Information-Constrained Coordination of Economic Behavior

Guy Aridor, Rava Azeredo da Silveira, Michael Woodford

Information-Constrained Coordination of Economic Behavior*

Guy Aridor† Rava Azeredo da Silveira‡ Michael Woodford§

April 30, 2023

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). The VAE is trained over a finite sample of past experiences, and is optimized to trade off the average payoff obtained over a distribution of possible games against the reconstruction error involved in predicting game payoffs on the basis of the latent state. Our use of the average payoff in the training objective requires us to generalize the basic VAE framework of Kingma and Welling (2014); we also compare our proposal to rationally inattentive decision making as modeled by Sims (2003) and Yang (2015). We apply our model to the coordination game in the experiment of Frydman and Nunnari (2022), and show that it offers an explanation for two salient features of the experimental evidence: both the gradual adjustment of 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, Cars Hommes, Tyler Malloy, Chris R. Sims, Mycal Tucker and participants at the Columbia Cognition and Decision Lab Workshop, the Workshop on Information-Theoretic Principles in Cognitive Systems (NeurIPS 2022) and at the Computational and Experimental Economics Workshop (Simon Fraser University, February 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 the Ambreen Chaudhri and Will Thompson from Kellogg Research Support for computational 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 (2022); 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, 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 (2022) 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.[2] We believe that these are additional reasons for

[2] Frydman and Nunnari (2022) 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 encoding to be the same for all values of x, so that the kind of internal representation required for perfectly accurate

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 $\{q_\theta(j)\}$ for each of the latent categories, and a learned distribution of external states $\tilde{p}_\theta(x\,|j)$ for each of

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.

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 4.3), 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.[3] 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 VAEs, the latent state is "decoded" using the generative model, treating the distribution $\tilde{p}_\theta(x \,|j)$ as

[3]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.

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

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

where $q_\phi(j)$ is the frequency with which the label j occurs in the joint distribution p_ϕ, 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. (See further discussion in Section 4.5 below.) 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.[4] Then for any recognition model ϕ and generative model θ, we can compute

$$L \equiv \mathrm{E}[\mathcal{L}(a(j);\, x)], \tag{2}$$

the expected loss when decisions are made using this VAE. Here a is the optimal decision rule

[4]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.

implied by the generative model,

$$a(j) \;=\; \arg\min_{a \in A} \mathrm{E}_{\tilde{p}_\theta}[\mathcal{L}(a;\, x)\,|j],$$

and $\mathrm{E}[\cdot]$ 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 model of information-constrained choice assumes that the parameters of the VAE are adjusted so as to solve the problem

$$\min_{\phi,\theta} \; D_{KL}(p_\phi || \tilde{p}_\theta) \;+\; \beta L, \tag{3}$$

where L is the expected loss measure defined in (2) 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 (3), and hence the way in which predicted behavior varies depending on the size of β.

While our primary interest in decision processes that minimize (3) 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

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

where

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

is the entropy of the distribution from which x is drawn,

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

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

$$R \;\equiv\; \mathrm{E}[D_{KL}(p_\phi(\cdot|x)||q_\theta)] \tag{4}$$

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_x is independent of the VAE parameters, the traditional VAE objective is equivalent to minimizing the value of $D_x + R$.

Alemi et al. (2018) propose that, rather than requiring that equal weight be put on the minimization of the terms D_x and R, as in the objective proposed by Kingma and Welling (2014), one can train a VAE by minimizing a generalized objective $D_x + \beta R$, 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

$$\min_{\phi,\theta} D_x + \beta_1 R + \beta_2 L, \tag{5}$$

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 $\beta_2 = 0$), and our basic model as another special case (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.

2.3 Comparison with Rational Inattention

The Sims (2003) model of rational inattention can be viewed as a model of this general type. It assumes, however, that for any choice of the encoding model ϕ, the generative model is given by

$$q_\theta = q_\phi, \qquad \tilde{p}_\theta(x|j) = \pi_\phi(x|j) \quad \forall j. \tag{6}$$

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 (5) requires, if (i) we rewrite (5) in the form

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

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 (6) holds, regardless of the choice of the encoding model. In such a limiting case, the optimal generative model θ is given by (6), 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

$$\min_{\phi} L + \psi I, \tag{8}$$

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 J 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 (5), 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 (5) 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 (8) 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.

Nor does it follow, however, that only versions of the problem (5) in which $\beta_1 < 1$ can be of interest, despite the popularity of this parameter assumption in the machine learning literature. In the problem considered by Alemi et al. (2018), $\beta_2 = 0$. In that case, if $\beta_1 \geq 1$, each of the other two terms in (7) achieves its theoretical lower bound (of zero) when the internal representation is perfectly uninformative ($p_\phi(j|x)$ is independent of x); hence one obtains non-trivial solutions only if $\beta_1 < 1$. But in our theory, instead, it is generally no longer optimal to choose a perfectly uninformative internal representation, even when $\beta_1 \geq 1$, because this would imply a large value for L. In our numerical results below, we illustrate how we can obtain non-trivial solutions even when $\beta_1 = 1$, so that our problem reduces to (3).

2.4 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, that specifies probabilities $p_\phi(j|x)$ conditional on the external state, as in the individual decision problem discussed

above.

Decoding the internal representation in order to allow selection of a desirable action in the case of each possible latent state j requires that the DM learn a generative model of the joint distribution of values for (j, x, a'). We assume that the class of possible generative models is of the form

$$\tilde{p}_\theta(j, x, a') \;=\; q_\theta(j) \cdot \tilde{p}_\theta(x, a'|j), \tag{9}$$

for some parametric family of possible distributions $\tilde{p}_\theta(x, a'|j)$. It then follows that whenever the latent state is j, the DM's optimal action choice (according to the learned generative model) will be given by

$$a(j) \;=\; \arg\min_{a \in A} \mathrm{E}_{\tilde{p}_\theta}[\mathcal{L}(a, a';\, x)\,|j]. \tag{10}$$

With this generalization of the definition above of the decision rule implied by a generative model, we can define an expected loss measure

$$L \;\equiv\; \mathrm{E}[\mathcal{L}(a(j), a';\, x)], \tag{11}$$

where the expectation $\mathrm{E}[\cdot]$ now refers an expected value given the distribution from which x and a' are *actually* drawn (as opposed to what the DM's generative model might imply about that distribution).

We assume that the DM's internal representation is trained so as to reduce the value of L, but also so as to achieve as great as possible a degree of congruence between the generative model and the joint distribution of external states x, labels j (by this DM), and opponent actions a' that actually occur. The latter joint distribution depends on the distribution $\pi(x)$ from which the external states are drawn, and the DM's recognition model ϕ, but also upon the recognition model ϕ' of the other player, and the other player's decision rule a' (which takes as an input j', that other player's internal representation of the external state). If we define $p_{\phi'}(a'|x)$ as the conditional probability of different action choices a' when the external state is x, as a result of the recognition model and decision rule of the other player (the parameters of which are jointly summarized by a vector ϕ'), then we can express the actual joint distribution of (j, x, a') as

$$p_\phi(j, x, a') \;=\; \pi(x) \cdot p_\phi(j|x) \cdot p_{\phi'}(a'|x). \tag{12}$$

We can then again assume that the parameters of the DM's VAE are adjusted so as to the solve the problem (3) but now defining $\tilde{p}_\theta$ as in (9), defining p_ϕ as in (12), and L as in (11).

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 VAE, more

specifically upon its 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 (3). 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 (2022). 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.[5] 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 (2022) are of the form shown in Table 1.[6] 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.[7]

[5] Models of these phenomena generally assume a large 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)).

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

[7] 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.

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. 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 numerical values of a and b in any particular application.[8] 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. 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

[8] 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.

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 $(1 - q(-x),\, 1 - p(-x))$.

We shall confine our theoretical analysis to cases in which the distribution from which the value of x is drawn is given by $N(0,\, \omega^2)$, for some value of ω; the two treatments in the experiment of Frydman and Nunnari (2022) are both of this form, for different values of ω. Note that in such cases, the prior density function $\pi(x)$ also exhibits one of the symmetries discussed above: $\pi(-x) = \pi(x)$ for all x. Hence in the case of any equilibrium $(p(x),\, q(x))$, the four equivalent equilibria just discussed must all involve the same pair of expected payoffs for the two players.

Because of these symmetries, it is natural to focus on equilibria that reflect the symmetry of the game payoffs — that is, equilibria in which the probabilities of selecting different actions are also invariant under the same two transformations. This would mean equilibria with the property that $p(x) = q(x)$, and also such that $p(-x) = 1 - p(x)$ for all x. We find that such equilibria exist, under a variety of possible assumptions about the precision with which it is possible for players to respond to the current state.

3.2 The Imprecision of Observed Behavior

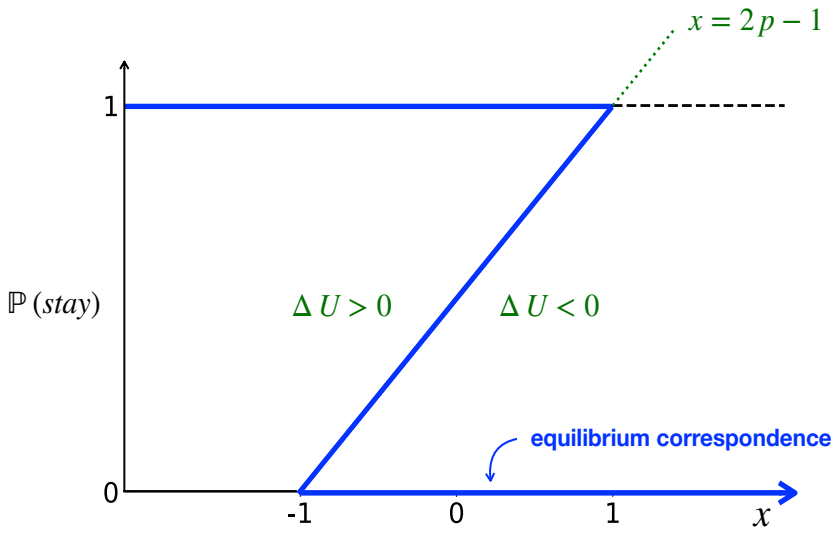
We first consider what equilibrium behavior should be like in the case that players are able to respond precisely to the value of x on a given trial (and to the way that their opponent plays in that state). To simplify the discussion, we consider only equilibria that are invariant under an exchange of 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

$$\Delta U \;=\; 2[(2p - 1) \;-\; x]. \tag{13}$$

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 dashed 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).

