of ω; note that the two treatments in the experiment of Frydman and Nunnari (2023) are both of this form, for different values of ω. In this case, the symmetry assumption is satisfied.

Second, we assume that there exist symmetry transformations, associating with each latent state j a corresponding state j^*, and with VAE (ϕ, θ) a corresponding VAE (ϕ^*, θ^*), such that (i) the transformations are invertible: $(j^*)^* = j$ for all j, and similarly $(\phi^*)^* = \phi$ and $(\theta^*)^* = \theta$; and (ii) $p_\phi(j\,|x) = p_{\phi^*}(j^*\,| - x)$, $q_\theta(j) = q_{\theta^*}(j^*)$, and $\tilde{p}_\theta(x\,|j) = \tilde{p}_{\theta^*}(-x\,|j^*)$ for all (j, x). (This assumption is a restriction upon the parametric families of possible recognition and generative models, and is satisfied by the parametric families assumed in our numerical examples below.) The transformed version of any decision rule is correspondingly specified by defining $a^*(j^*)$ to be the action opposite to $a(j)$, for any latent state j.

Under these additional assumptions, the game between VAEs possesses the second symmetry as well. We shall say that a specification of VAE (ϕ_1, θ_1) and decision rule a_1 for the row player, and VAE (ϕ_2, θ_2) and decision rule a_2 for the column player, represent an *equilibrium* of the game between VAEs if the strategy (ϕ_1, θ_1, a_1) is a best response to strategy (ϕ_2, θ_2, a_2) by the other player, and vice versa. Our double symmetry result states that if (ϕ_1, θ_1, a_1) for the row player and (ϕ_2, θ_2, a_2) for the column player are such an equilibrium, there must also be three additional equilibria: (ϕ_2, θ_2, a_2) for the row player and (ϕ_1, θ_1, a_1) for the column player; $(\phi_1^*, \theta_1^*, a_1^*)$ for the row player and $(\phi_2^*, \theta_2^*, a_2^*)$ for the column player; and $(\phi_2^*, \theta_2^*, a_2^*)$ for the row player and $(\phi_1^*, \theta_1^*, a_1^*)$ for the column player.

We can then define a *symmetric equilibrium* of the game between VAEs as an equilibrium that is invariant under both of the symmetry transformations: that is, an equilibrium in which (i) the strategies of the two players are identical: $\phi_1 = \phi_2$, $\theta_1 = \theta_2$, and $a_1 = a_2$; and (ii) the common strategy is invariant under the symmetry transformation: $\phi^* = \phi, \theta^* = \theta$, and $a^* = a$. Given the symmetry of the game between VAEs (under these additional assumptions), it is natural to conjecture the existence of a symmetric equilibrium; a unique equilibrium would have to be of this kind. In our numerical explorations reported below, we always find that a symmetric equilibrium exists, and that it appears to be the only possible kind of equilibrium. Note that an advantage of studying symmetric cases is that a smaller number of parameters are required to characterize equilibrium behavior.

4.2 Parameterization of the Generative and Recognition Models

We begin by specifying the class of generative models that we consider. We suppose that the set $\mathcal{J}$ of latent states has $2J$ elements (for some positive integer J), that we number as $j = 1, \ldots, J$ and $j = -1, \ldots, -J$. The reason for this notation is to allow us to define a symmetry transformation of the kind discussed above, by specifying that for each state j, we define $j^* = -j$. We further assume that each of the conditional distributions $\tilde{p}_\theta(x\,|j)$ is Gaussian: $x|j \sim N(\mu_j, \sigma_j^2)$. In the absence of any restrictions on the possible parameters of the Gaussian distributions, this class of generative models has the symmetry property proposed in section 4.1. For any generative model θ specified by parameters $\{q_\theta(j), \mu_j, \sigma_j\}$, there is a corresponding model θ^* specified by parameters $\{q_{\theta^*}(j), \mu_j^*, \sigma_j^*\}$, where $q_{\theta^*}(j) = q_\theta(-j)$, $\mu_j^* = -\mu_{-j}$, and $\sigma_j^* = \sigma_{-j}$ for all $j \in \mathcal{J}$.

We assume that the recognition model (encoding model) belongs to the parametric family

$$p_\phi(j\,|x) \;=\; \frac{\exp(\lambda_j x - \nu_j)}{\sum_{\ell\in\mathcal{J}}\exp(\lambda_\ell x - \nu_\ell)} \tag{10}$$

for certain coefficients $\{\lambda_j, \nu_j\}$. Only the differences between the values of λ_j for different latent states matter for the probabilities defined by (10), and similarly only the differences between the values of ν_j for different latent states matter; hence we can without loss of generality normalize the coefficients so that $\sum_{j\in\mathcal{J}}\lambda_j \;=\; 0$ and $\sum_{j\in\mathcal{J}}\nu_j \;=\; 0$.

This family of encoding rules implies that the log of the relative odds of choosing any two latent states j, j' is a linear function of x, with the log odds when $x = 0$ specified by the difference $\nu_{j'} - \nu_j$, and the rate at which the log odds change as x increases specified by the difference $\lambda_j - \lambda_{j'}$. Thus the parameters all have relatively simple interpretations in the specification given in (10). In addition, in the absence of any further parameter restrictions, this class of recognition models has the symmetry property proposed in section 4.1. For any recognition model ϕ specified by parameters $\{\lambda_j, \nu_j\}$, there is a corresponding model ϕ^* specified by parameters $\{\lambda^*_j, \nu^*_j\}$, where $\lambda^*_j = -\lambda_{-j}$ and $\nu^*_j = \nu_{-j}$ for all $j \in \mathcal{J}$.

Equation (10) implies that for any finite values of the parameters, $p_\phi(j\,|x) \in (0,\,1)$ for each j and all x, and that $p_\phi(j\,|x)$ will be a continuous function of x for each j. Thus it might seem that our model requires by assumption that a player's choices must be stochastic for all x, and that there will be no discontinuous jumps in the choice probabilities when plotted as functions of x — so that this feature of our model's predictions should be regarded as an assumption rather than a result. But in fact, the specification (10) allows for encoding rules that are arbitrarily close to deterministic encoding rules that jump discontinuously at certain critical values of x. In particular, it allows for encoding rules that are arbitrarily close to deterministic rules that would allow a decision rule based on the latent state to implement the choices associated with the single-jump symmetric equilibrium with perfectly precise action selection (discussed in section 3.2 above).[14]

That equilibrium requires that both players stay with probability 1 whenever $x < 0$, and that they both leave with probability 1 whenever $x > 0$. One can approach this pattern of play arbitrarily closely by assuming a recognition model in which there are two latent states ($J = 1$), letting the model parameters be $\lambda_1 = \lambda > 0, \lambda_{-1} = -\lambda$, and $\nu_1 = \nu_{-1} = 0$; and assuming a decision rule under which $a(j) = \mathit{stay}$ for all negative j and $a(j) = \mathit{leave}$ for all positive j. Then as $\lambda \to \infty$, the predicted pattern of play becomes arbitrarily close to the deterministic action selection required for perfect coordination. Since we do not impose any bounds on the $\{\lambda_j\}$ in our

[14] Note that this equilibrium achieves the maximum possible sum of expected payoffs for the two players, and thus can be regarded as representing the benchmark of perfect coordination between the two players, in addition to perfect optimization on the part of each player individually.

definition of equilibrium or our proposed learning rule, the fact that we do not obtain deterministic (or nearly deterministic) action selection, varying discontinuously (or so abruptly as to be nearly discontinuous), can be considered a consequence of the objective for which the parameters of the recognition model are optimized, rather than an assumption that has been built into the class of models that we consider.

In a symmetric equilibrium, both the generative and recognition models must take more special forms. Invariance under the first of the symmetries discussed in section 4.1 requires that the VAE parameters $\{q_\theta(j), \mu_j, \sigma_j, \lambda_j, \nu_j\}$ and decision rule $\{a(j)\}$ learned by each player are the same as those learned by the other. But in addition, each player's generative model, recognition model, and decision rule must be invariant if one reverses the sign of x for all external states, interchanges the labels of the two actions (both for the player and for their opponent), and interchanges the labels of latent states j and $-j$ (for each $j = 1, \ldots, J$). This latter symmetry will hold if and only if $q_\theta(-j) = q_\theta(j)$, $\mu_{-j} = -\mu_j$, $\sigma_{-j} = \sigma_j$, $\lambda_{-j} = -\lambda_j$, $\nu_{-j} = \nu_j$, and $a(-j)$ is the opposite action to $a(j)$, for all j.

4.3 Reduction to a Game Between Recognition Models

In the equilibrium definition above, we require each player's choice of parameter vectors (ϕ, θ, a) to be a best response to the other player's corresponding vectors. It is fairly easy, however, to reduce the game to one in which each player's "strategy" is simply the choice of a parameter vector ϕ for their recognition (or encoding) model together with a decision rule a. This is because a player's choice of ϕ directly determines their optimal choice of θ as well.

Given the prior π from which the state x is drawn, and the specification ϕ of a player's own recognition model, the joint distribution $p_\phi(j, x)$ is completely determined. The player's optimal generative model θ is then the one that minimizes $D_{KL}(p_\phi || \tilde{p}_\theta)$, just as in the basic VAE model of Kingma and Welling (2014), since the additional term L defined by (8) is independent of θ.[15] This means (as discussed in section 2.2) that the optimal θ will be the one that maximizes the likelihood of data drawn from the distribution p_ϕ. The optimal choice of q_θ will again be given by q_ϕ (defined in (2)), and for each latent state j, the optimal parameters (μ_j, σ_j) will be given by the mean and standard deviation (respectively) of the conditional distribution $p_\phi(x\,|j)$.

It is then possible to define an equilibrium as a situation in which the parameter vectors (ϕ, a) chosen by one player are a best response to the parameter vectors (ϕ', a') chosen by the other player, and vice versa. We can further simplify the characterization of equilibrium if we restrict

[15] Here we assume our baseline model in which the VAE training objective is (4). In the case of the generalized objective (7), the term R defined in (5) also depends on the probabilities q_θ, an aspect of the generative model. However, as already discussed in section 2.2, even when $\beta_1 \neq 1$, it remains the case that the optimal choice of q_θ is given by (2). And given this choice, the term R reduces to I, which depends only on p_ϕ. Thus the optimal generative model θ continues to be the one that minimizes $D_{KL}(p_\phi || \tilde{p}_\theta)$.

ourselves to the case of a game between recognition models that each involve only two latent states, as we do in the numerical results reported here.

4.4 The Case of Only Two Latent States

In our numerical illustrations of the coordination game between VAEs, we further specialize to the case in which $J = 1$, so that there are only two latent states ($j = -1, +1$) for each player. This is the minimum number of latent states required for the latent state to matter for observed behavior. On the other hand, two latent states suffice for the class of behavioral models that we consider to include arbitrarily close approximations to the behavioral rules required for perfect coordination. Thus our restriction to VAEs with only two latent states does not in itself imply that coordination must be imperfect. Indeed, in the equilibria that we display below, for all large enough values of β, equilibrium behavior involves perfect coordination.[16] For the sake of computational tractability, we here consider only the case with $J = 1$, leaving for further analysis elsewhere the consequences of allowing for a larger number of latent states.[17]

The restriction to the case of only two latent states simplifies the set of possible choices for the decision rules (a, a') that we must consider in order to fully characterize the set of possible equilibria. There are four logically possible decision rules for either player: (i) stay if $j = -1$ and leave if $j = 1$; (ii) leave if $j = -1$ and stay if $j = +1$; (iii) leave in either latent state; and (iv) stay in either latent state. The set of possible behavioral rules associated with decision rule (ii) is the same as in the case of decision rule (i): one simply changes the role of the two latent states in the recognition model and generative model in order to represent a given behavioral rule using one of these decision rules or the other. Hence we can without loss of generality exclude rule (ii) from consideration.

But all of the behavioral patterns associated with decision rules (iii) or (iv) can also be obtained as limiting cases of behavioral patterns that can be obtained using decision rule (i). (If one wants to model behavior under which the player always leaves, one can simply choose an encoding rule under which the latent state is $j = 1$ with probability 1, regardless of the value of x, and still use decision rule (i); and similarly if one wants to model behavior under which the player always stays.) Thus there is no advantage to a player of using a decision rule other than (i), as far as the minimum achievable value of L is concerned, regardless of the other player's VAE and decision

[16] See Figures 6 and 8 below.

[17] Such an analysis should reasonably take account of computational costs of allowing for more complex families of generative and recognition models. If one does, it may well sometimes be the case that it is optimal to have no more latent states than the number of available actions, even when the costs of additional latent states are modest. For example, the rational inattention analysis of Yang (2015) imposes no a priori restriction on the number of possible latent states, but finds that the optimal information structure for each player involves only two latent states, for any positive value of the information-cost parameter.

rule. But there is also no advantage of using decision rules (iii) or (iv) from the standpoint of reducing the value of $D_{KL}(p_\phi || \tilde{p}_\theta)$. When one chooses an encoding rule under which one of the latent states j is used with probability 1, regardless of the value of x, then the optimal generative model θ will assign probability $q_\theta(j) = 1$ to that latent state, and (if the prior π is Gaussian, as we assume here) will assign it a conditional distribution $\tilde{p}_\theta(x\,|j)$ equal to the prior $\pi(x)$. In this case, we will have $D_{KL}(p_\phi || \tilde{p}_\theta) = 0$, which is the lowest value that could be obtained using any other VAE.[18] Hence we can without loss of generality assume that the decision rule is rule (i) for each player.

An equilibrium of the game between two-state VAEs is then a pair of recognition models (ϕ, ϕ') such that ϕ is the optimal recognition model for a player facing an opponent who uses recognition model ϕ', and vice versa. Here the expected payoffs for the two players associated with a pair of recognition models (ϕ, ϕ') are defined by assuming (i) that each player's generative model is the one that is optimal for their recognition model, and (ii) that each player's decision rule specifies that they stay if their latent state is -1 and leave if their latent state is +1. Under these assumptions, we can define a function $V(\phi;\, \phi')$ as the value of the training objective (4) for the player with recognition model ϕ when the opponent uses recognition model ϕ'. Because of the symmetry of the game, the value of the other player's training objective will be $V(\phi';\, \phi)$. An *equilibrium* is then a pair $(\bar{\phi}, \bar{\phi}')$ such that

$$\bar{\phi} \in \text{argmin}_\phi V(\phi;\, \bar{\phi}'), \qquad \bar{\phi}' \in \text{argmin}_{\phi'} V(\phi';\, \bar{\phi}). \tag{11}$$

5 Equilibrium Behavior: Numerical Results

We proceed now to numerical analysis of the equilibria of the coordination game between two-state VAEs defined in the previous section. Note that in the case of our baseline training objective (4), the game is completely specified by assigning numerical values to two parameters: the value of ω (indicating the amount of prior uncertainty, in terms of our normalized units), and the value of β (indicating the relative weight placed on accurate action selection as opposed to congruence between the generative and recognition models). This makes it relatively straightforward to explore the complete parameter space numerically.

5.1 The Best-Response Mapping

We begin by considering the optimal recognition model ϕ for a player, for a given specification of the recognition model ϕ' of their opponent. In the case $J = 1$, we can parameterize each player's

[18] Note that this argument does not require that the prior be symmetric, only that it be Gaussian. Thus it continues to be valid in the case of asymmetric Gaussian priors of the kind considered in section 7.2 below.

