Intuitive Priors*

Jawwad Noor

August 9, 2021

Abstract

A probability measure over a multi-dimensional state space is an Intuitive Belief if it is an aggregation of pairwise associations. Associations are shown to correspond to an analog of pointwise mutual information, and a separability property in beliefs is shown to characterize the model. Associations are optimized so as to match the data, given by an objective probability distribution over observed states. The resulting beliefs exhibit inductive inference, generically placing positive probability on states that they have not observed. The model accommodates the classic belief biases documented in Psychology and explains correlation neglect, experience effects, the disposition effect and empirical patterns in overconfidence.

JEL classification: C45, D01, D90

Keywords: Belief formation, intuition, associations, neural networks, Boltzmann machine, machine learning, inductive inference, belief biases, correlation neglect, experience effect, overconfidence, disposition effect.

1 Introduction

Consider a researcher that is reading a paper and suddenly gets the feeling that something is not quite right in the results, but is unable to articulate why. Only subsequent thought determines whether her feeling is incorrect or whether it is in fact a flash of brilliance. This example illustrates the intuitive process: it is an automatic form of information processing that takes place beyond our awareness, the results of which we become aware of only through a feeling (Simon 1995, Betch 2007).

Real-world investors claim to often be empowered by their gut-level assessments when making decisions (Salas et al 2010, Hensman and Sadler-Smith 2011, Huang and Pearce 2015, Huang 2018). Anecdotally, investors speak of the importance of having a "feel for the market", and investors may not pursue an opportunity if they are "not feeling it". The idea that decisions may be based on a feeling without articulated reasons is also reminiscent of the conviction of Keynes (1936, pp 161-162) that investors are driven by a "spontaneous urge to action", which he referred to as "animal spirits".

*Noor is at the Department of Economics, Boston University, 270 Bay State Road, Boston MA 02215. Email: jnoor@bu.edu. This paper supercedes an earlier paper titled "Intuitive Beliefs". The author acknowledges valuable comments from Roland Benabou, Yi-Hsuan Lin and Bart Lipman and helpful feedback from seminar audiences at University of Paris 2, Yale, Duke, Hong Kong University, UC-Davis, UCLA, NYU-AD, Erasmus University, Princeton, UC-Riverside and Northwestern (Econ-CS workshop), and conference participants at Bounded Rationality in Choice Conference (Aarhus), Asian Meeting of the Econometric Society (Xiamen), Time Uncertainties and Strategies Conference (Paris), Decisions: Theory, Experiments, Applications (Paris), Risk Uncertainty and Decisions (Heidelberg), the Manchester Economic Theory Workshop, and North American Meeting of the Econometric Society (Montreal). Expert research assistance has been provided by Yuzhao Yang, Alonso Ahumada and Yi-Hsuan Lin. The author is also grateful for the hospitality of the Cowles Foundation and the Department of Economics at Yale where part of this research was completed. The usual disclaimer applies.

A role for intuition in choice under complexity may arise as a natural consequence of the fact that economic environments are too complex to be fully navigated with analytical reasoning alone.

This paper models the intuitive process and its formation. A key motivation is that there is now a large literature on the disparate deviations from rational behavior, and it is desirable to have a small number of models that unify as much of the evidence as possible. If a range of non-rational behaviors is in fact a product of intuitive reasoning, then a theory of intuition may serve such a purpose.

Our theory conceptualizes the phenomenon of intuition in terms of the phenomenon of *associations* (Tversky and Kahneman 1974, Betch 2007, Morewedge and Kahneman 2010).[1] An association is the brain's natural tendency to form *connections* between observations in the world.[2] Associations are, for instance, at play in memory retrieval cues (such as when a song on the radio evokes a memory), default thinking patterns (such as systematically interpreting things negatively) and phobias (such as when the sight of a spider evokes fear). Associations are acquired from the environment through the process of associative learning, and are continually strengthened or weakened by experience through factors such as frequency, repetition, similarity and salience (see Wasserman and Miller 1997 for a review of the psychology literature). Thus, an investor's gut response to news will be driven by associations she has formed between different variables through her experience in the environment. Similarly, a researcher's gut reactions when reading a paper arise from the connections formed through prior education and research that are triggered by the paper.

Restricting attention to the context of beliefs over a multi-dimensional state space, we model intuitive beliefs as being constructed from a network of associations. Our modelling inspiration comes from AI. Artificial neural networks are models of thinking, and machine learning is a model of how thinking patterns are learnt from the environment. We model beliefs as a stochastic neural network that is "trained" by the environment: associations are weights in a network, aggregated to produce a belief, and associations are determined in an environment so as to try to match an objective probability distribution.

More formally, we take as our primitive a probability measure p over some finite product set $\Omega = \prod_{i=1,..,N} \Omega_i$, consisting of multi-dimensional states of the world with generic element $x = (x_1, .., x_N)$. *Intuitive Beliefs* take the following form for any state $x = (x_1, .., x_N) \in \Omega$,

$$p(x_1, .., x_N) = \frac{1}{Z} exp \left[\sum_{i<j} a(x_i, x_j) + \sum_i b(x_i) \right].$$

In assessing the likelihood of state $x = (x_1, .., x_N)$, the model sums the associations $a(x_i, x_j)$ between each pair x_i, x_j in distinct dimension i, j, and the background association $b(x_i)$ between each x_i and unmodelled background information. The model and our notion of training of beliefs (outlined shortly) is directly inspired by the *Boltzmann machine* (Hinton and Sejnowski 1983, Ackley et al 1985), an energy-based stochastic neural network used in AI. However, the functional form is also reminiscent of the density of a multivariate Gaussian distribution, which aggregates pairwise relationships defined indirectly via the covariance matrix. In studying the structure of Intuitive

[1] Betch (2007) defines intuition in the following terms: "Intuition is a process of thinking. The input to this process is mostly provided by knowledge...primarily acquired via associative learning. The input is processed automatically and without conscious awareness. The output of the process is a feeling that can serve as a basis for judgments and decisions". Morewedge and Kahneman (2010) posit that intuitive judgements are made through automatic, non-deliberative "System 1" processing, which makes use of heuristics and associative memory.

[2] Associationism, a philosophical school with early expositors such as David Hume, recognized the creation of associations as the most basic function of the mind and sought to reduce all mental life to associations. It served as the foundation of behavioral psychology in the early 20th century until it gave way to the cognitive revolution in the mid 20th century. In its particular manifestation as conscious memory, assocations have been modelled as networks in cognitive psychology using spreading activation networks (Collins and Loftus 1975, Anderson 1983). More advanced modelling of associative memory was taken up in the study of artificial neural networks in AI (for instance Hopfield 1982), which also had an influence in psychology (Kahana 2020).

Beliefs, we establish an expression for a and b in terms of p (Theorem 1), provide a characterization result for the model (Theorem 2), and delineate its uniqueness properties (Theorem 3).

To model how Intuitive Beliefs are formed, we suppose that there is an objective probability distribution q over Ω, and that the agent's experience involves observing some subset of states $\phi \neq D \subset \Omega$. The agent's experience is therefore given by the conditional distribution $q(.|D)$ over D. We posit that associations are optimized by the intuitive process to best match the agent's experience. More formally we say that Intuitive Beliefs p are trained by q and D if they solve

$$min_{p' \in \Delta_{IB}(\Omega)} KL(q(.|D)||p'(.|D)),$$

where $\Delta_{IB}(\Omega)$ denotes the set of Intuitive Beliefs and $KL(.)$ denotes KL-divergence. Notably, we *permit* Intuitive Beliefs to be formed over all of Ω even though the agent only observes $\phi \neq D \subset \Omega$. Under various richness conditions on D, we characterize the set of trained beliefs (Theorem 4). We find that trained Intuitive Beliefs generically possess an *inductive inference* property, defined by $D \subset supp(p)$. That is, even though the agent only observes states in D, she may nevertheless place strictly positive probability on states outside D.

In one application of the training problem, we consider an extended environment where the limited observations D arise not due to constraints in the data available in the environment, but rather cognitive constraints on how many dimensions of a state the agent can perceive at a time. In this context, we show that trained Intuitive Beliefs take the following reduced form,

$$p(x) = \frac{1}{Z}\left[\prod_{i<j} \frac{q(x_i x_j)}{q(x_i)q(x_j)}\right] \times \prod_{i \in \Gamma} q(x_i), \tag{1}$$

that is, beliefs aggregate objective marginals $q(x_i)$ and a particular measure of correlation $\frac{q(x_i x_j)}{q(x_i)q(x_j)}$ (known as *pointwise mutual information* in Information theory). While Theorem 1 provides the general empirical meaning of a and b in terms of p in the abstract, the above reduced form provides the empirical meaning of a and b in terms of the empirical distribution q within which the agent is operating.

Finally, we show that our model of trained Intuitive Beliefs can accommodate a range of disparate empirical findings. In fact, most of these findings are generated by trained Intuitive Beliefs without any new assumptions:

- The model endogenously produces *correlation neglect*. Subjects in experiments appear unable to understand the set of payoffs generated by two correlated assets, nor reconstruct a true sequence of signals from one with a simple known bias (Eyster and Weiszacker 2016, Enke and Zimmerman 2013), behaving as if they allow for more possibilities than are logically implied by the correlation structure offered to them. The intuitive inference property noted above directly gives rise to this: the observed states shape the associations, and the aggregation of these associations by intuitive process generates a belief in new states, even if they are objectively impossible.

- While the inductive inference property can lead to a positive belief about impossible states, it can also lead to an exaggerated belief about the likelihood of states that have small objective probability. As a result, the model produces patterns observed in the overconfidence literature (Moore and Healy 2008) in Psychology and the disposition effect (Shefrin and Statman 1985) studied in Behavioral Finance.

- The model endogenously produces *experience effects*, whereby experience affects behavior even where it is not objectively relevant (Malmendier and Nagel 2016, Bordalo et al 2019). Observe that the reduced form (1) involves marginal beliefs $q(x_i x_j)$. Since these are computed using all y_{-ij}, it follows that the agent's experience with states $y \neq x$ enters into her belief about

x. So, for instance, a doctor may prescribe a treatment to a patient not just on the basis of adequate data on how the treatment has fared for this particular patient in the past, but also on the basis of her experience with the treatment on other patients.

- The model can explain classic belief biases studied in Psychology that reveal that subjects' judgments under uncertainty bear little semblance to the Bayesian model (Tversky and Kahneman 1974, Benjamin 2019). Unlike the preceding, this requires a substantive hypothesis: the agent's intuition about probabilities is shaped entirely by the sampling distribution. That is, we do not presume that the agent engages in any probabilitistic reasoning, but rather constructs beliefs based on associations shaped by the sampling distribution. The model gives rise to the Law of Small Numbers, Non-Belief in the Law of Large Numbers, the Gambler's fallacy and the Hot Hand Effect. In contrast to the literature that accommodates such findings by assuming non-Bayesian updating, we show that the experimental findings can be viewed as properties of the *prior* alone, and do not require a deviation from Bayesian updating.

Related literature: This paper intersects with several strands of literature.

Gilboa and Schmeidler (2003) and Billot et al (2005) study case-based beliefs, where a (presumably consciously deliberating) agent employs a subjective assessment of "similarity" between cases to convert a database of past observed cases into a belief over the possible outcomes of a current case. Argenziano and Gilboa (2019) take inspiration from classification tasks in AI to model the formation of the similarity function. Our model shares the inductive inference and experience effects properties with them, albeit formulated differently. See the discussion in Sections 4.1 and 5.3.

Within the literature on belief biases (see Benjamin 2019 for a review), Gennaioli and Shleifer (2010) modify Bayes Rule to capture the idea that learning an event may invite only particular "representative" states to come to the agent's mind, and Mullainathan (2002) and Bodoh-Creed (2020) explicitly incorporate properties of associative memory (such as the fact that a memory is easier to recall the more frequently it is recalled) in to a dynamic model. These models share with us the view that associations play a role in beliefs, but while we study the role of associations in construction of the prior (and invoke Bayesian updating where needed), these models study the role of associations in giving rise to Non-Bayesian updating (with an arbitrary prior).

Bordalo et al (2019) model associative recall in a choice context rather than beliefs. Past observations are stored in the agent's memory, and a new observation leads to (stochastic) recall of past observations by means of a similarity function. The recalled observations define a "norm" (which acts as a reference point), and the agent maximizes a norm-dependent utility. The model gives rise to experience effects, since experience determines memory and thus the norm and subsequent choice. In our model, q similarly determines beliefs, but experience effects arise also in a different form: for any event $E \subset \Omega$, Bayesian-conditional Intuitive Beliefs $p(\centerdot|E)$ can depend on the objective probability q of states outside E.

The preceding models are not based on neural networks. Spiegler (2016) provides a model of belief formation inspired by Bayesian networks (which are very different from the energy-based neural network that we use): beliefs are derived from some objective distribution q over a multi-dimensional state space Ω and "causal structure" (modelled as a directed acyclic graph). The agent rationally uses her causal structure, but is bounded in that she naively assumes the truth of her causal structure, which may well be wrong. Beliefs in this model combine q and the causal structure by means of the chain rule, and take the form of some product of conditional uni-dimensional marginals of q. As is clear from (1), the two models are orthogonal to each other: Intuitive Beliefs constitute a normalized product of uni-dimensional marginals and pointwise mutual information. The applications of the two models are also different.

The remainder of this paper proceeds as follows. Section 2 presents our model. Section 3 explores the empirical content of Intuitive Beliefs in the abstract while Section 4 explores it for Intuitive Beliefs trained by an environment. Section 5 presents applications. All proofs are collected in the Appendix.

2 Model

2.1 Primitives

For $1 < N < \infty$, the set $\Gamma = \{1, ..., N\}$ of *sources or elements of uncertainty* is a finite subset of $\mathbb{N}_+$ with generic elements $i, j, k, ...$ For each source of uncertainty $i \in \Gamma$, the finite abstract set $\Omega_i = \{x_i, y_i, z_i ...\}$ consists of all possible *elementary states* of source i, and is referred to as the *elementary state space* for source i. The (*full*) *state space* is the product space given by

$$\Omega := \prod_{i \in \Gamma} \Omega_i,$$

with generic element $x = (x_1, ..., x_N)$. An *event* in the full state space is, as usual, an element of some algebra on Ω:

$$\phi \neq \Sigma \subset 2^{\Omega},$$

with generic elements E, F, G. For any $I \subset \Gamma$, we adopt the notation $\Omega_I := \prod_{i \in I} \Omega_i$ and $\Omega_{-I} := \prod_{i \notin I} \Omega_i$. Similarly, a state $x_I z_{-I}$ specifies $x_I \in \Omega_I$ and $z_{-I} \in \Omega_{-I}$.

To illustrate the setup, consider the quarterly earnings announcements of a set of companies $\Gamma = \{1, .., N\}$. The possible earnings of company i are given by $\Omega_i = \{x_i, y_i, ..\}$. A state $x = (x_1, ..., x_N)$ is a vector of earnings announcements. The event that companies $I \subset \Gamma$ announce earnings $x_I \in \Omega_I$ is given by $E = \{x_I z_{-I} \in \Omega : z_{-I} \in \Omega_{-I}\}$.

Beliefs p over Ω are given by a standard probability measure over (Ω, Σ), and constitutes the primitive of our model. We treat beliefs as observable behavioral objects since they can be derived from betting preferences (Savage 1954) and indeed are routinely elicited in experiments (Schotter and Trevino 2014). It should be acknowledged that experiments routinely show that subjects' beliefs are non-probabilistic in that they violate additivity and do not even monotonically assign a higher probability to larger events (Tversky and Kahneman 1974). Nevertheless, we restrict attention to probability measures in this paper as part of a systematic development of our theory. An earlier version of this paper (Noor 2019) presented a non-probabilistic version of the model.

2.2 Abstract Intuitive Beliefs

An association is a psychological connection between observations. We distinguish between two kinds of associations: the connection between any pair of elementary states x_i and x_j, and that between an elementary state x_i and (unmodelled) background information.

Definition 1 *An associative network* (a, b) *on* Ω *is defined by*

(i) an association function that assigns to each distinct $i, j \in \Gamma$ *and* $x_i, x_j \in \cup_{k \in \Gamma} \Omega_k$ *a symmetric*[3] *associative weight* $a(x_i, x_j) \in \mathbb{R} \cup \{-\infty\}$, *which we write as* $a(x_i x_j)$,

(ii) a background association function b *that assigns* $b(x_i) \in \mathbb{R} \cup \{-\infty\}$ *to each* $x_i \in \cup_{k \in \Gamma} \Omega_k$.

Associations are pairwise and undirected in the model, and serve as the building blocks for beliefs over multi-dimensional states $x \in \Omega$ in the model. The fact that associations are undirected is without loss of generality, as directedness lacks empirical meaning in this static setting. Associations can take on positive or negative real-values, or a value of $-\infty$. The meaning of positive or negative associations will be explored in the sequel. The interpretation of $a(x_i x_j) = -\infty$ is that the occurrence of x_i is maximally associated with the non-occurrence of x_j, and vice versa (due to symmetry). A state of the world involving such $x_i x_j$ will be viewed as impossible by our agent.

Consider a state $x = (x_1, .., x_N) \in \Omega$. Imagine that the intuitive consideration of x triggers the subjective feeling $a(x_i x_j)$ of connection of between each pair x_i, x_j and the feeling $b(x_i)$ of

[3] That is, $a(x_i x_j) = a(x_j x_i)$.

connection between each x_i and background information. The total *associative energy* generated by these associations triggered by state $x = (x_1, .., x_N)$ is given by their additive aggregation $\sum_{i<j} a(x_i x_j) + \sum_{i \in \Gamma} b(x_i)$. We model the intuitive belief $p(x)$ as a normalized exponential function of this associative energy.

Definition 2 *Beliefs p over Ω are Intuitive Beliefs (IB) if there exists an associative network (a, b) and a real number $Z > 0$ such that for any $x \in \Omega$,*

$$p(x) = \frac{1}{Z} exp \left[\sum_{i<j} a(x_i x_j) + \sum_{i \in \Gamma} b(x_i) \right].$$

The set of all Intuitive Beliefs over Ω is denoted $\Delta_{IB}(\Omega)$.

The exponential function has no meaning beyond implying that $p(x)$ can be described alternatively as multiplicatively aggregating subjective building blocks: $p(x) = \frac{1}{Z} \prod_{i<j} \xi(x_i x_j) \times \prod_{i \in \Gamma} \zeta(x_i)$, where $\xi(x_i x_j) = exp[a(x_i x_j)]$ and $\zeta(x_i) = exp[b(x_i)]$. It is worth noting that statistical independence between dimensions i and j corresponds to a zero association $a(x_i x_j) = 0$ between any x_i and x_j. To illustrate, suppose we partition the dimensions Γ into two nonempty sets G_1, G_2 and suppose that there is zero association across the dimensions in these sets, that is, $a(x_i x_j) = 0$ for any $x \in \Omega$ and $i \in G_1$ and $j \in G_2$. Then it is straightforward to show that

$$p(x) = p(x_{G_1}) p(x_{G_2})$$

where $p(x_{G_k}) := p(x_{G_k} \Omega_{G_{-k}})$ denotes the marginal probability of the state taking values x_{G_k} on the dimensions G_k, for $k = 1, 2$.

The functional form is familiar from other literatures. When elementary states Ω_i consist of real numbers, the model subsumes the density of the multivariate Gaussian distribution:

$$f(x) = \frac{1}{\sqrt{2\pi^N |\Lambda|}} exp[-(x - \mu)^T \Lambda^{-1} (x - \mu)]$$

with variance-covariance matrix Λ. Indeed, just as the Gaussian distribution describes probabilities that are built from a variance-covariance matrix, Intuitive Beliefs are Gaussian in spirit in that they envisages likelihoods as arising from pairwise relationships between variables.[4]

The functional form is also related to the *Boltzmann machine* (Hinton and Sejnowski 1983, Ackley et al 1985), which belongs to the class of energy-based stochastic neural networks inspired by the Ising model of ferromagnetisim in Statistical Physics (Hopfield 1982). Imagine that Γ is a set of nodes, and each node $i \in \Gamma$ can take values in Ω_i (in typical AI applications, a node is either "on" or "off", captured by $\Omega_i = \{0, 1\}$ or $\Omega_i = \{-1, 1\}$). A state $x = (x_1, .., x_N) \in \Omega$ is therefore a configuration of the nodes in the network. Just as the respective spins of atoms interact to create energy, at each configuration $x \in \Omega$, each node $i \in \Gamma$ produces "energy" $b(x_i)$ and each pair of nodes produces "energy" $a(x_i x_j)$. The value $-\sum_{i<j} a(x_i x_j) + \sum_{i \in \Gamma} b(x_i)$ is known as the *energy* of the network (the negative quantity is a convention that comes from Physics). In energy-based networks, the configuration $x = (x_1, .., x_N)$ evolves dynamically because activation $x_i \in \Omega_i$ of each node $i \in \Gamma$ is presumed to depend on the total energy in the network. The steady state of the network is a steady state probability distribution over configurations. The Boltzmann machine posits that activation of any particular node is a logistic function of energy, and the resulting steady state distribution corresponds exactly to the Boltzmann-Gibbs distribution we used to define Intuitive Beliefs.