Figure 1: The Equilibrium Correspondence with Precise Action Choice

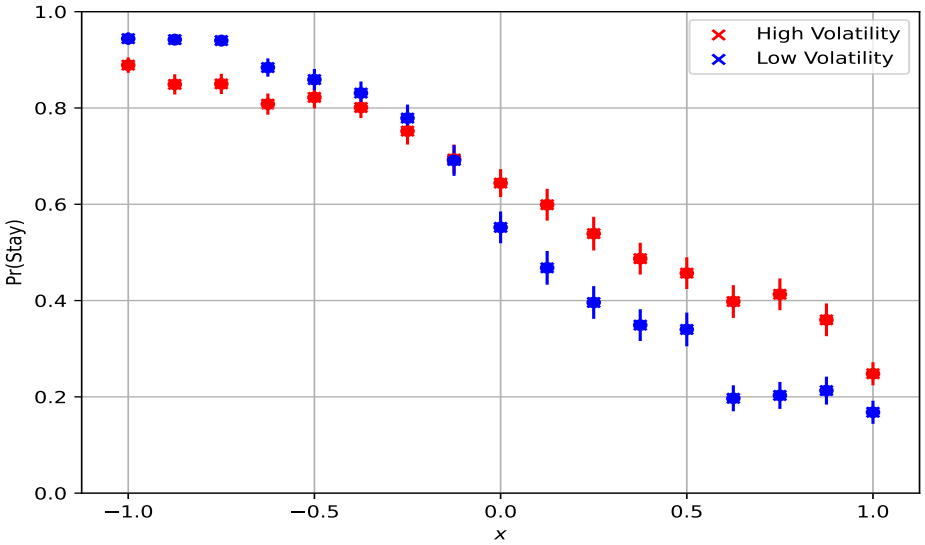


A function $p(x)$ specifying the state-contingent (and probabilistic) behavior of each of the players then represents a symmetric equilibrium of the game specified by the payoff matrix in Table 2 if and only if $p(x)$ is a single-valued function defined for all x on the real line, the graph of which 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 (2022), 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.[9] 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).

[9] In their "low volatility" treatment, $\omega = \sqrt{5}/2$, while in the "high volatility" treatment, $\omega = 5$ in our normalized units. The two alternative priors are shown graphically in top left panel of Figure 3 below.

Figure 2: State-Contingent Behavior in the Experiment of Frydman and Nunnari (2023)



3.3 Consequences of Rational Inattention

Subjects' behavior appears to respond less accurately to the exact state x than is required for an equilibrium with perfect optimization. But neither is subjects' behavior 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 (8), 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 ($j = 1$) leads the DM to choose action 1 and the other of which ($j = 2$) leads them to choose action 2. The implied probabilities of action choice are then of the form

$$\mathrm{P}(a\,|x) \;=\; \frac{\bar{p}(a)\exp[-\psi^{-1}\bar{\mathcal{L}}(a,x)]}{\sum_{\tilde{a}\in A}\bar{p}(\tilde{a})\exp[-\psi^{-1}\bar{\mathcal{L}}(\tilde{a},x)]}, \tag{14}$$

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 above, 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 (14), we conclude that

$$p(x) \;\equiv\; \mathrm{P}(stay\,|x) \;=\; [1 \,+\, \exp(-2\psi^{-1}\Delta U)]^{-1}, \tag{15}$$

where we again use the notation

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

for the net expected reward from staying rather than leaving,[10] given the state x and the implied probability of 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 (15) implies that a player should stay with exactly probability p. Using (13) to substitute for ΔU in (15), we conclude that (x, p) belong to this correspondence if and only if

$$x \;=\; (2p-1) \;-\; \frac{\psi}{2}\log\Big(\frac{p}{1-p}\Big). \tag{16}$$

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.[11] Note that there is a single value of x consistent with (16), for each value $p \in (0,\,1)$. This defines a function $x(p)$ that is continuous; with 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 (15) implies that $x(p) \to 2p - 1$, the function characterizing the equilibrium correspondence in Figure 1, in the limit as $\psi \to 0$.

[10] Here we identify the "loss" in the RI problem (8) with the negative of the reward shown in the game payoff matrix (2).

[11] 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.

At the same time, for any $\psi > 0$, equation (15) 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 (2022) 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 (16) 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.

We shall show instead 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 that should be flatter when the variance of the prior is greater — all features of the experimentally observed behavior displayed in Figure 2.

4 Equilibria of the VAE Model

We now consider the implications of applying the VAE model from Section 2 to the coordination game discussed in Section 3. In order to apply the model, we must choose specific parametric families of recognition and generative models, and then discuss the way in which the model parameters can be learned from a training dataset.

4.1 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

$j = -1, \ldots, -J$.[12] The conditional distribution associated with each of the latent states is assumed to be of the form

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

In addition, each of the conditional distributions $\tilde{p}_\theta(x\,|j)$ is assumed to be Gaussian: $x|j \;\sim N(\mu_j,\, \sigma_j^2)$. Hence the generative model is specified by the probabilities $\{q_\theta(j)\}$ of the latent states ($2J-1$ independent parameters); a probability $\tilde{p}_\theta(stay\,|j)$ for each latent state, indicating the probability that $a' = stay$ when the latent state is j ($2J$ additional parameters); and a mean μ_j and standard deviation σ_j for each latent state ($4J$ additional parameters). There are thus $8J-1$ free parameters for the generative model (in the absence of symmetry restrictions).

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

$$p_\phi(j\,|x) \;=\; \frac{A_j \exp(\lambda_j x)}{\sum_{\ell \in \mathcal{J}} A_\ell \exp(\lambda_\ell x)}. \tag{17}$$

Here the coefficients $\{A_j\}$ are all assumed to be positive, and we can without loss of generality normalize them so that $\sum_{j\in\mathcal{J}} A_j \;=\; 1$. The coefficients $\{\lambda_j\}$ may be of either sign, and without loss of generality we can normalize these so that $\sum_{j\in\mathcal{J}} \lambda_j \;=\; 0$. (Note that only the ratios of the coefficients A_j affect the encoding probabilities specified by (17), and similarly, only the differences of the coefficients λ_j matter. Hence we can multiply all of the $\{A_j\}$ by a positive factor so as to ensure that $\sum_j A_j = 1$, and add a constant to all of the $\{\lambda_j\}$ so as to ensure that $\sum_j \lambda_j = 0$, without affecting the implied encoding probabilities.) There are thus $4J-2$ free parameters for the recognition model (in the absence of symmetry restrictions), and hence a total of $12J-3$ parameters for the complete VAE of each player.

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 ratio $A_j/A_{j'}$ of the multiplicative factors associated with the two latent states, and the rate at which the log odds change as x increases specified by the difference $\lambda_j - \lambda_{j'}$ between the exponential factors. Thus both the ratios of the As and the differences of the λs have relatively simple interpretations in the specification given in (17).

Equation (17) 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.

[12] The reason for this notation is to allow us to study symmetric equilibria in which for each latent state j, there is a corresponding latent state $-j$ that is decoded using a conditional distribution $\tilde{p}_\theta(x, a'|-j)$ with the property that $\tilde{p}_\theta(x, a'|-j) = \tilde{p}_\theta(-x, -a\prime|j)$.

But in fact, the specification (17) 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).[13]

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 $A_j = 1/(2J)$ for all j, $\lambda_j = \lambda > 0$ for all positive j and $\lambda_j = -\lambda$ for all negative j, and assuming a decision rule under which $a(j) = stay$ for all negative j and $a(j) = leave$ 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 proposed training procedure, 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.

4.2 Symmetric Equilibria

We have noted above that the payoff matrix specified in Table 2 is invariant under two different transformations; this leads us to be interested in equilbria of the VAE model that are similarly invariant under both of these transformations. This requires first that the parameters

$$\{q_\theta(j), \tilde{p}_\theta(stay\,|j), \mu_j, \sigma_j, A_j, \lambda_j\}$$

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 $q_\theta(-j) = q_\theta(j)$, $\mu_{-j} = -\mu_j$, $\sigma_{-j} = \sigma_j$, $A_{-j} = A_j$, and $\lambda_{-j} = \lambda_j$ for all j; $a(-j) = -a(j)$ for all j; and $\tilde{p}_\theta(a\,|-j) = \tilde{p}_\theta(-a\,|j)$ for both a and all j. If we are interested only in the possibility of symmetric equilibria, then, instead of $24J - 6$ parameters to solve for ($12J - 3$ for each player), there will only be $6J - 3$ parameters.

[13] 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.