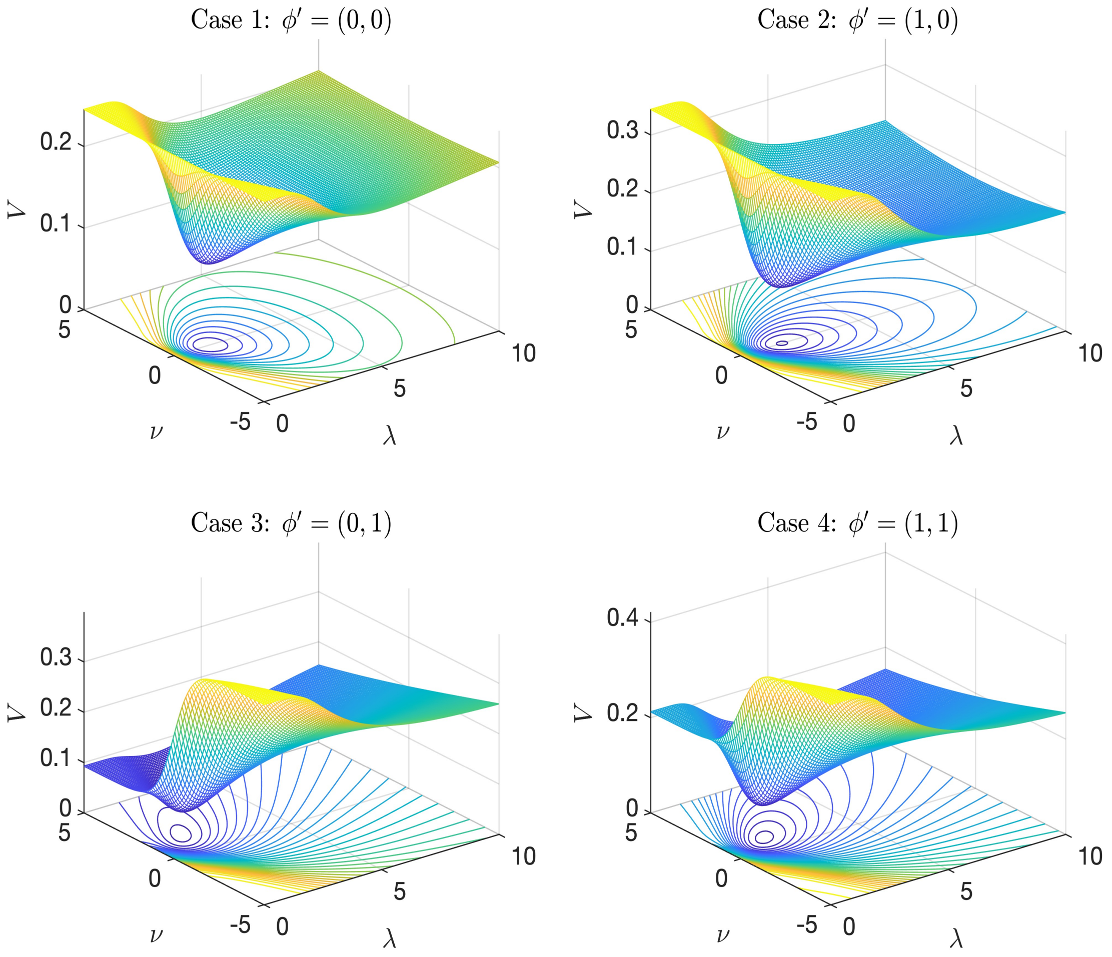


Figure 3: A player's training objective, plotted (on the vertical axis) as a function of their recognition model parameters ϕ (the two dimensions of the horizontal plane), for each of several different specifications (the four different panels) of the opponent's recognition model $\phi' = (\lambda', \nu')$. In all cases, $\beta = 0.2$ and $\omega = \omega_{low}$.

recognition model using only two numbers (λ, ν), corresponding to a recognition model of the form (10) with $\lambda_{-1} = -\lambda$, $\nu_{-1} = -\nu$, $\lambda_1 = \lambda$, $\nu_1 = \nu$. Thus we wish to determine the optimal parameters $\phi \equiv (\lambda, \nu)$ for a player given arbitrary choices $\phi' \equiv (\lambda', \nu')$ by their opponent.

The best-response mappings defined in (11) depend on the shape of the function $V(\phi;\, \phi')$. In Figure 3, we illustrate what the function $V(\lambda, \nu;\, \lambda', \nu')$ looks like, plotted in each panel as a function of the parameters (λ, ν), for given values of the parameters $\phi' = (\lambda', \nu')$, that are different in each panel of the figure. The top row of the figure shows two cases in which the opponent's recognition model is symmetric, meaning that $\nu' = 0$.[19] We observe in this case that regardless of the value of the opponent's λ', the function V is a symmetric function of the player's own recognition model parameters, in the sense that

$$V(\lambda, \nu;\, \lambda', 0) \;=\; V(\lambda, -\nu;\, \lambda', 0).$$

This symmetry implies that if the function has a unique minimum, it will necessarily be associated with $\nu = 0$. Our numerical results indicate that this seem always to be the case (as illustrated by the two cases shown in the figure). Comparison of the upper right panel with the upper left panel shows that increasing λ' (the right panel) causes V to increase less sharply for large positive values of λ: a given degree of sensitivity of action selection to the value of x is more valuable when the opponent's action is also more sensitive to the value of x (in the same direction).

The bottom row of the figure instead provides examples in which the opponent has an asymmetric recognition model: cases with $\nu' > 0$, implying a bias toward use of the latent state $j = -1$ (and hence toward staying) when x is near zero. Such a bias on the part of the opponent increases one's own incentive to stay, and hence (given the assumed decision rule) the incentive toward choosing the latent state $j = -1$ in the case of any external state x. The function V therefore achieves its lowest value in these cases at a point where $\nu > 0$, implying a bias toward the latent state -1 and hence toward staying.

An important technical issue is whether the minimization problems in (11) have unique solutions, so that the best-response mapping is single-valued (a function rather than a correspondence). The function $V(\lambda, \nu;\, \lambda', \nu')$ is not a convex function of the parameters (λ, ν); indeed, it is easily seen that the function has a finite upper bound, no matter how extreme the parameters of the recognition model may be. Nonetheless, numerical exploration suggests that the function has a unique

[19] When $\nu = 0$ for a player, their recognition model is symmetric, in the sense that $p_\phi(+1\,|x) = p_\phi(-1\,|-x)$ for all x, and the joint distribution p_ϕ is symmetric, in the sense that $p_\phi(j, x) = p_\phi(-j, -x)$ for all (j, x). This in turn implies that their optimal generative model will also be symmetric, in the sense that $\tilde{p}_\theta(j, x) = \tilde{p}_\theta(-j, -x)$ for all (j, x).

(and interior) minimum. This allows us to define a function

$$\Phi(\lambda') \;\equiv\; \text{argmin}_{\phi} V(\phi;\, \phi'), \tag{12}$$

so that $\phi = \Phi(\phi')$ is a player's optimal recognition model when their opponent's recognition model is parameterized by ϕ'.

We can then define an equilibrium of the game between recognition models as a fixed point of the iterated function Φ^2. If there exists a recognition model $\bar{\phi}$ such that $\Phi^2(\bar{\phi}) = \bar{\phi}$, then the pair $\phi = \bar{\phi}$, $\phi' = \Phi(\bar{\phi})$, constitute an equilibrium, since we have both

$$\phi' \;=\; \Phi(\bar{\phi}) \;=\; \Phi(\phi), \qquad \phi \;=\; \Phi^2(\bar{\phi}) \;=\; \Phi(\phi'),$$

and so both conditions of (11) are satisfied. (The converse is easily seen to be true as well.)

5.2 Symmetry of Equilibrium

It is logically possible for a fixed point $\bar{\phi}$ of the iterated mapping Φ^2 not to be a fixed point of Φ itself, i.e., for one to have $\Phi(\bar{\phi}) \neq \bar{\phi}$. In such a case, the equilibrium would be asymmetric (it would involve different VAEs, and hence different action frequencies, for the two players). If so, equilibrium cannot be unique. First of all, if (ϕ, ϕ') is an equilibrium, then (ϕ', ϕ) must also be an equilibrium, and if $\phi' \neq \phi$, these are two different equilibria.

But as discussed in section 4.1 above, the game between VAEs has another symmetry as well. This implies that if

$$(\lambda, \nu) \;=\; \Phi(\lambda', \nu'),$$

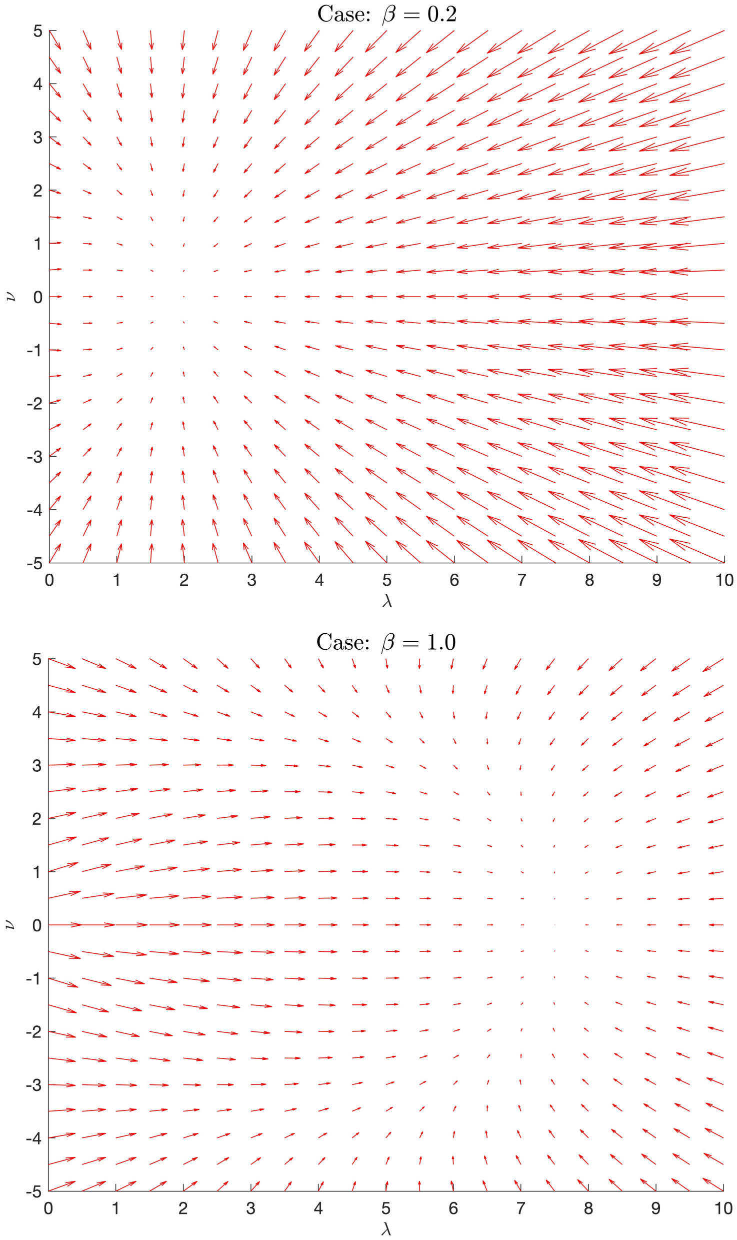
it must equally be true that[20]

$$(\lambda, -\nu) \;=\; \Phi(\lambda', -\nu').$$

This symmetry implies that if the mapping Φ has a unique fixed point, it must be of the symmetric form $\bar{\phi} = (\bar{\lambda}, 0)$. A fixed point of this kind defines a *symmetric equilibrium:* one in which each player has the same recognition model (and hence the same generative model as well), and it is of the symmetric form (i.e., involves zero bias). We thus conclude that if there is any equilibrium in which both players' VAEs are the same, it must involve recognition models with $\nu = 0$.

The symmetry just discussed also implies that if (λ, ν) is a fixed point of the iterated mapping Φ^2, the parameter vector $(\lambda, -\nu)$ must be a fixed point as well. This means that if there exists an asymmetric equilibrium (ϕ, ϕ'), with $\phi' \neq \phi$ and ν, ν' non-zero (the two players' recognition mod-

[20] This is the symmetry that implies that if the best response to a symmetric model $(\lambda', 0)$ is uniquely defined, it must also be symmetric (i.e., involve $\nu = 0$), as discussed above.



Case: $\beta = 0.2$

Case: $\beta = 1.0$

Figure 4: The vector field Ψ, plotted as a function of recognition-model parameters λ and ν. Equilibria correspond to zeroes of the vector field. In both panels, $\omega = \omega_{low}$.

els are different, and each of them is biased), then there must exist at least four distinct equilibria, corresponding to recognition models

$$(\lambda, \nu), \quad (\lambda, -\nu), \quad (\lambda', \nu'), \quad (\lambda', -\nu')$$

for the first player. It follows that in order for equilibrium to be unique, these four parameter vectors must coincide. This means that the unique equilibrium must be a symmetric equilibrium.

We examine numerically whether any non-symmetric equilibria exist, by plotting in the (λ, ν) plane the direction at each point of the vector field

$$\Psi(\phi) \;\equiv\; \Phi^2(\phi) \,-\, \phi.$$

Equilibria then correspond to zeroes of the vector field Ψ. Figure 4 displays two examples of what this vector field looks like, for two different values of β (but a common parameter value $\omega = \omega_{low}$). For each of the parameter configurations that we have examined, the vector field has a unique zero, which is located on the horizontal line $\nu = 0$ (as must be the case, given the symmetry of the vector field). There is thus a unique equilibrium, and it is a symmetric equilibrium, $\bar{\phi} = (\bar{\lambda}, 0)$, though the equilibrium parameter value $\bar{\lambda}$ depends on the size of β.[21]

Because of the symmetry discussed above, the restriction of the mapping $\Phi(\cdot)$ to the line $\nu = 0$ can be described by a scalar-valued function $\bar{\Phi}$, such that $\lambda = \bar{\Phi}(\lambda')$ if and only if $(\lambda, 0)$ is the best response to $(\lambda', 0)$. Then the symmetric equilibrium value of λ is given by the fixed point of this mapping, i.e., the quantity $\bar{\lambda}$ such that $\bar{\Phi}(\bar{\lambda}) = \bar{\lambda}$. Thus the problem of computing the symmetric equilibrium corresponding to any parameters (β, ω) can be reduced to one of finding the zero of a nonlinear function of one variable.

5.3 The Symmetric Equilibrium under Alternative Values of β

We now discuss how the quantitative properties of this equilibrium depend on the relative weight β used in the VAE training objective. We begin by considering how the value of β affects the best-response mapping Φ. Figure 5 plots the parameters (λ, ν) that represent a best response to a recognition model of the form $(\lambda', 0)$ on the part of the opponent. (We need only consider recognition models ϕ' of this symmetric form, because we have shown in the previous subsection that the only equilibrium corresponds to a fixed point of this kind.) The graphs of both functions

[21] Note that the conclusion that the unique equilibrium is symmetric depends on our assumption here of a symmetric prior. In section 7.2, we consider cases in which the prior is (at least initially) asymmetric (though still Gaussian). In these cases, the equilibrium of the game between VAEs still corresponds to the zero of the vector field Ψ, and the equilibrium is still a fixed point of Φ (so that both players use the same VAE in equilibrium), but the players' common recognition model involves $\nu \neq 0$.