[4] To illustrate, when $N = 2$, the bi-variate Gaussian distribution has density $p(x_1 x_2) = \frac{1}{Z} exp \left[a(x_1 x_2) + b(x_1) + b(x_2) \right]$ where the associative network (a, b) is defined by

$$a(x_1 x_2) = \frac{2\rho}{2(1 - \rho^2)} \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \text{ and } b(x_i) = -\frac{1}{2(1 - \rho^2)} \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2,$$

where μ_i and σ_i are the mean and standard deviation of x_i, $i = 1, 2$, and $\rho \sigma_1 \sigma_2$ is the covariance between x_1 and x_2.

2.3 Trained Intuitive Beliefs

We have modelled a belief as a probability measure that is constructed from building blocks, given by an underlying network of associations. We now augment the model with the natural hypothesis that associations are shaped by the environment. This extends the abstract model to include the formation of Intuitive Beliefs.

Begin by recalling that, for any probability measures p, q over Ω, the notion of *KL-divergence* is defined by

$$KL(q||p) := \sum_{x\in\Omega} q(x) ln \frac{q(x)}{p(x)},$$

with the convention that $0 \times ln[\frac{0}{p(x)}] = 0$.[5] For any subset of states $\phi \neq D \subset \Omega$, define the D-conditional distributions in the usual way by $p(x|D) = \frac{p(x)}{\sum_{y\in D} p(y)}$ and $q(x|D) = \frac{q(x)}{\sum_{y\in D} q(y)}$ for all $x \in D$.

While the primitives of the abstract model was a belief p over Ω, we now consider instead the primitives consisting of (a) an *objective distribution* q over Ω, and (b) a subset of objectively-possible states $\phi \neq D \subset supp(q)$ that the agent has in fact observed. Using these primitives, we endogenously derive an Intuitive Belief p:

Definition 3 *Intuitive Beliefs p over Ω are trained by q over Ω if p solves*

$$min_{p\in\Delta_{IB}}\ KL(\ q(\textbf{.}|D)\ ||\ p(\textbf{.}|D)\ ).$$

That is, while the agent experiences only the conditional distribution $q(\textbf{.}|D)$ over a subset of states, $\phi \neq D \subset \Omega$, the intuitive process seeks an Intuitive Belief p over the *full* state space Ω that has a conditional $p(\textbf{.}|D)$ that matches $q(\textbf{.}|D)$ as closely as possible on D, in the sense of minimizing KL-divergence. Observe that training utilizes only the conditional distribution $q(\textbf{.}|D)$ over D and not q over Ω per se. When $D = \Omega$, training corresponds to a standard machine learning problem for the Boltzmann machine where q over Ω describes the training data.

2.4 Perspective

Limited Intelligence. A key distinction between intuitive reasoning and deliberative reasoning is that the former lacks the rigor and logical sophistication of the latter. This is amply demonstrated by anecdotal examples of common reactions to the Monty Hall problem but also, most notably, the central exploration of the Heuristics and Biases program in Psychology is precisely to document how the intuitive assessments of likelihood are flawed (Tversky and Kahneman 1974). In what sense does our model of Intuitive Beliefs exhibit limited intelligence?

Intelligence can be gauged by one's capacity to recognize or learn patterns in the data. We observe that Intuitive Beliefs always satisfy the following property: for any given $x \in \Omega$,

$$p(x_i x_j) > 0 \text{ for all } i, j \in \Gamma \implies p(x) > 0, \tag{2}$$

that is, if each $x_i x_j$ is deemed possible, then x must also be deemed possible. This arises from the additive nature of pairwise associations, where $a(x_i x_j) = -\infty$ or $b(x_i) = -\infty$ for some $i, j \in \Gamma$ forces $p(x) = 0$. While property (2) can be possessed by a prior in any rational model, what makes it peculiar here is that Intuitive Beliefs possess this property *regardless of the data* $q(\textbf{.}|D)$. That is, Intuitive Beliefs exhibit limited intelligence because they cannot learn a pattern in the data whereby

[5] The *Kullback-Leibler (KL) divergence* or *relative entropy* is used ubiquitously in information theory. It is well known that KL-divergence is non-negative and strictly convex. Although it fails the triangle inequality, it is pervasively used as a notion of distance between distributions. Our theory is not necessarily tied to KL-divergence. Our sharpest result on training (Theorem 3) can be established with any notion of distance $d \geq 0$ that satisfies $d(p, q) = 0 \iff p = q$.

each $x_i x_j$ is objectively possible but their joint realization $x = (x_1, .., x_N)$ is not. It is as if Intuitive Beliefs are based on a simplistic kind of intuitive reasoning, one that is incapable of comprehending full vectors $x = (x_1, .., x_N)$, and where observations of each pair $x_i x_j$ causes it to admit the possibility of their joint occurrence. We will see in Section 5 how this form of limited intelligence enables the model to explain empirical evidence.

Higher Intelligence. We show now how higher degrees of intelligence can be achieved in the model without departing from the pairwise nature of associations. The discussion also clarifies that Intuitive Beliefs correspond to a special case of the Boltzmann machine.

The nodes in the Intuitive Belief model correspond to what are known as *visible nodes* in the AI literature, that is, nodes whose values correspond to observable variables (elementary states). However, in addition to visible nodes, neural networks in AI pervasively and instrumentally make use of *hidden nodes*, that is, nodes whose values are completely internal to the network, in contrast to visible nodes which form the network's interface with the environment. To illustrate, suppose that $\Gamma = \{1, .., N\}$ describe the visible nodes as before, but now also suppose that there are hidden nodes $G = \{N+1, .., N+H\}$ and each hidden node $h \in G$ can take values in some space Ω_h. (One can interpret $\{\Omega_i\}_{i\in\Gamma}$ as the objective elementary state spaces and $\{\Omega_h\}_{h\in G}$ as subjective ones). In the Boltzmann machine, there exists an associative network (a, b) on $\Omega \times \prod_{h\in G} \Omega_h$, so that there are associations connecting all nodes, visible or hidden. It will aid our exposition to consider a "Restricted Boltzmann machine with a single layer of hidden nodes", which is a parsimonious special case where connections are 0 between all visible nodes, and similarly between all hidden nodes: $a(x_i x_j) = 0$ for all $x_i x_j$ where $i, j \in \Gamma$ or $i, j \in G$. While the probability of a state $(x_1, .., x_N; y_{N+1}, ..., y_{N+H})$ describing both visible and hidden nodes takes the Boltzmann-Gibbs form as before,

$$p(x_1, .., x_N; y_{N+1}, ..., y_{N+H}) = \frac{1}{Z} exp\left[\sum_{i\le N<h} a(x_i y_h) + \sum_{i\in\Gamma} b(x_i) + \sum_{h\in G} b(y_h)\right],$$

the Restricted Boltzmann machine is defined by the *marginal distribution over the visible nodes*:

$$p(x_1, .., x_N) = \frac{1}{Z} \sum_{(y_{N+1},..,y_{N+H})\in\prod_{h\in G}\Omega_h} exp\left[\sum_{i\le N<h} a(x_i y_h) + \sum_{i\in\Gamma} b(x_i) + \sum_{h\in G} b(y_h)\right].$$

Given the flexibility afforded by hidden nodes, this network can be trained to learn a wide range of patterns in the data, and subsequently recognize patterns when given new but partial data and accomplish a range of tasks.

3 Results: Abstract Intuitive Beliefs

In this section we explore the empirical meaning of associations in the set up where we observe the agent's beliefs p only, abstracting from the environment. We make the simplifying assumption that beliefs have *full support*: $p(x) > 0$ for all $x \in \Omega$. It is easy to see that Intuitive Beliefs p have full support if and only if p is represented by a real-valued network (a, b).

3.1 Identification

Take any probability measure $p \in \Delta(\Omega)$ (not necessarily an Intuitive Belief) with full support and let K_I be the cardinality of Ω_I for any subset of dimensions $\phi \neq I \subset \Gamma$. Since Ω has a product structure, $K_I = \prod_{i\in I} K_i$. We begin with the simple observation that the marginal belief

$p(x_I) = \sum_{z_{-I} \in \Omega_{-I}} p(x_I z_{-I})$ on $x_I \in \Omega_I$ can in fact be viewed as a *normalized mean* of $p(x_I z_{-I})$ averaged over all $z_{-I} \in \Omega_{-I}$:

$$p(x_I) = \frac{1}{Z_I} \sum_{z_{-I} \in \Omega_{-I}} \frac{1}{K_{-I}} p(x_I z_{-I})$$

with a normalizing constant Z_I (which is easily shown to equal $\frac{1}{K_{-I}}$). It is trivial to show that this can equivalently be written as $p(x_I) = \frac{1}{Z_I} \sum_{z \in \Omega} \frac{1}{K} p(x_I z_{-I})$ where we sum across all $z \in \Omega$ instead, and divide by the cardinality K of Ω. Define a *geometric marginal* (*geo-marginal* for short) similarly by the normalized geometric mean of $p(x_I z_{-I})$ over all $z_{-I} \in \Omega_{-I}$:

$$p^g(x_I) := \frac{1}{Z_I} \prod_{z \in \Omega} p(x_I z_{-I})^{\frac{1}{K_{-I}}} \qquad x_I \in \Omega_I.$$

We show in Appendix A that geo-marginals are attractive mathematical objects since their properties mirror those of arithmetic marginals. We also show that marginals and geo-marginals coincide under statistical independence (that is, $p(x) = \prod_{i \in \Gamma} p(x_i)$ for all $x \in \Omega$).

Next, recall the measure of correlation $\varphi(x_i, x_j)$ between two variables known as *pointwise mutual information* (PMI) familiar from information theory:

$$\varphi(x_i, x_j) = ln \frac{p(x_i x_j)}{p(x_i) p(x_j)}.$$

If p exhibits statistical independence then $\frac{p(x_i x_j)}{p(x_i)p(x_j)} = 1$ and there is no correlation, $\varphi(x_i, x_j) = 0$. Positive and negative correlations are defined accordingly. For any $x_i, x_j \in \cup_{k \in \Gamma} \Omega_k$, define *geometric PMI* (*geo-PMI* for short) by

$$a_g(x_i x_j) = ln \frac{p^g(x_i x_j)}{p^g(x_i) p^g(x_j)},$$

where marginals have been replaced with geo-marginals in the definition of PMI. Since marginals and geo-marginals coincide under statistical independence, if p exhibits statistical independence then $a_g(x_i x_j) = 0$, as in the arithmetic case.

We are now ready to state our first result. In the context of Intuitive Beliefs, the notion of an associative network can be given empirical expression in terms of geo-marginals and geo-PMI:

Theorem 1 *A belief p with full support is an Intuitive Belief if and only if it is represented by an associative network (a_g, b_g) over Ω where, for all $x_i, x_j \in \cup_{k \in \Gamma} \Omega_k$,*

$$a_g(x_i x_j) = ln \frac{p^g(x_i x_j)}{p^g(x_i) p^g(x_j)} \text{ and } b_g(x_i) = ln p^g(x_i).$$

A proof outline is provided in Section 3.4. The Theorem tells us that the notion of association can be understood as a notion of correlation, and background association as a notion of marginal. Treating these as behavioral definitions, we can give meaning to a positive (resp. negative) association between x_i and x_j as $\frac{p^g(x_i x_j)}{p^g(x_i)p^g(x_j)} > 1$ (resp. < 1). More generally, the Theorem tells us that the model lends itself to the same kind of analysis as in standard models in Economics, where components of the representation can be measured empirically and be used to compare across individuals (such as risk aversion in Expected Utility theory). For instance, it can be used to compare if one agent has a stronger association $a_g(x_i x_j)$ than another. It can also be used to estimate (a_g, b_g) for an agent on the basis of data p on a limited sample $D \subset \Omega$, allowing us to predict the agent's beliefs about states that are out-of-sample (just as risk aversion estimated in the lab can be taken as an estimate for the population and can be used for to suggest parameter values in macro-finance models).

It is instructive to note that inserting this network into the representation yields a reduced form of the model: for all $x \in \Omega$,

$$p(x) = \frac{1}{Z} \times \left[\prod_{i<j} \frac{p^g(x_i x_j)}{p^g(x_i) p^g(x_j)} \right] \times \prod_{i \in \Gamma} p^g(x_i). \tag{3}$$

Thus, Intuitive Beliefs are constructed from its uni- and bi-dimensional geo-marginals, and more specifically, they combine uni-dimensional marginals with correlation (geo-PMI).

3.2 Characterization

Associations are the building blocks of Intuitive Beliefs. In this section we explore the observable expression of the separability between these building blocks that is presumed in the model. Although updating of beliefs is not the focus of this paper, our characterization result lends itself to better interpretation in a dynamic setup where the prior belief is an Intuitive Belief and subsequent posteriors are its Bayesian conditionals. A purely static characterization can be found in Appendix E.2.

For any full support belief p over Ω, a family of conditional beliefs $\{p(.|E)\}_{E\in\Sigma}$ over Ω is *Bayesian* if, for each event $E \in \Sigma$, the conditional belief satisfies $p(x|E) = \frac{p(x)}{\sum_{y\in E} p(x)}$ for all $x \in E$. Denoting the cardinality of any event $E \in \Sigma$ by $K(E)$ and taking any set of dimensions $\phi \neq I \subset \Gamma$, the conditional geo-marginal $p^g(x_I|E)$ is naturally defined by taking a normalized geometric mean of $p(x_I z_{-I}|E)$ over all z_{-I} that appear in E:

$$p^g(x_I|E) := \frac{1}{Z_I(E)} \prod_{z\in E} p(x_I z_{-I}|E)^{\frac{1}{K(E)}} \qquad x_I \in \Omega_I.$$

This in turn allows us to define conditional geo-PMIs $\frac{p^g(x_i x_j|E)}{p^g(x_i|E)p^g(x_j|E)}$. It is readily determined that a Bayesian family of Intuitive Beliefs can be described by conditional versions of (3) using conditional geo-PMI.[6]

When some states are ruled out by an event E, the prior association $\frac{p^g(x_i x_j)}{p^g(x_i)p^g(x_j)}$ between x_i and x_j is revised to a new value $\frac{p^g(x_i x_j|E)}{p^g(x_i|E)p^g(x_j|E)}$. We identify a restriction that the model implies for this revision. Consider more specifically any event $E \in \Sigma$ that is i,j-*unrestricted* in that it only restricts elementary states outside dimensions i,j: there exists a restriction $\phi \neq S_k \subset \Omega_k$ for all $k \neq i,j$ such that

$$E = \Omega_{ij} \times \prod_{i,j\neq k\in\Gamma} S_k. \tag{4}$$

The following property permits such an event to cause the association $\frac{p^g(x_i x_j)}{p^g(x_i)p^g(x_j)}$ to move to a different value $\frac{p^g(x_i x_j|E)}{p^g(x_i|E)p^g(x_j|E)}$, but requires that the *relative* associations on dimension i,j do not change.

---

[6]For completeness, we state this formally, relegating the proof to the appendix:

Proposition 1 *If $\{p(.|E)\}_{E\in\Sigma}$ over Ω is a Bayesian family of conditional beliefs derived from a full support Intuitive Belief p then for each $E \in \Sigma$ and $x \in E$,*

$$p(x|E) = \frac{1}{Z(E)} \times \left[\prod_{i<j} \frac{p^g(x_i x_j|E)}{p^g(x_i|E) p^g(x_j|E)} \right] \times \prod_{i\in\Gamma} p^g(x_i|E).$$

Definition 4 *The Bayesian family of conditional beliefs* $\{p(.|E)\}_{E\in\Sigma}$ *over* Ω *derived from a full support prior* p *satisfies* *Relative Associative Separability* *(RAS) if for any distinct* $i,j\in\Gamma$ *and any* $(x_i,x_j),(y_i,y_j)\in\Omega_{ij}$*, the ratio*

$$\frac{p^g(x_ix_j|E)}{p^g(x_i|E)p^g(x_j|E)}/\frac{p^g(y_iy_j|E)}{p^g(y_i|E)p^g(y_j|E)}$$

is the same across all i,j*-unrestricted events* $E\in\Sigma$*.*

As a statement of how the realization of elementary states outside i,j impact beliefs along dimensions i,j, RAS is a separability property: in determining the belief $p(x)$, for any distinct i,j and k, the x_i,x_j association is aggregated with the x_i,x_k associations and the x_jx_k in a separable way. We find that this property characterizes Intuitive Beliefs.

Theorem 2 *A Bayesian family of conditional beliefs* $\{p(.|E)\}_{E\in\Sigma}$ *over* Ω *derived from a full support prior* p *satisfies RAS if and only if* p *is an Intuitive Belief.*

A proof outline is provided below in Section 3.4. Note that when $N=2$, any probability measure satisfies RAS trivially and thus, by the Theorem, can be written as an Intuitive Belief. A direct demonstration is that if $N=2$ then the model takes the form $p(x_ix_j)=\frac{1}{Z}exp\,[a(x_ix_j)+b(x_i)+b(x_j)]$, and any p can be replicated by an appropriately chosen a while setting $b=0$. Intuitive Beliefs have peculiar content only when $N>2$.

3.3 Uniqueness Theorem

Having characterized the class of beliefs that admit an Intuitive Belief representation, we now characterize the class of all representations of a given Intuitive Belief in the following uniqueness theorem. (We are not aware of similar results in the AI literature, where the network (a,b) and p that define a Boltzmann machine are both treated as observable). Say that $\overline{z}\in\Omega$ is a *reference state* if $p(\overline{z})>0$ and for any $x_i\in\Omega_i$,

$$p(x_i)>0\implies p(x_i\overline{z}_i)>0,$$

that is, $\overline{z}$ occurs with positive probability and if we replace $\overline{z}_i$ with any x_i that has a positive marginal then the new state $x_i\overline{z}_i$ also has positive probability. The existence of a reference state implies some richness in the support of p. For instance, a belief p with binary support $\{x,y\}$ where x,y are non-overlapping (in the sense that $x_i\neq y_i$ for each i) does not admit any reference state. When there exist overlapping states, then the representation has sharp uniqueness properties.

Theorem 3 *(i) Intuitive Beliefs* p *are represented by* (a,b) *if and only if there they are represented by* (α,β) *where* $\beta=0$ *and, for all* $x\in\Omega$ *and distinct* $i,j\in\Gamma$*,*

$$\alpha(x_ix_j)=a(x_ix_j)+\frac{1}{N-1}[b(x_i)+b(x_j)].$$

(ii) Suppose there exists a reference state. Then Intuitive Beliefs p *are represented by* $(a,0)$ *and* $(\alpha,0)$ *iff for each* $i,j\in\Gamma$ *and* $x_i\in\Omega_i$ *there exist scalars* $\gamma_j(x_i)$ *and* k_i *satisfying* $\sum_{i\neq j\in\Gamma}\gamma_j(x_i)=k_i$*, such that for any* $x\in\Omega$ *and distinct* $i,j\in\Gamma$*,*

$$\alpha(x_ix_j)=a(x_ix_j)+\gamma_j(x_i)+\gamma_i(x_j). \tag{5}$$

The first claim shows that background associations can in fact be normalized to $b=0$.[7] The second shows how any two such b-normalized representations $(a,0)$ and $(\alpha,0)$ must be related (assuming the existence of a reference state, which holds trivially if we were to assume that beliefs

[7]Therefore, strictly speaking, background associations are irrelevant for the representation. They remain useful, however. The reduced form (3) is easier to interpret with the notion of background associations, and the canonical representation in Theorem 1 is harder to interpret if a is adjusted so as to absorb b, in which case it is no longer defined by geo-PMI. Furthermore, while not explored here, background associations can be employed in a dynamic extension of the model to model Non-Bayesian updating, by making them a function $b(.|E)$ of the event that an agent learns (see the earlier version of this paper, Noor 2019).

have full support). The relationship is described by a "budget" k_i for every dimension $i \in \Gamma$, from which one can allocate a portion $\gamma_j(x_i)$ to each j distinct from i, in a manner that balances the budget: $\sum_{i \neq j \in \Gamma} \gamma_j(x_i) = k_i$. For any representation $(a, 0)$, shifting $a(x_i x_j)$ by $\gamma_j(x_i) + \gamma_i(x_j)$ as in (5) generates a new representation $(\alpha, 0)$. Conversely, any two representations must be related by such budget-balanced shifts.

The non-uniqueness of the representation gives rise to the foundational challenge of finding a unique canonical representation. This is in fact the contribution of Theorem 1. It is interesting to note that the uniqueness issue changes when Intuitive Beliefs are determined endogeneously through the training problem (Definition 3). As we will see in Section 4, Trained Intuitive Beliefs are a function of q, and that function serves as a unique canonical representation in terms of q.

3.4 Proof Outline for Theorems 1 and 2

The proof of Theorem 1 is based on studying various normalizations of the model that are permitted by Theorem 3. Assume that p has full support so that every state is a reference state. Of particular interest are "z-normalized representations" (Lemma 9) where we fix any reference state $z \in \Omega$ and normalize the representation so that $a(z_i z_j) = b(z_i) = 0$ for all distinct $i, j \in \Gamma$ and $a(x_i z_j) = 0$ for all x_i. A key result is that such representations can be identified using p by the relationships $a(x_i x_j) = \frac{p(x_i x_j z_{-ij}) p(z)}{p(x_i z_{-i}) p(x_j z_{-j})}$ and $b(x_i) = \frac{p(x_i z_{-i})}{p(z)}$. These normalized representations are not adequate canonical representations because the choice of the reference state $z \in \Omega$ is arbitrary. The construction of the canonical representation comes from observing that the ratio $\frac{p(x_i x_j z_{-ij}) p(z)}{p(x_i z_{-i}) p(x_j z_{-j})}$ is *independent of* z_{-ij} in the model. Thus, the geometric mean of any set of z-normalized representations is still a representation for p. The canonical representation is obtained from the geometric mean of all the z-normalized representations, where z is varied over all of Ω. While each normalized representation yields the same p, the canonical representation stands out as readily interpretable, and derives value from the fact that geo-marginals are attractive mathematical objects.

