Proceedings of the 22nd International Conference on Cognitive Modelling (ICCM 2023)

Rational Compression in Choice Prediction

Max Taylor-Davies (s2227283@ed.ac.uk)
School of Informatics
University of Edinburgh

Christopher G. Lucas
School of Informatics
University of Edinburgh

Abstract

To successfully navigate its social environment, an agent must construct and maintain representations of the other agents that it encounters. Such representations are useful for many tasks, but they are not without cost. As a result, agents must make decisions regarding how much information they choose to store about the other agents in their environment. Using choice prediction as an example task, we illustrate the problem of finding agent representations that optimally trade off between downstream utility and information cost, before presenting the results of two behavioural experiments designed to examine this tradeoff in human social cognition. We find that people are sensitive to the balance between representation cost and downstream value, while still deviating from optimality.

Keywords: social cognition, resource rationality, decision-making, information theory

Introduction

In order to produce adaptive behaviour, an agent must acquire and maintain an internal representation of its environment (Craik, 1943; Tolman, 1948; Wilson et al., 2014). For instance, a foraging animal should have some representation of which areas of its environment are most likely to provide food, as well as which might contain sources of danger to avoid. But this is true not only for the inanimate features of the world—unless condemned to an entirely solitary existence, we can expect that many environments encountered by a hypothetical agent will contain *other agents*. Much as an agent should represent the rest of the environment, we expect that it ought also to represent these other agents. Humans do this, of course—in fact, it seems we automatically form mental representations of the other people we encounter (Dennett, 1987; Malle, 2008; Baker et al., 2017). We use these representations for a variety of different purposes: understanding the strengths and weaknesses of a colleague to effectively collaborate with them; determining whether a stranger should be treated as friend or foe; or predicting the plays of a chess opponent in order to defeat them. In general, a detailed representation of the world is more useful than a coarse one, in the sense of allowing greater predictive power or insight. But real agents, whether biological or artificial, inevitably have to contend with limits on their cognitive or computational resources. We therefore do not typically expect an agent to hold within their mind a 1:1 lossless model of the world; instead, they will employ a representational system that involves some degree of approximation or compression. Indeed, the argument for compression is perhaps especially clear in the specific case of representing other agents. As soon as we allow for the fact that this process goes both ways (i.e. as I represent agent X, agent X in turn represents me) then we have to contend with some level of recursion: my representation of agent X must contain within it some representation of myself. For these representations to involve no loss of information, my mind would have to contain within it a number of perfect copies of itself, which cannot be possible. This line of thinking motivates us to consider two related questions. First, how much information *should* an optimal agent represent about the other agents in its environment? And second, is the answer to this question reflected in the choices that people actually *do* make in response to this problem?

Over recent years, there has been a growing body of work in cognitive science that seeks to understand human cognition through the lens of *resource rationality* (Lieder & Griffiths, 2019; Bhui et al., 2021; Icard, 2023). As a framework, resource rationality extends the classic ideas of decision theory (Neumann & Morgenstern, 1953; Jeffrey, 1965) and rational analysis (Anderson, 1990) to account for the notion that agents do not possess infinite capacity for acquiring, storing or processing information. It can also be seen as building on the concept of *bounded* rationality popularised by Simon (among others), while being more explicit in its focus on the idea of *resourcefulness*, i.e. of agents making the most effective use of the cognitive resources available to them. Various formalisations of this idea are possible (Icard, 2023); we will adopt a version of what Icard terms the 'cost-theoretic approach', which considers a continuous tradeoff between the *utility* of a given cognitive or behavioural strategy and the cost of carrying it out. Note that this still leaves considerable flexibility via the choice of how both sides of this tradeoff are defined. As far as cost is concerned, our focus in this paper is specifically on information cost; i.e. the cost of acquiring and storing the representations (of other agents) that support a particular strategy. This is distinct from the computational cost of converting those representations into decisions or behaviour. While a complete analysis should account for both, we leave this for future work, and will focus in this paper on a task setting in which the optimal decision strategy is extremely simple given an appropriate representation.