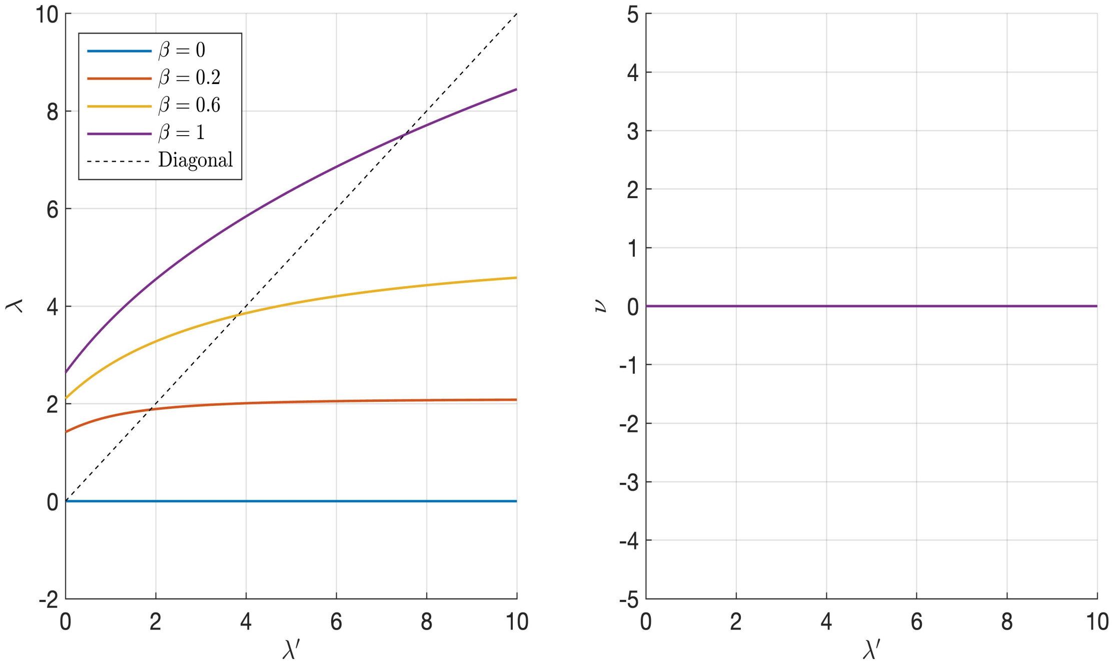


Figure 5: Best-response values of the parameters (λ, ν) to a symmetric recognition model $(\lambda', 0)$ on the part of the opponent, in the case of several different possible values of β. In each case, $\omega = \omega_{low}$.

are shown for each of several alternative possible values of β, for a given prior (the symmetric prior with $\omega = \omega_{low}$).[22]

For any value of β, the symmetric equilibrium corresponds to the value of λ' at which the graph of $\bar{\Phi}(\lambda')$, shown in the left panel of the figure, crosses the diagonal. As could also be seen from Figure 4, there is a unique such fixed point for each value of β. Figure 5, however, provides additional information about the function that determines the location of this fixed point. It can be seen from the behavior of the vector field Ψ along the horizontal line $\nu = 0$ in Figure 4 that $\bar{\Phi}^2(\lambda')$ is larger than λ' for all λ' less than the equilibrium value $\bar{\lambda}$, and smaller than λ' for all $\lambda' > \bar{\lambda}$. Figure 5 shows something stronger: that $\bar{\Phi}(\lambda')$ is an increasing function, with $\lambda' < \bar{\Phi}(\lambda') < \bar{\lambda}$ for any λ' less than the equilibrium value,[23] and each of these inequalities reversed for any λ' greater than $\bar{\lambda}$. This implies not only that there must be a unique fixed point $\bar{\lambda} > 0$, but that it can be computed by simply iterating the map $\bar{\Phi}$ to convergence from an arbitrary initial guess.

Figure 5 also shows that for any λ' chosen by the opponent, the value of $\bar{\Phi}(\lambda')$ is greater the larger the value of β. This then implies that the fixed point $\bar{\lambda}$ will be increasing as a function of β, as had already been shown for two particular values of β in the two panels of Figure 4. The

[22] The effect on the graph of instead varying the value of ω for a given value of β is discussed in section 6.2 below.

[23] The graph only shows the function for values $\lambda' \geq 0$ (the region of greatest interest, since the equilibrium value is always positive), but one can show that the claim also holds for negative values of λ'.

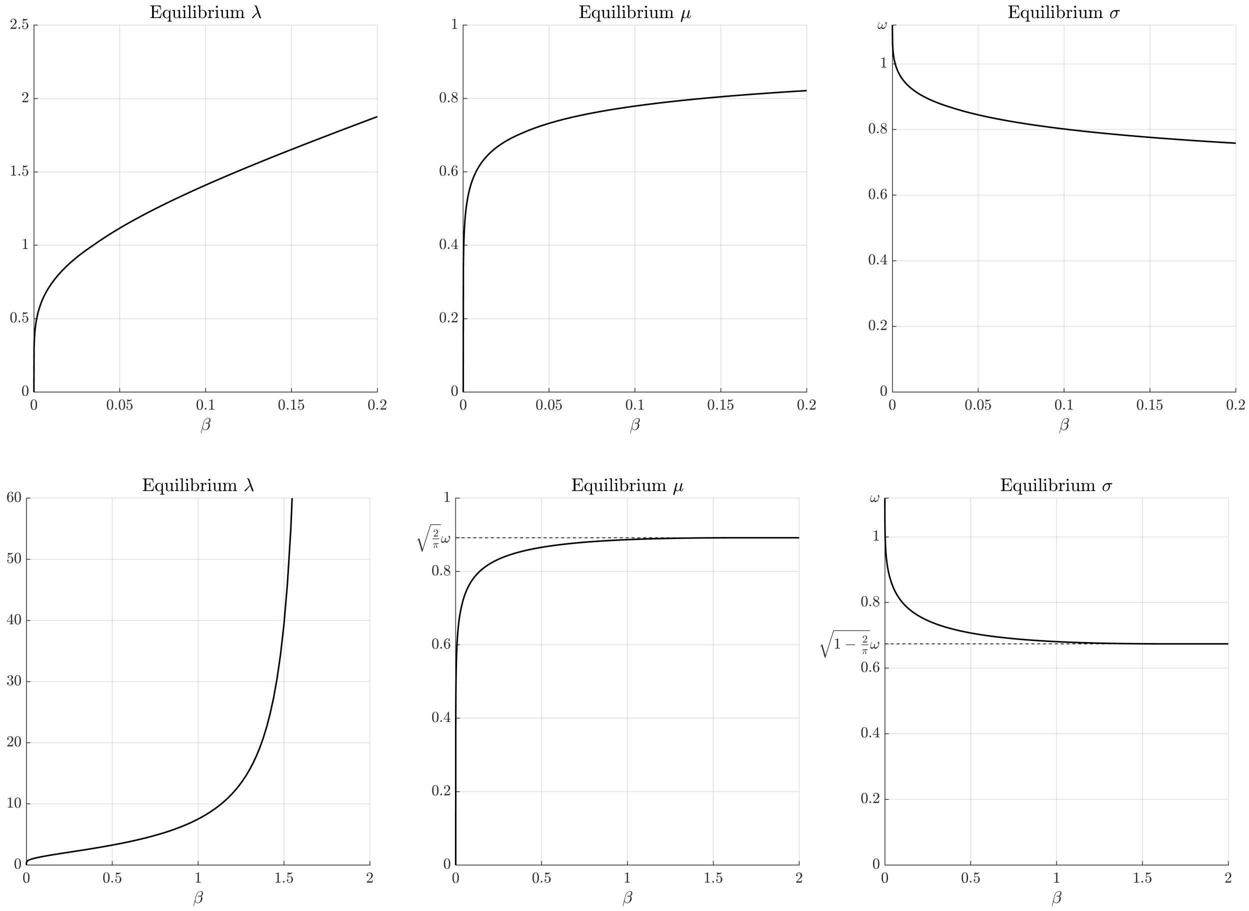


Figure 6: Solutions for the equilibrium values of the parameters λ, μ and σ of both players' symmetric VAEs, as functions of the value of β. Top row: solutions for small positive β. Bottom row: solutions for a larger range of values of β. In all cases, $\omega = \omega_{low}$.

equilibrium value of λ is plotted as a function of β (again, fixing the value $\omega = \omega_{low}$) in the panels in the left column of Figure 6.

The solution for λ is plotted for relatively small positive values of β in the top row; this figure shows how the equilibrium value of λ grows very sharply with even a small increase in β for values of β near zero, but grows more slowly for larger values of β. The solution is instead plotted for even larger values of β in the bottom row; this figure shows that for large enough values of β, the equilibrium value of λ again increases sharply with even a small further increase in β. Indeed, the optimal value of λ becomes unboundedly large (implying perfect coordination) for all values of β greater than a finite critical value $\bar{\beta}$ (about 1.645).

Given the solution for the optimal λ for any value of β (and thus for the optimal symmetric recognition model), one can then solve for the associated optimal generative model. Because the optimal generative model in the case of a symmetric recognition model is also symmetric, it must involve $q_\theta(j) = 1/2$ for both latent states, regardless of the value of β. Symmetry also implies that the other parameters of the optimal generative model must be of the form $\mu_{-1} = -\mu$, $\sigma_{-1} = \sigma$, $\mu_1 = \mu$, $\sigma_1 = \sigma$, for some coefficients μ, σ. The optimal values of μ and σ are also plotted as functions of β in Figure 6.

We see that the optimal value of μ is non-negative (and positive for all $\beta > 0$), corresponding to the fact that the optimal λ is non-negative (and also positive for all $\beta > 0$). Moreover, the optimal value of μ is an increasing function of β (just as λ is), up until the critical value $\bar{\beta}$ is reached, beyond which the optimal μ asymptotes to the finite value $\sqrt{2/\pi\omega}$. The optimal value of σ is always positive; it is equal to ω when $\beta = 0$, but falls monotonically as β (and hence λ) increases, asymptoting to the positive lower bound $\sqrt{1 - (2/\pi)}$ for all values of β larger than $\bar{\beta}$. The ways in which the equilibrium generative and recognition models vary with β are illustrated visually in Figure 12 in the Appendix.

5.4 Trade-offs Between Alternative Training Objectives

We have seen that regardless of the relative weight placed on the two terms in the hybrid objective (4), the unique equilibrium will be a symmetric equilibrium; and the quantitative properties of this equilibrium are completely determined once we know the value of the parameter λ of the symmetric encoding rule used by both players. Thus regardless of the training objective, there is only a one-parameter family of possible equilibrium behavioral patterns, indexed by λ. Since the equilibrium value of λ is furthermore found always to non-negative, the set of achievable behavioral patterns corresponds to the ones associated with parameter values $\lambda \geq 0$. (Note that all of those patterns can be achieved through a suitable choice of the training objective, since the equilibrium value of λ increases continuously from 0 to $+\infty$ as β is increased from 0 to $\bar{\beta}$, as

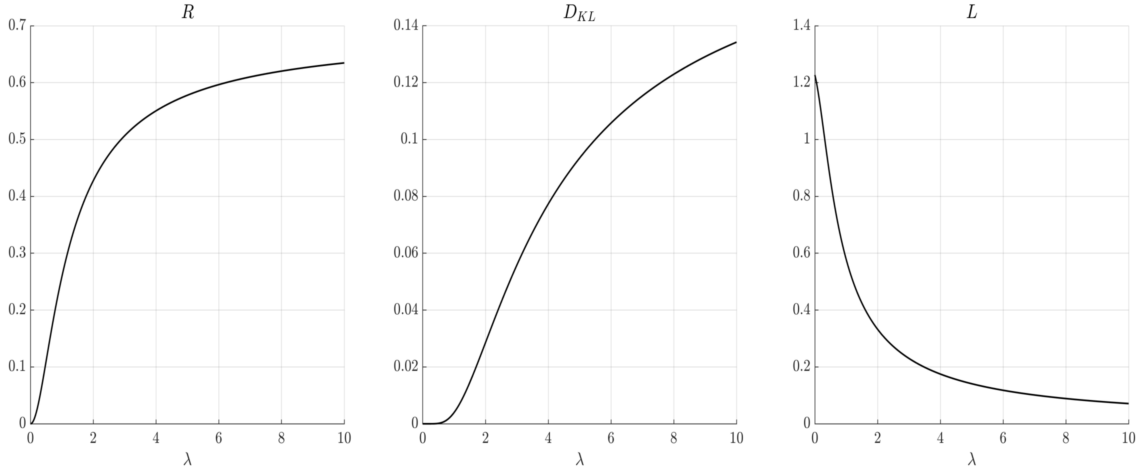


Figure 7: The values of three alternative training objectives in the stationary equilibria associated with different common values of λ for the two players. Each graph is shown for the case $\omega = \omega_{low}$.

shown in the lower left panel of Figure 6.)

The available trade-off between the two objectives — greater congruence between the generative model and the recognition model (i.e., lower D_{KL}) on the one hand, or more accurate action selection (i.e., lower L) on the other — can thus be studied by computing the achieved values of these alternative objectives for each of the possible values of $\lambda \geq 0$. The results of these calculations are plotted in the second and third panels of Figure 7. We see that increasing λ increases the value of D_{KL} (from a minimum value of zero, when $\lambda = 0$, to a finite positive asymptotic value as $\lambda \to \infty$) and reduces the value of L (which begins at its largest positive value when $\lambda = 0$, and approaches zero asymptotically as $\lambda \to \infty$). There is thus a trade-off between the two objectives, with lower L requiring a larger value of D_{KL}, and vice versa; any point on the trade-off can be achieved by choosing a suitable value of β between the extremes of $\beta = 0$ (which minimizes D_{KL} and maximizes L) and $\beta = \bar{\beta}$ (leading to the minimal value of L but a maximal value for D_{KL}).

The β-VAE proposal of Alemi et al. (2018) instead minimizes a weighted average of D and R, or alternatively, an objective with a positive weight on D_{KL} and a *negative* weight on R. Figure 7 also shows how the value of R differs across symmetric equilibria associated with different values of λ. Since D_{KL} and R are both monotonically increasing functions of λ, we see that there is a tension between the objectives of reducing D_{KL} and of increasing R: one requires λ to be kept as small as possible while the other requires it to be made as large as possible. By varying the relative weight on the two objectives (alternative values of the "β" of Alemi et al. (2018)), one can sweep out the entire one-parameter family of possible symmetric equilibria, from $\lambda = 0$ (when all of the weight is on reducing D_{KL}) to λ unboundedly large (when all of the weight is on maximizing R). But this is the same one-parameter family of equilibria as are already obtained in our baseline model with training objective (4); we have no need to assume an intrinsic interest in more informative latent states, as proposed by Alemi et al. (2018), once we penalize models that result in higher values of L.

The rational inattention model of Sims (2003) instead assumes that information structures minimize a weighted sum of L and the Shannon mutual information, or equivalently (since $I = R$ for all of the symmetric equilibria considered here), a weighted sum of L and R. Figure 7 shows that there is also a tension between these two objectives: reducing L requires making λ larger, while reducing R requires that it be made larger. Depending on the relative weight on the two terms, the symmetric equilibrium obtained will involve a larger or smaller value of λ; but again the theory would lead us to the same one-parameter family of possibilities.[24]

Finally, we have mentioned above a more general possible training objective (7), with weights on D_{KL}, R, and L all three. But in the current application, different values of the weights β_1, β_2 in the generalized objective can still only lead to the same one-parameter family of possible symmetric equilibria. Increasing the weight on reduction of L (through a higher value of β_2) and increasing the weight on R maximization (or reducing the weight on reduction of R , by making β_1 lower) are simply two different ways of inducing a larger equilibrium value of λ, with the same effects on the equilibrium pattern of behavior in either case. For this reason, we omit figures showing the quantitative effects of varying the parameter β_1 in (7).

6 Consequences for Observable Behavior

We turn now to the predictions of our VAE model for observed play in a coordination game of the kind described in section 3, at least to the extent that we think that subjects' cognitive processing should have sufficient opportunity to adapt to the statistics of play in a specific experimental situation. (The implications of incomplete learning are deferred until section 7.) We discuss two aspects in particular of experimentally observed behavior, as illustrated in Figure 2. First, we consider the reason for subjects' decisions to be stochastic, conditional on the current state x, and for the probability of selecting a particular action to vary only gradually as x changes, rather than switching discontinuously. And second, we consider the predicted effects on the conditional probability of action selection of a change in the prior distribution from which x is drawn on separate trials.