4.3 Computational Approach

Suppose that the training objective in our VAE model is specified by (3) for some weight $\beta > 0$, or more generally, by (7) for some weights $\beta_1, \beta_2 > 0$. We can train a pair of models, representing the two players in a game of the kind specified above. A sample of values $\{x_i\}$ for the external state is drawn from the prior distribution $\pi(x)$. Given conjectured recognition models and decision rules for each of the two players, we can simulate random draws of the latent state classification j_i and action $a_i = a(j_i)$ for each player in the case of each draw x_i of the external state. (Here the randomness in the latent state classifications by the two players is independent, whether they use the same recognition model or not.) In this way, we obtain a training sample $\{x_i, j_i, a_i'\}$ for each of the players. (Each player's action choice a_i becomes the opponent action choice a_i' for the other player.)

The parameters θ of each player's generative model can then be optimized for the training sample $\{x_i, j_i, a_i'\}$ for that player. Given this generative model for each player, we can next optimize the player's recognition model and decision rule for their generative model. The process can then be repeated using these new recognition models and decision rules to compute new training samples. This process is iterated until we achieve convergence of the parameters of the two VAEs to a sufficient tolerance; the outcome of this joint process of adaptation of each player's VAE decision process to the behavior of the other is what we mean by an equilibrium of the VAE model.

We are interested in characterizing the equilibrium of this process of mutual adaptation in the case of large enough training samples for there to be only a negligible degree of sampling error involved in optimizing against a finite sample rather than the theoretical joint distribution of (x, j, a') for each player given the prior, the recognition rules, and the decision rules of the two players. In that asymptotic limit, we expect the equilibrium to be symmetric (in the sense just explained), because of the symmetry of the game payoffs that is also respected by the VAE training objective. Hence we can simplify our computations (and obtain faster convergence to something close to a symmetric equilibrium of the limiting two-player game) by directly imposing the expected symmetry in our computational algorithm.

The algorithm used to produce the numerical results presented here (sketched as Algorithm 1) imposes symmetry in the following way. We iteratively adjust the $6J - 3$ parameters of a single VAE (imposing a priori the symmetry restrictions stated in the previous section) that is intended to represent the decision process of either of the players. To begin, a sample of values $\{x_i\}$ for the external state is again drawn from the prior distribution $\pi(x)$. Given a (single) conjectured recognition model and decision rule, we simulate *two independent* random draws j_i, j_i' of the latent state classification for each of the observations x_i, using the same probabilistic recognition model, one representing the DM's own encoding of the state, and the other the encoding by the other player. The decision rule is used to convert the other player's encoded value into a simulated

action choice $a'_i = a(j'_i)$ by the other player. In this way, we obtain a training sample $\{x_i, j_i, a'_i\}$ for the single DM, which can then be used to train the generative model. The recognition model and decision rule can then be optimized for this generative model, and the entire process repeated.

Algorithm 1 Equilibrium Computation Algorithm

function COMPUTE_EQUILIBRIUM
  Conjecture a decision rule, $a(j)$
  Draw sample $\{x_i\}_{i=1}^N$ of values from π
  Set $tol = 0.005$
  while Decision rule not validated **do**
    Initialize $\phi^{t=0}$ randomly, $\theta^{t=0} = 0$
    while $err < tol$ **do**
      Given ϕ^t, draw classifications $\{j_i\}$ and $\{j'_i\}$ for each x_i
      Using a, convert classifications $\{j_i\}_{i=1}^N$ to $\{a'_j\}_{i=1}^N$
      Optimize the generative model θ^t given the dataset $\{x_i, j_i, a'_i\}_{i=1}^N$.
      Optimize the recognition model ϕ^{t-1}, given θ^t and the dataset $\{x_i, a'_i\}_{i=1}^N$.
      $err = \frac{1}{|\theta^{t+1}|+|\phi^{t+1}|} \cdot (d(\theta^{t+1}, \theta^t)) + d(\phi^{t+1}, \phi^t)))$
      Increment t
    end while
    if Validate the optimality of the conjectured decision rule **then return** ϕ^{t+1}, θ^{t+1}
    else
      Reverse decision rule, $a(j)$
    end if
  end while
end function

If, in the case of a large enough sample, this process converges, the VAE parameter values to which it converges should represent a symmetric equilibrium of the two-player game. It is solutions of this kind that we display numerically below. (The possibility of additional asymmetric equilibria, even when, as here, the model specification is symmetric, may be of interest; but we leave this for study elsewhere.)

4.4 The Optimal Decision Rule, Given a Generative Model

We can provide further insight into the nature of symmetric equilibria by discussing some of the requirements for optimality of the parameters. We first consider the optimal decisions $\{a(j)\}$, given a generative model θ of the kind specified above. When the external state is classified by the latent state j, the DM chooses an action $a(j)$ so as to solve the problem (10), using the distribution $\tilde{p}_\theta(x, a')$ to model the likely external state and play by the opponent. The payoff matrix in Table 2 implies that the net expected gain from staying rather than leaving, assessed on the basis of the

generative model, will equal

$$2\Big[(2\tilde{p}_\theta(stay\,|j) - 1) \;-\; \mu_j\Big], \tag{18}$$

generalizing expression (13) to the case in which both p and x must be estimated using the generative model. Thus the DM should choose $a(j) = stay$ if and only if j is a state for which the generative model parameters are such that (18) is positive.

Note that if the parameters of the generative model satisfy the symmetry conditions stated above, (18) will be positive for state j if and only if it is negative for state $-j$; hence the optimal decision rule will in this case necessarily satisfy the symmetry conditions as well. In fact, in the equilibria that we report below, it takes a simple form: $a(j) = stay$ for all negative j, and $a(j) = leave$ for all positive j. Given that we assume a symmetric generative model, we can further assume without loss of generality that the latent states are numbered so that $\mu_j > 0$ for all positive j and $\mu_j < 0$ for all negative j. But then since $q_\theta(j) = q_\theta(-j)$ and $\sigma_j = \sigma_{-j}$, for any $j > 0$, the likelihood of latent state j (according to the generative model) is greater than the likelihood of latent state $-j$ if and only if $x > 0$. It therefore should be expected that the recognition model (when optimized for the generative model) will classify state x more often as j than as $-j$ if and only if $x > 0$. If this is true for all positive j, then under the conjectured decision rule, the opponent should choose to leave (because $j' > 0$) more often than to stay (because $j' < 0$) if and only if $x > 0$. And since $x > 0$ more often than $x < 0$ whenever j is positive, one should expect the generative model (when optimized in the way discussed below) to learn a value $\tilde{p}_\theta(stay\,|j) < 1/2$ (i.e., should learn to predict that the opponent is more likely to leave) whenever j is positive. If these conjectures hold, we should find that (18) is negative whenever j is positive (and correspondingly positive whenever j is negative), so that the conjectured decision rule will be optimal according to the generative model.

In practice, our algorithm starts by conjecturing this decision rule, and then checks whether the generative model that is learned under this assumption does indeed imply that the decision rule is optimal. In all of the numerical examples displayed below, we find that it is.

4.5 Optimization of the Generative Model

Another key step in the algorithm involves optimizing the parameters of the generative model for a given training sample $\{x_i, j_i, a'_i\}$, for $i = 1, \ldots, N$. Using p^{emp} to denote the empirical joint distribution of (x, j, a'), we consider the choice of parameters θ to minimize the training objective

$$D_{KL}^{emp} \;\equiv\; D_{KL}(p^{emp}||\tilde{p}_\theta).$$

This is equivalent to choosing the parameters θ to maximize the log likelihood of the training sample according to the model θ,

$$LL \equiv \sum_{i=1}^{N} \log q_\theta(j_i) + \sum_{i=1}^{N} \log \tilde{p}_\theta(x_i \,|j_i) + \sum_{i=1}^{N} \log \tilde{p}_\theta(a'_i \,|j_i). \tag{19}$$

It is easily seen that the log likelihood LL is maximized by choosing parameters

$$q_\theta(j) = \frac{N_j + N_{-j}}{2N},$$

where N_j is the number of times that $j_i = j$ in the training sample;

$$\mu_j = \frac{\sum_{i \in I_j} x_i - \sum_{i \in I_{-j}} x_i}{N_j + N_{-j}},$$

where I_j is the set of i for which $j_i = j$;

$$\sigma_j^2 = \frac{\sum_{i \in I_j} (x_i - \mu_j)^2 + \sum_{i \in I_{-j}} (x_i + \mu_j)^2}{N_j + N_{-j}};$$

and

$$\tilde{p}_\theta(stay \,|j) = \frac{N_{stay,j} + N_{leave,-j}}{N_j + N_{-j}},$$

where for either action $a' \in A$, $N_{a',j}$ is the number of elements of the training sample for which $j_i = j$ and $a'_i = a'$. Note that these expressions are more complex than the usual ones for maximum-likelihood estimation of a Gaussian mixture model, because of our imposition of our symmetry assumptions: we optimize only over symmetric specifications of the generative model.

These solutions are optimal if the training objective is simply the minimization of D_{KL}^{emp}, as in the classic VAE proposal of Kingma and Welling (2014). Our proposed training objective (3) includes an additional term, the average loss L. However, this second term depends on the generative model parameters θ only through the effect of the generative model on the DM's choice for the decision rule $a(j)$. Among generative models with the property that (18) is negative whenever j is positive (the conjecture discussed above), different choices of θ all lead to the same decision rule $a(j)$. Hence if the generative model specified by the equations above is one with this property (as we verify in our numerical solutions), then it is the optimal generative model (at least among generative models consistent with the conjecture), even when $\beta > 0$ in (3).