Task

General objective

In its most general form, the cost-theoretic approach to resource rationality is concerned with maximising an objective function that looks like this:

$$R := S - \lambda C \quad (1)$$

where S is some measure success or performance on our task of interest, C is some measure of the cost(s) we want to minimise, and λ is a tradeoff parameter that governs the relative weight assigned to each quantity. For our purposes, we make R a function of some chosen social representation χ:

$$R(\chi) := S(\chi) - \lambda C(\chi). \quad (2)$$

For any choice of (S, C, λ), the optimal representation is then given by $\chi^* = \arg\max_\chi R(\chi)$. This optimality criterion is similar to the objectives used within work on capacity-limited Bayesian decision-making and RL, such as Arumugam et al. (2024). The key difference (beyond our explicit focus on social representations) is that we are interested not so much in the cognitive cost of converting representations into behaviour, but in the cost the representations themselves. In general, we expect this be a combination of the cost involved in acquiring a representation (i.e. inferring it from observation), and the cost involved in storing it—for now we adopt a simplistic definition of $C(\chi)$ as the *number of bits* in χ, assuming that representations which require a greater number of bits to store or transmit will impose a higher cognitive cost.

Pairwise choice prediction

As for S, we construct a minimal social cognition task where one agent (Alice) tries to predict the choices made by a second agent (Bob). First, let $\mathcal{S}$ be some choice space. $\mathcal{S}$ can in general contain any sort of thing that an agent could make choices over; we will say here that it is the space of possible states of the environment. At trial t, we sample a random pair of states (s_1, s_2) uniformly from $\mathcal{S}$, and Alice makes a prediction c_{pred} about which state Bob will choose. Bob then makes his choice c_{actual}—if Alice's prediction was correct ($c_{\text{pred}} = c_{\text{actual}}$), she earns a reward. Alice's goal is to maximise her total reward earned over some large number of trials. This task is attractive in its conceptual simplicity—but it does also bear a relation to more realistic problems faced by people navigating social environments, such as predicting the lane choice of other drivers on the road, or which of two possible gifts your partner would prefer.

Of course, how well Alice can in principle do on this task depends on how Bob makes his choices. We will assume that Bob is a noisily rational agent whose decisions are described by a Boltzmann choice rule:

$$\Pr[\text{choose } s_1] = \frac{\exp\left(\frac{\mathrm{u}(s_1)}{\beta}\right)}{\exp\left(\frac{\mathrm{u}(s_1)}{\beta}\right) + \exp\left(\frac{\mathrm{u}(s_2)}{\beta}\right)} \quad (3)$$

where $\mathrm{u} : \mathcal{S} \to \mathbb{R}$ is Bob's utility function, which maps elements of $\mathcal{S}$ to scalar utilities, and β quantifies his 'decision noise' (i.e. the extent to which he deviates from optimal choice behaviour). Given this, and assuming access to some approximate representation $\hat{\mathrm{u}}$ of Bob's true utility function (defined over the same state space), the optimal strategy is clearly to make predictions as

$$c_{\text{pred}} \,|\, \hat{\mathrm{u}}, (s_1, s_2) = \underset{s \in (s_1, s_2)}{\arg\max}\, \hat{\mathrm{u}}(s) \quad (4)$$

Using a 0-1 loss, the objective function for a single trial is given by

$$R_{\text{trial}}(\hat{\mathrm{u}}) := \mathbb{1}(c_{\text{pred}} \,|\, \hat{\mathrm{u}} = c_{\text{actual}}) - \lambda n_{\text{bits}}(\hat{\mathrm{u}}) \quad (5)$$

To obtain the general objective function, over both trials and different instances of Bob (with different u), we will treat $\hat{\mathrm{u}}$ as a random variable resulting from the application of some 'representation scheme' to the true utility function u. We can then take the expectation over both state-pairs and $\hat{\mathrm{u}}$ to write

$$R_{\text{expected}}(\hat{\mathrm{u}}) := \underset{\mathcal{S}^2}{\mathbb{E}}[\mathbb{1}(c_{\text{pred}} \,|\, \hat{\mathrm{u}} = c_{\text{actual}})] - \lambda H[\hat{\mathrm{u}}] \quad (6)$$

where H denotes the differential entropy.