[24] The one-parameter family of solutions is only the same as in the case of our baseline model if we assume that the optimization is over the parametric families of generative and recognition models defined in section 4.2. RI theory as formulated by Sims includes no such restrictions; the theory discussed here can be viewed instead as a sort of variational approximation to RI theory.

6.1 Imprecise Action Selection

A notable feature of the observed behavior in Figure 2 is that subjects randomize between the two possible actions in the case of all of the values of x between -1 and 1, rather than always staying in the case of low values of x and always leaving when x is higher. Our model implies that behavior of this kind can be a stable outcome of learning by a VAE.[25] As shown above, the equilibrium recognition model is stochastic (i.e., λ is finite) for all values of β below the critical level $\bar{\beta}$; the implied probabilities of classifying a given external state with one or the other latent state are shown for several values of β in Figure 12 in the Appendix. Given the equilibrium decision rule, this immediately translates into a prediction of random choice conditional on the value of x, as illustrated in Figure 9 below.

It might be thought that we have "hard-wired" our model to yield this conclusion, but this is not the case: as noted above, it is possible for a VAE to yield discrete switching between staying with certainty and leaving with certainty. Indeed, this would represent equilibrium behavior, for some parameter values. Figure 13 in the Appendix shows, as an example, that if $\beta = 2$, the optimal recognition model involves making λ an unboundedly large positive quantity, for either of the two values of ω used in the experiment of Frydman and Nunnari (2023). Thus the VAE learns to classify the state as $j = -1$ with certainty whenever $x < 0$ and as $j = 1$ with certainty whenever $x > 0$. Since the learned decision rule $a(j)$ is also deterministic, this leads to deterministic behavior, and indeed to perfect coordination on mutually advantageous state-contingent behavior. (The same results are obtained for all large enough values of β, as shown for the case $\omega = \omega_{low}$ in Figure 6.)

But this is not the kind of behavior that should be learned in the case of a training objective that places a greater relative weight on congruence between the generative model and the statistics of the training sample produced by classification of external states using the recognition model (corresponding to a lower value of β). This is because a very large value of λ (implying more nearly deterministic action selection) implies a larger value of $D_{KL}(p_\phi || \tilde{p}_\theta)$, as shown in Figure 7, implying less congruence. A sufficiently greater relative weight on congruence therefore requires that λ remain of modest size; indeed, if the weight on the other objective is sufficiently small ($\beta \to 0$), the equilibrium value of λ can be made arbitrarily close to zero, implying minimal sensitivity of action selection to the circumstances on a particular trial.

The fact that we obtain this result even when we place no additional penalty on the degree of informativeness of the latent state (i.e., when we do not assume $\beta_1 > 1$, as in rational inattention models) depends critically on the fact that we limit the class of VAEs that we consider to a restricted

[25] A variety of other approaches have also been proposed that allow for stochastic behavior and consequent imperfect coordination in a game like the one that we discuss; these include quantal response equilibrium (Goeree et al. (2016), Friedman (2020)), rational inattention (Yang (2015)), Thompson sampling (Mauersberger (2022)), and evolutionary models (Arifovic et al. (2013)), in addition to the kind of efficient coding model that Frydman and Nunnari (2023) propose as an explanation of their data. We compare our model with these alternatives in Appendix section A.

parametric family — more specifically, on the fact that the class of possible *generative models* is restricted. Even granting the restriction to only two latent states, and requiring the recognition model to be a symmetric member of the family (10), if it were possible to choose the conditional distributions $\tilde{p}_\theta(x\,|j)$ with sufficient flexibility, it would always be possible to choose a generative model such that $D_{KL}(p_\phi||\tilde{p}_\theta) = 0$, regardless of the value of λ chosen for the recognition model. It would then be possible to choose λ simply with a view to increasing the accuracy of action selection, making it optimal to choose an unboundedly large value of λ in the case of any relative weight $\beta > 0$.

The result is different from this in our model (when $0 < \beta < \bar{\beta}$) because of our restriction on the class of possible generative models that can be considered by the VAE. The result that stochasticity can chosen in equilibrium does not, however, depend on such special assumptions as requiring the distributions $\tilde{p}_\theta(x\,|j)$ to be Gaussian, or assuming that there can be no more than two latent states; it simply requires that some conditional distributions can be better approximated by a member of the admissible family than others, and that the prior distribution $\pi(x)$ be a "simple" distribution that can be particularly well approximated in this way.

Given the restricted parametric family of recognition models considered here, our model predicts that in equilibrium, Prob[leave $|x$] should be a logistic function of x. More general shapes are possible, however, if we allow for other kinds of recognition models.[26] Moreover, the data shown in Figure 2 aggregate the choices of many different subjects; our model predicts a logistic curve of the kind illustrated in Figure 9 only if we assume that the weight β is the same for all of the subjects. In fact, as Frydman and Nunnari (2023) report, the curves fit to individual subjects' data are somewhat heterogeneous. If we assume a population of subjects that includes both higher and lower values of β, our model predicts that one should observe both relatively steep variation in aggregate choice frequencies for values of x around zero, and at the same time choice frequencies that remain relatively far from either zero or one even for relatively extreme values of x.

6.2 Effects of the Degree of Prior Uncertainty

Another notable feature of the behavior shown in Figure 2 is that the probability of staying, conditional on the value of x, varies depending on the variance of the distribution from which x is drawn in that experimental treatment. Our model predicts this as well, owing to the endogeneity of the equilibrium recognition rule (encoding rule), depending on the statistics of the observed values x in the environment to which the players' VAEs are adapted.

The intuition for this can be simply explained. Minimization of the discrepancy measure $D_{KL}(p_\phi||\tilde{p}_\theta)$ requires (among other things) that the marginal distribution $\tilde{p}_\theta(x)$ implied by the

[26] Note that it is the restrictiveness of the class of generative models, rather than the special class of recognition models, that is important for the prediction of equilibrium stochasticity under the argument just given.

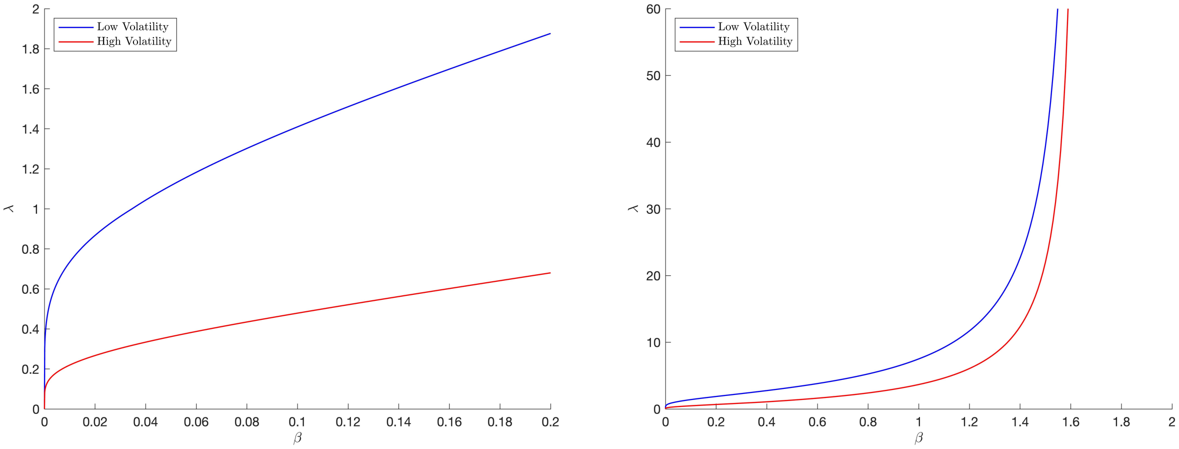


Figure 8: The equilibrium value of λ as a function of β, now shown for two different values of ω. (The left panel provides greater detail for small values of β.)

generative model not be too different from the prior distribution $\pi(x)$. When the prior distribution is made wider (passing from the "low volatility" to the "high volatility" condition), the degree to which D_{KL} increases is reduced by making the conditional distributions $\tilde{p}_\theta(x\,|j)$ correspondingly wider, and the means of the two distributions corresponding to the two latent states farther apart.[27] But such a change in the generative model also implies a change in the recognition model that will minimize the discrepancy between the joint distributions p_ϕ and $\tilde{p}_\theta$.

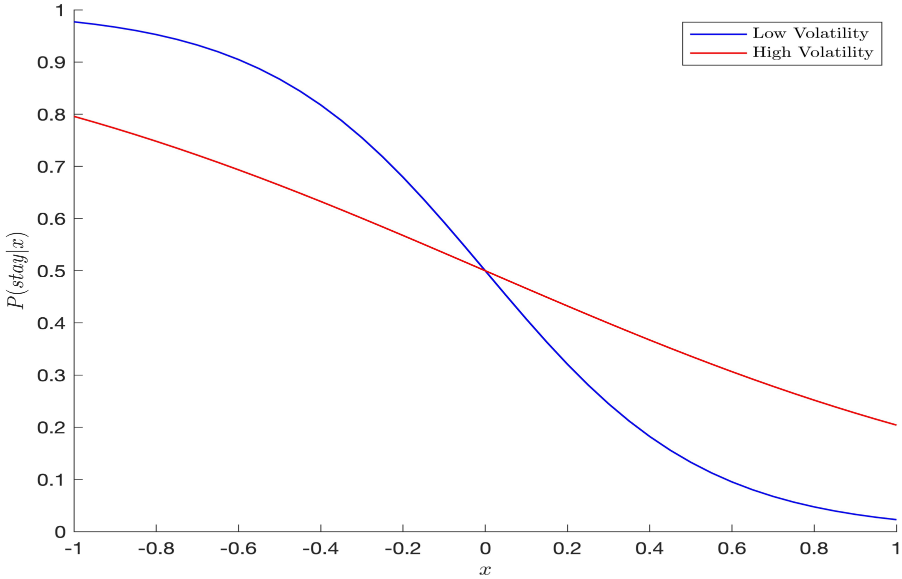


Figure 9: Predicted equilibrium choice frequencies under the two treatments of Frydman and Nunnari (2023). Results are shown for a training objective with $\beta = 0.2$.

If the VAE training objective were to make λ as large as possible (in order to minimize L), sub-

[27] For numerical illustrations, see Figure 12 in Appendix section B.

ject to an upper-bound constraint on the acceptable value of D_{KL}, the solution would be to increase both μ and σ in exact proportion to the increase in ω, while reducing λ in inverse proportion to the increase in ω. (These offsetting changes would result in no change in the value of D_{KL}, despite the change in the prior.) When the objective is instead to minimize (4) for some relative weight $\beta > 0$ that is not too large, this is not exactly the optimal adjustment, but (for any fixed recognition model ϕ' of the opponent) it continues to be optimal to reduce λ in response to an increase in ω. This means that the curve shown in the left panel of Figure 5 for a given value of β will shift downward for each value of λ', resulting in a lower value of λ^*, the value of λ' for which the curve $\lambda = \Phi(\lambda')$ crosses the diagonal.

The consequences for the equilibrium value of λ for each possible value of β are shown in Figure 8. The left panel of the figure shows how the curve shown in the top left panel of Figure 6 shifts if one increases the value of ω from ω_{low} to ω_{high}; the right panel correspondingly shows how the curve shown in the bottom left panel of Figure 6 shifts. We observe that λ is lower for each value of β. Thus if we assume that the VAE used in each experimental condition is one that is optimally adapted to that condition, but that the same training objective is used in both conditions, the players' choices should be less sensitive to the value of x on each trial under the "high volatility" condition. This is illustrated in Figure 9, for the case of VAE training objective (4) with weight $\beta = 0.2$. Comparison of this figure with Figure 2 shows that our model can account (at least qualitatively) for the central finding of Frydman and Nunnari (2023).

6.3 Greater Precision in Play Against an Automaton

Frydman and Nunnari (2023) also compare the experimental treatment in which two human subjects play one another (discussed above) to an alternative treatment in which a single human subject plays against a computer that chooses deterministically, according to a rule that the human subject is told: the computer will leave if and only if $x \geq 0$. Given this knowledge, the optimal strategy for the human player when playing against the computer is to also leave if and only if $x \geq 0$; and it is also an equilibrium for two perfectly-rational human players to follow this decision rule. Thus under an assumption of perfectly precise play by human players, there need be no difference between the play observed when two humans play each other and when a human plays against the computer. Instead Frydman and Nunnari (2023) find random action selection conditional on x in both cases, but considerably more random behavior when two humans play each other.

Our model also predicts this difference between the two treatments (at least qualitatively). Suppose that $0 < \beta < \bar{\beta}$ for the VAE players. As discussed above, when two VAEs play each other, in equilibrium each should learn a recognition model with parameter $\lambda = \bar{\lambda}$, the fixed point of the mapping $\bar{\Phi}$ graphed in the left panel of Figure 5. If instead a VAE plays against a computer that implements the deterministic choice rule, then the VAE should learn the recognition model that is the best response to an opponent with a recognition model with $\lambda' \to +\infty$, $\nu' = 0$. As shown above, this is a symmetric recognition model $(\hat{\lambda},\, 0)$, where $\hat{\lambda}$ is the limit of $\bar{\Phi}(\lambda')$ as $\lambda' \to +\infty$.

Thus in both cases, a VAE player should learn a symmetric recognition model, but the value of λ should be larger in the case of the VAE that plays against a computer, since it follows from the monotonicity of $\bar{\Phi}(\lambda')$ that

$$\bar{\lambda} \;=\; \bar{\Phi}(\bar{\lambda}) \;<\; \lim_{\lambda' \to +\infty} \bar{\Phi}(\lambda') \;=\; \hat{\lambda}.$$

(See Appendix section B.2 for numerical illustration of the difference.) Since the value of λ directly determines the steepness of the logistic curve Prob[leave $|x$] for a VAE player, our model predicts more random behavior when two players with endogenously imprecise behavior play against one another, than when one of them plays a deterministic automaton. Note that in our theory, what is important is not whether a player is human or a machine, nor whether a player attributes intentionality to the other player, but simply whether both players endogenously adjust their decision procedures to minimize a training objective of the form (4) with only a moderate weight β on accuracy of action selection.

7 Learning from a Finite Body of Experience

Thus far we have only sought to characterize play in an equilibrium after both players' internal models have come to be perfectly adapted to both the statistics of the external environment (the prior π) and to the statistics of each others' behavior. But an important strength of the VAE approach is that it also provides a model of the way in which decision algorithms can be trained on the basis of a finite body of experience. An important advantage of adaptive learning models of this kind is that — in addition to addressing questions about whether a state of equilibrium should actually be reached, even in a stationary environment — they allow us to analyze the dynamics of adaptation when the statistics of the environment shift.

The experimental results of Arifovic et al. (2013) indicate that in coordination games there is inertia in the strategies played as the underlying environment shifts. In particular, they conduct an experiment in which the initial situation is one in which players tend always to choose one action, followed by a change in the environment to a symmetric game (like the one analyzed above) in which coordination on either action represents an equally plausible equilibrium. They find that players continue to be more likely to choose one action over the other one even in the symmetric environment, with the direction of the bias determined by the direction of the asymmetry in the original environment. Thus experimental behavior exhibits history-dependence: behavior continues to be biased by the incentives that players faced in a past environment, long after the environment has changed. We show that our model (like the evolutionary model of Arifovic et al. (2013)) predicts behavior of this kind; indeed, we regard this as one of its strengths. But to show this, we must consider how the VAE parameters should be adjusted as additional experience accrues, rather than simply describing the equilibrium that should eventually be reached after long enough experience within a stationary environment.

7.1 A VAE Training Algorithm

We now introduce an explicit model of how the players' VAEs can be trained on the basis of a finite body of experience. This allows us both to study the history-dependence of behavior during a transition from one environment to a new one, and the convergence of the learning dynamics to an eventual equilibrium of the kind characterized above.