The proof of sufficiency of Theorem 2 uses RAS to build an expression for $p(x)$ for any given $x \in \Omega$ by starting with $E = \Omega_i \times \Omega_j \times \{z_{-ij}\}$ for some fixed $z \in \Omega$ and inductively changing z_{-ij} one dimension at a time $(z_{-ij}, x_k z_{-ijk}, x_k x_l z_{-ijkl}, ...)$ until we reach x_{-ij}, invoking RAS each step of the way. We obtain $p(x)$ in its reduced form (3), and that in turn implies that p has an Intuitive Belief representation. The proof of necessity of RAS is based on the observation above that the ratio $\frac{p(x_i x_j z_{-ij}) p(z)}{p(x_i z_{-i}) p(x_j z_{-j})}$ is independent of z_{-ij}. For any i, j-unrestricted event E, take a geometric mean of $\frac{p(x_i x_j z_{-ij}) p(z)}{p(x_i z_{-i}) p(x_j z_{-j})}$ across all $z \in E$ and invoke the noted independence property to conclude that ratios of such terms (where $p(z)$ drops out) do not depend on E. This yields RAS.

4 Results: Trained Intuitive Beliefs

In this section we explore the training problem (Definition 3) to establish properties of Trained Intuitive Beliefs.

4.1 Characterization and Inductive Inference

Say that a pair of elementary states $x_i x_j$ *appear* in D if there exists a state of the form $x_i x_j z_{-ij} \in D$. While training of beliefs p determines the agent's associations between $x_i x_j$ that appear in D, it does not restrict associations between pairs of elementary states that do not appear in D. To remove this source of non-uniqueness, our uniqueness result will focus on solutions p to the training problem that are *minimal* in that they satisfy $p(x) = 0$ for any $x \in \Omega$ that is not entirely composed of $x_i x_j$ that appear in D, that is, $p(x) = 0$ if there is $i, j \in \Gamma$ s.t. $x_i x_j$ does not appear in D. Thus, the agent's intuitive process is viewed as creating connections only for those $x_i x_j$ for there is evidence: it treats as impossible any pairs of elementary states that have never been observed, and consequently any state x that contains them. In the context of a conscious deliberating agent, minimality might correspond to a sort of prudence. But in our context of non-conscious intuitive processes, it could be interpreted as an inability to imagine possibilities too far beyond the realm of experience.

Our results will utilize some richness conditions, the first of which is:

Assumption 1 *There exists $\overline{z} \in D$ such that $x \in D \implies x_i\overline{z}_{-i} \in D$ for all $i \in \Gamma$.*

That is, D contains some state $\overline{z}$ and its one-dimensional deviations $x_i\overline{z}_{-i}$ defined by each $x \in D$ and $i \in \Gamma$. In the extended environment of Section 4.3, this state has a concrete interpretation as "no news", and $x_i\overline{z}_{-i}$ has the interpretation of "only x_i is observed", in which case assumption 1 states that for any state $x \in D$ that may be revealed to the agent and for each $i \in \Gamma$, there is a state $x_i\overline{z}_{-i} \in D$ that reveals the i^{th} dimension alone.

We now state our main result.

Theorem 4 *For any objective distribution q over Ω and any $\phi \neq D \subset supp(q)$ which satisfies assumption 1, the following hold:*

(i) (Characterization) An Intuitive Belief p is trained by q if and only if for all distinct $i, j \in \Gamma$,

$$p(x_ix_j|D) = q(x_ix_j|D), \qquad x_i, x_j \in \cup_{k\in\Gamma}\Omega_k.$$

(ii) (Inductive Inference) For any solution p,

$$D \subset \Omega_D \subset supp(p),$$

where

$$\Omega_D := \{x \in \Omega : \forall i, j \in \Gamma \text{ there exists } z \text{ s.t. } x_ix_jz_{-ij} \in D\}.$$

(iii) (Minimal Support) If p is a minimal solution, then

$$supp(p) = \Omega_D.$$

The first claim is that a solution p to the training problem exists if and only if it correctly matches q in terms of the two-dimensional marginals conditional on the observed states D. Indeed, it is very convenient that although p and q are defined over multi-dimensional vectors $x \in \Omega$, they are closest only when their conditional two-dimensional marginals match. Theorem 4(i) is an extension to our setting of a central result for Boltzmann machines in the AI literature (see Hinton and Sejnowski 1983, Ackley et al 1985) where training is defined for $D = \Omega$.[8]

The second claim in the Theorem states that the support of any solution p is weakly larger than D. It is natural that the support of p must include observed states D (indeed, KL divergence takes an infinite value if $q(x) > 0$ and $p(x) = 0$). However, the Theorem further claims that the support of p extends at least to the set Ω_D that consists of all states $x \in \Omega$ which are composed entirely of elementary states x_ix_j that appear in D. The reason was expressed in the property (2) of Intuitive Beliefs observed earlier where, for any $x \in \Omega$, if each x_ix_j is deemed possible, then so is x. In particular, if the observations in D establish the possibility of each x_ix_j, then Intuitive Beliefs entertain $p(x) > 0$ despite never observing x. The third claim in the Theorem establishes that the support of any minimal solution p is exactly Ω_D.

[8]In the machine learning problem applied to the Boltzmann machine (corresponding to taking $D = \Omega$ in our definition of training), the noted result in AI states that p is a minimizer if and only if $p(x_ix_j) = q(x_ix_j)$ for all $x_i, x_j \in \cup_{k\in\Gamma}\Omega_k$. This is the basis of a standard training algorithm ("gradient descent") which, roughly speaking, follows these steps: starting with any weights in the network, the i^{th} iteration of the algorithm adjusts each weight in the network in accordance with the distance between the two-dimensional marginals (more precisely, the partial derivative of KL-divergence wrt the weight), and the algorithm terminates when the marginals are within a pre-specified threshold.

The fact that the agent may place strictly positive probability $p(x) > 0$ on a state x outside D is a key property and we refer to it as *intuitive inductive inference.* To illustrate, if based on her observations D, the agent has formed an association between "animal" and "alibinism" and between "animal" and "swans" then the agent intuitively comes to believe in the existence of an albino swan even if she has never seen one: if $D = \{(animal, dog, alibino), (animal, swan, \neg alibino)\}$ then she assigns strictly positive probability to the state $(animal, swan, albino)$ even though this state does not lie in her observations D. Some of our applications in the sequel exploit the inductive inference property.

It is worth recalling that case-based beliefs (Gilboa and Schmeidler 2003, Billot et al 2005) also exhibit inductive inference, though it is formulated differently: if the agent has seen a case $(dog, albino)$, then the agent may believe in $(swan, albino)$ because there exists a subjective "similarity function" relating dogs and swans. In the case-based model, the similarity function does not specify the source of the similarity. In contrast, in our model, there is an explicit chain of states connecting swan with albino (namely, one in which $swan, animal$ appears and another in which $animal, albino$ appears) using the agent's observations. Also, the notions of similarity and associations are not identical. Associations are a general notion of psychological connection (which are influenced by similarity, frequency of co-occurrence, repetition, salience, etc) but our particular model of belief formation uses frequency of co-occurrence (under $q(\cdot|D)$) to determine the strength of association. It may be that dog and swan are considered sufficiently dissimilar that the case-based agent assigns a low probability to seeing an albino swan, but depending on how often our agent has seen albinos she may assign a high probability of finding an albino swan.

The proof of Theorem 4 proceeds as follows. While it is known that KL-divergence is a strictly convex function in its second argument, the space of Intuitive Beliefs is not generally convex, and so the first order conditions are generally not sufficient for a solution. We instead identify a convex Euclidean subspace A that serves as a space of normalized representations for Intuitive Beliefs that respect the minimality property.[9] For any $a \in A$ and Intuitive Belief p_a that it represents, we say that the "distance" between q and a is $g_D(a) := KL(q(\cdot|D)||p_a(\cdot|D))$. We prove that this function is convex in a, and so, the first order conditions are sufficient. These first order conditions imply the characterization result in Theorem 4. For the uniqueness result (Proposition 2) below, we show that, under assumption 2, $g_D(a)$ is strictly convex in a, and therefore the minimizer $a \in A$ of $g(a)$, if it exists, must be unique.

4.2 Uniqueness and Existence

If we strengthen assumption 1 to include two-dimensional deviations $x_i x_j \overline{z}_{-ij}$ then we obtain:

Assumption 2 *There exists $\overline{z} \in D$ such that $x \in D \implies x_i \overline{z}_{-i} \in D$ and $x_i x_j \overline{z}_{-ij} \in D$ for all $i, j \in \Gamma$.*

Under this strengthening, we obtain:

Proposition 2 *(Uniqueness) Under assumption 2, the training problem admits at most one minimal solution.*

Without imposing minimality, there may be many solutions that assign arbitrary associations to $x_i x_j$ that do not appear in D. Proposition 2 guarantees that, under the noted richness condition on D, there is no more than one minimal solution. While uniqueness is a convenient property for applications, we note that a multiplicity of solutions is also interesting in that it provides a basis for different agents forming different intuitions despite identical experiences.

[9] Assumption 1 yields the "reference state" used in Theorem 3, thereby permitting us to exploit our results on normalized representations in Appendix C.1.

The preceding results notwithstanding, the existence of a solution to the training problem is generally not ensured due to the limited number of parameters in our model.[10] There are at least two routes one can take to ensure existence. One route, which we leave to future research, is to extend Intuitive Beliefs to includes hidden nodes as in Section 2.4. Indeed, in AI, hidden nodes provide free parameters that help guarantee the existence in the problem of training the Boltzmann machine. The second route, which we illustrate next, is to find restrictions on D that are sufficient for the existence of a solution.

Consider the following richness condition – stronger than assumptions 1 and 2 – where D contains a state $\overline{z}$ and some set of its one- and two-dimensional deviations only.

Assumption 3 *D is not a singleton and there exists $\overline{z} \in D$ such that each $x \in D$ takes the form $x_i\overline{z}_{-i}$ or $x_ix_j\overline{z}_{-ij}$ for some $i, j \in \Gamma$ and $(x_i, x_j) \in \Omega_i \times \Omega_j$.*

Under this restriction on D we can show that:

Proposition 3 *(Existence) For any objective distribution q over Ω and any $\phi \neq D \subset supp(q)$ which satisfies assumption 3, there exists a unique minimal solution p. This solution p satisfies*

$$p(\centerdot|D) = q(\centerdot|D),$$

and has a reduced form s.t. for each $x \in supp(p) = \Omega_D$,

$$p(x) = \frac{1}{Z} \times \left[\prod_{i<j} \frac{q(x_ix_j\overline{z}_{-ij})}{q(x_i\overline{z}_{-i})q(x_j\overline{z}_{-j})}\right] \times \left[\prod_{i\in\Gamma} q(x_i\overline{z}_{-i})\right].$$

Thus, assumption 3 is a sufficient condition for existence and uniqueness (among minimal solutions). In this solution, the agent's belief p over Ω is such that it exactly matches q on the subdomain D. Moreover, we can provide a reduced form of the model in terms of q, which reveals that the trained network satisfies

$$a(x_ix_j) = ln\frac{q(x_ix_j\overline{z}_{-ij})}{q(x_i\overline{z}_{-i})q(x_j\overline{z}_{-j})} \text{ and } b(x_i) = lnq(x_i\overline{z}_{-i}),$$

for any (x_i, x_j) such that $(x_i\overline{z}_{-i}), (x_j\overline{z}_{-j}) \in D$, and take a value $-\infty$ otherwise. It is instructive to compare with Section 3.1 where our primitive consisted of a given Intuitive Belief p and we obtained a means of identifying a network in terms of p. Here our primitive is defined by the agent's environment – a distribution q and a set of observed states D – and the theory implies the formation of an Intuitive Belief defined by a network that is identified in terms of q and D.

4.3 Application: Extended Environment

The notion of training requires the intuitive process to match the conditional objective likelihood of states $x \in D$. However, given that cognition is bounded, it is plausible to hypothesize that the intuitive process is unable to assimilate all dimensions of an observed state simultaneously. Rather it assimilates only a few of its dimensions at a time. That is, instead of thinking of D as a constraint on objective data, let us think of it as arising from some cognitive constraint on the agent.

To model this, extend the environment so that not only is there a state $x \in \Omega$ that appears in the environment, but there is a (subjective) "news source" that reports only the elementary states on a

[10] For instance, when $D = \Omega$, the first order conditions define a system of nonlinear equations that is generically over-identified: there are $\prod_{i<j}[K_iK_j] + \frac{N(N-1)}{2}$ equations (generated by the first order conditions defined for each distinct $x_i, x_j \in \cup_{k\in\Gamma}\Omega_k$, and the fact that marginals along any i, j must sum to 1). We can normalize $b = 0$ (by the uniqueness result proved in Theorem 3) and so p is determined by a network $(a, 0)$ that has $\sum_{i<j}[K_iK_j]$ parameters.

random subset of dimensions. For any dimension i, use z_i^* to denote "no news about dimension i". For any $I \subset \Gamma$, let $x_I z_{-I}^*$ denote "elementary states on I are revealed to be x_I but there is no news on the remaining dimensions". Use $z^* = (z_1^*, .., z_N^*)$ to denote no news in all dimensions. Therefore, we are in an extended environment where the elementary news space is $\Omega_i^* = \Omega_i \cup \{z_i^*\}$, and the full news space is:

$$\Omega^* = \prod_{i \in \Gamma} \Omega_i^*.$$

Note that $\Omega \subset \Omega^*$.

Fix a probability distribution q over states Ω as before. The extended probability distribution q^* over news Ω^* is derived as follows: The probability of receiving news on any dimension i is $\sigma \in (0,1)$ and this is independent of the state. Then the probability of receiving news $(x_I z_{-I}^*) \in \Omega^*$ is

$$q^*(x_I z_{-I}^*) = \sigma^{\#I}(1-\sigma)^{N-\#I} q(x_I),$$

where $\#I$ is the cardinality of $I \subset \Gamma$ and where we adopt the convention that $q(x_\phi) = 1$.

Next we specify the subset $D^* \subset \Omega^*$ of news the agent in fact receives. We assume that D^* consists only of news of the form $z^*, (x_i z_{-i}^*)$ and $(x_i x_j z_{-ij}^*)$ where up to two dimensions are revealed at a time:

$$D^* := \{(x_i x_j z_{-ij}^*) \in \Omega^* : i, j \in \Gamma,\ x_k \in \Omega_k^* \text{ for } k = i, j, \text{ and } q(x_i x_j) > 0\}.$$

This corresponds to assumption 3. While this assumption will generate a parsimonious model that will be exploited below in applications (Section 5), it is interesting to note that the idea of attending to up to two dimensions at a time is somewhat reminiscent of findings in experiments using eye tracking data that show that subjects primarily engage in binary comparisons in a multi-alternative choice context (Russo and Rosen 1975, Krajbich and Rangel 2011).

In this setting we obtain the following corollary of Proposition 3.

Proposition 4 *In the extended environment, assuming that D is not a singleton, the following holds:*

(i) There exists a unique minimal Intuitive Belief p over Ω^* that is trained by q^*.

(ii) The trained Intuitive Belief p satisfies $p(.|D) = q(.|D)$.

(iii) The reduced form of a minimal trained Intuitive Belief conditional on $\Omega \subset \Omega^*$ is given by: for any $x \in supp(p) = \Omega_{D^*} \cap \Omega$,

$$p(x|\Omega) = \frac{1}{Z}\left[\prod_{i<j} \frac{q(x_i x_j)}{q(x_i)q(x_j)}\right] \times \prod_{i \in \Gamma} q(x_i).$$

Relative to Proposition 3, the novelty is in the third claim. While the news source was constructed to operationalize the notion of "training by low-dimensional observations", the key interest in applications is understanding the structure of p conditioned on the objective state space $\Omega \subset \Omega^*$. By the inductive inference property (Theorem 4), this conditional $p(.|\Omega)$ has support $\Omega_{D^*} \cap \Omega$. The Proposition characterizes this belief. It is as if the trained associative network *imbibes objective PMI* $\frac{q(x_i x_j)}{q(x_i)q(x_j)}$ *and objective marginals* $q(x_i)$. This yields a sharp representation of beliefs in terms of q with the added benefit that it eschews the need to specify D in applications.

5 Applications

We now present applications of Trained Intuitive Beliefs. The first two applications (correlation neglect and overestimation of small probabilities) arise from the inductive inference property and the second two (experience effects and classic belief biases) are not related to beliefs about unobserved states.

5.1 Inductive Inference and Correlation Neglect

Recent literature has documented that people behave as if they do not understand what is objectively ruled out by the correlations embodied in assets, signals, etc. See Eyster and Weizsäcker (2016) and Enke and Zimmerman (2013) for experimental evidence as well as Brunnermeier (2009) for research suggesting that the 2008 financial crisis can be attributed to such "correlation neglect". We show that correlation neglect arises naturally as a property of trained Intuitive Beliefs: it is an expression of inductive inference.

To illustrate, suppose there are three states $S = \{s', s'', s'''\}$ and three assets a_1, a_2, a_3 that can each have either a high h or low l payoff as defined in the following matrix.

$asset\backslash state$	s'	s''	s'''
a_1	h	h	l
a_2	h	l	h
a_3	l	h	h

Denote the probability distribution over S by π. Then we obtain an induced distribution q on outcome profiles $\Omega = \{h, l\}^3$ (given by the columns of the matrix), defined by

$$q(x_1, x_2, x_3) = \pi(\{s \in S : a_i(s) = x_i \text{ for } i = 1, 2, 3\}).$$

For instance, the outcome profile (h, h, l) has probability $q(h, h, l) = \pi(s_1)$. Clearly, the support of q is $supp(q) = \{(h_1, h_2, l_3), (h_1, l_2, h_3), (l_1, h_2, h_3)\}$.

Correlation neglect occurs when $supp(q) \subset supp(p)$, that is, the agent behaves as if she admits as possible something that is impossible given the correlation structure (see Eyster and Weizsäcker 2016 and Ellis and Piccione 2017 for models with this feature). To see that this arises in trained Intuitive Beliefs, consider the model characterized in Proposition 3. Taking $D = supp(q)$, each $x \in D$ serves as a state $\overline{z}$ that satisfies assumption 3. Therefore, by Proposition 3, $supp(q) = D \subset supp(p)$, and in particular the support of the unique minimal trained p is

$$supp(q) \cup \{(h_1, h_2, h_3), (h_1, l_2, l_3), (l_1, l_2, h_3), (l_1, h_2, l_3)\}.$$

Thus, $supp(q) \subset supp(p)$, as was to be shown. The reason that the agent perceives, say, (h_1, h_2, h_3) as possible is that the agent sees that assets 1 and 2 can have the pairwise realization (h_1, h_2) and assets 2 and 3 can similarly yield (h_2, h_3). Due to the intuitive aggregation of pairwise associations, her intuition connects these and admits (h_1, h_2, h_3) as a possibility.

5.2 Inductive Inference and Small Probabilities

The psychology literature studying overconfidence finds that subjects tend to be overconfident about their performance on difficult tasks and underconfident about their performance on easy tasks (Moore and Healy 2008). A similar pattern in beliefs is suggested by the Disposition Effect, where investors sell winning stock too soon and hold on to losing stock for too long (see Shefrin and Statman, 1985, and subsequent literature). Benjamin (2019) reviews evidence that beliefs tend to have thicker tails than the objective distribution.

Such findings can be explained by the hypothesis that subjects overestimate small probabilities. We show that an overestimation of low probability states can arise as a consequence of the inductive inference property in our model. As a property of the support of p, inductive inference implies that impossible states can be deemed possible. By a simple continuity argument, inductive inference can also lead the agent to overestimate the probability of a very unlikely state.

To illustrate, suppose there are $N > 0$ companies and that Ω_i denotes the possible earnings of company i. For each company $i \in \Gamma$ there is an elementary state $l_i \in \Omega_i$ representing a low earning. Let $l = (l_1, ..., l_N) \in \Omega$ denote the "bad state" where all companies perform badly. Let $E = \Omega \backslash \{l\}$

denote all other states. Start with any objective distribution q that has full support, and for each $\alpha \in (0,1]$ consider the objective distribution q_α defined by

$$q_\alpha(x) = \alpha q(x) + (1-\alpha) q(x|E),$$

where $q(x|E)$ is the conditional objective probability of x given E. Thus, $q_\alpha(x)$ has full support for all $\alpha > 0$, and as $\alpha \to 0$ it must be that the objective probability $q_\alpha(l)$ of the bad state $l = (l_1, ..., l_N)$ goes to 0. When $\alpha \to 0$ let us say that the market is improving.

Suppose p_α is an Intuitive Belief trained by q_α in the sense of Proposition 4. While the objective probability of the bad state satisfies $lim_{\alpha \to 0} q_\alpha(l) = 0$, we verify in Appendix I that

$$lim_{\alpha \to 0} p_\alpha(l) > 0.$$

That is, the agent overestimates the likelihood of a bad state. The reason is that each pair $l_i l_j$ is possible according to the limit distribution $q(.|E)$ (and thus $p(.|E)$) and thus inductive inference implies that l will be considered possible in the limit. By continuity of p_α in α, the agent will exhibit $lim_{\alpha \to 0} p_\alpha(l) > 0 = lim_{\alpha \to 0} q_\alpha(l)$.