Suppose instead that we use a more general training objective of the form (7). In this case, we must also consider the effect of the parameters θ on the value of the "rate" measure R. It follows from (4) that R is the difference between the average value of $\log p_\phi(j_i \,|x_i)$ and the average value

of $\log q_\theta(j_i)$, in the training sample. The former quantity is independent of the choice of generative model, so that minimization of R for the training sample is equivalent to maximization of the second quantity, i.e.,

$$\frac{1}{N}\sum_{i=1}^{N}\log q_\theta(j_i) \;=\; \frac{1}{N}LL_1,$$

where LL_1 is the first term on the right-hand side of (19). Thus minimization of $D_{KL}+(\beta_1-1)R$ is equivalent to maximization of $LL+(\beta_1-1)LL_1$, or the quantity

$$\beta_1\sum_{i=1}^{N}\log q_\theta(j_i) \;+\; \sum_{i=1}^{N}\log \tilde{p}_\theta(x_i\,|j_i) \;+\; \sum_{i=1}^{N}\log \tilde{p}_\theta(a'_i\,|j_i). \tag{20}$$

But the solution given above does not merely minimize LL; it minimizes each of the three terms in (19) separately. Hence it also minimizes (20), for any value of $\beta_1 > 0$. If we add an additional term $\beta_2 L$ to the training objective, this does not affect the optimal choice of θ, as just discussed. Hence the solution specified by the above equations gives closed-form expressions for the optimal generative model, for training objectives of the form (7), regardless of the values assumed for the weights β_1 and β_2.

4.6 Optimization of the Recognition Model

Finally, we consider the optimal choice of the recognition model ϕ for a given generative model θ. For a given sample of values $\{x_i, a'_i\}$ for the external state and the other player's action choice, and a given recognition model ϕ, let p_ϕ^{emp} be the joint distribution for (x, j, a') such that the marginal distribution for the variables (x, a') is given by the empirical distribution, and the conditional distribution for $(j\,|x, a')$ is given by $p_\phi(j\,|x)$, independently of the value of a'. Then for a given generative model θ, the step of the algorithm that optimizes the recognition model chooses parameters ϕ so as to minimize $D_{KL}(p_\phi^{emp}||\tilde{p}_\theta)$.

This can be computed as a function of ϕ, and when the parametric family of recognition models that we consider is simple enough, it is straightforward to optimize over the parameters ϕ. For example, in the symmetric case with $J = 1$, discussed below, the vector ϕ consists of a single parameter λ, so that this step of the algorithm only requires solving for the minimum of a nonlinear function of a single variable. But even in the case of a much more complex family of recognition models (say, a neural network), the parameters can be optimized using standard methods such as back-propagation.

5 Numerical Results

We illustrate the implications of our model by presenting examples of numerical solutions under alternative parameterizations. Our numerical examples consider two possible prior distributions, corresponding to the "low volatility" and "high volatility" treatments in the experiment of Frydman and Nunnari (2022). The probability density functions associated with these two priors are shown in the upper left panel of Figure 3 below.

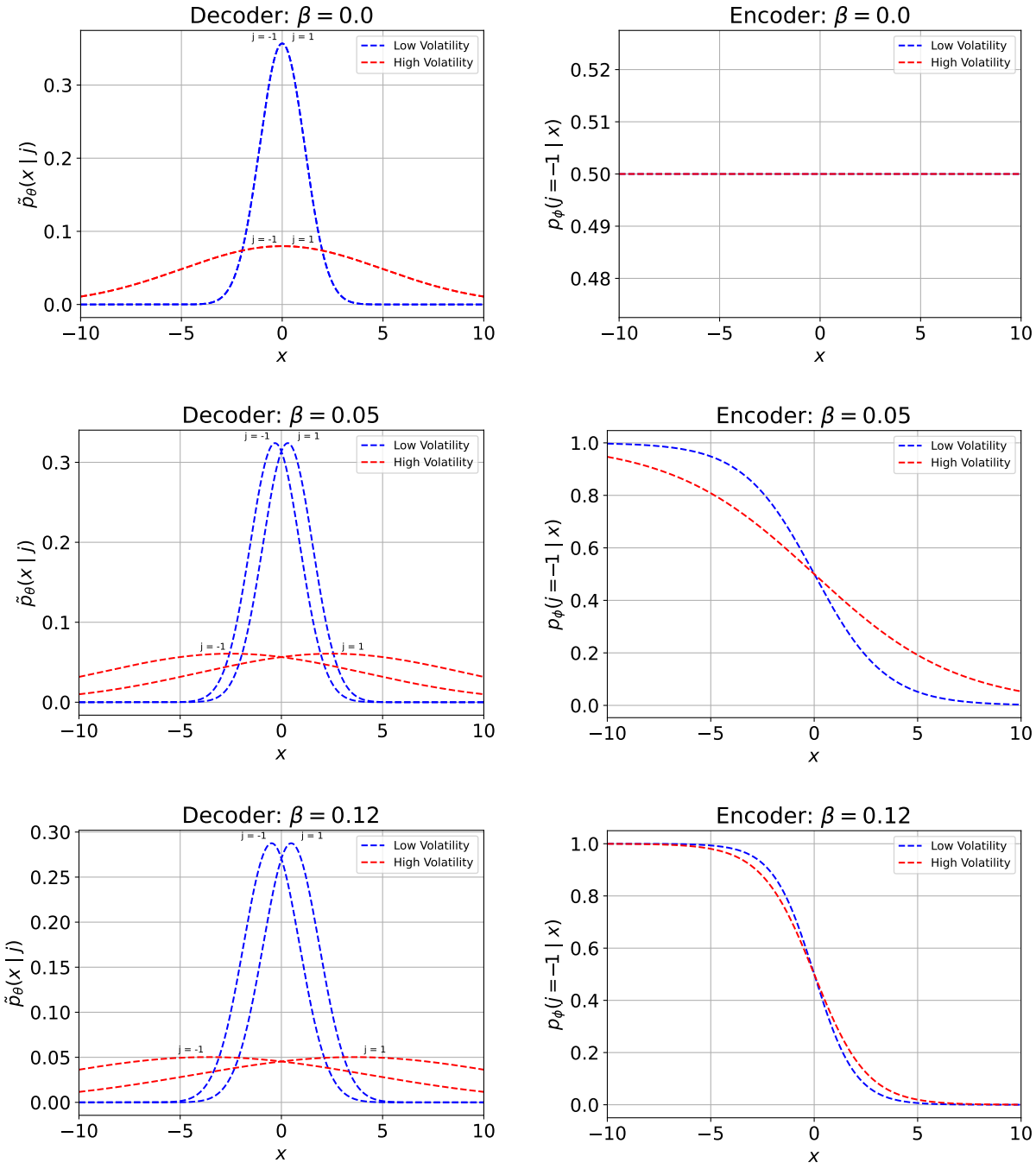
5.1 Alternative Values of β: The Case of Two Latent States

We first consider a game in which both players' decisions are made by a symmetric VAE with two latent states (i.e., the case $J = 1$). The VAE is trained using a training objective of the form (3), for a variety of choices of the relative weight β. Figure 3 shows the optimized generative (decoding) and recognition (encoding) models for each of several possible values of β. When there are only two latent states, $j = -1$ and $j = 1$, our symmetry assumption requires that $q_\theta(j) = 1/2$ for each latent state. The generative model for the joint distribution of the variables (j, x) is therefore specified by displaying the density functions $\tilde{p}_\theta(x\,|j)$ for each of the two values of j (left column of the figure). The recognition model is fully specified by plotting $p_\phi(-1\,|x)$ as a function of x (right column of the figure). Each panel of the figure shows the optimized models both for the low-variance prior (in blue) and for the high-variance prior (in red).

The first row shows the optimal model for the objective with $\beta = 0$ (which reduces to the standard VAE model of Kingma and Welling (2014)). In this case, 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 $\pi(x)$.[14] The optimal recognition model is the same for both priors, and is one in which the probability of classifying the state using latent state j is independent of the external state x, so that the latent state is completely uninformative. (This in turn explains why the optimal generative model is one in which the two latent states are both associated with the same distribution of external states.) 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 (3) with a weight $\beta > 0$.

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

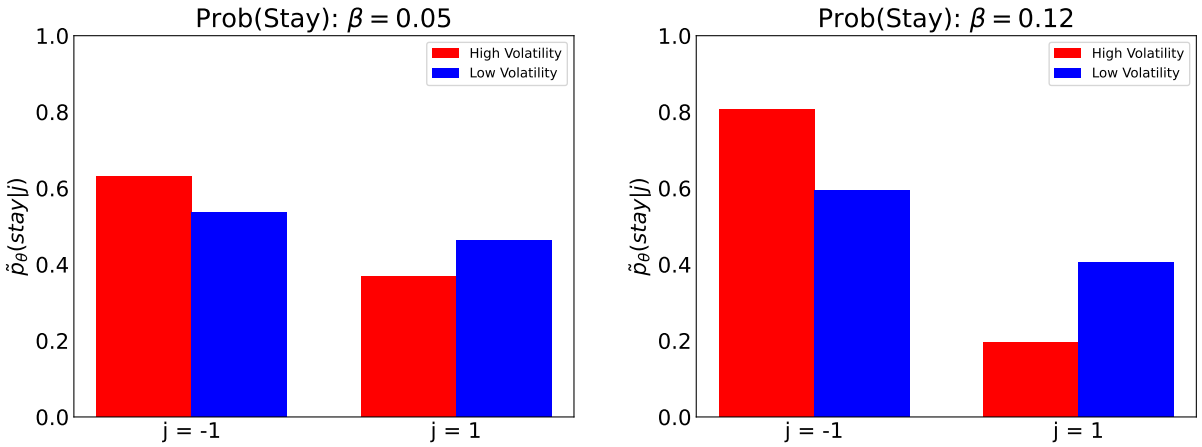
Figure 3: Learned Generative and Recognition Models



The second row of Figure 3 shows the corresponding results when $\beta = 0.05$. With a positive value of β, the distributions $\tilde{p}_\theta(x\,|j)$ are now different between the two latent states (left panel), 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 panel). The optimal encoding rule is now different depending on the value of ω: drawing x from a more dispersed distribution means that x has to be farther away from zero to get the classification probability to be asymmetric to a given degree. When one combines this result with the optimal decision rule (explained above, in Section 4.4), the probability of classifying the state using latent state $j = -1$ becomes the probability that the DM chooses to stay. Thus we see in this case that the decision whether to stay is probabilistic (conditional on the external state), and that the probability of staying is a continuously decreasing function of x, as in the experimental data (Figure 2).