If we only have to represent a very small number of agents, or a small state space $\mathcal{S}$, then it may be feasible to represent utility functions exactly (i.e. use $\hat{\mathrm{u}} = \mathrm{u}$), even for $\lambda > 0$. But if the agent population or state space is large, or if $\lambda \gg 0$, then the optimal representation in terms of Equation 6 will likely be an approximation $\hat{\mathrm{u}}$ that discards some information for the sake of lower entropy. A nice consequence of the simplicity of our prediction task is that we can write out an analytical expression for the expected success (i.e. prediction accuracy) given an arbitrary $\hat{\mathrm{u}}$:

$$\underset{\mathcal{S}^2}{\mathbb{E}}\left[\mathbb{1}(c_{\text{pred}} \,|\, \hat{\mathrm{u}} = c_{\text{actual}})\right] = \frac{1}{2} + \frac{1}{2}\underset{\mathcal{S}^2}{\mathbb{E}}\left[\operatorname{sign}(\Delta\mathrm{u}\,\Delta\hat{\mathrm{u}}) \tanh\left(\frac{\Delta\mathrm{u}}{2\beta}\right)\right] \quad (7)$$

where $\Delta\mathrm{u} = \mathrm{u}(s_1) - \mathrm{u}(s_2)$ and $\Delta\hat{\mathrm{u}} = \hat{\mathrm{u}}(s_1) - \hat{\mathrm{u}}(s_2)$. A derivation for this expression is given in Appendix B, but the intuition here is that the prediction accuracy given $\hat{\mathrm{u}}$, relative to the prediction accuracy given u, depends on the probability that $\hat{\mathrm{u}}$ can correctly resolve the 'polarity' of a pair of states resolved by u. The objective function in Equation 6 can then be written as

$$R_{\text{expected}}(\hat{\mathrm{u}}) = \frac{1}{2} + \frac{1}{2}\underset{\mathcal{S}^2}{\mathbb{E}}\left[\operatorname{sign}(\Delta\mathrm{u}\,\Delta\hat{\mathrm{u}}) \tanh\left(\frac{\Delta\mathrm{u}}{2\beta}\right)\right] - \lambda H[\hat{\mathrm{u}}] \quad (8)$$

Compression through state aggregation

So far, we have just considered the idea of approximate representations in the abstract. But what might these approximate representations actually look like? One straightforward way to approximate a utility function is through state aggregation—i.e. group all states within a given-sized 'patch' of $\mathcal{S}$ under a single value (Sutton & Barto, 2018; Abel et al., 2019). It is important here to note that we do not take this to be an optimal (or even particularly strong) compression strategy for any given state space $\mathcal{S}$—but its simplicity and generality makes it an attractive choice for illustrating the tradeoff dynamics that we are concerned with. To do this, we set up a simulation