In the above illustration, the improving market still allows the possibility that two or more companies could have a low earning simultaneously. If we consider instead a market that improves so much that at most one company can have a low earning at a time (that is, E now only includes states x where $x_i = l_i$ for at most one $i \in \Gamma$), then this overestimation of the probability of the bad state l does not hold anymore and indeed, $lim_{\alpha \to 0} p_\alpha(l) = 0$. This is because the inductive inference property only extends to those states wholly constructed using pairwise occurrences $x_i x_j$ that feature in E. Nevertheless, overestimation still exists for other vanishing states. For instance, since occurrences such as $l_i x_j$ and $x_j l_k$ appear in E, the agent still believes in the possibility of states that include $l_i x_j l_k$, even though these have vanishing objective probability.

5.3 Experience Effect

An agent who has lived in a highly competitive environment will carry her experience with her even when dealing with people in a different, more cooperative environment. This illustrates an *experience effect*. Malmendier and Nagel (2011) demonstrate empirically that subjective expectations about inflation are largely determined by investors' life-time experience of inflation rather than available historical data. Experience effects are also studied in Bordalo et al (2019) in a deterministic choice context. While not highlighted in Gilboa and Schmeidler (2001, 2003) and Billot et al (2005), case-based beliefs also exhibit experience effects.

Consider a doctor who is evaluating whether a treatment plan is appropriate for a patient. Let K denote a generic patient (identified with, say, demographic information and medical history), let θ denote the various possible outcomes of the treatment and let μ denote the current vital signs, etc. The doctor sees a patient K with characteristics μ and is assessing the probability of outcome θ of some given treatment.

Suppose that the relative frequency distribution over past cases (θ, μ, K) is given by:

$$q(\theta, \mu, K) = q(\theta|\mu, K) q(\mu) q(K),$$

so that patient's demographics and history K and her characteristics μ are independently determined, but they jointly determine the likelihood of the outcome θ of the treatment. We show that[11]

[11]Proof: Given the functional form for q and the independence assumption $\frac{q(\mu,K)}{q(\mu)q(K)} = 1$, by Proposition 4 trained beliefs take the form

$$p(\theta, \mu, K) = \frac{1}{Z} \frac{q(\theta, \mu)}{q(\theta)q(\mu)} \frac{q(\theta, K)}{q(\theta)q(K)} \frac{q(\mu, K)}{q(\mu)q(K)} q(\theta)q(\mu)q(K) = \frac{1}{Z} \frac{q(\theta, \mu)q(\theta, K)}{q(\theta)} = \frac{1}{Z} q(\theta, \mu) q(K|\theta).$$

Proposition 5 *If the agent's Intuitive Beliefs p are trained in the sense of Proposition 4 then the Bayesian conditional of the beliefs over θ given characteristics μ and identity K is given by*

$$p(\theta|\mu, K) = \frac{1}{W} q(K|\theta) \sum_{\hat{K}} q(\theta|\mu, \hat{K}) q(\hat{K})$$

for some normalizing constant W.

That is, conditional on observing patient K with characteristics μ, the agent's belief $p(\theta|\mu, K)$ about outcome θ involves several terms, of which the most relevant for our purposes is

$$\sum_{\hat{K}} q(\theta|\mu, \hat{K}) q(\hat{K}).$$

This is the objective conditional distribution $q(\theta|\mu, \hat{K})$ of outcome θ averaged over all patients $\hat{K}$ with vitals μ the agent has observed. This implies the existence of an experience effect: the belief $p(\theta|\mu, K)$ does not depend only on the objective data $q(\theta|\mu, K)$ that pertains to this patient, but also on the agent's experience of how well the treatment performed on all patients with the same characteristics.

In case-based beliefs (Gilboa and Schmeidler 2001, 2003, Billot et al 2005), the agent's beliefs about the outcome for a particular patient can depend on the outcomes observed for other (similar) patients. In Billot et al (2005), when the patient K is fixed and μ is eschewed, case-based beliefs can be written as $p_{CB}(\theta) = \frac{1}{W} \sum_{\hat{K}} r(\theta|\hat{K}) v(\hat{K}, K)$ where r denotes a subjective probability distribution over outcomes for patient $\hat{K}$ and $v(\hat{K}, K)$ denotes the subjective similarity between $\hat{K}$ and the fixed patient K. Writing our model as $p(\theta|\mu, K) = \frac{1}{W} \sum_{\hat{K}} q(\theta|\mu, \hat{K}) v(\hat{K}, K, \theta)$ where $v(\hat{K}, K, \theta) := q(\hat{K}) q(K|\theta)$ helps provide some comparison and contrast in this application.

5.4 Belief Biases

Research in psychology reveals that subjects' beliefs derived in experiments bear little semblance to the Bayesian model (Kahneman and Tversky 1972, Tversky and Kahneman 1974, Benjamin 2019). Subjects appear to believe that relative frequencies in small samples are similar to those in large samples, in contrast with what the Law of Large Numbers dictates. Thus, they believe that the proportion of heads in any sample of tosses of a fair coin will be approximately 50% even in small samples. This is known as the *Law of Small Numbers* (Tversky and Kahneman 1974). On the other hand, they believe that relative frequencies in large samples will have more spread than the Law of Large Numbers dictates. This is known as *Non-Belief in the Law of Large Numbers* (Benjamin et al 2016). Subjects believe that a fair coin is more likely to give rise to a tail than a head following 3 heads – this is known as the *gambler's fallacy.* Subjects also exhibit a belief in the *hot hand effect:* they believe the bias of a coin changes temporarily to make a streak more likely. We show that trained Intuitive Beliefs can produce these findings.

Denote the probability of heads by $\mu \in [0, 1]$. Let K denote the sample size of tosses. Then the objective probability of getting a proportion $\theta \in [0, 1]$ of heads in a sample of K tosses of a coin with bias μ is $q(\theta|\mu, K) = \frac{K!}{\theta K!(1-\theta)K!} \mu^{\theta K} (1 - \mu)^{(1-\theta)K}$, where this probability is zero if θK is not

---

Using this to compute the Bayesian conditional $p(\theta|\mu, K) = \frac{p(\theta,\mu,K)}{p(\mu,K)}$ yields the expression $p(\theta|\mu, K) = \frac{q(\mu,\theta)q(K|\theta)}{\sum_{\theta'} q(\mu,\theta')q(K|\theta')}$. Expanding the $q(\mu, \theta)$ term yields

$$p(\theta|\mu, K) = \frac{\left[\sum_{\hat{K}} q(\theta|\mu, \hat{K}) q(\mu) q(\hat{K})\right] q(K|\theta)}{\sum_{\theta'} \left[\sum_{\hat{K}} q(\theta'|\mu, \hat{K}) q(\mu) q(\hat{K})\right] q(K|\theta')} = \frac{\left[\sum_{\hat{K}} q(\theta|\mu, \hat{K}) q(\hat{K})\right] q(K|\theta)}{\sum_{\theta'} \left[\sum_{\hat{K}} q(\theta'|\mu, \hat{K}) q(\hat{K})\right] q(K|\theta')} = \frac{1}{W} \left[\sum_{\hat{K}} q(\theta|\mu, \hat{K}) q(\hat{K})\right] q(K|\theta).$$

as desired.■

an integer. Assume that the objective probability $q(\mu)$ of the coin having bias μ and the objective probability $q(K)$ of observing a sample of size K are independent. Therefore the agent's experience is given by the objective probability distribution:

$$q(\theta, \mu, K) = q(\theta|\mu, K)q(\mu)q(K).$$

Assume that q has a support with arbitrarily large but finite cardinality. The following proposition writes the expression in Proposition 5 differently:[12]

Proposition 6 *If the agent's Intuitive Beliefs p are trained in the sense of Proposition 4 then the Bayesian conditional of the beliefs over θ given bias μ and sample size K is given by*

$$p(\theta|\mu, K) = \frac{q(\mu|\theta)q(\theta, K)}{\sum_{\theta'} q(\mu|\theta')q(\theta', K)}.$$

Thus, when told that a coin with bias μ will be tossed K times, the intuitive assessment of the proportion of heads θ depends on (i) the objective marginal probability $q(\theta, K)$ indicating how likely it is that θ and K are observed and (ii) the objective marginal probability $q(\mu|\theta) = \frac{q(\theta,\mu)}{q(\theta)}$ that a coin has bias μ given that it yields a proportion θ of heads in the sample.

5.4.1 Experience Effect: Law of Small Numbers and Non-Belief in the Law of Large Numbers

Observe that $q(\mu|\theta)$ depends on the distribution $q(\hat{K})$ over the sample sizes that the agent has experienced:

$$q(\mu|\theta) = \sum_{\hat{K}} \frac{q(\mu, \hat{K}, \theta)}{q(\theta)} = \frac{q(\mu)}{q(\theta)} \times \sum_{\hat{K}} q(\theta|\mu, \hat{K})q(\hat{K}).$$

In particular, $q(\mu|\theta)$ is a (scaled) average of sampling distributions $q(\theta|\mu, \hat{K})$ across $\hat{K}$ (with respect to $q(\hat{K})$). Therefore, her belief $p(\theta|\mu, K)$ for a given sample size K will depend on what she has observed in other samples, including those of different sizes. Given that the sample distribution becomes less spread out as the sample size increases, we can expect that if K is smaller than many of the samples she has experienced, then she will exhibit the Law of Small Numbers, whereas if it is larger, then she will exhibit Non-Belief in the Law of Large Numbers.

5.4.2 Gambler's Fallacy

To simplify the exposition, suppose that the agent has only observed coins with a given bias μ. Thus we are effectively in a two-dimensional environment, and it is readily established that $p(\theta, \mu, K) = p(\theta, K) = q(\theta, K)$, that is, the agent learns the true distribution. Moreover, conditional beliefs satisfy

$$p(\theta|\mu, K) = q(\theta|K),$$

that is, $p(\theta|\mu, K)$ reflects the objective *sampling distribution* $q(\theta|K)$. This is worth emphasizing. While experimental evidence adequately reveals that subjects have a poor intuitive understanding of probability theory, we suggest a reason for why this may be so: intuition is trained by sample distributions, rather than the axioms of probability theory. Her intuitive understanding of what

[12] Proof: Following the proof for Proposition 5,

$$p(\theta, \mu, K) = \frac{1}{Z} \frac{q(\theta, \mu)}{q(\theta)q(\mu)} \frac{q(\theta, K)}{q(\theta)q(K)} \frac{q(\mu, K)}{q(\mu)q(K)} q(\theta)q(\mu)q(K) = \frac{1}{Z} \frac{q(\theta, \mu)q(\theta, K)}{q(\theta)} = \frac{1}{Z} q(\mu|\theta)q(\theta, K).$$

Using this to compute the Bayesian conditional $p(\theta|\mu, K) = \frac{p(\theta,\mu,K)}{p(\mu,K)}$ yields the result. ■

it means for a coin to have bias $\mu = \frac{1}{2}$ is that it generates samples that are more likely to have a balanced number of heads and tails than an unbalanced number. At an intuitive level she has no basis for understanding the analytical meaning of "the probability of L heads in K tosses". Consequently, if an experimenter asks this agent what is the probability of "3 heads in 3 tosses" and "2 heads and a tail in 3 tosses", she responds with $p(\theta = 1|\frac{1}{2}, 3) = \frac{1}{8}$ and $p(\theta = \frac{2}{3}|\frac{1}{2}, 3) = \frac{3}{8}$ respectively, thereby exhibiting the Gambler's Fallacy.

It is worth emphasizing that the literature has focused on models of non-Bayesian updating to capture the Gambler's Fallacy (see for instance Rabin 2002 and Benjamin et al 2019). In contrast, we posit that the experimental evidence can be accommodated in a model of Intuitive Beliefs with standard Bayesian updating. That is, we suggest that the Gambler's Fallacy is a property of prior beliefs, and not necessarily updating.

5.4.3 Hot-Hand Effect

As in the literature (for instance, see Rabin and Vayanos 2010), let us view the Hot-Hand effect as arising from a belief that the bias of the coin is not fixed, despite the information given to the agent. Imagine, in particular, that the agent's belief about the sample mean θ in K tosses is not conditioned on μ. This yields the expression:[13]

$$p(\theta|K) = \sum_{\mu} q(\theta|\mu, K)q(\mu),$$

that is, $p(\theta|K)$ is the sampling distribution $q(\theta|\mu, K)$ averaged over the different μ the agent has experienced. If there is a single μ experienced, then $p(\theta|K) = q(\theta|\mu, K)$ is bell-shaped with mean μ. Indeed, as above, the agent exhibits the Gambler's Fallacy, since sample means are considered more likely if they are closer to μ. However, if the agent has experienced various μ then $p(\theta|K)$ may no longer be bell-shaped. To illustrate, if the agent has only experienced three coins – a fair coin and two coins with bias $\mu = 0, 1$ – and these have been experienced in equal proportions, then $p(\theta|K)$ is a symmetric tri-modal distribution with modes at $\mu = 0, \frac{1}{2}, 1$. It follows that if the agent is told that the coin will be tossed thrice then she may consider 3 heads more likely than 2 heads and a tail, in line with the Hot Hand Effect.

A Appendix: Geometric Marginals

Consider any probability measure $p \in \Delta(\Omega)$, not necessarily an Intuitive Belief. Denote the cardinality of Ω_i by K_i. Then for any $\phi \neq I \subset \Gamma$, the cardinality of $\Omega_I = \prod_{i\in I} \Omega_i$ is $K_I := \prod_{i\in I} K_i$.

For any $\phi \neq I \subset \Gamma$, define the geometric I-marginal of $p \in \Delta(\Omega)$ by

$$p^g(x_I) := \frac{1}{Z_I} \prod_{z_{-I}\in\Omega_{-I}} p(x_I z_{-I})^{\frac{1}{K_{-I}}} \qquad x_I \in \Omega_I,$$

where $Z_I := \sum_{y_I\in\Omega_I} \prod_{z_{-I}\in\Omega_{-I}} p(y_I z_{-I})^{\frac{1}{K_{-I}}}$. The first lemma shows that we can equivalently, and conveniently, take a product over all $z \in \Omega$ instead. The simple proofs for the results are omitted.

Lemma 1 *For any $\phi \neq I \subset \Gamma$, and any $x_I \in \Omega_I$,*

$$p^g(x_I) = \frac{\prod_{z\in\Omega} p(x_I z_{-I})^{\frac{1}{K}}}{\sum_{y_I\in\Omega_I} \prod_{z\in\Omega} p(y_I z_{-I})^{\frac{1}{K}}}.$$

[13]Compute that

$$p(\theta|K) = \sum_{\mu} \frac{q(\theta, \mu)q(\theta, K)/q(\theta)}{q(K)} = \frac{q(\theta, K)}{q(K)} = \frac{\sum_{\mu} q(\theta|\mu, K)q(\mu)q(K)}{q(K)} = \sum_{\mu} q(\theta|\mu, K)q(\mu).$$

Recall that the marginal of a marginal defines a corresponding marginal of the original probability measure. The next lemma shows that the corresponding property holds for geometric marginals

Lemma 2 *For any $\phi \neq J \subset \Gamma$ and $j \in J$ let $I = J \backslash \{j\}$. Then for any $x_I \in \Omega_I$,*

$$p^g(x_I) = \frac{\prod_{z_j \in \Omega_j} p^g(x_I z_j)^{\frac{1}{K_j}}}{\sum_{y_I \in \Omega_I} \prod_{z_j \in \Omega_j} p^g(y_I z_j)^{\frac{1}{K_j}}} \qquad x_I \in \Omega_I.$$

Say that p exhibits statistical independence if $p(x) = \prod_{i \in \Gamma} p^m(x_i)$ for all $x \in \Omega$. Write $p^m(x_I) = \prod_{i \in I} p^m(x_i)$ for any $\phi \neq I \subset \Gamma$. The next lemma shows that marginals and geo-marginals coincide under statistical independence.

Lemma 3 *If p exhibits statistical independence then for any $\phi \neq I \subset \Gamma$ and $x \in \Omega$*

$$p^g(x_I) = p^m(x_I).$$

In particular

$$p^g(x_I) = \prod_{i \in I} p^g(x_i)$$

B Appendix: Proof of Theorem 3

B.1 Basic Identification Result

Consider any Intuitive Belief p and any representation (a, b). Let $\Omega^+ := supp(p)$. Without making it explicit, we make use of the fact that for any $x \in \Omega^+$ and $i, j \in \Gamma$, $a(x_i x_j) > -\infty$ and $b(x_i) > -\infty$.

Lemma 4 *For all $x, z \in \Omega$ and $y, y_i z_{-i} \in \Omega^+$,*
(i) $\frac{p(x)}{p(y)} = exp\left[\sum_{i<j}[a(x_i x_j) - a(y_i y_j)] + \sum_i [b(x_i) - b(y_i)]\right]$
(ii) $\frac{p(x_i z_{-i})}{p(y_i z_{-i})} = exp\left[\sum_{i \neq j \in \Gamma}[a(x_i z_j) - a(y_i z_j)] + [b(x_i) - b(y_i)]\right]$.

Proof. The first claim follows trivially from the model. The second follows just as easily: for any $y_i z_{-i} \in \Omega^+$,

$$\frac{p(x_i z_{-i})}{p(y_i z_{-i})} = exp[[\sum_{i \neq j \in \Gamma} a(x_i z_j) + \sum_{i \neq k < j \neq i} a(z_k z_j)] - [\sum_{i \neq j \in \Gamma} a(y_i z_j) + \sum_{i \neq k < j \neq i} a(z_k z_j)] + b(x_i) - b(y_i) + \sum_{i \neq j \in \Gamma} [b(z_i) - b(z_i)]$$

$$= exp\left[\sum_{i \neq j \in \Gamma} [a(x_i z_j) - a(y_i z_j) + b(x_i) - b(y_i)]\right],$$

as desired. ■

The next key result provides the connection between a and p in any representation.

Lemma 5 *For any Intuitive Belief p and any $z \in \Omega^+$ and $x_i x_j$ s.t. $x_i z_{-i}, x_j z_{-j} \in \Omega^+$,*

$$\frac{exp[a(x_i x_j) + a(z_i z_j)]}{exp[a(x_i z_j) + a(z_i x_j)]} = \frac{p(x_i x_j z_{-ij}) p(z)}{p(x_i z_{-i}) p(x_j z_{-j})}.$$

Proof. By the representation and the previous lemma, for any $z \in \Omega^+$ and $x_i x_j$ s.t. $x_i z_{-i}, x_j z_{-j} \in \Omega^+$,

$$\frac{p(x_i x_j z_{-ij})}{p(z_i z_j z_{-ij})} = exp[[a(x_i x_j) - a(z_i z_j)] + \left[\sum_{ij \neq k \in \Gamma} a(x_i z_k) - a(z_i z_k)\right] + \left[\sum_{ij \neq k \in \Gamma} a(x_j z_k) - a(z_j z_k)\right]$$
$$+ [b(x_i) - b(z_i) + b(x_j) - b(z_j)]]$$

$$= exp[[a(x_i x_j) - a(z_i z_j) - [a(x_i z_j) - a(z_i z_j)] - [a(x_j z_i) - a(z_j z_i)]]$$
$$+ \left[\sum_{i \neq k \in \Gamma} [a(x_i z_k) - a(z_i z_k)] + [b(x_i) - b(z_i)]\right] + \left[\sum_{j \neq k \in \Gamma} [a(x_j z_k) - a(z_j z_k)] + [b(x_j) - b(z_j)]\right]$$
$$= exp\left[a(x_i x_j) - a(x_i z_j) - a(x_j z_i) + a(z_j z_i)\right] \times \frac{p(x_i z_{-i})}{p(z_i z_{-i})} \frac{p(x_j z_{-j})}{p(z_j z_{-j})}$$
$$= \frac{exp[a(x_i x_j) + a(z_i z_j)]}{exp[a(x_i z_j) + a(z_i x_j)]} \times \frac{p(x_i z_{-i})}{p(z_i z_{-i})} \frac{p(x_j z_{-j})}{p(z_j z_{-j})}$$

which rearranges to

$$\frac{p(x_i x_j z_{-ij})}{p(x_i z_{-i}) p(x_j z_{-j})} = \frac{p(z_i z_j z_{-ij})}{p(z_i z_{-i}) p(z_j z_{-j})} \times \frac{exp[a(x_i x_j) + a(z_i z_j)]}{exp[a(x_i z_j) + a(z_i x_j)]}$$
$$= \frac{1}{p(z)} \times \frac{exp[a(x_i x_j) + a(z_i z_j)]}{exp[a(x_i z_j) + a(z_i x_j)]},$$

thereby establishing the claim. ∎

Below we will often use the following convenient fact without explicit reference to it. Adopt the notation $\prod_{i<j\leq m} := \prod_{i,j:i<j\leq m}$. Take any scalars c_{ij} satisfying $c_{ij} = c_{ji}$.