Finally, the bottom row shows the corresponding results when $\beta = 0.12$. With an even larger weight put on the accuracy of action selection, 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 4: Learned Generative Models (cont'd)



There is another aspect of the optimal generative model to determine, namely the learned values for the probabilities $\tilde{p}_\theta(stay\,|j)$, for each of the two latent states. In the case that $\beta = 0$, the latent states are distributed independently of the value of x, as implied by the encoding model in the upper right panel of Figure 3. The two latent states chosen by the two players will then be statistically independent of one another, and either player's generative model eventually learns that the other's action choice is uncorrelated with their own latent state. This means that $\tilde{p}_\theta(stay\,|j) = 1/2$ for either j, and regardless of the value of ω.

If instead $\beta > 0$, each player's latent state is correlated with the external state (j and j' are both more likely to take the value 1 when x is larger). This means that the other player's latent state —

and hence the other player's action — is to some extent predictable from one's own latent state; specifically, the other player is more likely to stay when one's own latent state is $j = -1$. This is shown in Figure 4 for the two cases $\beta = 0.05$ and $\beta = 0.12$. Each panel shows the results for a particular value of β. Within the panel, the value of $\tilde{p}_\theta(stay\,|j)$ is shown as a bar graph, for each of the two latent states (left vs. right bar graphs) and for each of the two different priors (red vs. blue bars within each bar graph).

One observes that in each case, $\tilde{p}_\theta(stay\,|j)$ is greater when $j = -1$. However, the differentiation between the two latent states is greater (for either prior) when β is larger; this is because the encoding rule varies more sharply as a function of x when β is larger, as shown in Figure 3. One also observes, for a given value of β, that the probabilities are more different across the two latent states when ω is larger (the "high volatility" case) than when it is smaller. This is because extreme values of x — values far enough from zero to make the latent state classification highly predictable — occur more often in the "high volatility" case, so that the latent states of the two players are more correlated in this case, despite the fact that the encoding rule varies more sharply with x (for values of x not far from zero) in the "low volatility" case.

5.2 A More General Training Objective

We can generalize our analysis by considering the more general training objective (7), exploring the effects on the optimal VAE of independent variation in each of the two relative weights β_1, β_2. However, in the symmetric case with only two latent states (as assumed above), we obtain no new phenomena by considering this additional dimension of variation in the training objective.

The reason is that in the symmetric case with only two latent states, the feasible recognition models vary with respect to the value of a single parameter (the parameter $\lambda > 0$ such that $\lambda_1 = \lambda, \lambda_{-1} = -\lambda$). Given a symmetric prior, the value of the parameter λ completely determines the joint distribution of the variables (x, j, j'), and hence the joint distribution of the variables (x, j, a') from which the training sample is sampled. We have further shown above that the generative model θ that is optimal for a given training sample is independent of the weights in the training objective, even in the case of the more general objective (7). Hence the parameters of the optimal generative model are completely determined by the specification of the value of λ.

The effects on the optimal VAE of a change in the weights in the training objective thus result entirely from the effects of a change in the training objective on the optimal choice of λ. But there is only a one-parameter family of VAE models that can be optimal for different training objectives, corresponding to the different values of λ. We have already observed what progression through this family of possibilities looks like in Figure 3: the first row displays the solution when $\lambda = 0$, the second row illustrates the consequences of choosing a positive value of λ, and the third row the

consequences of an even larger positive value. (Figure 4 illustrates additional consequences of a progressive increase in the value of λ.)

If we consider the two-parameter family of training objectives (7), varying β_1 and β_2 still simply moves one along this same one-parameter family of optimal VAEs. Increasing β_2 for fixed β_1 leads to progressively larger choices of λ, with the consequences shown in Figures 3 and 4. Alternatively, reducing β_1 for fixed β_2 (as in the "β-VAE models" of Alemi et al. (2018)) also leads to progressively larger choices of λ, with exactly these same consequences. Hence we do not show additional figures for cases in which $\beta_1 \neq 1$.

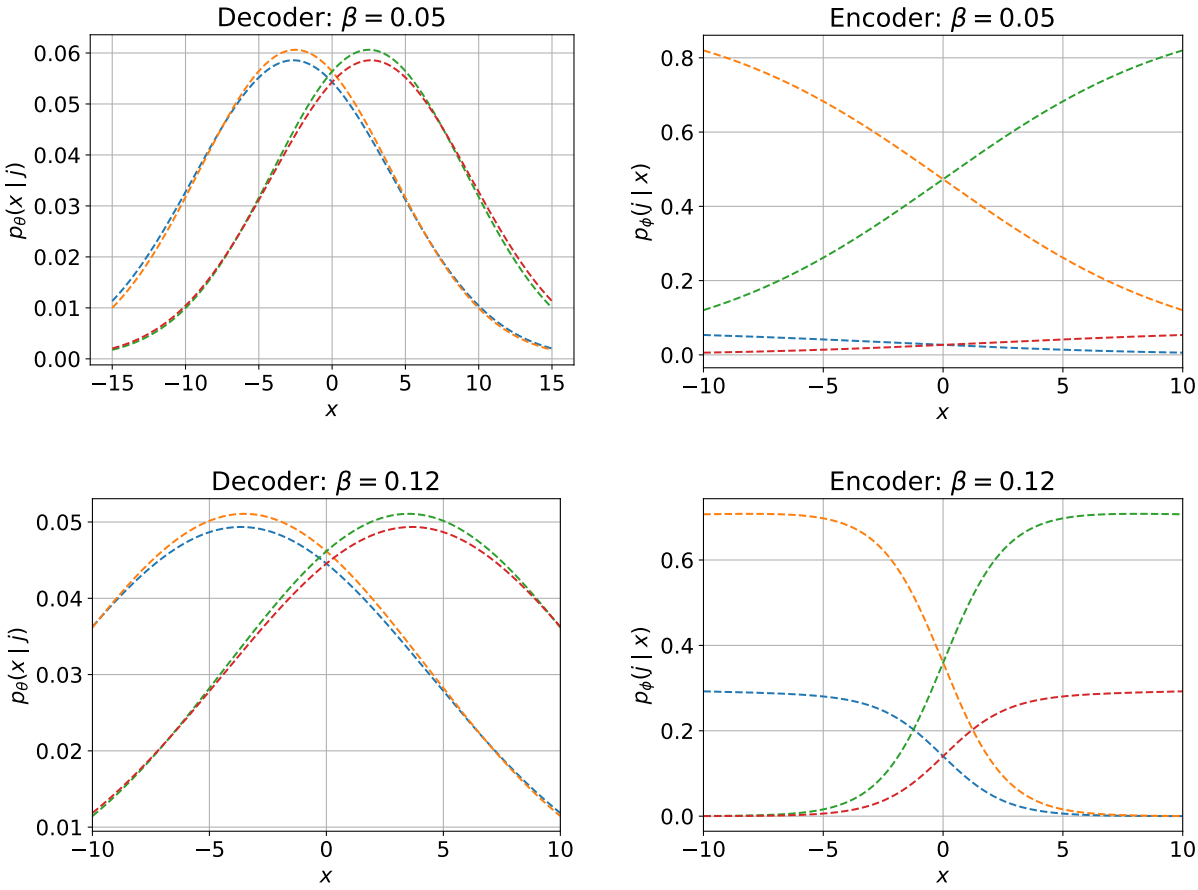
5.3 Allowing More than Two Latent States

The flexibility with which the generative model can approximate the statistics of the training sample is obviously increased by allowing for a larger number of latent states. However, it is not obvious that this additional flexibility matters much in the case of VAEs trained for the specific decision task faced by players in the coordination game studied here. As an example, Figure 5 shows what we obtain when we use the method explained in section to train a symmetric VAE with four latent states (the case $J = 2$). In the numerical examples shown here, the training objective is of the form (3), for two of the same values of β as are used in Figure 3, but we show the results only for the case of the "high-volatility" prior distribution.

In our numerical simulations, the VAE that is learned is one in which the two latent states with negative j are each associated with virtually the same distribution $\tilde{p}_\theta(x\,|j)$ of values for the external state, as are the two latent states with positive j. Thus the generative model that is learned is essentially equivalent to one that could be represented using only two latent states, and indeed essentially the same as the one that is learned in the $J = 1$ case for that value of β, shown in the corresponding rows of Figure 3.

Given this, it does not really matter which of the two latent states with negative j is used to classify a given external state, or which of the two states with positive j; it only matters with what probability the recognition model classifies the external state using a positive latent state as opposed to a negative latent state. With regard to this question, the encoding functions that can be learned when $J = 2$ are somewhat more flexible than those that are possible when $J = 1$: instead of the log odds of classification using a positive latent state having to be a linear function of x, it can be the logarithm of the sum of two exponential functions of x. But this additional flexibility has only a minor effect. In the $\beta = 0.05$ case (top row), the optimal recognition model almost always classifies the external state using one or the other of two latent states only (one positive and one negative). In the $\beta = 0.12$ case (bottom row), all four states are used; but the value of λ_j is similar for both positive values of j (and similarly for both negative values). Hence the log odds

Figure 5: The Optimal VAE with Four Latent States



of classification using a positive latent state are again a nearly-linear function of x.