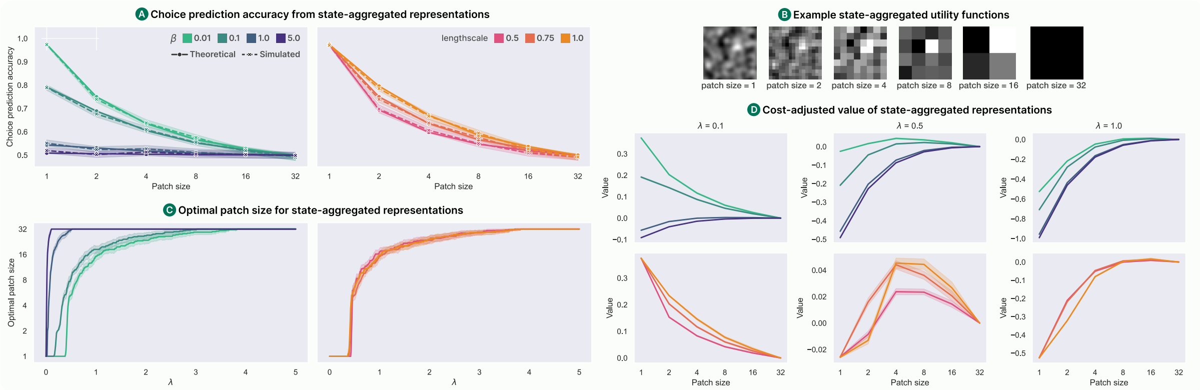

Figure 1: Results of simulating pairwise choice prediction with state-aggregated utility function representations, for noisily rational target agents with spatially correlated 2D utility functions. **(A)** theoretical (Equation 7) vs simulated prediction accuracy as a function of aggregation patch size for different values of utility function lengthscale and target agent β. **(B)** illustration of state aggregation levels for example 2D utility function. **(C)** optimal patch size as a function of increasing tradeoff parameter λ. **(D)** cost-adjusted return (Equation 8) as a function of patch size, with cost given by continuous entropy.

environment where agents make choices over pairs of tiles in a 32x32 2D grid. Each agent has a spatially correlated utility function drawn from a Gaussian Process (with RBF kernel), and a decision noise parameter β. Over a number of trials, we simulate a choice prediction strategy using different levels of state aggregation (where state aggregation is measured by 'patch size', i.e. the number of grid tiles grouped under a single value in the aggregated representation). For this simple setup, the specific relationship between patch size and prediction accuracy should be determined by both the lengthscale of the Gaussian Process (i.e. how smoothly u varies over $\mathcal{S}$), and the decision noise β of the agents making the choices—we therefore repeat the simulation for different values of each parameter (keeping the other constant).

The results of these simulations are shown in Figure 1. First, panel **(A)** shows that Equation 7 successfully captures the effect on simulated choice prediction accuracy of increasing patch size, across all simulated values of lengthscale and β. As we would expect, prediction accuracy decreases monotonically with increasing patch size. Furthermore, for any given patch size < 32, prediction accuracy decreases with increasing β (i.e. as agents become more unpredictable). We also see that the decrease in prediction accuracy with patch size is less steep at higher lengthscale (i.e. smoother u), where less information is lost for a given amount of aggregation. Panel **(D)** illustrates the expected cost-adjusted return (Equation 8) as a function of patch size, for the same set of β and lengthscale values, and for various values of the tradeoff parameter λ. We see that the optimal aggregation level is shifted to the right as we increase λ (and thus care more about information cost). This same trend is also seen in panel **(C)**, which shows directly how the optimal patch size changes as a function of λ.

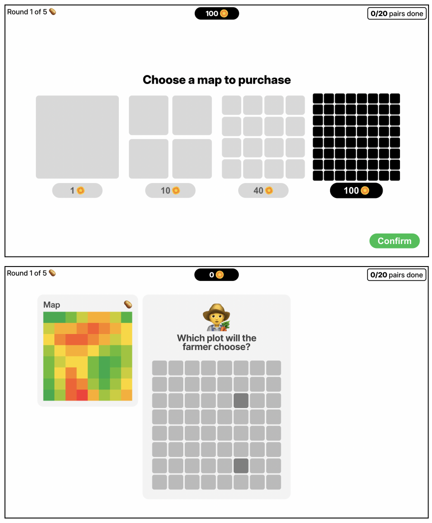

Figure 2: Interface for behavioural experiments

Experiment 1

In the preceding sections, we presented a theoretical and computational analysis of the tradeoff between information cost and predictive value faced by agents representing others' utility functions. We now seek to shed light on this tradeoff in *human* social cognition—that is, does our optimal analysis predict people's actual choices about how much information to represent about other agents' utility functions?