Lemma 6 *For any $m \geq 2$ and strictly positive scalars c_{ij} defined for each distinct $i, j \leq m$, it must be that*

$$\prod_{i<j\leq m} c_{ij} = \prod_{i \leq m} \prod_{i \neq j \leq m} c_{ij}^{0.5}.$$

Proof. The term $\prod_{i<j\leq m} c_{ij}$ takes each i and multiplies the terms c_{ij} where $j > i$, and then takes the product over all i. We obtain the same expression if we take each i and multiply the terms $c_{ij}^{0.5}$ with all $j \neq i$, and then take the product over all i. ∎

B.2 Proof of Uniqueness Theorem

Proof. Take any Intuitive Belief p and suppose there exists a reference state, that is, $\bar{z} \in \Omega$ s.t. $p(\bar{z}) > 0$ and for any $i \in \Gamma$ and $x_i \in \Omega_i$, $p(x_i) > 0 \implies p(x_i \bar{z}_{-i}) > 0$. Step 1 establishes the first desired claim in the Theorem, Step 2 and 3 establish sufficiency of the second desired claim and Step 4 establishes its necessity.

Step 1: Show that p is represented by (a, b) iff it is represented by $(\alpha, 0)$ where

$$\alpha(x_i x_j) = a(x_i x_j) + \frac{1}{(N-1)}[b(x_i) + b(x_j)].$$

Suppose p is represented by (a, b). By definition the normalizing constant Z is strictly positive. Therefore, given $\Sigma_{y \in \Omega} p(y) = 1$, it must be that $\Sigma_{y \in \Omega} exp[\sum_{i<j} a(y_i y_j) + \sum_i b(y_i)] = Z > 0$.

Take any $x \in \Omega$, and observe that
$\sum_{i<j} \alpha(x_i x_j) + \sum_i 0$
$= \sum_{i<j} a(x_i x_j) + \sum_{i<j} \frac{1}{(N-1)}[b(x_i) + b(x_j)]$
$= \sum_{i<j} a(x_i x_j) + \frac{1}{(N-1)} \frac{1}{2} \sum_i \sum_{i \neq j \in \Gamma} [b(x_i) + b(x_j)]$
$= \sum_{i<j} a(x_i x_j) + \frac{1}{(N-1)} \frac{1}{2} \sum_i [(N-1)b(x_i) + \sum_{i \neq j \in \Gamma} b(x_j)]$
$= \sum_{i<j} a(x_i x_j) + \frac{1}{(N-1)} \frac{1}{2} \sum_i [(N-2)b(x_i) + \sum_{j \in \Gamma} b(x_j)]$
$= \sum_{i<j} a(x_i x_j) + \frac{1}{(N-1)} \frac{1}{2} [(N-2) \sum_i b(x_i) + N \sum_{j \in \Gamma} b(x_j)]$
$= \sum_{i<j} a(x_i x_j) + \frac{1}{(N-1)} \frac{1}{2} [2(N-1) \sum_i b(x_i)]$
$= \sum_{i<j} a(x_i x_j) + \sum_i b(x_i)$.

Therefore $exp[\sum_{i<j} \alpha(x_i x_j)] = exp[\sum_{i<j} a(x_i x_j) + \sum_i b(x_i)]$. Moreover, $\Sigma_{y \in \Omega} exp[\sum_{i<j} \alpha(y_i y_j)] = \Sigma_{y \in \Omega} exp[\sum_{i<j} a(y_i y_j) + \sum_i b(y_i)] > 0$. Consequently $p(x) = \frac{exp[\sum_{i<j} a(x_i x_j) + \sum_i b(x_i)]}{\Sigma_{y \in \Omega} exp[\sum_{i<j} a(y_i y_j) + \sum_i b(y_i)]} = \frac{\sum_{i<j} \alpha(x_i x_j)}{\Sigma_{y \in \Omega} exp[\sum_{i<j} \alpha(y_i y_j)]}$, establishing the result.

Step 2: Show that, if p is represented by $(a, 0)$ and $(\alpha, 0)$ then there exists a function $(j, x_i) \mapsto \gamma_j^*(x_i)$ s.t. for any $x \in \Omega$ and distinct $i, j \in \Gamma$,

$$a(x_i x_j) = \alpha(x_i x_j) + \gamma_j^*(x_i) + \gamma_i^*(x_j).$$

Suppose p is represented by $(a, 0)$ and $(\alpha, 0)$. If $p(x_i) = 0$ then define $\gamma_j^*(x_i)$ arbitrarily. If $p(x_i) > 0$ then consider the reference state $\overline{z} \in \Omega$ and for any $j \neq i$ let

$$\gamma_j^*(x_i) = a(x_i \bar{z}_j) - \alpha(x_i \bar{z}_j) - \frac{1}{2}[a(\bar{z}_i \bar{z}_j) - \alpha(\bar{z}_i \bar{z}_j)].$$

Since $p(\overline{z}) > 0$ and $p(x_i \overline{z}_{-i}) > 0$, by the representation all the terms in the expression are real valued.

Now we show that the desired equality holds. Take any $x_i x_j$. If $p(x_i)p(x_j) = 0$ then $p(x_i x_j) = 0$ and by the representation, it must be that $a(x_i x_j) = \alpha(x_i x_j) = -\infty$. The desired equality holds trivially.

Next suppose $p(x_i)p(x_j) > 0$. Then by Lemma 5,

$$\frac{exp\,[a(x_i x_j) + a(\bar{z}_i \bar{z}_j)]}{exp\,[a(x_i \bar{z}_j) + a(\bar{z}_i x_j)]} = \frac{p(x_i x_j \bar{z}_{-ij}) p(\bar{z})}{p(x_i \bar{z}_{-i}) p(x_j \bar{z}_{-j})} = \frac{exp\,[\alpha(x_i x_j) + \alpha(\bar{z}_i \bar{z}_j)]}{exp\,[\alpha(x_i \bar{z}_j) + \alpha(\bar{z}_i x_j)]]},$$

and it follows that

$$a(x_i x_j) = \alpha(x_i x_j) + [a(x_i \bar{z}_j) - \alpha(x_i \bar{z}_j)] + [a(\bar{z}_i x_j) - \alpha(\bar{z}_i x_j)] - [a(\bar{z}_i \bar{z}_j) - \alpha(\bar{z}_i \bar{z}_j)]$$

$$= \alpha(x_i x_j) + [a(x_i \bar{z}_j) - \alpha(x_i \bar{z}_j)] + [a(\bar{z}_i x_j) - \alpha(\bar{z}_i x_j)]$$

$$-\frac{1}{2}[a(\bar{z}_i \bar{z}_j) - \alpha(\bar{z}_i \bar{z}_j)] - \frac{1}{2}[a(\bar{z}_j \bar{z}_i) - \alpha(\bar{z}_j \bar{z}_i)]$$

$$= \alpha(x_i x_j) + \gamma_j^*(x_i) + \gamma_i^*(x_j),$$

where we exploited the symmetry of a.

Step 3 : Show that for each $i \in \Gamma$ there is k_i s.t. for each $x_i \in \Omega_i$,

$$\sum_{i \neq j \in \Gamma} \gamma_j(x_i) = k_i,$$

Suppose as in Step 3 that p is represented by both (a, b) and (α, β). Begin by defining, for each x and i, the quantities

$$K(x_i) := \sum_{i \neq j \in \Gamma} \gamma_j(x_i), \text{ and } K(x) = \frac{1}{2} \sum_{i \in \Gamma} K(x_i).$$

Observe that for any $x \in \Omega$,
$\sum_{i<j} a(x_i x_j)$
$= \sum_{i<j} [\alpha(x_i x_j) + \gamma_j(x_i) + \gamma_i(x_j)]$
$= \sum_{i<j} [\alpha(x_i x_j)] + \sum_{i<j} [\gamma_j(x_i) + \gamma_i(x_j)]$
$= \sum_{i<j} [\alpha(x_i x_j)] + \frac{1}{2} \sum_{i \in \Gamma} \sum_{i \neq j \in \Gamma} \gamma_j(x_i)]$
$= \sum_{i<j} [\alpha(x_i x_j) + \frac{1}{2} \sum_{i \in \Gamma} K(x_i)$, that is,

$$\sum_{i<j} a(x_i x_j) = \sum_{i<j} [\alpha(x_i x_j)] + \frac{1}{2} K(x).$$

Given this observation, note that since p is represented by both (a, b) and (α, β) , we have that for all $x \in \Omega$,

$$\frac{exp\left[\sum_{i<j} \alpha(x_i x_j)\right]}{\sum_{y \in \Omega} exp\left[\sum_{i<j} \alpha(y_i y_j)\right]} = p(x) = \frac{exp\left[\sum_{i<j} a(x_i x_j)\right]}{\sum_{y \in \Omega} exp\left[\sum_{i<j} a(y_i y_j)\right]}$$

$$= \frac{exp\left[\sum_{i<j} \alpha(x_i x_j) + \frac{1}{2} K(x)\right]}{\sum_{y \in \Omega} exp\left[\sum_{i<j} \alpha(y_i y_j) + \frac{1}{2} K(y)\right]}$$

$$= exp[\frac{1}{2} K(x)] \times \frac{exp\left[\sum_{i<j} \alpha(x_i x_j)\right]}{\sum_{y \in \Omega} exp\left[\sum_{i<j} \alpha(y_i y_j) + \frac{1}{2} K(y)\right]}.$$

Looking at the first and last terms in this sequence of equalities, we see that for all $x \in \Omega$,

$$exp[\frac{1}{2} K(x)] = \frac{\sum_{y \in \Omega} exp\left[\sum_{i<j} \alpha(y_i y_j) + \frac{1}{2} K(y)\right]}{\sum_{y \in \Omega} exp\left[\sum_{i<j} \alpha(y_i y_j)\right]}.$$

Since the RHS is independent of x, it follows that $K(x)$ is independent of x. Therefore, $K(x) = K(y)$ for all x, y.

Finally, for any x and i, consider $x_i \bar{z}_{-i}$ and observe that
$K(x_i \bar{z}_{-i}) = K(\bar{z})$
$\implies K(x_i) + \sum_{i \neq j \in \Gamma} K(\bar{z}_j) = K(\bar{z}_i) + \sum_{i \neq j \in \Gamma} K(\bar{z}_j)$
$\implies K(x_i) = K(\bar{z}_i)$
$\implies \sum_{i \neq j \in \Gamma} \gamma_j(x_i) = \sum_{i \neq j \in \Gamma} \gamma_j(\bar{z}_i)$.
Define $k_i = \sum_{i \neq j \in \Gamma} \gamma_j(\bar{z}_i)$ to complete the step.

Step 5: Establish necessity.

Suppose that $(a, 0)$ represents p and consider $(\alpha, 0)$ as in the desired statement. Observe that
$\sum_{i<j} a(x_i x_j)$
$= \sum_{i<j} [\alpha(x_i x_j) + \gamma_j(x_i) + \gamma_i(x_j)]$
$= \sum_{i<j} \alpha(x_i x_j) + \sum_{i<j} [\gamma_j(x_i)] + \sum_{i<j} [\gamma_i(x_j)]$
$= \sum_{i<j} \alpha(x_i x_j) + \sum_{i \in \Gamma} \sum_{i \neq j \in \Gamma} [\gamma_j(x_i)]$

$= \sum_{i<j} \alpha(x_i x_j) + \sum_{i\in\Gamma} k_i$, that is, letting $K := \sum_{i\in\Gamma} k_i$, it must be that

$$\sum_{i<j} a(x_i x_j) = \sum_{i<j} \alpha(x_i x_j) + K.$$

It is then straightforward to see that

$$p(x) = \frac{exp\left[\sum_{i<j} a(x_i x_j)\right]}{\sum_{y\in\Omega} exp\left[\sum_{i<j} a(y_i y_j)\right]}$$

$$= \frac{exp\left[\sum_{i<j} \alpha(x_i x_j) + K\right]}{\sum_{y\in\Omega} exp\left[\sum_{i<j} \alpha(y_i y_j) + K\right]}$$

$$= \frac{exp\left[\sum_{i<j} \alpha(x_i x_j)\right]}{\sum_{y\in\Omega} exp\left[\sum_{i<j} \alpha(y_i y_j)\right]},$$

thereby establishing that $(\alpha, 0)$ also represents p, as was to be shown. ■

C Appendix: Proof of Theorem 1

We explore several classes of normalized representations using Theorem 3.

C.1 Normalization 1

Lemma 7 *Suppose there exists a reference state $\overline{z} \in \Omega$. An Intuitive Belief representation $(\alpha, 0)$ can be normalized by setting $\alpha(\overline{z}_i \overline{z}_j) = 0$ for all distinct $i, j \in \Gamma$, and $\alpha(x_i \overline{z}_j) = \alpha(x_i \overline{z}_{j'})$ for all $i \neq j, j' \in \Gamma$ and $x \in \Omega$. Such a normalized representation is unique, given $\overline{z}$. Moreover, $p(\overline{z}) \times Z = 1$ holds in the representation normalized wrt $\overline{z}$.*

Proof. Take any representation $(a, 0)$, and take any reference state $\overline{z} \in \Omega$ with $p(\overline{z}) > 0$. Take any $x_i \in \Omega_i$. If $p(x_i) = 0$ then define $\gamma_j(x_i)$ arbitrarily. If $p(x_i) > 0$ then define

$$\gamma_j(x_i) = -\frac{1}{N-1}[\zeta(x_i) - \zeta(\overline{z}_i)] + \frac{1}{2} a(\overline{z}_i \overline{z}_j) - a(x_i \overline{z}_j) + \frac{1}{N-1} \sum_{i\neq k\in\Gamma} [a(x_i \overline{z}_k) - a(\overline{z}_i \overline{z}_k)].$$

Note that
$\sum_{i\neq j\in\Gamma} \gamma_j(x_i)$
$= \frac{1}{2} \sum_{i\neq j\in\Gamma} a(\overline{z}_i \overline{z}_j) - \sum_{i\neq j\in\Gamma} a(x_i \overline{z}_j) + \sum_{i\neq k\in\Gamma} [a(x_i \overline{z}_k) - a(\overline{z}_i \overline{z}_k)]$
$= -\frac{1}{2} \sum_{i\neq j\in\Gamma} a(\overline{z}_i \overline{z}_j) = k_i$. In particular, γ along with $k_i := \zeta(\overline{z}_i) - \frac{1}{2} \sum_{i\neq j\in\Gamma} a(\overline{z}_i \overline{z}_j)$ satisfy the desired properties in Theorem 3.

Step 1: Show that p is represented by $(\alpha, 0)$ defined by $\alpha(x_i x_j) = -\infty$ if $p(x_i)p(x_j) = 0$ and

$$\alpha(x_i x_j) = [a(x_i x_j) + a(\overline{z}_i \overline{z}_j) - a(x_i \overline{z}_j) - a(x_j \overline{z}_i)]$$
$$+ \frac{1}{N-1} \left[\sum_{i\neq k\in\Gamma} [a(x_i \overline{z}_k) - a(\overline{z}_i \overline{z}_k)] + \sum_{j\neq k\in\Gamma} [a(x_j \overline{z}_k) - a(\overline{z}_j \overline{z}_k)] \right],$$

otherwise.

By Theorem 3 we obtain a new representation $(\alpha, 0)$ defined by:
$\alpha(x_i x_j) = a(x_i x_j) + \gamma_j(x_i) + \gamma_i(x_j)$
$= a(x_i x_j) - \frac{1}{N-1}[\zeta(x_i) - \zeta(\overline{z}_i)] + \frac{1}{2}a(\overline{z}_i\overline{z}_j) - a(x_i\overline{z}_j) + \frac{1}{N-1}\sum_{i\neq k\in\Gamma}[a(x_i\overline{z}_k) - a(\overline{z}_i\overline{z}_k)]$
$-\frac{1}{N-1}[\zeta(x_j) - \zeta(\overline{z}_j)] + \frac{1}{2}a(\overline{z}_j\overline{z}_i) - a(x_j\overline{z}_i) + \frac{1}{N-1}\sum_{j\neq k\in\Gamma}[a(x_j\overline{z}_k) - a(\overline{z}_j\overline{z}_k)]$
$= a(x_i x_j) - \frac{1}{N-1}[\zeta(x_i) - \zeta(\overline{z}_i) + \zeta(x_j) - \zeta(\overline{z}_j)] + a(\overline{z}_i\overline{z}_j) - a(x_i\overline{z}_j) - a(x_j\overline{z}_i)$
$+\frac{1}{N-1}\sum_{i\neq k\in\Gamma}[a(x_i\overline{z}_k) - a(\overline{z}_i\overline{z}_k)] + \frac{1}{N-1}\sum_{j\neq k\in\Gamma}[a(x_j\overline{z}_k) - a(\overline{z}_j\overline{z}_k)]$.
Step 2: Show that α satisfies for all distinct $i, j, k \in \Gamma$ and $x_i \in \Omega_i$,

$$\alpha(\overline{z}_i\overline{z}_j) = 0 \text{ and } \alpha(x_i\overline{z}_j) = \alpha(x_i\overline{z}_k).$$

Use the expression for α in Step 1 to obtain $\alpha(\overline{z}_i\overline{z}_j) = 2a(\overline{z}_i\overline{z}_j) - 2a(\overline{z}_i\overline{z}_j) + 0 = 0$. If $p(x_i) = 0$ then $\alpha(x_i\overline{z}_j) = 0$ for all j. If $p(x_i) > 0$ then
$\alpha(x_i\overline{z}_j) = a(x_i\overline{z}_j) + a(\overline{z}_i\overline{z}_j) - a(x_i\overline{z}_j) - a(\overline{z}_j\overline{z}_i)$
$+\frac{1}{N-1}\sum_{i\neq k\in\Gamma}[a(x_i\overline{z}_k) - a(\overline{z}_i\overline{z}_k)] + \frac{1}{N-1}\sum_{j\neq k\in\Gamma}[a(\overline{z}_j\overline{z}_k) - a(\overline{z}_j\overline{z}_k)]$
$= 0 + \frac{1}{N-1}\sum_{i\neq k\in\Gamma}[a(x_i\overline{z}_k) - a(\overline{z}_i\overline{z}_k)]$, which does not depend on j.
Step 3: Conclusion.

We have thus shown that there always exists a normalized representation as stated in the lemma. Note that in a normalized representation $(\alpha, 0)$, since $\alpha(\overline{z}_i\overline{z}_j) = 0$, it must be that $p(\overline{z}) = \frac{1}{Z}exp[\sum_{i<j}\alpha(\overline{z}_i\overline{z}_j)] = \frac{1}{Z}$, that is, $p(\overline{z}) = \frac{1}{Z}$.

In Lemma 8 below we show that α can be written in terms of p. Therefore the uniqueness of the representation is a corollary of that result. ■

Lemma 8 *A network $(\alpha, 0)$ is a $\overline{z}$-normalized representation for p if and only if for all distinct $i, j \in \Gamma$ and $(x_i, x_j) \in \Omega_{ij}$,*

$$exp[\alpha(x_i x_j)] = \begin{cases} p(\overline{z})^{\frac{N-3}{N-1}} \dfrac{p(x_i x_j \overline{z}_{-ij})}{[p(x_i\overline{z}_{-i})p(x_j\overline{z}_{-j})]^{\frac{N-2}{N-1}}} & \textit{if } p(x_i)p(x_j) > 0 \\ 0 & \textit{otherwise} \end{cases}.$$

Proof. If $(\alpha, 0)$ satisfies the noted expression then it is readily determined that $\alpha(\overline{z}_i\overline{z}_j) = 0$ for all distinct $i, j \in \Gamma$ and $\alpha(x_i\overline{z}_j) = \alpha(x_i\overline{z}_{j'})$ for all $i \neq j, j' \in \Gamma$ and $x \in \Omega$. Thus $(\alpha, 0)$ is a $\overline{z}$-normalized representation.