Thus predicted behavior in the game (with this training objective) is essentially the same is essentially the same in the case of a VAE with $J = 2$ (shown in Figure 5) and one with $J = 1$ (shown in Figure 3). Even larger values of J similarly make no difference, according to our numerical explorations. We accordingly use VAEs with $J = 1$ and a training objective (3) as the basis for our discussion of predicted behavior in the next section.

6 Predicted Behavior in the Coordination Game

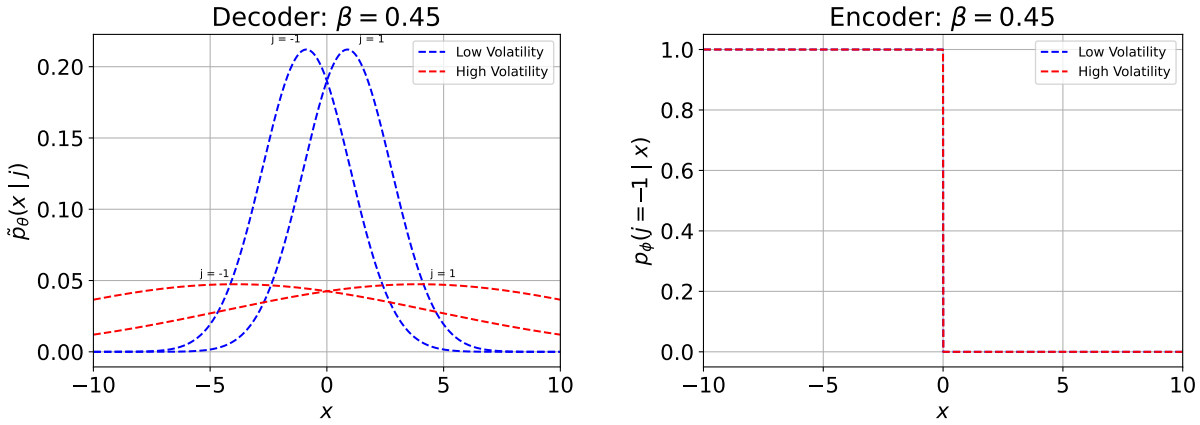
We turn now to the predictions of our VAE model for observed play in a coordination game of the kind described in Section 3. We discuss two aspects in particular of experimentally observed behavior, as illustrated in Figure 2. First, we consider the reason for subjects' decisions to be stochastic, conditional 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.

6.1 Learning Imprecise Action Selection

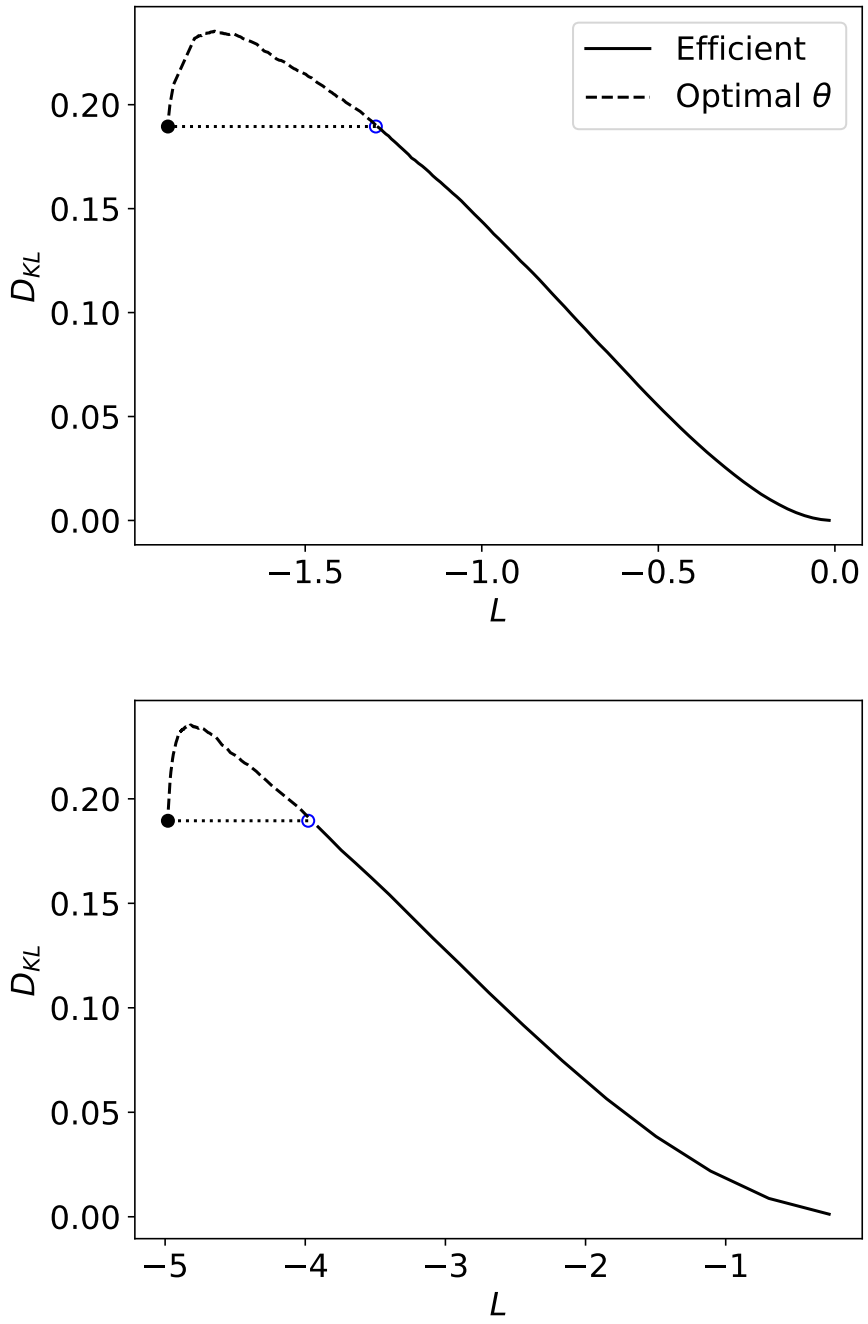
A notable feature of the observed behavior in Figure 2 is that subjects randomize between the two possible actions in the case of all of the values of x between -1 and 1, rather than always staying in the case of low values of x and always leaving when x is higher. Our model implies that behavior of this kind can be a stable outcome of learning by a VAE.

Figure 6: Learned Models With a Greater Weight on Action Selection



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 limiting values of our parameters) for a VAE to yield discrete switching between staying with certainty and leaving with certainty. Indeed, this is what will be learned, for some parameterizations. Figure 6 shows the learned generative and recognition models, using the same format as in Figure 3, but for the value $\beta = 0.45$. In this case, the optimal recognition model involves making λ_1 an unboundedly large positive quantity (and λ_{-1} correspondingly unboundedly negative). 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 exactly the same behavior as in the symmetric Nash equilibrium discussed in Section 3.1. (The same results are obtained for all large enough values of β.)

Figure 7: The Tradeoff Between Minimizing $D_{KL}(p_\phi||\tilde{p}_\theta)$ and Minimizing L



But this is not the kind of behavior that is 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. This is because of the tradeoff that exists between the goal of maximizing congruence (minimizing $D_{KL}(p_\phi||\tilde{p}_\theta)$) and the goal of maximizing the DM's expected payoffs in the game (minimizing L):

a low value of D_{KL} can be achieved only by allowing behavior to be quite random, despite the fact that this necessarily increases the value of L.

We can display this tradeoff by computing the set of possible symmetric equilibria of the game between VAEs, when the VAEs are trained to minimize L subject to a constraint that

$$D_{KL}(p_\phi||\tilde{p}_\theta) \ \leq \ \bar{D},$$

for some finite upper bound $\bar{D}$; our interest is in how the achievable value of L for both players varies with the tightness of the upper bound $\bar{D}$. The numerical tradeoffs between the two objectives are shown in Figure 7, for each of the prior distributions considered above. (The top panel shows the tradeoff in the case of the "low-volatility" prior, and the bottom panel the corresponding tradeoff in the case of the "high-volatility" prior.) Again we assume symmetric VAEs with only two latent states.

For any value of the constraint $\bar{D}$, a symmetric equilibrium in which both players use efficient VAEs will be one in which each player's VAE has a generative model that has been optimized for their recognition model, and each player will have a recognition model parameterized by some common value of λ. We thus begin by computing the symmetric equilibrium associated with each possible value of λ, letting λ vary from zero (the case of perfectly uninformative latent states, corresponding to the top right panel of Figure 3) to an unboundedly large positive value (the case of deterministic choice, shown in the right panel of Figure 6). For each value of λ, there is a unique equilibrium, given by the unique solution θ for the optimal generative model for this encoding model, described in Section 4.5 above. We can then compute the values of both D_{KL} and L associated with this equilibrium. Plotting the pairs $(L(\lambda),\ D_{KL}(\lambda))$ for increasing values of λ in the $L - D_{KL}$ plane, we obtain the curves shown by the dashed curves in either panel of Figure 7.

In either panel of the figure, the locus of equilibria with optimal generative models starts in the lower right corner, where the encoding model specified by $\lambda = 0$ results in $D_{KL} = 0$ (the minimum possible value) and $L = 0$ (the maximum possible value, in the asymptotic limit of an infinite training sample). As λ is increased, L decreases, but at the cost of increasing D_{KL}, up to a certain point; beyond a certain critical value of λ, further increases in λ lower both L and D_{KL}. As $\lambda \to \infty$, the optimal generative model approaches the one shown in Figure 6, and the values $(L(\lambda),\ D_{KL}(\lambda))$ approach (L^*, D^*), the point shown as a large filled dot on the vertical axis.