Procedure

To answer this question, we developed a behavioural experiment based on the simple pairwise choice prediction task outlined above. We recruited a total of $n = 90$ adults through the online platform Prolific, who were then directed to an online game consisting of 5 rounds (excluding an initial tutorial round), and instructed to try and maximise their final score. To incentivise performance, participants were rewarded with bonus payments for achieving scores above a certain threshold. The cover story for the task was that participants (playing the role of Alice) had to predict the choices made between different plots of land (tiles) in a field (8x8 2D grid) by a farmer (playing the role of Bob) trying to grow a particular crop. While making choice predictions, participants were given access to a (possibly aggregated) representation of the farmer's utility function in the form of a map of the different plots' 'quality' for growing the crop in question. Crucially, the tradeoff dynamics were introduced via a game mechanic were participants *spent* points to acquire this map, and *earned* points for correct predictions. At the start of each round, participants selected a level of state aggregation, paying a cost in points determined by the level of aggregation chosen. They were then presented with a series of randomly sampled tile pairs—for each pair they were instructed to select the tile that they thought agent would choose, while being able to see the (possibly aggregated) map they had just 'purchased'. For simplicity, the farmer was set to be perfectly rational, i.e. $\beta \to 0$ under the choice rule in Equation 3. Figure 2 shows the main components of the game interface. We varied two factors between participants in a 2x3 design (with 15 participants per condition): the texture of the 2D utility functions ('rough', 'smooth', corresponding to GP lengthscales of 0.5 and 2 respectively), and the absolute costs of the different maps ('low', 'medium', 'high'). All participants faced the same number and sequencing of rounds, regardless of condition, and the *relative* cost of the different maps was always the same. To maximise their overall score, a given participant would need to choose, at each round, the level of state aggregation that optimally balanced cost against expected predictive value (depending on their assigned condition). We recorded participants' choices of aggregation level at each round, as well as all of the pairwise choice predictions that they made.

Results

From our behavioural data, we compute participants' average prediction accuracy as a function of aggregation level (patch size), split by texture condition. We then compare these in Fig 3 **(A)** to Equation 7. We can see that participants' average prediction accuracy is fairly well captured by the model—that is, participants in general made effective use of the information contained in their chosen representations. Participants were also more accurate in the smooth utility function condition, reflecting the fact that less information is lost when aggregating spatially correlated functions with higher lengthscale. So, our model predicts how participants' prediction accuracy varies with aggregation level. But can it predict which aggregation levels participants will select? For each of the 6 conditions, we compare the recorded proportions of participants' patch size selections against the choice distribution given by three different variants of a noisily rational model

$$\Pr\{\text{select } \hat{u}\} \propto \exp\left(\frac{V_m(\hat{u})}{\beta}\right) \tag{9}$$

where $V_m(\hat{u})$ is set as either the expected accuracy, the negative cost, or the full cost-adjusted return (from Equation 8). This comparison is shown in Figure 3 (B), using $\beta = 0.25$. While none of these three models is able to capture participants' patch size selections perfectly, it is clear that the full resource-rational choice rule is a much better fit than either the accuracy-only or cost-only models—indicating that to at least some extent, participants are sensitive to the tradeoff between information cost and predictive value in selecting representations. For instance, participants' selection probability decreased monotonically with increasing patch size in the low-cost condition, and increased almost monotonically in the high-cost condition. However, for the medium cost condition, the resource-rational model predicts a greater difference in selection probabilities between the rough and smooth conditions than was reflected in participants' behaviour. This suggests that participants in our experiment, while sensitive in general to the balance of value and cost, were not *fully* resource-rational with respect to the specific parameters of their task environment.

Experiment 2

Procedure

We conduct a second behavioural experiment, as a small variation on Experiment 1. Rather than varying utility function texture, we now vary the decision noise of the target agent. Participants (total $n = 30$) were divided equally between a 'low noise' condition, where they encountered an agent with $\beta = 0.01$, and a 'high noise' condition, where $\beta = 1.0$. The game structure and mechanics were otherwise unchanged from Experiment 1. All utility functions were taken from the 'smooth' condition of Experiment 1 (lengthscale $= 2.0$), and absolute map costs were kept constant between all participants.