We posit an online algorithm for the iterative adjustment of the parameters of each of the players' VAEs as experience accrues. In each time period t, each player computes new estimates (θ_t, ϕ_t) of the parameters that specify their generative and recognition models; these parameter estimates are chosen to minimize the estimated value of the training objective (4), given the sample of experience to that point. As explained above, the generative model θ that minimizes (4) is the one that sets the parameters $\{q_\theta(j), \mu_j, \sigma_j^2\}$ equal to the corresponding first and second moments of the joint distribution for (j, x) implied by the prior and the player's recognition model. We assume that in each period, the estimated parameters θ_t are given by the moments of the *empirical* distribution to that point — that is, the moments of the sample $\{x_\tau, j_\tau\}$ for $\tau = 1, 2, \ldots, t$.[28]

Our online algorithm for estimating the parameter vector ϕ is based on the observation that the parameters that minimize (4), given a generative model θ and a joint distribution for the variables (x, a') implied by the prior and the behavior of the other player, must satisfy a vector of first-order conditions of the form[29]

$$\mathrm{E}[\Delta(x, a') \cdot \nabla(x)] \;=\; 0, \tag{13}$$

where the vector

$$\nabla(x) \;\equiv\; \nabla_\phi p_\phi(1\,|x)$$

is the gradient of the recognition rule evaluated at any state x (a function that depends on the parameters ϕ), and $\mathrm{E}[\cdot]$ denotes an expected value over the joint distribution for the variables (x, a'). The function $\Delta(x, a')$ depends on the parameters (θ, ϕ); it can be shown to be a function of

[28] Some of these sample moments will be undefined for low values of t; for example, the algorithm cannot estimate μ_j until there has been at least one period in which the state is classified using latent state j. But the way that we define the estimates in these early periods has no consequences for the predicted later evolution of the estimates.

[29] See Appendix section C for details.

the form

$$\Delta(x, a') \;\equiv\; v(x)^{\mathrm{T}}\phi \;-\; m(x, a')^{\mathrm{T}}\psi(\theta). \tag{14}$$

Here the vector-valued functions $v(x)$ and $m(x, a')$ are defined independently of the VAE parameters; ϕ is the vector whose elements are the parameters (λ, ν) and $\psi(\theta)$ is a vector of nonlinear functions of the parameters θ; and we use a superscript T to denote a transpose.

Hence the first-order conditions (13) can be written in the form

$$M\,\phi \;=\; N\,\psi(\theta), \tag{15}$$

where the moment matrices

$$M \;\equiv\; \mathrm{E}[\nabla v^{\mathrm{T}}], \qquad N \;\equiv\; \mathrm{E}[\nabla m^{\mathrm{T}}] \tag{16}$$

are expected values of functions of (x, a') that are independent of the parameters θ, and depend on ϕ only through its role in the calculation of $\nabla(x)$. We can thus express M and N as the expected values of functions of the triple of random variables (x, a', ∇), where the expected value is over the joint distribution for these three variables.

The learning algorithm goes through the following sequence of steps in each time period t:[30]

1. A new observation x_t is drawn from the prior distribution associated with the current environment. Each player's algorithm computes a value for the gradient vector $\nabla_t = \nabla(x_t)$, using the parameters ϕ_{t-1} of their existing recognition model.[31]

2. Each player's VAE selects a latent state j_t with which to represent the current state x_t; the stochastic encoding rule is determined by the player's existing parameter estimates ϕ_{t-1}. The two players' stochastic encoding rules are conditionally independent. Each player then selects the action $a_t = a(j_t)$ recommended by the decision rule specified in section 4.4.

3. New estimates M_t, N_t of the moment matrices defined in (16) are computed, taking into account the new observations (x_t, a'_t, ∇_t). The moments are computed on the assumption that the complete sample of observations $\{x_\tau, a'_\tau, \nabla_\tau\}$ for $\tau = 1, \ldots, t$ represents a succession of independent draws from the same joint distribution, the moments of which are to be estimated. This allows the new estimates M_t, N_t to be computed recursively on the basis of the previous estimates M_{t-1}, N_{t-1} and the new observations (x_t, a'_t, ∇_t).

[30] The explicit equations that define the algorithm are given in Appendix section C.

[31] In period $t = 1$, the algorithm starts with an arbitrary specification of the "existing" recognition model ϕ_0 for each player. In any later period, the parameters ϕ_{t-1} are the ones determined by the algorithm in the previous period.

4. Similarly, new estimates of the moments of the joint distribution of (j, x) are computed, taking into account the new observations (j_t, x_t). Again the moments are computed on the assumption that the complete sample of observations $\{j_\tau, x_\tau\}$ through period t represents a succession of independent draws from the same joint distribution, and again this allows recursive updating of the estimated moments. The new estimated parameters θ_t for each player's generative model are set equal to these moment estimates.

5. The new estimates ϕ_t of the parameters of each player's recognition model are the ones implied to be optimal by the first-order conditions (15), given the player's current estimates of both the moment matrices and the generative model parameters. Thus the new estimates are given by[32]

$$\phi_t \;=\; M_t^{-1} N_t \psi(\theta_t). \tag{17}$$

The same sequence of steps is then repeated in period $t+1$ (now using the vector ϕ_t to determine the "existing" recognition model), and so on.

Note that this algorithm has the property that, if the VAE parameter estimates (θ_t, ϕ_t) eventually converge, then the moment matrices M, N will converge as well, the resulting estimates will satisfy the FOCs (15) for the optimality of ϕ, and the parameters θ will be optimal as well. In our numerical experiments with the algorithm, it does converge, and to the VAE parameters associated with the symmetric equilibrium characterized above.

7.2 Convergence to Equilibrium and History-Dependence

Here we present a simple example of the learning dynamics implied by the algorithm just proposed. We compare two scenarios: in one, for the first 1000 periods, the state x is drawn from the asymmetric prior distribution L, while in the other, for the first 1000 periods it is drawn from asymmetric prior distribution H. In either case, from period $t = 1001$ onward, the state is instead drawn from the symmetric distribution M. Each of the three priors is a Gaussian distribution with standard deviation $\omega_{high} = 5$, but their means are different: the mean is 5 in environment H, 0 in environment M, and -5 in environment L. We are interested in how the VAEs adjust to the change in the environmental distribution at time $t = 1001$, and in the extent to which behavior in periods $t > 1000$ continues to be influenced by which of the two environments was experienced in periods $t \leq 1000$.

We have characterized the symmetric equilibrium associated with a symmetric prior such as M above; here we consider whether such an equilibrium is eventually reached if the two VAEs use our

[32] As explained in Appendix section C, the matrix M_t will be non-singular (and positive definite) with probability 1 in all periods $t \geq 2$. An arbitrary rule can be used to select the vector ϕ_1 in period 1, with little consequence for the subsequent dynamics.

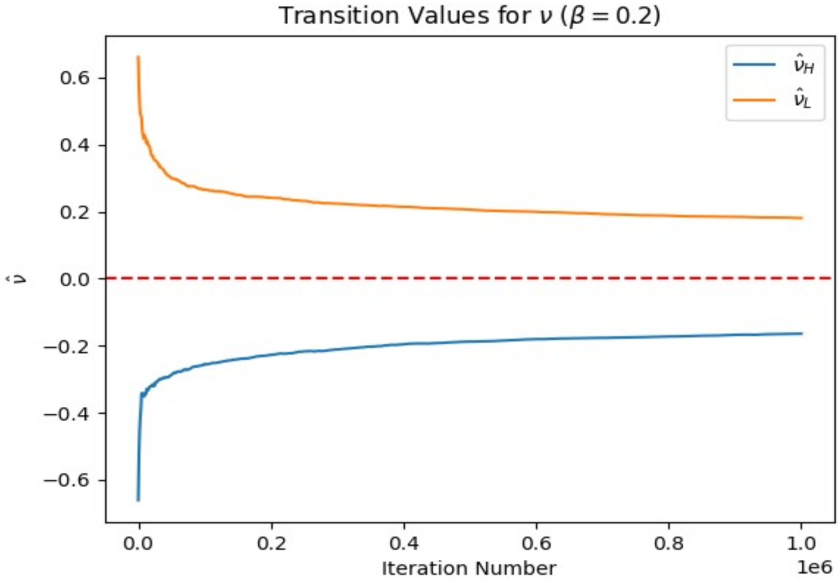


Figure 10: Evolution of the estimated value of the bias parameter ν_t, under simulated transition dynamics from two different initial conditions (H, L). Path shown is the average of the estimates by the two players. The dashed line shows the predicted long-run (equilibrium) parameter value for prior distribution M.

proposed training rules. The equations discussed above can also be used to define an equilibrium in the case of an asymmetric prior such as L or H; in this case, the equilibrium is necessarily asymmetric, but can again be computed as a zero of the mapping Ψ.

In the numerical simulations reported here, we start the algorithm in period $t = 1001$, assuming existing VAE parameters $\theta_{1000}, \phi_{1000}$, and corresponding moment matrices M_{1000}, N_{1000}, equal to the equilibrium predictions under the asymmetric prior (L or H as the case may be). We then simulate the learning dynamics given these initial conditions, with new states $\{x_t\}$ drawn from prior distribution M for all $t \geq 1001$. Results are shown in the case of training objectives of the form (4) with $\beta = 0.2$.

The evolution of the coefficient estimate $\hat{\nu}_t$ under each of the two scenarios is shown in Figure 10.[33] In fact, there are two values of this parameter each period, under either of the scenarios: an estimated value for each of the two VAEs. The two values are similar, in a given scenario, because both VAEs have the same training objective and observe the same sequence of external states; but they are slightly different because of the independent randomness in the way that the two recognition models choose the latent states j_t and j'_t. The figure plots a single trajectory for each scenario, corresponding to the average each period of the estimates by the two VAEs. Similar figures could be displayed for the evolution of the estimated sensitivity parameter $\hat{\lambda}_t$, but the path

[33] The horizontal axis measures the number of time periods since the transition to distribution M. Thus "iteration number" 1 corresponds to $t = 1001$, and so on.

of the bias parameter $\hat{\nu}_t$ is especially interesting because of how different the estimated values initially are under the two scenarios.

When the VAEs are initially exposed to draws from distribution L (so that $x < 0$ most of the time), they start the simulation with a large positive estimate of ν (that is, a strong bias toward staying), whereas when they are initially exposed to H (so that $x > 0$ most of the time), they start with a large negative estimate (a strong bias toward leaving). Once the state begins to be drawn instead from the symmetric distribution M, the size of the bias is reduced (regardless of its initial sign). During the first 1000 or so periods in the new environment, convergence is fairly rapid, but it becomes much slower as further experience is accumulated.

The equilibrium value of the parameter associated with the new environment M (i.e., $\nu = 0$) is shown as a horizontal dashed line in the figure. We present further evidence in Appendix section C.3 indicating that as the simulations are run farther, the estimates appear to converge to precisely those associated with the symmetric equilibrium. The same is true of other parameters, such as the VAE estimates of λ. Thus the equilibrium predictions discussed above are also the predictions of our model of learning dynamics, under the assumption that a long enough time is spent in a stationary environment.

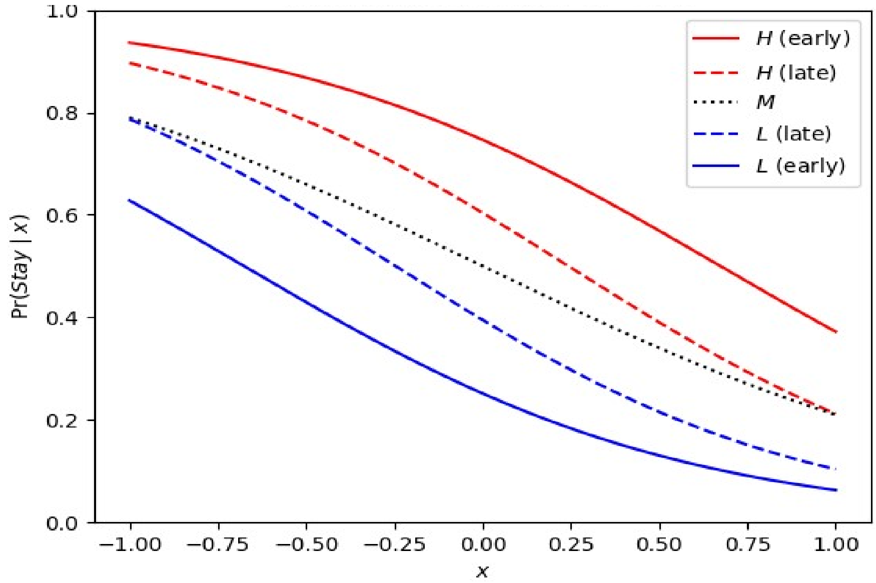


Figure 11: The probability of a player's choosing to stay, as a function of the next value of x that is drawn (the horizontal axis), given the parameters ϕ_t of their recognition model after a certain amount of experience with environment M, under each of the two scenarios depicted in Figure 10. The curves average the probabilities of staying for the two players (which are nearly identical). For each scenario, the implications of the learned recognition model are shown after 2000 draws from distribution M (the "early" curves), after 500,000 draws (the "late" curves), and in the long-run equilibrium (dashed curve).

But while our model predicts that convergence to the equilibrium estimates (and hence to equilibrium behavior) should occur eventually, the convergence can be relatively slow. Hence we should expect to see visible history-dependence in experiments like those of Arifovic et al. (2013), where there is a change in the distribution from which the payoff matrix is drawn on individual trials.

The predicted change in subjects' propensity to stay over the course of our two scenarios is shown in Figure 11. The solid curve labeled "H (early)" plots the behavioral function Prob$[\text{stay}\,|x]$ in period $t = 3000$, under the scenario where the states were initially drawn from distribution H; the curve labeled "L (early)" plots the corresponding function for the scenario where the states were initially drawn from L. One observes that these curves are quite different, despite the fact that in both cases the VAEs have at this point observed 2000 successive draws from distribution M. (In one case, the VAEs would choose to stay more than 70 percent of the time for values of x near zero, while in the other they would leave more than 70 percent of the time under the same circumstances.)

The dashed curve labeled "H (late)" plots the same behavioral function, again under the scenario where the states were initially drawn from H, but now for period $t = 501,000$. The dashed curve labeled "L (late)" shows the corresponding curve when the environment was initially L. These two curves remain visibly different, even after 500,000 periods of common experience, though less far apart than the two solid curves. As $t \to \infty$, both curves convergence to the dotted curve, representing equilibrium behavior in the case of the distribution M. (This curve is the same as the flatter curve in Figure 9.)

8 Conclusion

We offer a model of imprecise action selection in strategic environments, the implications of which are illustrated in the context of a simple (but much-studied) class of coordination games. Our approach is based on the supposition that players choose their actions on the basis of an algorithm with the structure of a VAE, the parameters of which have been trained on the basis of a finite sample of prior experience, both with the environment and their opponent's pattern of play. We show that our model can explain the persistence of randomness in players' action selection conditional on the payoffs available on a given trial, even after extensive training, under two crucial assumptions: (i) that individual decision situations are classified using one or another of a small set of latent states, and that only this latent state is available as a basis for action choice, and (ii) that there is a sufficient degree of concern to have an internal model (generative model), used to interpret the latent state and choose an appropriate action, with the property of fairly high congruence between the joint distribution of external states and latent states implied by the generative and the joint distribution that is experienced.

We show moreover that the random action selection predicted by our model captures important features of experimentally observed behavior in experiments like those of Frydman and Nunnari (2023). Notably, the probability of a player's choosing to $stay$ falls monotonically (but only gradually, rather than discontinuously) with increases in the value x of the outside option that can be obtained by choosing to $leave$. The sensitivity of choice probabilities to the value of x also varies with the range of variation in x across trials (i.e., with the variance of the prior distribution from which x is drawn on each trial): for a given value of x, choices are predicted to be (and observed to be) more random when there is greater prior uncertainty about$x$. Finally, our model predicts that choice probabilities in a given decision situation parameterized by x will depend on the history of past values of x that have been encountered — not on the distribution from which x is drawn in the *current* environment, but the distributions from which previously experienced values of x have been drawn. Thus our model can explain the path-dependence of the behavior upon which subjects coordinate in experiments like that of Arifovic et al. (2013).