The set of possible symmetric equilibria associated with different upper bounds $\bar{D}$ on the VAE optimization problem are then the points on the efficient frontier shown as a solid line in either of the panels of the figure. For any value of $\bar{D} \geq D^*$, the efficient solution is the one in which choice is perfectly deterministic, represented by the filled dot at (L^*, D^*). But for any value satisfying $0 < \bar{D} < D^*$, the efficient solution is instead the point on the solid curve at vertical height $\bar{D}$.

Any value of the bound in this range implies that the VAEs learned in equilibrium will involve randomization; indeed, the amount of randomization in equilibrium jumps discontinuously as $\bar{D}$ falls below the critical value D^*.

We also note that for any bound in the range $0 < \bar{D} < D^*$, either player's choice frequencies are predicted to be given by a function of the form

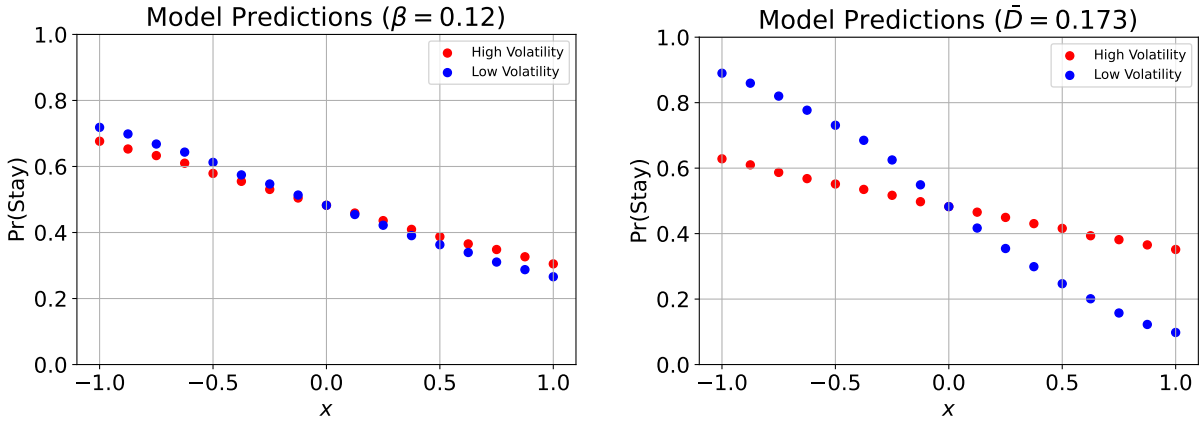
$$\mathrm{P}(stay\,|x) \;=\; \frac{e^{-\lambda x}}{e^{\lambda x} + e^{-\lambda x}}, \tag{21}$$

where the parameter λ takes the maximum value consistent with the bound $\bar{D}$. It follows that this probability should be a continuous, monotonically decreasing function of x, as observed in the experimental data shown in Figure 2.

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 recognition rule (encoding rule) that is learned, depending on the statistics of the observed values $\{x_i\}$ in the training sample.

Figure 8: Effect of Prior Uncertainty on Predicted Behavior



We can consider the predicted effects of the prior distribution $\pi(x)$ on the equilibrium value of λ in (21) in either of two ways. One would be to assume that the players' VAEs are trained using a training objective of the form (3), where the value of β remains the same regardless of the prior. The predicted probability of staying, as a function of x, is shown in a case of this kind in the left panel of Figure 8, under the assumption that $\beta = 0.12$ in both environments. These are the action

probabilities implied by the two encoding functions shown in the bottom right panel of Figure 3, using the optimal decision rule: $a(-1) = stay$, $a(1) = leave$. (The curves are different than in Figure 3 because here we show the predicted choice probabilities only for a discrete set of values of x between -1 and 1, to facilitate comparison with the experimental data shown in Figure 2.)

Alternatively, we might assume that the VAEs are trained so as to minimize L subject to an upper bound $\bar{D}$ on the acceptable degree of divergence measured by D_{KL}, with the bound $\bar{D}$ assumed to be the same regardless of the prior. The right panel of the figure shows the predicted probability of staying, as a function of x, in the case of a common upper bound $\bar{D} = 0.173$ for both players. Under either of these assumptions about the way in which the VAE is trained, we see that players' choices are predicted to vary less sharply with changes in x when x is drawn from a prior with greater variance.

It is simple to show analytically why this is true in the case of a fixed bound $\bar{D}$ that is independent of the prior. Under any prior of the Gaussian family considered here, the rescaled state variable $\tilde{x} \equiv x/\omega$ has a prior distribution that is independent of ω, namely: $\tilde{x} \sim N(0, 1)$. In terms of this state variable, we can write the behavioral rule (21) as

$$\mathrm{P}(stay\,|\tilde{x}) \;=\; \frac{e^{-\tilde{\lambda}\tilde{x}}}{e^{\tilde{\lambda}\tilde{x}} + e^{-\tilde{\lambda}\tilde{x}}},$$

where $\tilde{\lambda} \equiv \omega \cdot \lambda$. We can parameterize the class of possible generative models in terms of the rescaled state variable as well, writing the distribution $\tilde{p}_\theta(\tilde{x}\,|j)$ for any j as $N(\tilde{\mu}_j,\, \tilde{\sigma}_j^2)$. We can then show that the optimal generative model (in terms of the rescaled state variable) for any specified recognition model will depend only on the value of $\tilde{\lambda}$, independent of the value of ω. From this it follows that the value of D_{KL} implied by assuming that the generative model is optimized for the recognition model depends only on the value of $\tilde{\lambda}$, in a way that is independent of the value of ω.

We can then see how the constrained-optimal value of λ associated with a given value of $\bar{D}$ should vary across the true treatments. The maximum value of $\tilde{\lambda}$ allowed by a given bound $\bar{D}$ is the same across environments corresponding to different values of ω. Hence our model predicts that the parameter λ in (21) should be given by $\tilde{\lambda}/\omega$, where the value of $\tilde{\lambda}$ is an increasing function of $\bar{D}$ (that does not depend on ω). If we measure the degree of sensitivity of players' behavior to the value of x by a "discrimination threshold" — say, the amount by which x must be increased in order for $\mathrm{P}(stay\|x)$ to fall from 0.75 to 0.25 — then we obtain a simple prediction: in an environment with greater prior uncertainty about the value of x, the discrimination threshold should increase precisely in proportion to the value of ω, the standard deviation of the prior distribution.[15]

These two alternative assumptions about what aspect of the training objective remains fixed

[15]Note that this is exactly the same quantitative prediction as Frydman and Nunnari (2022) obtain from their model of efficient coding, discussed further below.

across environments agree in predicting that choice frequencies should be less sensitive to the value of x in a more volatile environment, but they make different quantitative predictions about exactly how much less sensitivity there should be, as one can see in Figure 8. Nor are these the only two possibilities. We can obtain quantitative predictions intermediate between these two extremes if we suppose that players' VAEs are trained so as to minimize an objective of the form

$$\min_{\phi,\theta} [D_{KL}(p_\phi || \tilde{p}_\theta)/\bar{D}]^\gamma + \beta L,$$

for some $\bar{D} > 0$ and $\gamma \geq 1$. If $\gamma = 1$, this is equivalent to a training objective of the form (3), the case illustrated in the left panel of Figure 8. In the limit as $\gamma \to \infty$, this training objective approaches the one assumed in the right panel of the figure. Intermediate values of γ result in intermediate degrees of dependence of the learned value of λ on the value of ω in a particular environment. We leave for future work further study of which specification of the training objective best matches observed behavior.

6.3 Alternative Models of Imprecise Coordination

Thus far we have compared our approach with the analysis of imprecise coordination by Yang (2015), based on RI theory. We now briefly compare our approach with others that also model subjects' choices as stochastic conditional on the current value of x.

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 (2022) 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 (14) 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 (14)) 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 (2022).

Frydman and Nunnari (2022) 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 the rescaled state $\tilde{x}$ across environments ("range normalization"). Our model also has this property (in the notation used above, $\tilde{\lambda}$ is the same across environments), in the case that the bound $\bar{D}$ is assumed to be the same across environments. 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 (2022).

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 (2022) 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 $\mu(1)$ 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 $\mu(j)$ 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.[16] Our VAE approach instead provides an explanation for randomness 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. While the architecture that we propose is fairly different from the one that they assume, our model 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.

7 Adaptation to a Change in the Statistics of the Environment

Another feature that our model shares with that of Arifovic et al. (2013) is that rather than simply defining equilibrium as a situation in which each player's decision algorithm is optimal for them given the decision algorithms used by the others, we provide a model of the way in which decision algorithms are 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

[16] 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 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.

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.