Results

The results of Experiment 2 are shown in Figure 4. Participants' selection of representations is compared to the same three models as used for Experiment 1. In this setting, the resource-rational model captures the idea that the cost-usefulness tradeoff is affected by target agent β. As an agent's decision-making gets noisier, the marginal predictive value of information about their utility function decreases—therefore the representation strategy of an agent seeking to optimise this tradeoff should be shifted towards higher aggregation as decision noise increases. Looking at Figure 4, we can see that this trend is indeed reflected in participants' behaviour, at least to some degree: for instance, the lowest aggregation level was chosen more in the 'low noise' condition, and the highest aggregation level was chosen more in the 'high noise' condition. As in Experiment 1, the resource-rational model

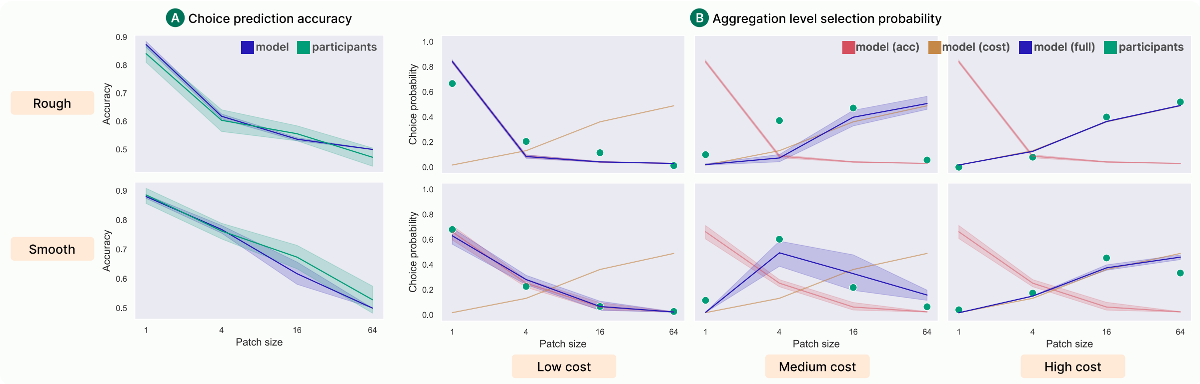

Figure 3: Results of Experiment 1. **(A)** participants' average choice prediction accuracy as a function of patch size, split by texture condition and compared to the theoretical accuracy predicted by Equation 7. **(B)** empirical patch size selection probabilities for participants from each condition, compared to those given by Boltzmann-rational models (Equation 9) based on only expected accuracy, only information cost and the full cost-adjusted return objective (Equation 8), with $\beta = 0.25$.

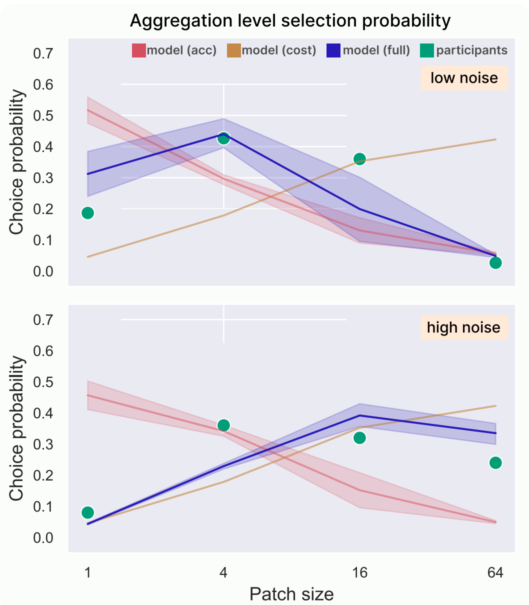

Figure 4: Results of Experiment 2: empirical patch size selection probabilities for participants from low and high noise conditions, compared to those given by Boltzmann-rational models (Equation 9) based on only expected accuracy, only information cost and full cost-adjusted return (Equation 8), with model $\beta = 0.45$.

gives a better fit than either of the accuracy-only or cost-only models, but again we see points of noticeable deviation (at patch size $= 16$ in the 'low noise' condition and patch size $= 4$ in the 'high noise' condition).