Conversely, take any $\overline{z}$-normalized representation $(\alpha, 0)$ for p. By Lemma 7, $Z \times p(\overline{z}) = 1$. Since $\alpha(\overline{z}_j\overline{z}_k) = 0$ for all distinct $j, k \in \Gamma$ and that $a(x_i\overline{z}_k)$ is independent of k, we see that
$p(x_i\overline{z}_{-i}) = \frac{1}{Z} \times exp\left[\sum_{i\neq k\in\Gamma}\alpha(x_i\overline{z}_k) + \sum_{i\neq j<k\neq i}\alpha(\overline{z}_j\overline{z}_k)\right]$
$= p(\overline{z}) \times exp\left[\sum_{i\neq k\in\Gamma}\alpha(x_i\overline{z}_k)\right]$
$= p(\overline{z}) \times exp\left[(N-1)\alpha(x_i\overline{z}_j)\right]$ for any $i \neq j \in \Gamma$, we obtain the property that

$$\alpha(x_i\overline{z}_j) = \frac{1}{N-1}ln[\frac{p(x_i\overline{z}_{-i})}{p(\overline{z})}],$$

for all $i \neq j \in \Gamma$. Finally, observe that, by Lemma 5, when $p(x_i)p(x_j) > 0$, the representation must satisfy:

$$\frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z})}{p(x_i\overline{z}_{-i})p(x_j\overline{z}_{-j})} = \frac{exp[\alpha(x_i x_j) + \alpha(\overline{z}_i\overline{z}_j)]}{exp[\alpha(x_i\overline{z}_j) + \alpha(\overline{z}_i x_j)]} = \frac{exp[\alpha(x_i x_j)]}{[\frac{p(x_i\overline{z}_{-i})}{p(\overline{z})}]^{\frac{1}{N-1}}[\frac{p(x_j\overline{z}_{-j})}{p(\overline{z})}]^{\frac{1}{N-1}}},$$

and so

$$exp[\alpha(x_i x_j)] = \frac{p(x_i x_j \overline{z}_{-ij})}{p(x_i\overline{z}_{-i})p(x_j\overline{z}_{-j})} \times p(\overline{z}) \times [\frac{p(x_i\overline{z}_{-i})}{p(\overline{z})}\frac{p(x_j\overline{z}_{-j})}{p(\overline{z})}]^{\frac{1}{N-1}}$$
$$= p(\overline{z})^{\frac{N-3}{N-1}}\frac{p(x_i x_j \overline{z}_{-ij})}{[p(x_i\overline{z}_{-i})p(x_j\overline{z}_{-j})]^{\frac{N-2}{N-1}}},$$

as desired. ■

C.2 Normalization 2

Lemma 9 *Suppose there exists a reference state. A belief p with full support is an Intuitive Belief if and only if, for any reference state $\overline{z} \in \Omega$, it is represented by an associative network $(a_{\overline{z}}, b_{\overline{z}})$ with the property that for any $x_i, x_j \in \cup_{k \in \Gamma} \Omega_k$,*

$$exp[a_{\overline{z}}(x_i x_j)] = \begin{cases} \frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z})}{p(x_i \overline{z}_{-i})p(x_j \overline{z}_{-j})} & \text{if } p(x_i)p(x_j) > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$exp[b_{\overline{z}}(x_i)] = \frac{p(x_i \overline{z}_{-i})}{p(\overline{z})}.$$

Indeed, for any reference $\overline{z} \in \Omega$, there exists a representation $(a_{\overline{z}}, b_{\overline{z}})$ where $a_{\overline{z}}(\overline{z}_i \overline{z}_j) = a_{\overline{z}}(x_i \overline{z}_j) = b_{\overline{z}}(\overline{z}_i) = 0$. Moreover, this representation implies the existence of the reduced form where for any $x \in \Omega$,

$$p(x) > 0 \implies p(x) = p(\overline{z}) \times \left[\prod_{i<j} \frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z})}{p(x_i \overline{z}_{-i})p(x_j \overline{z}_{-j})} \right] \times \prod_{i \in \Gamma} \frac{p(x_i \overline{z}_{-i})}{p(\overline{z})},$$

which can also be written as $p(x) > 0 \implies p(x) = p(\overline{z})^{\frac{(N-2)(N-1)}{2}} \times \frac{\prod_{i<j} p(x_i x_j \overline{z}_{-ij})}{\prod_i p(x_i \overline{z}_{-i})^{N-2}}$.

Proof. If there exists such a representation then p is trivially an Intuitive Belief. Conversely, suppose p is an Intuitive Belief. Consider a normalized representation for p and re-write the expression in Lemma 8 for the case $p(x_i)p(x_j) > 0$ as:

$$exp[\alpha(x_i x_j)] = \frac{\frac{p(x_i x_j \overline{z}_{-ij})}{p(\overline{z})}}{\frac{p(x_i \overline{z}_{-i})}{p(\overline{z})} \frac{p(x_j \overline{z}_{-j})}{p(\overline{z})}} [\frac{p(x_i \overline{z}_{-i})}{p(\overline{z})}]^{\frac{1}{N-1}} [\frac{p(x_j \overline{z}_{-j})}{p(\overline{z})}]^{\frac{1}{N-1}}.$$

Observe that for any $x \in \Omega$ such that $p(x) > 0$, it must be that $p(x_i) > 0$ for all i. Therefore, for any $x \in \Omega$ such that $p(x) > 0$, inserting the above expression into the representation and redefining the scaling factor as needed yields:

$p(x) = \frac{1}{Z} \times \prod_{i<j} exp[a(x_i x_j)]$
$= \frac{1}{Z} \times \prod_{i<j} \frac{\frac{p(x_i x_j \overline{z}_{-ij})}{p(\overline{z})}}{\frac{p(x_i \overline{z}_{-i})}{p(\overline{z})} \frac{p(x_j \overline{z}_{-j})}{p(\overline{z})}} [\frac{p(x_i \overline{z}_{-i})}{p(\overline{z})}]^{\frac{1}{N-1}} [\frac{p(x_j \overline{z}_{-j})}{p(\overline{z})}]^{\frac{1}{N-1}}$
$= \frac{1}{Z} \times \left[\prod_{i<j} \frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z})}{p(x_i \overline{z}_{-i})p(x_j \overline{z}_{-j})}\right] \times \prod_{i<j} [\frac{p(x_i \overline{z}_{-i})}{p(\overline{z})}]^{\frac{1}{N-1}} [\frac{p(x_j \overline{z}_{-j})}{p(\overline{z})}]^{\frac{1}{N-1}}$
$= \frac{1}{Z} \times \left[\prod_{i<j} \frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z})}{p(x_i \overline{z}_{-i})p(x_j \overline{z}_{-j})}\right] \times \prod_{i \in \Gamma} \prod_{i \neq j \in \Gamma} [\frac{p(x_i \overline{z}_{-i})}{p(\overline{z})}]^{\frac{1}{2(N-1)}} [\frac{p(x_j \overline{z}_{-j})}{p(\overline{z})}]^{\frac{1}{2(N-1)}}$
$= \frac{1}{Z} \times \left[\prod_{i<j} \frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z})}{p(x_i \overline{z}_{-i})p(x_j \overline{z}_{-j})}\right] \times \prod_{i \in \Gamma} \left[[\frac{p(x_i \overline{z}_{-i})}{p(\overline{z})}]^{\frac{N-1}{2(N-1)}} \times \left[\prod_{i \neq j \in \Gamma} [\frac{p(x_j \overline{z}_{-j})}{p(\overline{z})}]^{\frac{1}{2(N-1)}} \right] \right]$
$= \frac{1}{Z} \times \left[\prod_{i<j} \frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z})}{p(x_i \overline{z}_{-i})p(x_j \overline{z}_{-j})}\right] \times \prod_{i \in \Gamma} \left[[\frac{p(x_i \overline{z}_{-i})}{p(\overline{z})}]^{\frac{N-1}{2(N-1)}} \times [\frac{p(x_i \overline{z}_{-i})}{p(\overline{z})}]^{\frac{-1}{2(N-1)}} \left[\prod_{j \in \Gamma} [\frac{p(x_j \overline{z}_{-j})}{p(\overline{z})}]^{\frac{1}{2(N-1)}} \right] \right]$
$= \frac{1}{Z} \times \left[\prod_{i<j} \frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z})}{p(x_i \overline{z}_{-i})p(x_j \overline{z}_{-j})}\right] \times \left[\prod_{j \in \Gamma} [\frac{p(x_j \overline{z}_{-j})}{p(\overline{z})}]^{\frac{N}{2(N-1)}} \right] \times \prod_{i \in \Gamma} \left[[\frac{p(x_i \overline{z}_{-i})}{p(\overline{z})}]^{\frac{N-2}{2(N-1)}} \right]$
$= \frac{1}{Z} \times \left[\prod_{i<j} \frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z})}{p(x_i \overline{z}_{-i})p(x_j \overline{z}_{-j})}\right] \times \left[\prod_{i \in \Gamma} [\frac{p(x_i \overline{z}_{-j})}{p(\overline{z})}]^{\frac{N}{2(N-1)} + \frac{N-2}{2(N-1)}} \right]$
$= \frac{1}{Z} \times \left[\prod_{i<j} \frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z})}{p(x_i \overline{z}_{-i})p(x_j \overline{z}_{-j})}\right] \times \left[\prod_{i \in \Gamma} \frac{p(x_i \overline{z}_{-j})}{p(\overline{z})} \right]$. By Lemma 7, $\frac{1}{Z} = p(\overline{z})$. This proves that if p is an Intuitive Belief, then it must admit the desired reduced form. It is clear from the reduced form that there exists a representation $(a_{\overline{z}}, b_{\overline{z}})$ as in the statement of the Lemma.

Finally, to verify the alternative reduced form, note that for all $x \in \Omega$ with $p(x) > 0$,

$$p(x) = p(\overline{z}) \times \left[\prod_{i<j} \frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z})}{p(x_i \overline{z}_{-i})p(x_j \overline{z}_{-j})} \right] \times \prod_{i \in \Gamma} \frac{p(x_i \overline{z}_{-i})}{p(\overline{z})}$$

$$= p(\bar{z}) \times \left[\prod_i \prod_{i \neq j \in \Gamma} \frac{p(x_i x_j \bar{z}_{-ij}) p(\bar{z})}{p(x_i \bar{z}_{-i}) p(x_j \bar{z}_{-j})} \right]^{0.5} \times \prod_{i \in \Gamma} \frac{p(x_i \bar{z}_{-i})}{p(\bar{z})}$$

$$= p(\bar{z}) \times \left[\prod_i p(\bar{z})^{N-1} \times \prod_i \left[\frac{1}{p(x_i \bar{z}_{-i})} \right]^{N-1} \times \prod_i \left[\prod_{i \neq j \in \Gamma} \frac{p(x_i x_j \bar{z}_{-ij})}{p(x_j \bar{z}_{-j})} \right] \right]^{0.5} \times \prod_{i \in \Gamma} \frac{p(x_i \bar{z}_{-i})}{p(\bar{z})}$$

$$= p(\bar{z})^{1+\frac{N(N-1)}{2}} \times \left[\prod_i \frac{1}{p(x_i \bar{z}_{-i})} \right]^{\frac{N-1}{2}} \times \left[\prod_j \frac{1}{p(x_j \bar{z}_{-i})^{N-1}} \right]^{0.5} \times \left[\prod_i \prod_{i \neq j \in \Gamma} p(x_i x_j \bar{z}_{-ij}) \right]^{0.5} \times \prod_{i \in \Gamma} \frac{p(x_i \bar{z}_{-i})}{p(\bar{z})}$$

$$= p(\bar{z})^{1+\frac{N(N-1)}{2}-N} \times \prod_i \left[\frac{1}{p(x_i \bar{z}_{-i})} \right]^{N-1} \times \left[\prod_{i<j} p(x_i x_j \bar{z}_{-ij}) \right] \times \prod_{i \in \Gamma} p(x_i \bar{z}_{-i})$$

$$= p(\bar{z})^{\frac{(N-2)(N-1)}{2}} \times \frac{\prod_{i<j} p(x_i x_j \bar{z}_{-ij})}{\prod_i p(x_i \bar{z}_{-i})^{N-2}},$$

as desired. ■

C.3 Proof of Theorem 1

Proof. By the full support assumption, every state is a reference state. Collecting the $p(\bar{z})$-terms in the reduced form established in Lemma 9 and defining $\frac{1}{Z_{\bar{z}}} = p(\bar{z})^{1+\frac{N(N-1)}{2}-N} = p(\bar{z})^{\frac{(N-1)(N-2)}{2}}$, we see that p is an Intuitive Belief if and only if for each $\bar{z} \in \Omega$, it admits the reduced form

$$p(x) = \frac{1}{Z_{\bar{z}}} \times \left[\prod_{i<j} \frac{p(x_i x_j \bar{z}_{-ij})}{p(x_i \bar{z}_{-i}) p(x_j \bar{z}_{-j})} \right] \times \prod_{i \in \Gamma} p(x_i \bar{z}_{-i}), \quad x \in \Omega.$$

Since the reduced form for each $\bar{z} \in \Omega$ expresses the same p, so will the geometric mean of all these reduced forms. Consequently, for all $x \in \Omega$,

$$p(x) = \prod_{\bar{z} \in \Omega} \left[\frac{1}{Z_{\bar{z}}} \times \left[\prod_{i<j} \frac{p(x_i x_j \bar{z}_{-ij})}{p(x_i \bar{z}_{-i}) p(x_j \bar{z}_{-j})} \right] \times \prod_{i \in \Gamma} p(x_i \bar{z}_{-i}) \right]^{\frac{1}{K}}$$

$$= \frac{1}{Z'} \times \left[\prod_{i<j} \frac{p^h(x_i x_j)}{p^h(x_i) p^h(x_j)} \right] \times \prod_{i \in \Gamma} p^h(x_i),$$

where $\frac{1}{Z'} := \frac{1}{\prod_{\bar{z} \in \Omega} [Z_{\bar{z}}^{\frac{1}{K}}]}$ and for any $I \subset \Gamma$,

$$p^h(x_I) := \prod_{\bar{z} \in \Omega} p(x_I z_{-I})^{\frac{1}{K}}, \qquad x \in \Omega.$$

Given Lemma 1, dividing this by the constant $\sum_{y_I \in \Omega_I} p^h(y_I)$ converts p^h into a geo-marginal. Inserting suitable constants into the expression for $p(x)$ and defining an appropriate scaling factor $\frac{1}{Z}$, we see that p admits a representation of the desired form. ■

D Proof of Proposition 1

Proof. As noted in the proof of Theorem 1, Lemma 9 yields that p is an Intuitive Belief if and only if for each $z \in \Omega$, it admits the reduced form

$$p(x) = \frac{1}{Z_z} \times \left[\prod_{i<j} \frac{p(x_i x_j z_{-ij})}{p(x_i z_{-i}) p(x_j z_{-j})} \right] \times \prod_{i \in \Gamma} p(x_i z_{-i}), \quad x \in \Omega.$$

Taking a geometric average over z$\in E \in \Sigma$ yields

$$\begin{aligned} p(x) &= \prod_{z \in E} \left[\frac{1}{Z_z} \times \left[\prod_{i<j} \frac{p(x_i x_j z_{-ij})}{p(x_i z_{-i}) p(x_j z_{-j})} \right] \times \prod_{i \in \Gamma} p(x_i z_{-i}) \right]^{\frac{1}{K(\mathbf{z})}} \\ &= \frac{1}{Z'(E)} \times \left[\prod_{i<j} \frac{p^g(x_i x_j | E)}{p^g(x_i|E) p^g(x_j|E)} \right] \times \prod_{i \in \Gamma} p^g(x_i|E) \end{aligned}$$

for any appropriate scalar $Z'(E)$. Letting $Z(E) := p(E) Z'(E)$, the Bayesian update can then be written as

$$p(x|E) = \frac{p(x)}{p(E)} = \frac{1}{Z(E)} \times \left[\prod_{i<j} \frac{p^g(x_i x_j | E)}{p^g(x_i|E) p^g(x_j|E)} \right] \times \prod_{i \in \Gamma} p^g(x_i|E),$$

as desired. ■

E Proof of Theorem 2

We provide a characterization result involving just the prior p and then obtain the desired result for the dynamic setup as a corollary.

E.1 Associative Separability Conditions

Consider a data restriction on each dimension $i \in \Gamma$ given by a set of elementary states $\phi \neq S_i \subset \Omega_i$, and let $S = \prod_{i \in \Gamma} S_i \subset \Omega$. For any dimensions $i, j \in \Gamma$, consider the subset of $\Omega = \Omega_{ij} \times \Omega_{-ij}$ that applies the data restriction to all dimensions outside i, j:

$$\Omega_{ij} \times S_{-ij},$$

the cardinality of which is denoted $K(\Omega_{ij} \times S_{-ij})$. Compute a normalized geometric mean of $p(x_I z_{-I})$ using only $z \in \Omega_{ij} \times S_{-ij}$, that is, for any $I \subset \Gamma$, define the *estimated geo-marginal* by:

$$\hat{p}^{S_{-ij}}(x_I) := \frac{1}{Z_I} \prod_{z \in \Omega_{ij} \times S_{-ij}} p(x_I z_{-I})^{\frac{1}{K(\Omega_{ij} \times S_{-ij})}} \qquad x_I \in \Omega_I.$$

Consider the following relationship between geo-PMI and estimated geo-PMI:

Definition 5 *Beliefs p over Ω with full support satisfy Prior Relative Associative Separability (PRAS) if for any $\phi \neq S = \prod_{i \in \Gamma} S_i$, any distinct $i, j \in \Gamma$ and all $(x_i, x_j), (y_i, y_j) \in \Omega_{ij}$,*

$$\frac{p^g(x_i x_j)}{p^g(x_i) p^g(x_j)} \Big/ \frac{p^g(y_i y_j)}{p^g(y_i) p^g(y_j)} = \frac{\hat{p}^{S_{-ij}}(x_i x_j)}{\hat{p}^{S_{-ij}}(x_i) \hat{p}^{S_{-ij}}(x_j)} \Big/ \frac{\hat{p}^{S_{-ij}}(y_i y_j)}{\hat{p}^{S_{-ij}}(y_i) \hat{p}^{S_{-ij}}(y_j)}.$$

The interpretation is similar to RAS except that it applies only to the prior p and the data restriction lies in calculating (estimated) geo-marginals.

We show that PRAS is related to a simpler condition. For any given $\overline{z} \in \Omega$, distinct $i, j \in \Gamma$ and $x_i, x_j \in \cup_{k\in\Gamma}\Omega_k$, consider the ratio:

$$a_{\overline{z}}(x_i x_j) = \frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z})}{p(x_i \overline{z}_{-i})p(x_j \overline{z}_{-j})}, \tag{6}$$

Consider the following property:

Definition 6 *A belief p on Ω with full support satisfies* Weak Prior Relative Associative Separability *(WPRAS) if for all $\overline{z}, \overline{w} \in \Omega$ and each distinct $i, j \in \Gamma$ and $x_i \in \Omega_i$, $x_j \in \Omega_j$*

$$\frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z}_i \overline{z}_j \overline{z}_{-ij})}{p(x_i \overline{z}_j \overline{z}_{-ij})p(\overline{z}_i x_j \overline{z}_{-ij})} = \frac{p(x_i x_j \overline{w}_{-ij})p(\overline{z}_i \overline{z}_j \overline{w}_{-ij})}{p(x_i \overline{z}_j \overline{w}_{-ij})p(\overline{z}_i x_j \overline{w}_{-ij})},$$

that is, $a_{\overline{z}_{ij}\overline{z}_{-ij}}(x_i x_j) = a_{\overline{z}_{ij}\overline{w}_{-ij}}(x_i x_j)$.

We first show that WPRAS is equivalent to PRAS.

Lemma 10 *Any belief p on Ω with full support satisfies WPRAS if and only if it satisfies PRAS.*

Proof. $\Longrightarrow$: Suppose that WPRAS is satisfied. Then for any distinct $i, j \in \Gamma$ and $x_i, x_j \in \cup_{k\in\Gamma}\Omega_k$,

$$\begin{aligned}
\frac{p^g(x_i x_j)}{p^g(x_i)p^g(x_j)} &= \prod_{z\in\Omega}\left[\frac{p(x_i x_j z_{-ij})}{p(x_i z_{-i})p(x_j z_{-j})}\right]^{\frac{1}{K}} \quad \text{by definition of } p^g \\
&= \frac{1}{\prod_{z\in\Omega} p(z)^{\frac{1}{K}}} \times \prod_{z\in\Omega}\left[\frac{p(x_i x_j z_{-ij})p(z)}{p(x_i z_{-i})p(x_j z_{-j})}\right]^{\frac{1}{K}} \\
&= \frac{1}{\prod_{z\in\Omega} p(z)^{\frac{1}{K}}} \times \prod_{z\in\Omega}\left[\frac{p(x_i x_j y_{-ij})p(z_i z_j y_{-ij})}{p(x_i z_j y_{-ij})p(z_i x_j y_{-ij})}\right]^{\frac{1}{K}} \quad \text{for any fixed } y_{-ij} \in \Omega_{-ij} \\
&= \frac{1}{\prod_{z\in\Omega} p(z)^{\frac{1}{K}}} \times \prod_{(z_i,z_j)\in\Omega_{ij}} \prod_{z_{-ij}\in\Omega_{ij}}\left[\frac{p(x_i x_j y_{-ij})p(z_i z_j y_{-ij})}{p(x_i z_j y_{-ij})p(z_i x_j y_{-ij})}\right]^{\frac{1}{K}} \\
&= \frac{1}{\prod_{z\in\Omega} p(z)^{\frac{1}{K}}} \times \prod_{(z_i,z_j)\in\Omega_{ij}} \left[\frac{p(x_i x_j y_{-ij})p(z_i z_j y_{-ij})}{p(x_i z_j y_{-ij})p(z_i x_j y_{-ij})}\right]^{\frac{K_{-ij}}{K}} \\
&= \frac{1}{\prod_{z\in\Omega} p(z)^{\frac{1}{K}}} \times \prod_{(z_i,z_j)\in\Omega_{ij}} \left[\frac{p(x_i x_j y_{-ij})p(z_i z_j y_{-ij})}{p(x_i z_j y_{-ij})p(z_i x_j y_{-ij})}\right]^{\frac{1}{K_{ij}}},
\end{aligned}$$

that is, for any fixed $y_{-ij} \in \Omega_{-ij}$,

$$\frac{p^g(x_i x_j)}{p^g(x_i)p^g(x_j)} = \frac{1}{\prod_{z\in\Omega} p(z)^{\frac{1}{K}}} \times \prod_{(z_i,z_j)\in\Omega_{ij}} \left[\frac{p(x_i x_j y_{-ij})p(z_i z_j y_{-ij})}{p(x_i z_j y_{-ij})p(z_i x_j y_{-ij})}\right]^{\frac{1}{K_{ij}}}$$

But since this holds for any $y_{-ij} \in S_{-ij}$, we preserve the equality if we take a geometric mean of the RHS expression wrt $y_{-ij} \in S_{-ij}$. Therefore, writing $K_{S_{-ij}}$ for the cardinality of S_{-ij},

$$\frac{p^g(x_i x_j)}{p^g(x_i)p^g(x_j)} = \prod_{y_{-ij}\in S_{-ij}}\left[\frac{1}{\prod_{z\in\Omega} p(z)^{\frac{1}{K}}} \times \prod_{(z_i,z_j)\in\Omega_{ij}} \left[\frac{p(x_i x_j y_{-ij})p(z_i z_j y_{-ij})}{p(x_i z_j y_{-ij})p(z_i x_j y_{-ij})}\right]^{\frac{1}{K_{ij}}}\right]^{\frac{1}{K_{S_{-ij}}}}$$

$$= \frac{1}{\prod_{z\in\Omega} p(z)^{\frac{K_{S_{-ij}}}{K}}} \times \prod_{(z_i z_j y_{ij})\in\Omega_{ij}\times S_{-ij}} \left[\frac{p(x_i x_j y_{-ij}) p(z_i z_j y_{-ij})}{p(x_i z_j y_{-ij}) p(z_i x_j y_{-ij})}\right]^{\frac{1}{K_{ij} K_{S_{-ij}}}}$$

$$= \frac{\prod_{(z_i z_j y_{ij})\in\Omega_{ij}\times S_{-ij}} p(z_i z_j y_{-ij})}{\prod_{z\in\Omega} p(z)^{\frac{K_{S_{-ij}}}{K}}} \times \prod_{(z_i z_j y_{ij})\in\Omega_{ij}\times S_{-ij}} \left[\frac{p(x_i x_j y_{-ij})}{p(x_i z_j y_{-ij}) p(z_i x_j y_{-ij})}\right]^{\frac{1}{K(\Omega_{ij}\times S_{-ij})}}$$

$$= \left[\frac{\prod_{(z_i z_j y_{ij})\in\Omega_{ij}\times S_{-ij}} p(z_i z_j y_{-ij})}{\prod_{z\in\Omega} p(z)^{\frac{K_{S_{-ij}}}{K}}} \frac{Z_{ij}}{Z_i Z_j}\right] \frac{\hat{p}^{S_{-ij}}(x_i x_j)}{\hat{p}^{S_{-ij}}(x_i)\hat{p}^{S_{-ij}}(x_j)},$$

where Z_i, Z_j, Z_{ij} are the scalars required to convert the geo-means into geometric marginals. Defining the constant in the bracket in the last expression by $\zeta_{S_{-ij}} > 0$, we obtain the result that for all $(x_i, x_j) \in \Omega_{ij}$,

$$\frac{p^g(x_i x_j)}{p^g(x_i)p^g(x_j)} = \zeta_{S_{-ij}} \frac{\hat{p}^{S_{-ij}}(x_i x_j)}{\hat{p}^{S_{-ij}}(x_i)\hat{p}^{S_{-ij}}(x_j)}.$$

It follows that PRAS holds.