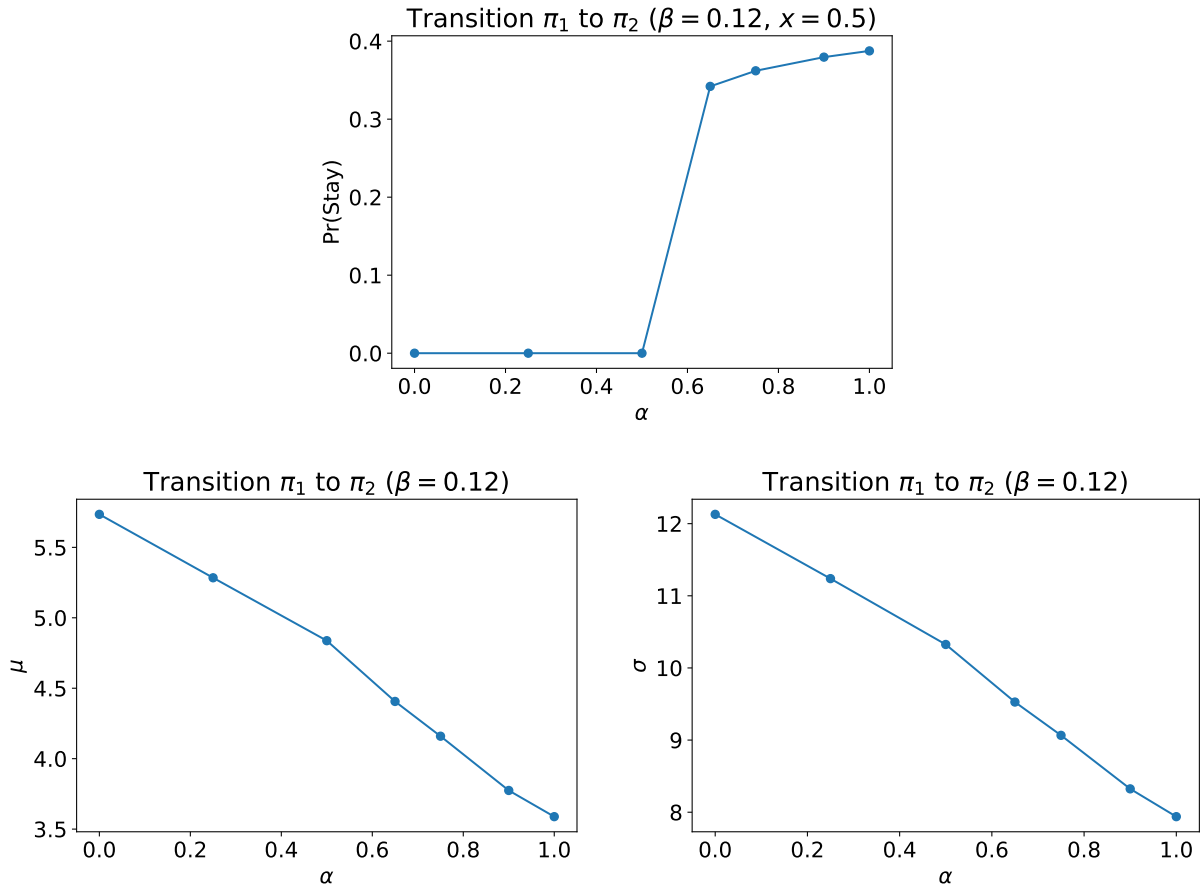
The experimental results of Arifovic et al. (2013) show that in coordination games there is inertia in the strategies played as the underlying environment shifts. In particular, Arifovic et al. (2013) show that if the initial state of the world is such that everyone always plays $stay$ and then the underlying environment shifts to a region where it is optimal in equilibrium to play $leave$ or to only play $stay$ with some probability, there is a transition period between these two equilibria. In this transition period, players will play $stay$ with higher probability relative to the new equilibrium, and vice versa if transitioning from an initial state where in equilibrium they were playing $leave$. Thus (at least during a transition period that may last for quite some time) the observed pattern of play should be history-dependent. We show that our model (like the evolutionary model of Arifovic et al. (2013)) predicts behavior of this kind.

Until now, we have assumed that a DM's training sample is based on some large number of draws from a fixed prior distribution π. However, in order to consider adaptation, we need a model of memory that dictates the training sample at any given point in time as well as how this evolves over time. Furthermore, a desirable property of such a model would be that the decision-maker is never explicitly aware that the underlying statistics of the environment have shifted and rather that the optimal encoding model naturally shifts as more samples from the new environmental distribution are observed. In order to meet these requirements we consider that the training set at a given point in time is determined by a temporally biased reservoir sampling procedure and is provided in Algorithm 2. This algorithm closely follows the algorithm described in Aggarwal (2006), which guarantees that the probability that the r-th observed stimuli value is still present in the sample after the t-th observation is given by $f(r,t) = e^{-\lambda(t-r)}$, where $\lambda = \frac{1}{m}$ for a fixed memory size m. Thus, as the number of samples from the new environmental sample increases the fraction of samples remaining from the original distribution decays exponentially (i.e. eventually converging to the new distribution), the algorithm requires $O(1)$ memory, and the decision-maker never explicitly has to be aware of a shift in the distribution.

We consider the transition from $\pi_1 = x \sim \mathcal{N}(\mu_1 = 5.0, \omega^2)$ to $\pi_2 = x \sim \mathcal{N}(\mu_2 = 0.0, \omega^2)$ and with $\omega^2 = 5.0$. In Appendix A we demonstrate numerically the empirically observed distributions during the transition between π_1 and π_2 and further empirically display the rate of convergence of the training sample to π_2 for different values of m. These exercises highlight that there is a region where the samples in the training data will be drawn from a mixture of π_1 and π_2. In order to build intuition for the transition period, instead of directly considering values of m, we instead focus on characterizing the model implications for an exogenously provided α fraction of samples in the training set coming from π_1.[17]

[17] We can map α to a number of samples given a fixed memory size m.

Figure 9: Model Predictions and Estimates in Transition from π_1 to π_2



For concreteness, we focus on the results for $\beta = 0.12$ and consider the predictions of our model for $x = 0.5$ as we vary α:[18]

$$\alpha \in \{0.0, 0.25, 0.5, 0.6, 0.75, 0.9, 1.0\}$$

The first panel of Figure 9 plots the model predictions for $\Pr(stay)$ as α increases. Figure 9 shows that when $\alpha = 0.0$ we have that $\Pr(stay) = 0.0$. Without any path-dependence, we would expect that the predicted $\Pr(\text{Stay})$ should jump to the resulting value for $\alpha = 1.0$. However, it takes until $\alpha = 0.6$ for $\Pr(stay) > 0$. Furthermore, once α is sufficiently high such that $\Pr(stay) > 0$, it gradually converges to the predictions under the environmental distribution, π_2, but even for $\alpha = 0.9$, it still results in a lower predicted probability of $\Pr(stay)$. The second and third panels

[18]We focus on a concrete value of x for ease of illustration and visualization of the results, though qualitatively similar patterns hold across different values of x and for different values of μ_1, ω, and β.

of Figure 9 show how the model estimates for μ and σ further adapt as α increases. We observe that the model gradually converges towards the μ observed in our baseline cases, consistent with the empirical mean of the training sample converging towards zero, as well as that the volatility of the decoding model decreases over time. Combined, these results are qualitatively consistent with the experimental results of Arifovic et al. (2013), as we observe path-dependence that leads to initial over-playing of $leave$ while also eventually converging to the new equilibrium.

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 (2022). 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 recog-

nition 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

Aggarwal, C. C. (2006). On biased reservoir sampling in the presence of stream evolution. In *Proceedings of the 32nd international conference on Very large data bases*, pp. 607–618.

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 (2022). Cognitive imprecision and strategic behavior. *Available at SSRN 3939522*.

Goeree, J. K., C. A. Holt, and T. R. Palfrey (2016). Quantal response equilibrium. In *Quantal Response Equilibrium*. 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.

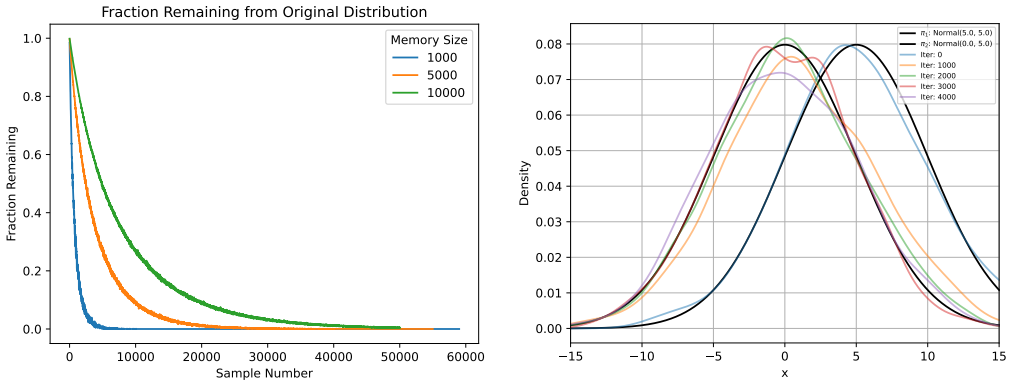
Appendix

A Sample Adaptation

Algorithm 2 Temporally Biased Reservoir Sampling

function BIASEDRESERVOIRSAMPLING(draws,memory_size)
    reservoir = list()
    for item $\in$ draws **do**
        $F = \frac{\text{length(reservoir)}}{\text{memory\_size}}$
        if randomFloat(0, 1) $< F$ **then**
            replaced_index = randomInteger(1, length(reservoir))
            reservoir[replaced_index] = item
        else
            reservoir.append(item)
        end if
    end for
    return reservoir
end function

Figure A1: Transition of Empirical Distribution



Notes: The figure on the left demonstrates the fraction of remaining samples in the training sample from π_1 as the number of samples from π_2 increases. We consider this for $m \in \{1000, 5000, 10000\}$. The figure on the right plots the kernel density of the observed empirical distribution in the transition from π_1 to π_2 with $m = 1000$ and between 0 to 4000 samples.

We investigate the rate of decay of samples from the original π_1 distribution in the transition from $\pi_1 = x \sim \mathcal{N}(\mu_1 = 5.0, \omega^2)$ to $\pi_2 = x \sim \mathcal{N}(\mu_2 = 0.0, \omega^2)$ and with $\omega^2 = 5.0$. The left panel of Figure A1 displays the fraction of remaining observations from π_1 as more samples are drawn from π_2 and confirms that there is an exponential decay of samples from π_1 as well as that the rate of decay is faster when m is lower. The second panel of Figure A1 displays the resulting empirical distribution between π_1 and π_2 for $m = 1000$ and demonstrates the empirical distribution during the transition period.