Discussion

In this paper, we have considered the relatively unexplored problem of how much information to represent about other agents in social cognition, through the example task of predicting an agent's choices over pairs of options. Specifically, we examined the tradeoff that an observer agent faces between information cost and predictive value in choosing how much information to represent about a target agent's utility function. We first presented some brief theoretical and computational analysis of how a simple state aggregation strategy can be used to navigate this tradeoff. We then conducted two behavioural experiments to compare people's choices of representation to a resource-rational state-aggregation model. Our findings were mixed: while for both experiments the resource-rational model fit our recorded data better than simpler decision rules based only on expected predictive accuracy *or* representation cost, participants still showed non-trivial deviations from optimality. This may be explained by the fact that our experimental setup is highly simplistic, and uses only a single *explicit* representation cost as stand-in for the real cognitive costs of information acquisition and storage. For instance, participants' behaviour may look closer to optimal under an extended model that accounts for additional constraints on e.g. attention and memory. Future experimental work should attempt to probe these nuances, and bridge the high-level computational view presented in this paper with more detailed and psychologically grounded notions of cognitive cost. An additional direction for future work is exploring strategies for obtaining resource-rational representations of agent utility functions that go beyond naive state aggregation—e.g. by representing individuals primarily in terms of their group affiliations or other social identity cues.

Appendix

A: Representation entropy

Here we provide some additional details on our use of representation entropy as a measure of cognitive cost. For a comprehensive introduction to entropy (and other related information-theoretic quantities) we direct the reader to Cover & Thomas (2006) or MacKay (2002). For a discrete random variable $X \in \mathcal{X}$ with probability mass function $p(x) := \Pr\{X = x\}$, the entropy of X is given by

$$H_b(X) = -\sum_{x \in \mathcal{X}} p(x) \log_b(p(x)) \tag{10}$$

When $b = 2$ (as it typically is), the entropy has units of bits. For a continuous random variable $Y \in \mathcal{Y}$ with probability density function $p(y)$, the differential entropy of Y is given by

$$H_b(Y) = -\int_{y \in \mathcal{Y}} p(y) \log_b(p(y)) \tag{11}$$

If Y follows a multivariate Gaussian distribution with covariance Σ, then $H(Y)$ is computed as

$$H_2(Y) = \frac{1}{2}\log_2|\Sigma| + \frac{n}{2}\big(\log_2(2\pi e)\big) \tag{12}$$

where n is the dimensionality of Y (Rasmussen & Williams, 2006). This allows us to compute the entropy of our agent utility functions u, since each is a continuous random variable distributed according to a multivariate Gaussian with known Σ. To compute the entropy of a state-aggregated utility function estimate û, we can treat û as a new continuous RV with a lower-dimensional multivariate Gaussian distribution whose covariance Σ_{agg} is determined entirely by Σ and the level of state aggregation. Determining Σ_{agg} is then sufficient to compute $H(\hat{u})$.

B: Expected prediction accuracy from approximate utility functions

Let û be an arbitrary approximation to the utility function u. We want to find an expression for $\mathbb{E}_{\mathcal{S}^2}\left[\mathbb{1}(c_{\text{pred}} \,|\, \hat{u} = c_{\text{actual}})\right]$—that is, the expected accuracy of an observer predicting the choices of a noisily rational agent over pairs of different states sampled independently from $\mathcal{S}$, given that the observer represents the target agent's utility function as û. For any given pair of states (s_1, s_2) we define $\Delta u = u(s_1) - u(s_2)$ and $\Delta\hat{u} = \hat{u}(s_1) - \hat{u}(s_2)$. Since the optimal prediction strategy is to predict the higher-value state, the prediction made for a given state pair, guided by representation û, depends only on $\text{sign}(\Delta\hat{u})$. For any particular pair (s_1, s_2), the sign product between u and û can take one of three values: $\text{sign}(\Delta u \Delta\hat{u}) \in \{-1, 0, 1\}$. Let p_u be the probability that an observer representing the *full* u would predict the choice correctly. We can then express the equivalent probability for û $p_{\hat{u}}$ in terms of p_u as

$$\begin{aligned} p_{\hat{u}} = &\Pr\big\{\text{sign}(\Delta u \Delta\hat{u}) = 1\big\} p_u \\ &+ \Pr\big\{\text{sign}(\Delta u \Delta\hat{u}) = -1\big\}(1 - p_u) \\ &+ \Pr\big\{\text{sign}(\Delta u \Delta\hat{u}) = 0\big\}\frac{1}{2} \end{aligned}$$