Our conclusions depend importantly on the restricted classes of possible generative and recognition models that are allowed by the VAE architecture. This raises an important question about the particular classes of models that should be assumed in our approach. While the topic clearly deserves further study, we have suggested that our restriction of attention to a particular parametric family of encoding models (recognition models) in our numerical results may not be crucial for our qualitative conclusions; our restriction to a particular parametric family of generative models (Gaussian mixture models with some finite number of components) seems instead to be more critical.

We have argued that it is reasonable to suppose that people learn using algorithms that search over only some finitely-parameterized space of possible generative models, on the ground that this makes it possible to learn a rule of behavior that is appropriate to an environment using only a sample of modest size of experience of that environment. But this still leaves open the question of which finitely-parameterized class of generative models it makes sense to expect the learning process to entertain. Consideration of possible grounds for selecting such a family, and investigation of the degree to which our conclusions are sensitive to the particular class that is used, will be important topics for further research.

References

Alemi, A., B. Poole, I. Fischer, J. Dillon, R. A. Saurous, and K. Murphy (2018). Fixing a broken elbo. In *International Conference on Machine Learning*, pp. 159–168.

Arifovic, J. and J. H. Jiang (2019). Strategic uncertainty and the power of extrinsic signals – evidence from an experimental study of bank runs. *Journal of Economic Behavior and Organization 167*, 1–17.

Arifovic, J., J. H. Jiang, and Y. Xu (2013). Experimental evidence of bank runs as pure coordination failures. *Journal of Economic Dynamics and Control 37*(12), 2446–2465.

Bowman, S., L. Vilnis, O. Vinyals, A. Dai, R. Jozefowicz, and S. Bengio (2016). Generating sentences from a continuous space. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pp. 10–21.

Chen, X., D. P. Kingma, T. Salimans, Y. Duan, P. Dhariwal, J. Schulman, I. Sutskever, and P. Abbeel (2017). Variational lossy autoencoder. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.

Friedman, E. (2020). Endogenous quantal response equilibrium. *Games and Economic Behavior 124*, 620–643.

Frydman, C. and S. Nunnari (2023). Coordination with cognitive noise.

Goeree, J. K., C. A. Holt, and T. R. Palfrey (2016). *Quantal response equilibrium: A stochastic theory of games*. Princeton University Press.

Heinemann, F., R. Nagel, and P. Ockenfels (2004). The theory of global games on test: experimental analysis of coordination games with public and private information. *Econometrica 72*(5), 1583–1599.

Heinemann, F., R. Nagel, and P. Ockenfels (2009). Measuring strategic uncertainty in coordination games. *Review of Economic Studies 76*(1), 181–221.

Kingma, D. P. and M. Welling (2014). Auto-encoding variational bayes. In Y. Bengio and Y. LeCun (Eds.), *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*.

Kingma, D. P., M. Welling, et al. (2019). An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning 12*(4), 307–392.

Malloy, T., T. Klinger, and C. R. Sims (2022). Modeling human reinforcement learning with disentangled visual representations. *Reinforcement Learning and Decision Making (RLDM)*.

Mauersberger, F. (2022). Thompson sampling: A behavioral model of expectation formation for economics. *Available at SSRN 4128376*.

Morris, S. and H. S. Shin (2003). Global games: Theory and applications. In M. Dewatripont, L. Hansen, and S. Turnovsky (Eds.), *Advances in Economics and Econometrics: Eighth World Congress*, pp. 56–114. Cambridge University Press.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Tucker, M., J. Shah, R. Levy, and N. Zaslavsky (2022). Towards human-agent communication via the information bottleneck principle. *arXiv preprint arXiv:2207.00088*.

Woodford, M. (2020). Modeling imprecision in perception, valuation and choice. *Annual Review of Economics 12*, 579–601.

Yang, M. (2015). Coordination with flexible information acquisition. *Journal of Economic Theory 158*, 721–738.

APPENDIX

A Alternative Models of Imprecise Coordination

Here we compare our VAE model with a variety of other models of imprecise decision making in strategic situations, some of which have also been applied to coordination games of the kind treated in this paper.

A.1 Rational Inattention

The model of rational inattention proposed by Sims (2003) has been applied to a coordination game of the kind discussed in section 3 by Yang (2015). Like our model, RI models predict that choice should be stochastic, as a result of its being based on a noisy internal representation of the situation (like the latent state in our model), rather than directly on the exact external state x; and as in our model, the conditional probabilities of different internal representations are optimized for the prior distribution associated with a given environment. Indeed, the kind of RI model analyzed by Yang (2015) can be considered a special case of the generalized version of our model presented in section 2.2.

A.1.1 RI as a Special Case of the VAE Framework

The Sims (2003) model of rational inattention can be viewed as a model of same general type as the models discussed in section 2.2, and more specifically, as a special case of the generalized VAE model with training objective (6). It assumes, however, that for any choice of the encoding model ϕ, the generative model is given by

$$q_\theta \;=\; q_\phi, \qquad \tilde{p}_\theta(x|j) \;=\; \pi_\phi(x|j) \quad \forall j. \tag{18}$$

That is, the decoding of the latent states is assumed to reflect correct Bayesian inference; no reference to a separately specified "generative model" is needed.

But this is also what the solution to the problem (6) requires, if (i) we rewrite (6) in the form (7), and consider the limiting case in which both $\beta_1 - 1$ and β_2 are negligible (but the ratio $(\beta_1 - 1)/\beta_2 = \psi > 0$ remains well-defined); and (ii) the class of possible generative models is necessarily flexible enough to allow θ such that (18) holds, regardless of the choice of the encoding model. In such a limiting case, the optimal generative model θ is given by (18), for any choice of ϕ, because of the identity (1). Then $D_{KL}(p_\phi||\tilde{p}_\theta) = 0$ regardless of the choice of ϕ, and R is equal to I, the mutual information between x and j implied by the encoding model ϕ. The problem then reduces to a choice of the encoding rule to solve

$$\min_{\phi} L + \psi I, \tag{19}$$

as assumed in the RI literature. The standard RI problem also assumes that there is no restriction on the class of possible encoding rules (including no limit on how large the set of latent states may be).

This yields a highly parsimonious theory, but arguably an unrealistic one, in that it assumes that a very difficult (high-dimensional) optimization problem is solved, and that the solution should depend on more information about the environment than can be obtained from any finite body of experience. The fact that the RI model represents a limiting case of the problem (6), however, means that the solution to a more computationally tractable version of this problem can be viewed as an approximation to an RI model.

Note that it does not follow, though, that versions of the general problem (6) in which $\beta_1 > 1$ must be the ones of economic interest. RI theory assumes that $\beta_1 > 1$ (i.e., that $\psi > 0$) because it is only in this case that the theory implies that decision making should be at all imprecise. (The mutual information term in problem (19) would otherwise be irrelevant.) In our theory, instead, it will in general not be possible to simultaneously achieve $D_{KL}(p_\phi || \tilde{p}_\theta) = 0$ and perfectly accurate choice; hence even when $\beta_1 \leq 1$, often VAE parameters will be learned that imply stochastic choice.

A.1.2 Implications of RI for the Coordination Game

We have seen (Figure 2) that subjects' behavior appears to respond less accurately to the exact state x than is required for an equilibrium with perfect optimization, and RI models with $\psi > 0$ provide one possible explanation for imprecision in this sense. But in fact the kind of behavior shown in the figure is not well captured by the kind of imprecise decision making assumed in RI models, and applied to this kind of coordination game by Yang (2015). The information structure that solves the problem (19), when applied to a binary choice problem like the one faced by players in this game, involves exactly two possible latent states ("signals" in the terminology of Sims), one of which leads the DM to leave and the other of which leads them to stay. The implied probabilities of action choice are then of the form

$$\mathrm{P}(a\,|x) \;=\; \frac{\bar{p}(a)\exp[-\psi^{-1}\bar{\mathcal{L}}(a,x)]}{\sum_{\tilde{a}\in A}\bar{p}(\tilde{a})\exp[-\psi^{-1}\bar{\mathcal{L}}(\tilde{a},x)]}, \tag{20}$$

for each of the two actions $a \in A$, where $\bar{p}(a)$ is the unconditional probability of choosing action a (or of receiving the signal that leads to this choice), and

$$\bar{\mathcal{L}}(a,x) \;\equiv\; \mathrm{E}[\mathcal{L}(a,a';\,x)\,|a,x]$$

is the expected loss from choosing action a in state x, given the stochastic choice rule (conditional on x) of the other player.

If we restrict attention to solutions that conform to both of the two symmetries of the game discussed in section 3.1 of the main text, then the choice probabilities conditional on x will be the same for both players, and in addition, for either player we will have $\mathrm{P}(a\,|x) \;=\; \mathrm{P}(-a\,|-x)$ for both a and all x, where $-a$ means the action opposite to a. The symmetry of the prior density then implies that in such a solution we must have $\bar{p}(a) \;=\; \bar{p}(-a) \;=\; 1/2$. Substituting this solution for $\bar{p}(a)$ into (20), we conclude that

$$p(x) \;\equiv\; \mathrm{P}(\mathit{stay}\,|x) \;=\; [1 \,+\, \exp(-2\psi^{-1}\Delta U)]^{-1}, \tag{21}$$

where we again use the notation

$$\Delta U \;\equiv\; \bar{\mathcal{L}}(\mathit{leave}\,|x) \,-\, \bar{\mathcal{L}}(\mathit{stay}\,|x)$$

for the net expected reward from staying rather than leaving,[34] given the state x and the implied probability of the other player's staying.

We can again define an equilibrium correspondence as the set of all pairs (x,p) with the property that if the state is x and the other player's probability of staying is p, the RI solution (21) implies that a player should stay with exactly probability p. Using (9) to substitute for ΔU in (21), we conclude that (x,p) belong to this correspondence if and only if

$$x \;=\; (2p-1) \,-\, \frac{\psi}{2}\log\left(\frac{p}{1-p}\right). \tag{22}$$

A symmetric RI equilibrium is then characterized by a function $p(x)$ such that (i) for any real number x, the pair $(x,p(x))$ belong to the equilibrium correspondence, and (ii) $p(-x) = 1-p(x)$ for all x.

It is easy to show that such an equilibrium exists.[35] Note that there is a single value of x consistent with (22), for each value $p \in (0,\;1)$. This defines a function $x(p)$ that is continuous; with

[34] Here we identify the "loss" in the RI problem (19) with the negative of the reward shown in the game payoff matrix (2).

[35] In fact, in the case of any small enough value of ψ, there are many possible solutions. See Yang (2015) for discussion of the indeterminacy of the RI solution.

a range equal to the entire real line; and with the symmetry $x(1-p) = -x(p)$ for all $p \in (0,\,1)$. Hence for any $x > 0$, one can find at least one $p \in (0,\,1)$ such that $x(p) = x$; let this be our solution for $p(x)$ when $x > 0$. Then setting $p(0) = 1/2$ and $p(x) = 1 - p(-x)$ for any $x < 0$, one obtains a solution for all x with the desired properties. Moreover, the solution necessarily involves $0 < p(x) < 1$ for all x, so that the theory predicts stochastic action selection over a range of values of x, as observed in the behavior of the experimental subjects shown in Figure 2.

However, it remains the case, at least for all small enough values of the information cost parameter ψ, that the predicted action probabilities do not decline gradually with increases in x, in the way seen in Figure 2. Note that for any $p \in (0,\,1)$, equation (21) implies that $x(p) \to 2p - 1$, the function characterizing the equilibrium correspondence in Figure 1, in the limit as $\psi \to 0$. At the same time, for any $\psi > 0$, equation (21) implies that $x(p) \to -\infty$ as $p \to 1$ and that $x(p) \to +\infty$ as $p \to 0$. Hence for any small enough value of $\psi > 0$, the graph of the equilibrium correspondence must be similar to the one shown in Figure 1, but with the corners smoothed (a backwards-S shape rather than a Z shape). It then follows that any single-valued function $p(x)$ that is a selection from the correspondence must involve a discontinuous jump at one or more values of x. In fact, one can easily show that $x(p)$ is a non-monotonic function in the case of any cost parameter $\psi < 1$, so that a symmetric RI equilibrium must involve a discontinuous jump in any of those cases. Yet no such discontinuity is visible in Figure 2, in the case of either of the two treatments.

There is another important respect in which Figure 2 is problematic for the RI model. As Frydman and Nunnari (2023) stress, the rate at which the probability of choosing to stay declines with increases in x is sharper when the state x is drawn from the low-variance distribution than when it is drawn from the high-variance distribution. Yet equation (22) describes the equilibrium correspondence for the RI model, regardless of the prior distribution from which x is drawn: we have only assumed that the prior is symmetric (in the sense that $\pi(-x) = \pi(x)$ for all x), as is true in both of the treatments of Frydman and Nunnari. Hence if $p(x)$ represents a symmetric RI solution in the case of one symmetric prior, it must also represent a symmetric RI solution in the case of every other symmetric prior; and the two choice curves shown in Figure 2 cannot both represent selections from the same equilibrium correspondence. Thus the treatment effect shown in Figure 2 is contrary to the prediction of the RI model.

In the main text, we show instead that our VAE model of imprecise action selection predicts stochastic choice for all values of x; a function $p(x)$ that decreases continuously and monotonically with increases in x; and one that should be flatter when the variance of the prior is greater — all features of the experimentally observed behavior displayed in Figure 2.

A.2 Other Models of Imprecise Coordination

There are a number of other approaches that also model subjects' choices as stochastic conditional on the current value of x, that we discuss more briefly.

Probably the most influential model of stochastic choice in strategic settings of the kind considered here is the theory of quantal response equilibrium (QRE: Goeree et al. (2016)). QRE assumes that each player's action selection probabilities are determined by the (correctly learned) average equilibrium payoffs associated with the alternative actions. But instead of assuming that the action with the higher average equilibrium payoff is chosen with certainty, the probability of staying is assumed to be a continuously increasing function of the difference in expected payoffs ΔU. As Frydman and Nunnari (2023) discuss, QRE predicts that there should be no change in $\mathrm{P}(stay\,|x)$ as a function of x across environments with different prior distributions, of the kind seen in Figure 2. The reason is closely related to our discussion above of why RI theory predicts no change, if the parameter ψ is the same across environments. If we restrict attention to symmetric equilibria, as in our discussion above, then we must have $\bar{p}(a) = 1/2$ for both actions; and in this case (20) reduces to a logistic function of the expected payoff difference, a commonly used specification in QRE models ("logit QRE").

This assumes that the QRE parameter specifying the sensitivity of choice to the expected payoff difference (analogous to the parameter ψ in (20)) should remain the same across environments. Friedman (2020) instead proposes a generalization of standard QRE modeling in which the precision parameter is endogenously determined for a particular game. However, endogenizing the precision parameter for each possible game (i.e., each possible payoff matrix) still implies that equilibrium action selection probabilities should be determinate functions of x, independently of the prior from which x is drawn; thus without introducing cognitive imprecision of additional sort (as in our approach), there would still be no reason for observed behavior to differ across the two treatments of Frydman and Nunnari (2023).