$\Longleftarrow$: Now suppose PRAS holds. Apply PRAS to the case where there is only one reference state: $S = \{\overline{z}\}$ for some $\overline{z} \in \Omega$. Then $K(\Omega_{ij} \times \{\overline{z}_{-ij}\}) = K_{ij}$. By definition,

$$\hat{p}^{S_{-ij}}(x_i x_j) := \frac{1}{Z_{ij}} \prod_{z\in\Omega_{ij}\times\{\overline{z}_{-ij}\}} p(x_i x_j z_{-ij})^{\frac{1}{K_{ij}}}$$

$$= \frac{1}{Z_{ij}} \prod_{(z_i z_j)\in\Omega_{ij}} p(x_i x_j \overline{z}_{-ij})^{\frac{1}{K_{ij}}} = \frac{1}{Z_{ij}} p(x_i x_j \overline{z}_{-ij}),$$

and similarly,

$$\hat{p}^{S_{-ij}}(x_i) := \frac{1}{Z_i} \prod_{z\in\Omega_{ij}\times\{\overline{z}_{-ij}\}} p(x_i z_j z_{-ij})^{\frac{1}{K_{ij}}}$$

$$= \frac{1}{Z_i} \prod_{z_j\in\Omega_j} \prod_{z_i\in\Omega_i} p(x_i z_j \overline{z}_{-ij})^{\frac{1}{K_{ij}}} = \frac{1}{Z_i} \prod_{z_j\in\Omega_j} p(x_i z_j \overline{z}_{-ij})^{\frac{1}{K_j}}.$$

Define $exp[\hat{a}_{\overline{z}_{-ij}}(x_i, x_j)] := \frac{p(x_i x_j \overline{z}_{-ij})}{[\prod_{z_j\in\Omega_j} p(x_i z_j \overline{z}_{-ij})^{\frac{1}{K_j}}][\prod_{z_i\in\Omega_i} p(z_i x_j \overline{z}_{-ij})^{\frac{1}{K_i}}]}$. Then,

$$\frac{exp[\hat{a}_{\overline{z}_{-ij}}(x_i, x_j)]}{exp[\hat{a}_{\overline{z}_{-ij}}(y_i, y_j)]} = \frac{\frac{p(x_i x_j \overline{z}_{-ij})}{p(y_i y_j \overline{z}_{-ij})}}{[\prod_{z_j\in\Omega_j}[\frac{p(x_i z_j \overline{z}_{-ij})}{p(y_i z_j \overline{z}_{-ij})}]^{\frac{1}{K_j}}] \times [\prod_{z_i\in\Omega_i} \frac{p(z_i x_j \overline{z}_{-ij})}{p(z_i y_j \overline{z}_{-ij})}]^{\frac{1}{K_i}}]}.$$

In particular, compute

$$\frac{exp[\hat{a}_{\overline{z}_{-ij}}(x_i, x_j)]}{exp[\hat{a}_{\overline{z}_{-ij}}(x_i, \overline{z}_j)]} = \frac{\frac{p(x_i x_j \overline{z}_{-ij})}{p(x_i \overline{z}_j \overline{z}_{-ij})}}{[\prod_{z_i\in\Omega_i} \frac{p(z_i x_j \overline{z}_{-ij})}{p(z_i \overline{z}_j \overline{z}_{-ij})}]^{\frac{1}{K_i}}]} \text{ and } \frac{exp[\hat{a}_{\overline{z}_{-ij}}(\overline{z}_i, \overline{z}_j)]}{exp[\hat{a}_{\overline{z}_{-ij}}(\overline{z}_i, x_j)]} = \frac{\frac{p(\overline{z}_i \overline{z}_j \overline{z}_{-ij})}{p(\overline{z}_i x_j \overline{z}_{-ij})}}{[\prod_{z_i\in\Omega_i} \frac{p(z_i x_j \overline{z}_{-ij})}{p(z_i \overline{z}_j \overline{z}_{-ij})}]^{\frac{1}{K_i}}]}.$$

Recalling the canonical representation (a_g, b_g) (Theorem 1), by PRAS, it follows that

$$\frac{exp[a_g(x_i, x_j)]}{exp[a_g(x_i, \overline{z}_j)]} \times \frac{exp[a_g(\overline{z}_i, \overline{z}_j)]}{exp[a_g(\overline{z}_i, x_j)]} = \frac{exp[\hat{a}_{\overline{z}_{-ij}}(x_i, x_j)]}{exp[\hat{a}_{\overline{z}_{-ij}}(x_i, \overline{z}_j)]} \times \frac{exp[\hat{a}_{\overline{z}_{-ij}}(\overline{z}_i, \overline{z}_j)]}{exp[\hat{a}_{\overline{z}_{-ij}}(\overline{z}_i, x_j)]} = \frac{p(x_i x_j \overline{z}_{-ij}) p(\overline{z}_i \overline{z}_j \overline{z}_{-ij})}{p(x_i \overline{z}_j \overline{z}_{-ij}) p(\overline{z}_i x_j \overline{z}_{-ij})}.$$

Since the left hand side expression, $\frac{exp[a_g(x_i,x_j)]}{exp[a_g(x_i,\overline{z}_j)]} \times \frac{exp[a_g(\overline{z}_i,\overline{z}_j)]}{exp[a_g(\overline{z}_i,x_j)]}$, does not depend on $\overline{z}_{-ij}$, it follows that the right hand side expression will not change if we replace $\overline{z}$ with $\overline{z}_i\overline{z}_j\overline{w}_{-ij}$ for any $\overline{w}_{-ij} \in \Omega_{-ij}$. Conclude that

$$\frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z}_i\overline{z}_j\overline{z}_{-ij})}{p(x_i\overline{z}_j\overline{z}_{-ij})p(\overline{z}_i x_j\overline{z}_{-ij})} = \frac{p(x_i x_j \overline{z}_{-ij})p(\overline{z}_i\overline{z}_j\overline{w}_{-ij})}{p(x_i\overline{z}_j\overline{z}_{-ij})p(\overline{z}_i x_j\overline{w}_{-ij})},$$

which establishes WPRAS. ■

E.2 Static Characterization Result

Theorem 5 *A belief p on Ω with full support satisfies PRAS if and only if it is an Intuitive Belief.*

Proof. If p is an Intuitive Belief, then by Lemma 5, for any representation (a,b) it must be that

$$\frac{exp[a(x_i x_j) + a(z_i z_j)]}{exp[a(x_i z_j)] \times exp[a(x_j z_i)]} = \frac{\frac{p(x_i x_j z_{-ij})}{p(z)}}{\frac{p(x_i z_{-i})}{p(z)}\frac{p(x_j z_{-j})}{p(z)}}.$$

Conclude that $\frac{\frac{p(x_i x_j z_{-ij})}{p(z)}}{\frac{p(x_i z_{-i})}{p(z)}\frac{p(x_j z_{-j})}{p(z)}}$ is independent of z_{-ij}, and so, WPRAS holds. By Lemma 10, PRAS is satisfied, as desired.

Conversely, suppose that PRAS holds. By Lemma 10, WPRAS holds. If $N = 2$ then obtain a representation simply by defining $exp[a(x_i x_j)] = p(x_i x_j)$. Let $Z = 1$ and we obtain an Intuitive Belief representation. Henceforth assume $N > 2$. Fix some $z \in \Omega$ throughout.

Step 1: Show that, for $Z = \frac{1}{p(z)}$, and any x_i, x_j, x_k,

$$p(x_i x_j x_k z_{-ijk}) = \frac{1}{Z} \prod_{l,m \in \{i,j,k\}: i<j} \frac{p(x_l x_m z_{-lm})}{p(x_l z_{-l})p(x_m z_{-m})} \times \prod_{l \in \{i,j,k\}} p(x_l z_{-l})$$

Take any x_i, x_j, x_k, z and consider $w_{-ij} = x_k z_{-ijk}$. By WPRAS,

$$\frac{p(x_i x_j z_{-ij})p(z_i z_j z_{-ij})}{p(x_i z_j z_{-ij})p(z_i x_j z_{-ij})} = \frac{p(x_i x_j, x_k z_{-ijk})p(z_i z_j, x_k z_{-ijk})}{p(x_i z_j, x_k z_{-ijk})p(z_i x_j, x_k z_{-ijk})},$$

Rearranging this expression yields:

$p(x_i x_j x_k z_{-ijk}) = p(z_i z_j z_{-ij})\frac{p(x_i x_j z_{-ij})p(x_i z_j x_k z_{-ijk})p(z_i x_j x_k z_{-ijk})}{p(x_i z_j z_{-ij})p(z_i x_j z_{-ij})p(z_i z_j x_k z_{-ijk})}$

$= p(z)\frac{p(x_i x_j z_{-ij})p(x_i x_k z_{-ik})p(x_j x_k z_{-jk})}{p(x_i z_{-i})p(x_j z_{-j})p(x_k z_{-k})}$.

$= p(z)\frac{p(x_i x_j z_{-ij})}{p(x_i z_{-i})p(x_j z_{-j})}\frac{p(x_i x_k z_{-ik})}{p(x_i z_{-i})p(x_k z_{-k})}\frac{p(x_j x_k z_{-jk})}{p(x_j z_{-j})p(x_k z_{-k})}p(x_i z_{-i})p(x_j z_{-j})p(x_k z_{-k})$, which yields the desired expression.

Step 2: Show that for any $x \in \Omega$,

$$p(x) = \frac{1}{Z}\left[\prod_{i,j \in \{1,..,N\}, i<j} \frac{p(x_i x_j z_{-ij})}{p(x_i z_{-i})p(x_j z_{-j})} \times \prod_{i \in \{1,..,N\}} p(x_i z_{-i})\right].$$

Assume the induction hypothesis that for any n elements of Γ, which we abuse notation for and label as $1,..,n$,[14] we have the functional form

$$p(x_1..x_n z_{-1,...,n}) = \frac{1}{Z} \prod_{i,j \in \{1,..,n\}: i<j} \frac{p(x_i x_j z_{-ij})}{p(x_i z_{-i})p(x_j z_{-j})} \times \prod_{i \in \{1,..,n\}} p(x_i z_{-i}),$$

[14] This is an abuse of notation since by definition, $\Gamma := \{1,...,N\}$, whereas we will use $1,...,n$ to denote generic elements in Γ.

where $Z = \frac{1}{p(z)}$. To prove the induction step, take any $1,..,n,n+1 \in \Gamma$, any $x_{n+1} \in \Omega_{n+1}$. Define

$A = p(x_1...x_{n-1}z_n x_{n+1} z_{-1,....,n+1})$, $B = p(x_1...x_{n-2}z_{n-1}x_n x_{n+1} z_{-1,...,n+1})$, $C = p(x_1...x_{n-2}z_{n-1}z_n x_{n+1} z_{-1..n+1})$.

Adopt the simplifying notation $y_l^{l+m} := (y_l..., y_{l+m})$ for any state y. Letting $w_{-n-1,.n} := x_1^{n-2} x_{n+1} z_{n+2}^N$, WPRAS implies
$\frac{p(x_{n-1}x_n, z_{-n-1,n})p(z_{n-1}z_n, z_{-n-1,n})}{p(x_{n-1}z_n, z_{-n-1,n})p(z_{n-1}x_n, z_{-n-1,n})}$
$= \frac{p(x_{n-1}x_n, x_1^{n-2}x_{n+1}z_{n+2}^N)p(z_{n-1}z_n, x_1^{n-2}x_{n+1}z_{n+2}^N)}{p(x_{n-1}z_n, x_1^{n-2}x_{n+1}z_{n+2}^N)p(z_{n-1}x_n, x_1^{n-2}x_{n+1}z_{n+2}^N)}$
$= \frac{p(x_1...x_{n+1}z_{-1..n+1})p(x_1...x_{n-2}z_{n-1}z_n x_{n+1}z_{-1..n+1})}{p(x_1...x_{n-1}z_n x_{n+1}z_{-1..n+1})p(x_1...x_{n-2}z_{n-1}x_n x_{n+1}z_{-1..n+1})}$
$= \frac{p(x_1...x_{n+1}z_{-1..n+1})C}{AB}$, which yields

$$p(x_1...x_{n+1}z_{-1..n+1}) = \left[\frac{AB}{C}\right]\left[\frac{p(x_{n-1}x_n z_{-n-1,n})p(z)}{p(x_{n-1}z_{-n-1})p(x_n z_{-n})}\right]. \tag{7}$$

By the induction step, the terms A, B, C are given by

$$A = \frac{1}{Z} \times \prod_{i,j\in\{1,..,n-1,n+1\}:i<j} \frac{p(x_i x_j z_{-ij})}{p(x_i z_{-i})p(x_j z_{-j})} \times \prod_{i\in\{1,..,n-1,n+1\}} p(x_i z_{-i}),$$

$$B = \frac{1}{Z} \times \prod_{i,j\in\{1,..,n-2,n,n+1\}:i<j} \frac{p(x_i x_j z_{-ij})}{p(x_i z_{-i})p(x_j z_{-j})} \times \prod_{i\in\{1,..,n-2,n,n+1\}} p(x_i z_{-i}),$$

$$C = \frac{1}{Z} \times \prod_{i,j\in\{1,..,n-2,n+1\}:i<j} \frac{p(x_i x_j z_{-ij})}{p(x_i z_{-i})p(x_j z_{-j})} \times \prod_{i\in\{1,..,n-2,n+1\}} p(x_i z_{-i}).$$

We see that

$$\frac{B}{C} = \left[\prod_{j\in\{1,..,n-2,n+1\}} \frac{p(x_n x_j z_{-nj})}{p(x_n z_{-n})p(x_j z_{-j})}\right] \times p(x_n z_{-n})$$

$$= \frac{1}{\frac{p(x_{n-1}x_n z_{-n-1,n})}{p(x_{n-1}z_{-n-1})p(x_n z_{-n})}} \left[\prod_{i\in\{1,..n-1,n+1\}} \frac{p(x_i x_n z_{-in})}{p(x_i z_{-i})p(x_n z_{-n})}\right] \times p(x_n z_{-n})$$

and therefore

$$\frac{AB}{C} = \frac{1}{Z} \frac{p(z)}{\frac{p(x_{n-1}x_n z_{-n-1,n})}{p(x_{n-1}z_{-n-1})p(x_n z_{-n})}} \times \prod_{i,j\in\{1,..,n+1\}:i<j} \frac{p(x_i x_j z_{-ij})}{p(x_i z_{-i})p(x_j z_{-j})} \times \prod_{i\in\{1,..,n+1\}} p(x_i z_{-i}).$$

Inserting this into the equality (7) and using $Z = \frac{1}{p(z)}$ yields:

$$p(x_1...x_{n+1}z_{-1..n+1}) = \frac{1}{Z}\left[\prod_{i,j\in\{1,..,n+1\}:i<j} \frac{p(x_i x_j z_{-ij})}{p(x_i z_{-i})p(x_j z_{-j})} \times \prod_{i\in\{1,..,n+1\}} p(x_i z_{-i})\right],$$

completing the induction step. Conclude that $p(x)$ can be written in the desired way.

Step 3: Conclude the proof of sufficiency.

The expression for p in Step 2 is an Intuitive Belief with $a(x_i x_j) = \frac{p(x_i x_j z_{-ij})}{p(x_i z_{-i})p(x_j z_{-j})}$ and $b(x_i) = p(x_i z_{-i})$, where $z \in \Omega$ is fixed, as desired. ■

E.3 Proof of Theorem 2

Proof. Take any distinct $i, j \in \Gamma$ and i, j-unrestricted $E \in \Sigma$ of the form $E = \Omega_{ij} \times S_{-ij}$. Then there are scalars $\lambda_{ij}(E) > 0$ and $\lambda_i(E) > 0$ such that for any $x \in E$,

$$p^g(x_i x_j|E) := \frac{1}{Z_{ij}(E)} \prod_{z \in E} p(x_i x_j z_{-ij}|E)^{\frac{1}{K(E)}} = \frac{1}{p(E) Z_{ij}(E)} \prod_{z \in E} p(x_i x_j z_{-ij})^{\frac{1}{K(E)}} = \lambda_{ij}(E) \hat{p}^{S_{-ij}}(x_I)$$

and similarly,

$$p^g(x_i|E) := \frac{1}{p(E) Z_i(E)} \prod_{z \in E} p(x_i w_j z_{-ij})^{\frac{1}{K(E)}} = \lambda_i(E) \hat{p}^{S_{-ij}}(x_i).$$

These expressions imply

$$\frac{p^g(x_i x_j|E)}{p^g(x_i|E) p^g(x_j|E)} \Big/ \frac{p^g(y_i y_j|E)}{p^g(y_i|E) p^g(y_j|E)} = \frac{\hat{p}^{S_{-ij}}(x_i x_j)}{\hat{p}^{S_{-ij}}(x_i) \hat{p}^{S_{-ij}}(x_j)} \Big/ \frac{\hat{p}^{S_{-ij}}(y_i y_j)}{\hat{p}^{S_{-ij}}(y_i) \hat{p}^{S_{-ij}}(y_j)}.$$

Given this equality, we see that RAS is equivalent to PRAS. The desired result follows from Theorem 5. ■

F Appendix: Proof of Theorem 4 and Proposition 2

Begin with an observation. Consider the D-training problem and say that $x_i x_j$ *appears* in D if there exists z s.t. $x_i x_j z_{-ij} \in D$. Denote the set of such pairs by

$$E_D = \{(x_i x_j) \in \cup_{i<j} (\Omega_i \times \Omega_j) : x_i x_j \text{ appears in } D\}.$$

Since $q(x) > 0$ for all $x \in D \subset supp(q)$, it follows that $KL(q||p) = \infty$ if $p(x) = 0$ for any $x \in D$. Therefore, it must be that any solution p to the D-training problem satisfies $p(x) > 0$ for all $x \in D$. Taking any representation (α, β) with the normalization $\beta = 0$, it must be that $\alpha(x_i x_j) > -\infty$ for all $x_i x_j$ that appear in D. The representation implies that $p(x) > 0$ for any x constructed using $x_i x_j$ that appear in D.

Moreover, minimality requires $p(x) = 0$ for any x for which there exists i, j s.t. $x_i x_j$ does not appear in D (that is, $\alpha(x_i x_j) = -\infty$ for any such $x_i x_j$). Thus, any solution p to the training problem must satisfy

$$supp(p) = \Omega_D := \{x \in \Omega : x_i x_j \in E_D \text{ for all } i, j \in \Gamma\}.$$

Then assumption 1 implies that there is $\overline{z} \in D$ s.t. for any such p, $p(x_i) > 0 \implies p(x_i \overline{z}_{-i}) > 0$. That is, $\overline{z} \in D$ is a reference state for all p satisfying $supp(p) = \Omega_D$.