Using the fact that $\mathbb{E}_{\mathcal{S}^2}\left[\text{sign}(\Delta u \Delta\hat{u})\right] = \Pr\big(\text{sign}(\Delta u \Delta\hat{u}) = 1\big) - \Pr\big(\text{sign}(\Delta u \Delta\hat{u}) = -1\big)$, we can then write the expected prediction accuracy using û as

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathcal{S}^2}\left[\mathbb{1}(c_{\text{pred}} \,|\, \hat{u} = c_{\text{actual}})\right] = &\frac{\mathbb{E}\left[\text{sign}(\Delta u \Delta\hat{u})\right] + 1}{2} p_u \\ &+ \frac{1 - \mathbb{E}_{\mathcal{S}^2}\left[\text{sign}(\Delta u \Delta\hat{u})\right]}{2}(1 - p_u) \end{aligned} \tag{13}$$

Substituting

$$p_u = \mathop{\mathbb{E}}_{\mathcal{S}^2}\left[\frac{1}{\exp\left(\frac{-|\Delta u|}{\beta}\right) + 1}\right] \tag{14}$$

(from the definition of the Boltzmann-rational choice rule), and using the identity

$$\begin{aligned} &(z+1)\frac{1}{\exp(-x)+1} \\ &+ (1-z)\left(1 - \frac{1}{\exp(-x)+1}\right) \\ &= z\tanh\left(\frac{x}{2}\right) + 1 \end{aligned} \tag{15}$$

we obtain

$$\mathop{\mathbb{E}}_{\mathcal{S}^2}\left[\mathbb{1}(c_{\text{pred}} \,|\, \hat{u} = c_{\text{actual}})\right] = \frac{1}{2} + \frac{1}{2}\mathop{\mathbb{E}}_{\mathcal{S}^2}\left[\text{sign}(\Delta u \Delta\hat{u})\tanh\left(\frac{\Delta u}{2\beta}\right)\right] \tag{16}$$

References

Abel, D., Arumugam, D., Asadi, K., Jinnai, Y., Littman, M. L., & Wong, L. L. (2019, Jul.). State abstraction as compression in apprenticeship learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 3134-3142. doi: 10.1609/aaai.v33i01.33013134

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Arumugam, D., Ho, M. K., Goodman, N. D., & Van Roy, B. (2024, 04). Bayesian reinforcement learning with limited cognitive load. *Open Mind*, *8*, 395-438. doi: 10.1162/opmi_a_00132

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 1–10. doi: 10.1038/s41562-017-0064

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, *41*, 15-21. doi: 10.1016/j.cobeha.2021.02.015

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory (wiley series in telecommunications and signal processing)* (Second ed.). USA: Wiley-Interscience.

Craik, K. (1943). *Hypothesis on the nature of thought.*

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: The MIT Press.

Icard, T. F. (2023). *Resource rationality.* (Unpublished manuscript)

Jeffrey, R. C. (1965). *The logic of decision*. New York, NY, USA: University of Chicago Press.

Lieder, F., & Griffiths, T. L. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*. doi: 10.1017/S0140525X1900061X

MacKay, D. J. C. (2002). *Information theory, inference & learning algorithms*. USA: Cambridge University Press.

Malle, B. F. (2008). Chapter 12 - the fundamental tools, and possibly universals, of human social cognition. In R. M. Sorrentino & S. Yamaguchi (Eds.), *Handbook of motivation and cognition across cultures* (p. 267-296). San Diego: Academic Press. doi: 10.1016/B978-0-12-373694-9.00012-X

Neumann, J. V., & Morgenstern, O. (1953). *Theory of games and economic behavior* (Third ed.). Princeton, NJ: Princeton University Press.

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Cambridge, Mass.: MIT Press.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (Second ed.). The MIT Press.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, *55 4*, 189-208.

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*, 267-279.