Frydman and Nunnari (2023) instead propose a model in which (as in our model) decisions are assumed to be based on an imprecise internal representation of the external state; they further posit a model in which the precision of encoding varies endogenously with the statistics of the environment, based on models of efficient coding from the neuroscience literature on neural coding of sensory features. In their model, the internal representation is specified by a real number r (rather than a discrete latent state, as assumed in our model), and the encoding probabilities $p(r\,|x)$ are assumed to be the same functions of a rescaled state variable ("range normalization"), regardless of the variance of the prior distribution. Our model also has this property, if instead of minimizing a weighted sum of D_{KL} and L, we consider the closely related problem of minimizing L subject to an upper bound $\bar{D}$ on the admissible size of D_{KL}. If we assume that the bound $\bar{D}$ is invariant across different prior distributions for x, the VAE model will also predict range normalization. Because our model also assumes an imprecise encoding rule, the precision of which varies endogenously in order to minimize a loss function, it can be viewed as a type of efficient coding model, and in this respect our explanation for context-sensitive behavior is closely related to that of Frydman and Nunnari (2023).

There is however a subtle difference between our formulation and theirs. Efficient coding models emphasize the existence of computational constraints on the class of possible encoding models (often taken to represent neurobiological constraints), but frequently assume optimal decoding of the information contained in the imprecise internal representation (as Frydman and Nunnari (2023) do). Our model assumes restrictive parametric families for both encoding and decoding (as is standard in the VAE literature), but we would argue that it is the restricted class of possible generative models (decoding models) that is crucial for our conclusions. Given our assumed class of possible generative models (with only two latent states), even in the case of a fully flexible recognition model it will not be possible to achieve a value of $D_{KL}(p_\phi || \tilde{p}\theta)$ lower than a specified bound (for a low enough value of the bound $\bar{D}$), except by choosing a value of μ that is not too large and a value of σ that is not too small — so that the marginal distribution for x that is implied by the generative model is not too different from the prior distribution $\pi(x)$ — and by choosing a recognition model that is stochastic, so that the joint distribution for (x, j) implied by the recognition model is not too different from the one implied by the generative model. Moreover, the optimized values of μ and σ will be proportional to the value of ω, as in our discussion above, and because of this, the optimal recognition model will again imply a discrimination threshold proportional to the value of ω, even when it is chosen with complete flexibility. Thus it is the coarseness of the class of possible generative models — which we justify on the ground that it allows the generative model to be learned on the basis of even a small sample of experience — that plays the crucial role in our analysis.

Mauersberger (2022) offers another model of imprecise action selection in strategic settings, based on the idea of "Thompson sampling" from the machine learning literature on bandit problems. He proposes that at each decision point, a DM samples one element from a correct Bayesian posterior over possible decision situations, and then chooses the action that would be optimal for that situation; sampling a single element (rather than optimizing against the entire posterior distribution) introduces intrinsic randomness into the DM's behavior. And since an environment with greater prior uncertainty should lead to more dispersed posteriors as well, this model provides an explanation for noisier behavior in more uncertain environments. (Indeed, Mauersberger stresses this result, and contrasts it with the QRE prediction.) But the motivation for optimizing against a single sample remains unclear.[36] Our VAE approach instead provides an explanation for ran-

[36] In the literature on Thompson sampling, sampling is typically argued to provide a desirable degree of experimentation with options that might turn out to be better than had been believed on the basis of incomplete evidence. But in

domness in the classification of an objective situation summarized by the value of x, as discussed above. The stochastic recognition rule can be viewed as similar to sampling from a posterior (over the possible latent states that may have given rise to the current situation, according to the generative model); the decision rule $a(j)$ then selects an action that is optimal under the beliefs about the situation implied by that latent state (according to the generative model). Thus our model can be viewed as providing a justification for something similar to posterior sampling.

Finally, Arifovic et al. (2013) propose an evolutionary model to explain the variability of observed behavior in a multi-player version of the kind of coordination game that we study here. In this kind of model, predicted play is random as a result of randomness in both the mutations through which new strategies enter the population of active strategies and the randomness of the process of selection through which less-successful strategies are culled. While the architecture that we propose is fairly different from the one that they assume, we also model the way in which the strategies used by players evolve over time to be better adapted to their environment (including the effects of the evolving strategies of other players), and our learning algorithm (described in section 7 and in Appendix C below) also involves an element of random sampling. Our model also shares with theirs the feature that players optimize their behavioral models over a restrictive class of possible algorithms, in response to observed play by others. The use of a restrictive class of possible algorithms may well be important for explaining why the degree of imprecision observed in the behavior of human subjects should persist even with experience.

B Additional Figures

B.1 Symmetric Equilibrium of the Game Between VAEs

In section 5.3 of the main text, we discuss how the properties of the symmetric equilibrium vary with the value of β, for some fixed value of ω. Here we provide additional visual illustration of what the equilibrium VAEs of the players are like, and how they vary as the value of β is increased.

The optimal encoding and decoding models are shown for a few values of β, and two different values of ω (corresponding to the two experimental conditions in the experiment of Frydman and Nunnari (2023)) in Figure 12. The panels in the left column display the joint distribution for (j, x) implied by the generative model by plotting the conditional density function for x associated with each of the two latent states $j = -1, 1$, which states are assigned equal unconditional probability. (Each panel actually shows four conditional distributions, because two different generative models are shown in each panel: the optimal generative model when $\omega = \omega_{low}$ (in blue) and the one when

the coordination game discussed here, there is no need to choose a particular action in order to learn more about the distribution of payoffs associated with that action, and hence no benefit from experimentation.

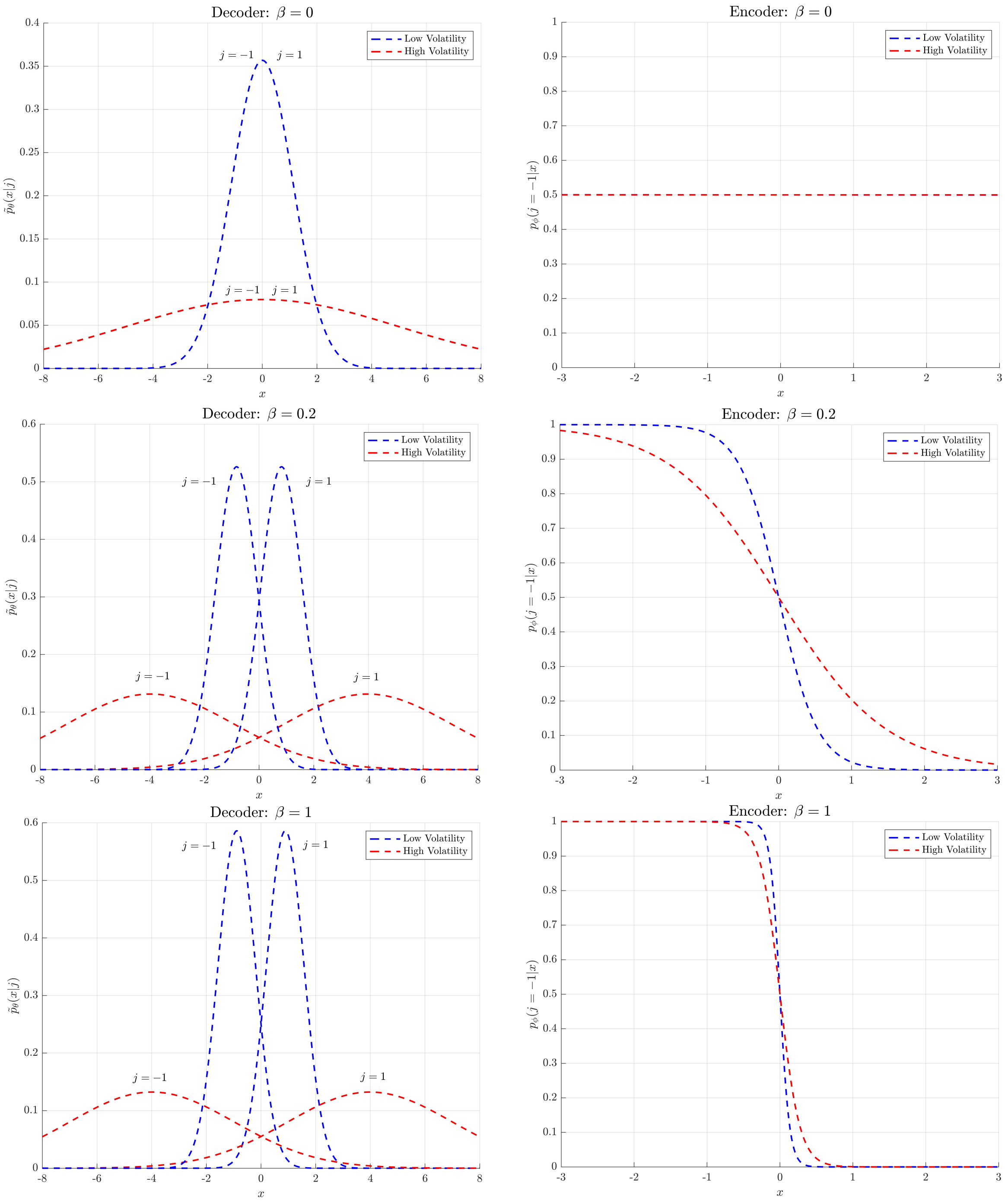


Figure 12: Equilibrium generative and recognition models (common to both players), for each of several values of β (the different rows). Each panel shows the results for two different symmetric priors, the case $\omega = \omega_{low}$ (blue) and the case $\omega = \omega_{high}$ (red).

$\omega = \omega_{high}$ (in red).) The panels in the right column display the corresponding recognition (or encoding) models. Note that when there are only two latent states, as assumed here, the recognition model is completely specified by plotting the function $p_\phi(-1\,|x)$ as a function of x. (Each panel contains two curves, though, again corresponding to the two different values of ω.)

The first row shows the optimal VAE for the training objective with $\beta = 0$ (which corresponds to the standard VAE model of Kingma and Welling (2014)). In this case, the optimal encoder (recognition model) fails to distinguish at all between different external states ($\lambda = 0$), making the latent state completely uninformative about the latent state. As a consequence, the optimal generative model is one in which the distribution $\tilde{p}_\theta(x\,|j)$ is the same for both latent states, and equal to the prior distribution $\pi(x)$.[37] Because the joint distributions p_ϕ and $\tilde{p}_\theta$ are identical in this case (both are equal to the product of a uniform distribution over the two latent states and the prior distribution over x), the objective $D_{KL}(p_\phi||\tilde{p}_\theta)$ achieves its minimum possible value, zero.

This collapse of the optimal model to one in which the latent states are completely uninformative illustrates the problem with the classic VAE formulation that leads Alemi et al. (2018) to propose their "β-VAE" alternative. We instead show that we can obtain informative latent states without assigning a direct premium to informativeness as such, by using a training objective (4) with a weight $\beta > 0$. The second and third rows of the figure illustrate the consequences of progressively increasing β. When $\beta > 0$, the distributions $\tilde{p}_\theta(x\,|j)$ are different between the two latent states (left panels), and the encoding rule is one in which the probability of classifying the state using latent state $j = -1$ is higher the more negative the external state x is (right panels). With an even larger weight put on the accuracy of action selection (as in the bottom row), the optimal encoding rule differentiates even more sharply between higher and lower values of x, and the optimal generative model correspondingly associates more sharply differentiated distributions of values of x with the two latent states.

Figure 12 also illustrates the effects of a change in ω for a given value of β. When $\beta = 0$, the optimal recognition model is one that implies that the latent state will be completely uninformative, regardless of the value of ω. But for values of $\beta > 0$ (but still small enough to be below the upper bound $\bar{\beta}$), the degree to which it is optimal for the recognition model to discriminate between nearby values of x depends on the size of ω, for reasons already discussed in the main text (see section 6.2). Note that the two recognition models shown in the right panel of the second row directly correspond to the two curves shown in Figure 9 of the main text, since the probability of assigning latent state $j = -1$ is also (given the assumed decision rule) the probability of choosing to stay.

For large enough values of β, neither the exact value of β nor the size of ω matters for the form

[37] Thus the two curves in the left panel of the first row illustrate the difference between the priors associated with the two experimental treatments.

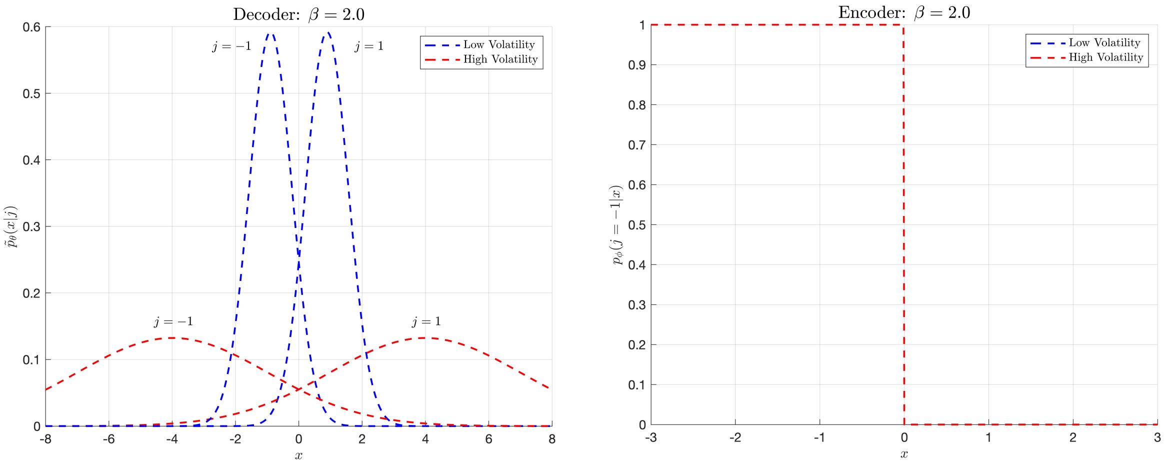


Figure 13: Equilibrium generative and recognition models (common to both players) when $\beta = 2$, shown for the same two values of ω as in Figure 12. (Here the optimal encoding rules for the two values of ω coincide.)

of the optimal recognition model, because perfect discrimination between positive and negative values of x comes to be optimal. This is illustrated in Figure 13, which shows the equilibrium generative and recognition models, using the same format as in Figure 12, but for the value $\beta = 2$. (This is a value large enough to exceed the critical value $\bar{\beta}$, as shown in Figure 8.)

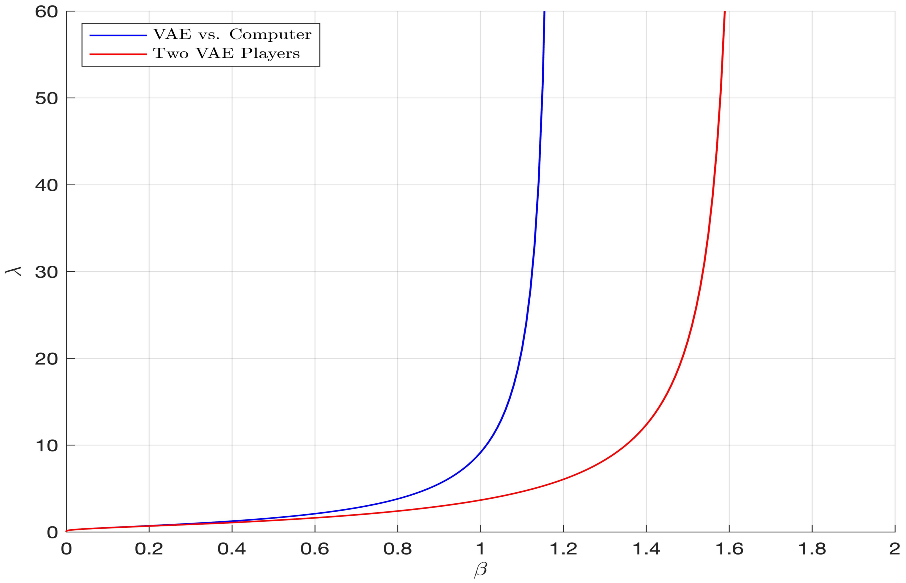


Figure 14: The equilibrium value of λ in the recognition models learned by VAE players, under two different assumptions about the opponent. In both cases, $\omega = \omega_{high}$.

B.2 Equilibrium of a Game Between a VAE and a Deterministic Automaton

In section 6.3 of the main text, we discuss the equilibrium that should result if a VAE player plays against a deterministic automaton, that always leaves if and only if $x \geq 0$. Here we illustrate numerically the difference in the predicted play of the VAE player (in equilibrium, i.e., once adaptation to the environment is complete) in this case relative to the one in which two VAEs play each other (treated in the previous subsection).

As explained in the main text, in both cases the VAE should learn a symmetric recognition model $\phi = (\lambda, 0)$; the only difference is in the value of λ in the two cases. Figure 14 shows the equilibrium value of λ, plotted as a function of the value of β in the VAE training objective, for each of the two cases. (The figure is drawn under the assumption that $\omega = \omega_{high}$, as in the "Algorithm" condition of the experiment of Frydman and Nunnari (2023).) In either case, the value of λ that should eventually be learned by any of the VAEs is a monotonically increasing function of β (with $\lambda \to 0$ as $\beta \to 0$). However, for any value of β in the interval $0 < \beta < \bar{\beta}$, the equilibrium value of λ is higher (implying less-random action selection) in the case of play against a computer than in the case of play against another VAE with the same value of β.

The difference is particularly important if the value of β is neither too small, nor too close to $\bar{\beta}$. (For example, Figure 15 compares the implied choice curves in the two cases when $\beta = 1.0$. Here we again assume that $\omega = \omega_{high}$.) If instead the value of β is small, the two choice curves are more similar (choice is nearly insensitive to the value of x, in either case); and this is also true if the value of β is near $\bar{\beta}$, or larger (choice is nearly, or even completely deterministic, in either case).

C Details of the Learning Algorithm

Here we provide additional details about the learning algorithm proposed in section 7

C.1 FOCs for Optimal Parameterization of the Recognition Model

Taking as given a joint distribution for (x, a') and generative model θ, we consider the problem of choosing recognition model parameters ϕ to minimize the training objective (4). Note that in the two-latent-state case considered here, (10) implies that

$$p_\phi(1\,|x) \;=\; \frac{\exp(\phi^{\mathrm{T}} v(x))}{\exp(\phi^{\mathrm{T}} v(x)) + \exp(-\phi^{\mathrm{T}} v(x))}, \tag{23}$$

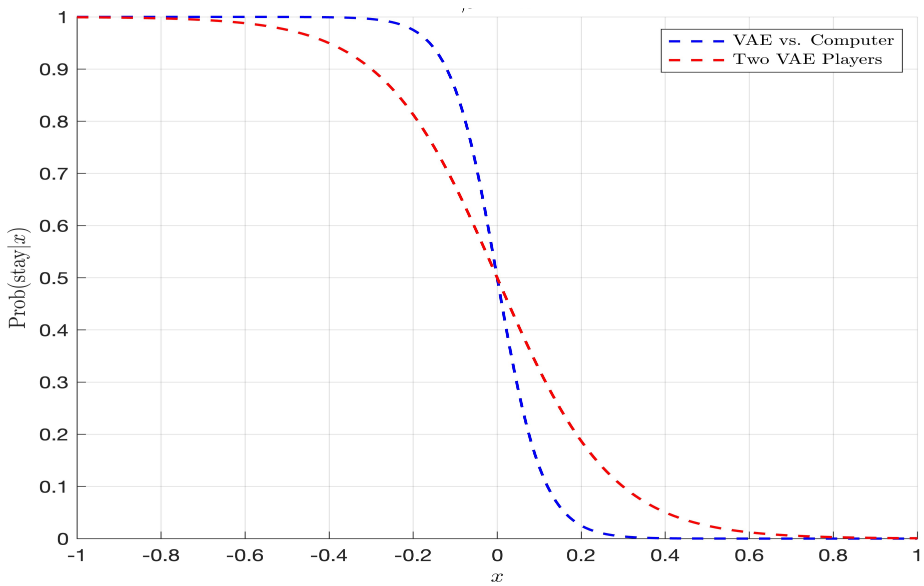


Figure 15: The probability of staying as a function of the external state x, in the equilibria of games between different types of players. In both games, the state x is drawn from a prior with standard deviation ω_{high}, and VAEs are trained with an objective in which $\beta = 1.0$.

using the notation

$$\phi \;\equiv\; \begin{bmatrix} \lambda \\ \nu \end{bmatrix}, \qquad v(x) \;\equiv\; \begin{bmatrix} x \\ -1 \end{bmatrix}.$$

Hence the gradient is the vector

$$\nabla(x) \;\equiv\; \nabla_\phi p_\phi(1\,|x) \;=\; w(x)v(x), \tag{24}$$

where

$$w(x) \;\equiv\; \frac{2}{[\exp(\phi^{\mathrm{T}}v(x)) + \exp(-\phi^{\mathrm{T}}v(x))]^2}$$

is a positive scalar (that depends on the value of ϕ as well as x).

The first-order conditions for the optimal choice of ϕ can be written in the form

$$\mathrm{E}\left[\sum_j V_j(x,a')\cdot\nabla_\phi p_\phi(j\,|x)\right] \;=\; 0, \tag{25}$$

where for each latent state j,

$$V_j(x,a') \;\equiv\; \beta\mathcal{L}(a(j),a';x) \;+\; \log\frac{\pi(x)p_\phi(j|x)}{q_\theta(j)\tilde{p}_\theta(x|j)}.$$

Because

$$\nabla_\phi p_\phi(-1\,|x) \;=\; -\nabla_\phi p_\phi(1\,|x),$$

the system (25) can be written more simply as (13), where[38]

$$\Delta(x,a') \;\equiv\; -\frac{1}{2}\,[V_1(x,a') \;-\; V_{-1}(x,a')].$$

If we define sign(a') as -1 if $a' =$ stay, and +1 if $a' =$ leave, then the differential payoffs in the coordination game can be written as

$$\mathcal{L}(\text{leave},\, a'; x) \;-\; \mathcal{L}(\text{stay},\, a'; x) \;=\; -2\text{sign}(a') \;-\; 2x.$$

It then follows that $\Delta(x,a')$ can be written in the form (14), using the definitions

$$m(x,a')^{\mathrm{T}} \;\equiv\; [\text{sign}(a') \quad 1 \quad x \quad x^2]$$

and

$$\begin{aligned}
\psi_1 &\equiv \beta, \\
\psi_2 &\equiv \frac{1}{2}\log\frac{q_\theta(1)}{q_\theta(-1)} - \frac{1}{2}\log\frac{\sigma_1}{\sigma_{-1}} - \frac{1}{4}\left[\left(\frac{\mu_1}{\sigma_1}\right)^2 - \left(\frac{\mu_{-1}}{\sigma_{-1}}\right)^2\right], \\
\psi_3 &\equiv 1 + \frac{1}{2}\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_{-1}}{\sigma_{-1}^2}\right), \\
\psi_4 &\equiv -\frac{1}{4}\left(\sigma_1^{-2} - \sigma_{-1}^{-2}\right)
\end{aligned}$$

for the elements of the vector $\psi(\theta)$.

C.2 Dynamical Equations

In each period t, pre-existing values for the VAE parameters θ_{t-1}, ϕ_{t-1} for each of the two players are given, and pre-existing values for the moment matrices M_{t-1}, N_{t-1} for each player as well. A new external state x_t is drawn, as an independent draw from the prior distribution corresponding to the environment at time t. (Note that the recursive algorithm continues to be well-defined even if at some point the environment shifts.) Each of the VAEs classifies this state as an example of latent state j_t (an independent choice by each of the VAEs), a draw from the conditional distribution (23) defined by parameters $\phi = \phi_{t-1}, x = x_t$. Each player then chooses the action $a_t = a(j_t)$

[38] The prefactor $-1/2$ in this definition does not change the implications of equation system (25); it is present solely in order to allow some expressions to be written more simply in the formulas given below for $\Delta(x,a')$.

determined by their VAE's latent state assignment.

The moment matrices and VAE parameters are then updated by each player in the following way. After observing the state x_t and the other player's action choice a'_t, each player can compute the current value of the vectors $v_t \equiv v(x_t)$ and $m_t \equiv m(x_t, a'_t)$. Each player also computes a current value for the gradient vector ∇_t using (24), in which we again set $\phi = \phi_{t-1}, x = x_t$. Based on the history of the algorithm to that point, the current estimates of the moment matrices are given by

$$M_t \equiv \frac{1}{t}\sum_{\tau=1}^{t} \nabla_\tau v_\tau^{\mathrm{T}}, \qquad N_t \equiv \frac{1}{t}\sum_{\tau=1}^{t} \nabla_\tau m_\tau^{\mathrm{T}}.$$

(Note that these are just the moments defined in (16), evaluated using the empirical joint distribution $\{(x_\tau, a'_\tau, \nabla_\tau)\}$.)

Equivalently, we can define the current estimates of the moment matrices using the recursions

$$M_t = \frac{1}{t}\nabla_t v_t^{\mathrm{T}} + \frac{t-1}{t} M_{t-1}, \tag{26}$$

$$N_t = \frac{1}{t}\nabla_t m_t^{\mathrm{T}} + \frac{t-1}{t} N_{t-1}, \tag{27}$$

starting from initial conditions $M_0 = N_0 = 0$. Equations (26)–(27) allow M_t and N_t to be calculated purely on the basis of the pre-existing values M_{t-1}, N_{t-1} and the new observations x_t, a'_t.

Next, each player's estimates of the parameters θ for their VAE are updated. As explained in the main text, the estimate θ_t is the parameter vector that maximizes the likelihood of the data $\{x_\tau, a'_\tau\}_{t=1}^{t}$, under the assumption that the observation (x_τ, a'_τ) in each period is an independent draw from the same joint distribution parameterized by θ. And given the parametric family of generative models assumed here, these MLE estimates are just the sample moments of the data. The current estimate of $q_\theta(j)$, which we denote q_{jt}, will just be the fraction of periods (among the first t periods) in which the latent state is j:

$$q_{jt} = \frac{1}{t}\sum_{\tau=1}^{t} \mathbb{I}_t(j),$$

where the indicator function $\mathbb{I}_t(j)$ is equal to 1 if $j_t = j$, and to zero otherwise. This element of θ_t can be updated recursively:

$$q_{jt} = \frac{1}{t}\mathbb{I}_t(j) + \frac{t-1}{t} q_{j,t-1}. \tag{28}$$

The parameters of each of the Gaussian conditional distributions $\tilde{p}_\theta(x\,|j)$ are similarly estimated on the basis of the corresponding sample moments:

$$\mu_{jt} \;=\; \frac{1}{q_{jt}} \cdot \frac{1}{t}\sum_{\tau=1}^{t} \mathbb{I}_\tau(j) x_\tau,$$

$$\sigma^2_{jt} \;=\; \frac{1}{q_{jt}} \cdot \frac{1}{t}\sum_{\tau=1}^{t} \mathbb{I}_\tau(j) x_\tau^2 \;-\; (\mu_{jt})^2.$$

Hence these parameters can be updated recursively as well:

$$\mu_{jt} \;=\; \frac{1}{q_{jt}} \cdot \frac{1}{t}\mathbb{I}_t(j) x_t \;+\; \frac{q_{j,t-1}}{q_{jt}} \cdot \frac{t-1}{t}\mu_{j,t-1}, \tag{29}$$

$$\sigma^2_{jt} \;=\; \frac{1}{q_{jt}} \cdot \frac{1}{t}\mathbb{I}_t(j) x_t^2 \;+\; \frac{q_{j,t-1}}{q_{jt}} \cdot \frac{t-1}{t}[\sigma^2_{j,t-1} + \mu^2_{j,t-1}] \;-\; (\mu_{jt})^2. \tag{30}$$

Thus the new estimates θ_t can all be computed using only the previous values θ_{t-1} and the new observations x_t, j_t.

Once these are obtained, the new estimates ϕ_t can be computed using (17). Thus the recursive algorithm is completely specified by the prior distribution from which x_t is drawn each period; the conditional probabilities (23) for drawing a latent state j_t corresponding to the external state x_t; the decision rule $a_t = a(j_t)$; equation (24) for computation of the gradient vector; recursions (26)–(30) for computing updated estimates of the moment matrices M_t, N_t, and the generative model parameters θ_t; and finally equation (17) for updating the recognition model parameters ϕ_t.

C.3 Convergence

Figure 16 provides additional detail about the longer-run trajectories of the two simulations shown in Figure 10 of the main text. Rather than ν_t, the vertical axis now plots the logarithm of the absolute distance of ν_t from ν^*, the predicted equilibrium value of the parameter in the symmetric equilibrium associated with prior distribution M. And rather than t (or more precisely, the iteration number $t - 1000$), the horizontal axis now plots the logarithm of t. We observe that after a large number of iterations, the path of the estimate ν_t becomes nearly deterministic, and furthermore closely follows a negatively-sloped straight line when shown as a log-log plot. This means that for all large enough t,

$$\log|\nu_t - \nu^*| \;\sim\; -\gamma \log t,$$

for some constant $\gamma > 0$. Alternatively (and recalling that $\nu^* = 0$, because of the symmetry of distribution M), it means that

$$\nu_t \;\sim\; t^{-\gamma}$$

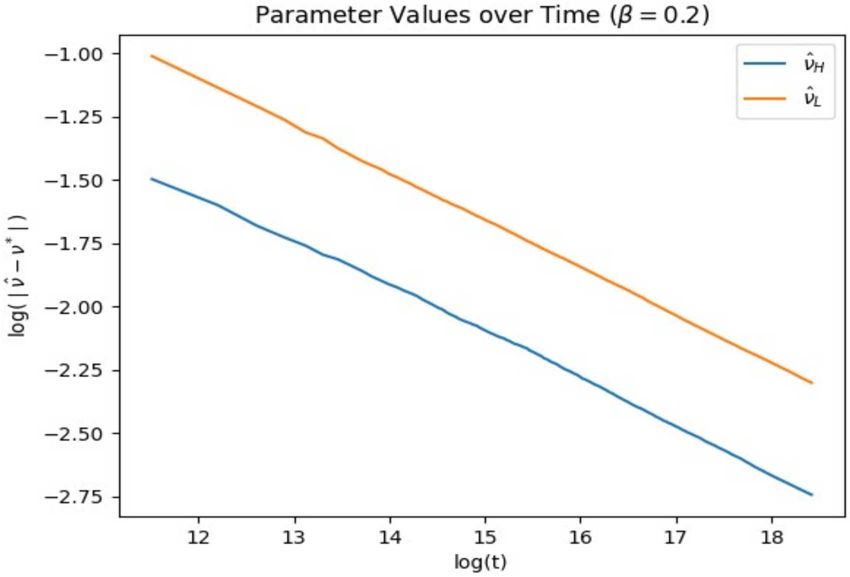


Figure 16: An alternative (log-log) plot of longer-run behavior under the two trajectories shown in Figure 10. Extrapolation of these paths to larger numbers of iterations implies that one should observe geometric convergence.

for some constant $\gamma > 0$. Moreover, the value of the γ appears to be the same in both simulations — and thus is evidently independent of the initial conditions — though the intercepts of the two lines in Figure 16 are different.

If we assume that the same relationship between ν_t and t continues to hold for even larger numbers of iterations (and our numerical results indicate that the relationship holds even more closely the larger the number of iterations), this implies that ν_t converges eventually to zero (i.e., to the equilibrium value ν^*) almost surely. The same is true for the other elements of the vectors ϕ_t and θ_t as well, though we do not show the corresponding figures. This means that under our model of VAE training, the equilibrium predictions derived in the earlier sections of the paper represent predictions about the outcome of learning dynamics as $t \rightarrow \infty$, assuming that the prior distribution eventually ceases to change.