Since Δ_{IB} is not generally a convex set (though we will see that it is closed under geometric mixtures), it will prove useful to find a convex set of representations to work with. For this purpose, we define a class A of normalized representations next. Given Lemmas 7 and 8, any belief $p \in \Delta_{IB}$ that is minimal and for which $\overline{z}$ is a reference state can be represented by a $\overline{z}$-normalized network $(a, 0)$ where a is defined for all $x_i, x_j \in \cup_{k \in \Gamma} \Omega_k$ by

$$exp[a(x_i x_j)] = \begin{cases} p(\overline{z})^{\frac{N-3}{N-1}} \frac{p(x_i x_j \overline{z}_{-ij})}{[p(x_i \overline{z}_{-i}) p(x_j \overline{z}_{-j})]^{\frac{N-2}{N-1}}} & \text{if } p(x_i) p(x_j) > 0 \\ 0 & \text{otherwise} \end{cases}. \quad (8)$$

Identify $(a, 0)$ with the association function a restricted to the subdomain $E_D \subset \cup_{i<j} (\Omega_i \times \Omega_j)$ (given that $a(x_i x_j) = -\infty$ outside E_D by minimality). Therefore a is an element of Euclidean space with dimensionality equal to the cardinality of E_D. By Lemma 8, this representation is characterized by

the restriction that $a(\overline{z}_i\overline{z}_j) = 0$ and $a(x_i\overline{z}_j) = a(x_i\overline{z}_k)$ for all distinct $i, j, k \in \Gamma$ and any $x \in \Omega$. Let A denote the set of all possible minimal $\overline{z}$-normalized networks:

$$A = \{a : E_D \to \mathbb{R}|\ a(\overline{z}_i\overline{z}_j) = 0 \text{ and } a(x_i\overline{z}_j) = a(x_i\overline{z}_k) \text{ for all distinct } i, j, k \in \Gamma \text{ and any } x \in \Omega\}.$$

Note that A depends on $\overline{z} \in \Omega$ but we are suppressing this in the notation. Moreover, for any $a \in A$, the corresponding $p_a \in \Delta_{IB}$ is defined by $p_a(x) = \frac{1}{Z_a} exp[\sum_{i<j} a(x_ix_j)]$. Some useful facts are established next:

Lemma 11 *The following hold:*

(i) For any $a_1, a_2 \in A$ and the respective Intuitive Beliefs $p_1, p_2 \in \Delta_{IB}$ that they represent, it holds that $a_1 = a_2 \iff p_1 = p_2$.

(ii) A is convex. In particular, for any $a_1, a_2 \in A$ and the respective Intuitive Beliefs $p_1, p_2 \in \Delta_{IB}$ that they represent, and for any $\theta \in [0, 1]$, the network $a = \theta a_1 + (1-\theta)a_2$ represents Intuitive Beliefs given by the following normalized geometric mixture: for all $x \in \Omega$,

$$p_a(x) = \frac{p_1^{\theta}(x)p_2^{1-\theta}(x)}{\sum_{y\in\Omega} p_1(y)^{\theta}p_2(y)^{1-\theta}}.$$

Proof. (i) Necessity is obvious given (8). Conversely, suppose that $a_1 = a_2$. By (8), $p_1(x) \geq 0 \iff p_2(x) \geq 0$. Moreover, for all $x_i, x_j \in \cup_{k\in\Gamma}\Omega_k$ such that $p_1(x_i)p_1(x_j) > 0$ (equivalently, $p_2(x_i)p_2(x_j) > 0$), $a_1(x_ix_j) = a_2(x_ix_j)$ implies

$$p_1(\overline{z})^{\frac{N-3}{N-1}} \frac{p_1(x_ix_j\overline{z}_{-ij})}{[p_1(x_i\overline{z}_{-i})p_1(x_j\overline{z}_{-j})]^{\frac{N-2}{N-1}}} = p_2(\overline{z})^{\frac{N-3}{N-1}} \frac{p_2(x_ix_j\overline{z}_{-ij})}{[p_2(x_i\overline{z}_{-i})p_2(x_j\overline{z}_{-j})]^{\frac{N-2}{N-1}}}.$$

Taking $x_j = \overline{z}_j$, this implies $\frac{p_1(x_i\overline{z}_{-i})}{p_1(\overline{z})} = \frac{p_2(x_i\overline{z}_{-i})}{p_2(\overline{z})}$. Using this to simplify the above equality yields

$$\frac{p_1(x_ix_j\overline{z}_{-ij})}{p_1(\overline{z})} = \frac{p_2(x_ix_j\overline{z}_{-ij})}{p_2(\overline{z})}.$$

But then, by the reduced form established in Lemma 9, for any $x \in \Omega$ such that $p(x) > 0$,

$$\begin{aligned}
\frac{p_1(x)}{p_1(\overline{z})} &= \left[\prod_{i<j} \frac{p_1(x_ix_j\overline{z}_{-ij})p_1(\overline{z})}{p_1(x_i\overline{z}_{-i})p_1(x_j\overline{z}_{-j})}\right] \times \prod_{i\in\Gamma} \frac{p_1(x_i\overline{z}_{-i})}{p_1(\overline{z})} \\
&= \left[\prod_{i<j} \frac{\frac{p_1(x_ix_j\overline{z}_{-ij})}{p_1(\overline{z})}}{\frac{p_1(x_i\overline{z}_{-i})}{p_1(\overline{z})}\frac{p_1(x_j\overline{z}_{-j})}{p_1(\overline{z})}}\right] \times \prod_{i\in\Gamma} \frac{p_1(x_i\overline{z}_{-i})}{p_1(\overline{z})} \\
&= \left[\prod_{i<j} \frac{\frac{p_2(x_ix_j\overline{z}_{-ij})}{p_2(\overline{z})}}{\frac{p_2(x_i\overline{z}_{-i})}{p_2(\overline{z})}\frac{p_2(x_j\overline{z}_{-j})}{p_2(\overline{z})}}\right] \times \prod_{i\in\Gamma} \frac{p_2(x_i\overline{z}_{-i})}{p_2(\overline{z})} = \frac{p_2(x)}{p_2(\overline{z})}.
\end{aligned}$$

In particular, for all $x, y \in \Omega$ such that $p_1(y) > 0$ (equivalently, $p_2(y) > 0$) we have $\frac{p_1(x)}{p_1(y)} = \frac{p_2(x)}{p_2(y)}$ and therefore it must be that $p_1 = p_2$.

(ii) Take any $a_1, a_2 \in A$ and the respective Intuitive Beliefs $p_1, p_2 \in \Delta_{IB}$ with respective normalizing constants Z_1, Z_2. Consider some $a = \theta a_1 + (1-\theta)a_2$. The normalizing constant Z_a for a must satisfy

$Z_a = \sum_{y\in\Omega} exp[\sum_{i<j}[a(y_iy_j)]$

$= \sum_{y\in\Omega} exp[\sum_{i<j}[\theta a_1(y_iy_j) + (1-\theta)a_2(y_iy_j)]$

$= \sum_{y\in\Omega} \left[exp[\sum_{i<j} a_1(y_iy_j)]\right]^{\theta} \left[exp[\sum_{i<j}[a_2(y_iy_j)]\right]^{1-\theta}$, and so

$$Z_a = Z_1^{\theta} Z_2^{1-\theta} \times \sum_{y\in\Omega} p_1(y)^{\theta} p_2(y)^{1-\theta}.$$

Therefore $(a, 0)$ represents an Intuitive Belief p given by: for all $x \in \Omega$,
$p(x) = \frac{1}{Z_a} exp[\sum_{i<j}[a(x_ix_j)]$
$= \frac{1}{Z_1^{\theta} Z_2^{1-\theta} \sum_y p_1(y)^{\theta} p_2(y)^{1-\theta}} exp[\sum_{i<j}(\theta a_1(x_ix_j) + (1-\theta)a_2(x_ix_j))]$
$= \frac{1}{\sum_y p_1(y)^{\theta} p_2(y)^{1-\theta}} \frac{exp[\theta \sum_{i<j} a_1(x_ix_j)]}{Z_1^{\theta}} \times \frac{exp[(1-\theta)\sum_{i<j} a_2(x_ix_j)]}{Z_2^{1-\theta}}$
$= \frac{p_1^{\theta}(x) p_2^{1-\theta}(x)}{\sum_y p_1(y)^{\theta} p_2(y)^{1-\theta}}$, as desired. It also follows that A is convex. ■

Now fix $\phi \neq D \subset \Omega$ and a reference state $\overline{z} \in D$, and consider the corresponding set of representations A. For any $a \in A$ let $Z_a(D) := \sum_{y\in D} exp[\sum_{i<j} a(y_iy_j)]$. Define a function $KL_D : \Delta(\Omega) \times A \to \mathbb{R}$ by

$$KL_D(q||a) = q(D) ln Z_a(D) + \sum_{x\in D} q(x) \left[ln q(x) - [\sum_{i<j} a(x_ix_j)]\right]. \tag{9}$$

Lemma 12 *For any a and corresponding p_a, the belief p_a is a minimizer of the D-training problem if and only if a solves*

$$min_{a\in A} KL_D(q||a). \tag{10}$$

Proof. For any $a \in A$ and corresponding $p_a \in \Delta_{IB}$, denote the D-conditional Intuitive Belief by $p_a^D(x) = \frac{p_a(x)}{\sum_{y\in D} p_a(y)} = \frac{1}{Z_a(D)} exp[\sum_{i<j} a(x_ix_j)]$ for all $x \in D$, where $Z_a(D) = \sum_{y\in D} exp[\sum_{i<j} a(y_iy_j)]$. Denote the D-conditional objective distribution by q^D. The D-training problem involves $KL(q^D||p_a^D)$, given by. However we can define $KL : \Delta(\Omega) \times A \to \mathbb{R} \cup \{\infty\}$ by

$$KL(q^D||p_a^D) = \sum_{x\in D} q^D(x)[ln q^D(x) - ln p_a^D(x)]$$

$$= \sum_{x\in D} \frac{q(x)}{q(D)} [ln \frac{q(x)}{q(D)} - ln[\frac{1}{Z_a(D)} exp[\sum_{i<j} a(x_ix_j)]]]$$

$$= \sum_{x\in D} \frac{q(x)}{q(D)} [ln q(x) - [\sum_{i<j} a(x_ix_j)]] + \sum_{x\in D} \frac{q(x)}{q(D)} [-ln q(D) + ln Z_a(D)]$$

$$= \sum_{x\in D} \frac{q(x)}{q(D)} [ln q(x) - [\sum_{i<j} a(x_ix_j)]] + [-ln q(D) + ln Z_a(D)].$$

which is ordinally equivalent to $KL_D(q||a)$ when viewed as a function of a. Therefore choosing p_a to minimize $KL(q^D||p_a^D)$ is equivalent to choosing a to minimize $KL_D(q||a)$. ■

Lemma 13 *(i) The function $a \mapsto KL_D(q||a)$ is convex.*
(ii) The function is strictly convex under assumption 2.

Proof. (i) Consider networks $a_1, a_2 \in A$ with respective Intuitive Beliefs $p_1, p_2 \in \Delta_{IB}$. Take any $\theta \in [0, 1]$ and consider $a = \theta a_1 + (1-\theta)a_2$. Define $M_{\theta} := \frac{Z_a(D)}{Z_{a_1}(D)^{\theta} Z_{a_2}(D)^{1-\theta}}$ and observe that

$$KL_D(q||a) = q(D) ln Z_a(D) + \sum_{x\in D} q(x)[ln q(x) - [\sum_{i<j} a(x_ix_j)]]$$

$$= \sum_{x\in D} q(x)[lnq(x) - [\theta \sum_{i<j} a_1(x_i x_j) + (1-\theta)\sum_{i<j} a_2(x_i x_j)]] + q(D)lnZ_a(D)$$

$$= \sum_{x\in D} q(x)[\theta[lnq(x) - \sum_{i<j} a_1(x_i x_j)] + (1-\theta)[lnq(x) - \sum_{i<j} a_2(x_i x_j)]] + q(D)lnZ_a(D)$$

$$= \theta KL_D(q||a_1) + (1-\theta)KL_D(q||a_2) + [q(D)lnZ_a(D) - \theta q(D)lnZ_{a_1}(D) - (1-\theta)q(D)lnZ_{a_1}(D)]$$

$$= \theta KL_D(q||a_1) + (1-\theta)KL_D(q||a_2) + q(D)ln\frac{Z_a(D)}{Z_{a_1}(D)^{\theta} Z_{a_2}(D)^{1-\theta}}$$

$$= \theta KL_D(q||a_1) + (1-\theta)KL_D(q||a_2) + q(D)lnM_{\theta}.$$

So the desired convexity is established once we show that

$$M_{\theta} \leq 1.$$

To determine that this inequality holds, first compute that
$Z_a(D) := \sum_{y\in D} exp[\sum_{i<j} a(y_i y_j)]$
$= \sum_{y\in D} exp[\theta \sum_{i<j} a_1(y_i y_j) + (1-\theta)\sum_{i<j} a_2(y_i y_j)]$
$= \sum_{y\in D} exp[\theta \sum_{i<j} a_1(y_i y_j)] \times exp[(1-\theta)\sum_{i<j} a_2(y_i y_j)]$
$= Z_{a_1}(D)^{\theta} Z_{a_2}(D)^{1-\theta} \sum_{y\in D} p_1^D(y)^{\theta} p_2^D(y)^{\theta}$, and so

$$M_{\theta} = \frac{Z_a(D)}{Z_{a_1}(D)^{\theta} Z_{a_2}(D)^{1-\theta}} = \sum_{y\in D} p_1^D(y)^{\theta} p_2^D(y)^{\theta}.$$

Trivially, $M_{\theta} \leq 1$ if $\theta \in \{0,1\}$ or if $\theta \in (0,1)$ and $p_1^D = p_2^D$. If $\theta \in (0,1)$ and $p_1^D \neq p_2^D$ (in which case $\frac{p_1^D(y)}{p_2^D(y)}$ cannot be constant across $y \in D$) then applying Jensen's inequality for strictly concave functions yields:
$M_{\theta} = \sum_{y\in D} p_1^D(y)^{\theta} p_2^D(y)^{1-\theta} = \sum_{y\in D}[\frac{p_1^D(y)}{p_2^D(y)}]^{\theta} p_2^D(y)$
$< [\sum_y \frac{p_1(y)}{p_2(y)} p_2(y)]^{\theta} = [\sum_y p_1(y)]^{\theta} = 1$, and so, $M_{\theta} < 1$. This establishes that $M_{\theta} \leq 1$, and thus, $KL_D(q||a)$ is convex.

(ii) By the preceding, we see that $KL_D(q||a)$ is strictly convex in a if $M_{\theta} < 1$ for all $\theta \in (0,1)$ and $a_1 \neq a_2$. We also showed that for any $\theta \in (0,1)$ and $a_1 \neq a_2$, if we have $p_1^D \neq p_2^D$ then it must be that $M_{\theta} < 1$. It follows that a sufficient condition for $KL_D(q||a)$ to be strictly convex in a is that

$$a_1 \neq a_2 \implies p_1^D \neq p_2^D.$$

We show that this sufficient condition holds under assumption 2.

Consider any $a_1 \neq a_2$ and suppose by way of contradiction that $p_1^D = p_2^D$. For $r = 1, 2$, the corresponding Intuitive Beliefs for a_r are given for all $x \in D$ by

$$p_r^D(x) = \frac{1}{Z_{a_r}(D)} exp[\sum_{i<j} a_r(x_i x_j)],$$

where $Z_{a_r}(D) := \sum_{y\in D} exp[\sum_{i<j} a_r(y_i y_j)]$. Since $a_r(\overline{z}_i \overline{z}_j) = 0$ for all i, j, the hypothesis implies $\frac{1}{Z_{a_1}(D)} = p_1^D(\overline{z}) = p_2^D(\overline{z}) = \frac{1}{Z_{a_2}(D)}$, that is,

$$Z_{a_1}(D) = Z_{a_2}(D).$$

However, $a_1 \neq a_2$ implies that there exists $x_i x_j$ s.t. $a_1(x_i x_j) \neq a_2(x_i x_j)$. By definition, the domain of a_1, a_2 are E_D, the set of pairs (y_i, y_j) that appear in D. Therefore, there must exist some $x_i x_j z_{-ij} \in D$, and in particular by assumption 2, it must be that $x_i x_j \bar{z}_{-ij} \in D$ such that $a_1(x_i x_j) \neq a_2(x_i x_j)$. But then by the preceding,

$$p_1^D(x_i x_j \bar{z}_{-ij}) = \frac{1}{Z_{a_1}(D)} exp[a_1(x_i x_j)] \neq \frac{1}{Z_{a_2}(D)} exp[a_2(x_i x_j)] = p_2^D(x_i x_j \bar{z}_{-ij}),$$

a contradiction. ■

Lemma 14 *Intuitive Belief p is a solution to D-training problem if and only if $p(x_i x_j|D) = q(x_i x_j|D)$ for all distinct $i, j \in \Gamma$ and $x \in D$.*

Proof. By Lemma 12, it suffices to solve (10). Since the optimization problem is unconstrained, the Lagrangian is

$$\mathcal{L}(a) = KL_D(q||a) = q(D) ln Z_a + \sum_{x \in D} q(x) \left[ln q(x) - \sum_{i<j} a(x_i x_j) \right],$$

As noted in Lemmas 11 and 13, A is a convex set and the function $a \mapsto KL_D(q||a)$ is convex. Consequently, the first order conditions are both necessary and sufficient for optimality. It follows that a minimizer exists if and only if it satisfies the first order conditions, which we derive below.

For any $x_i, x_j \in \cup_{k \in \Gamma} \Omega_k$ denote by $D_{x_i x_j}$ all the states $z \in D$ that take on values $z_i = x_i$ and $z_j = x_j$ on dimension i and j respectively. The first order condition wrt any $a(x_i x_j)$, where $i, j \in \Gamma$ are distinct and $x_i, x_j \in \cup_{k \in \Gamma} \Omega_k$, yields:

$\frac{\partial \mathcal{L}}{\partial a(x_i x_j)} = 0$

$\iff q(D) \frac{\sum_{w \in D_{x_i x_j}} exp[\sum_{k<l} a(w_k w_l)]}{\sum_{y \in D} exp[\sum_{k<l} a(y_k y_l)]} + \left[\sum_{z \in D_{x_i x_j}} q(z)[0-1] \right] = 0$

$\iff q(D) \frac{\sum_{w \in D_{x_i x_j}} exp[\sum_{k<l} a(w_k w_l)]}{\sum_{y \in D} exp[\sum_{k<l} a(y_k y_l)]} = \sum_{z \in D_{x_i x_j}} q(z)$

$\iff \sum_{w \in D_{x_i x_j}} p(z|D) = \sum_{z \in D_{x_i x_j}} q(z|D)$

$\iff p(x_i x_j|D) = q(x_i x_j|D)$, which yields the conclusion that the 2-dimensional marginals must match, if a solution exists. This completes the proof. ■

Lemma 15 *Under assumption 2, if a solution to the D-training problem exists, then it is unique in the class of minimal solutions.*

Proof. Under assumption 2 and given Lemmas 11 and 13, we have a minimization problem (10) where A is convex and the objective function $a \mapsto KL_D(q||a)$ is strictly convex. It follows that if a solution exists, then there exists a unique minimal $p \in \Delta_{IB}$ that solves the D-training problem (10). ■

G Appendix: Proof of Proposition 3

Suppose assumption 3 holds and recall the set A of representations used in the proof of Theorem 4. We show that there exists a minimal network $a \in A$ (with corresponding Intuitive Belief p_a) such that

$$KL(q^D||p_a^D) = \sum_{x \in D} q^D(x)[ln q^D(x) - ln p_a^D(x)] = 0.$$

This is achieved by finding a network $a \in A$ such that $x \in D \implies \frac{p_a(x)}{p_a(D)} = \frac{q(x)}{q(D)}$. We in fact establish a stronger condition:

$$x \in D \implies exp[\sum_{i<j} a(x_i x_j)] = \frac{q(x)}{q(\overline{z})}. \tag{11}$$

To see that this is stronger, consider the normalizing constant $Z_a(D) = \sum_{y \in D} exp[\sum_{i<j} a(y_i y_j)]$ and observe that the stronger condition implies that $\frac{p_a(x)}{p_a(D)} := \frac{1}{Z_a(D)} exp[\sum_{i<j} a(x_i x_j)] = \frac{q(x)}{Z_a(D) q(\overline{z})}$, and summing across $x \in D$ yields $1 = \frac{q(D)}{Z_a(D) q(\overline{z})}$, that is, $Z_a(D) q(\overline{z}) = q(D)$. Combining this with $\frac{p_a(x)}{p_a(D)} = \frac{q(x)}{Z_a(D) q(\overline{z})}$ yields $\frac{p(x)}{p(D)} = \frac{q(x)}{q(D)}$ as desired.

We search for a network in the class A of minimal networks. Recall that each $a \in A$ is normalized so that $a(\overline{z}_i \overline{z}_j) = 0$ and $a(x_i \overline{z}_j) = a(x_i \overline{z}_k)$ for all distinct $i, j, k \in \Gamma$ and all $x \in \Omega$.

Lemma 16 *The desired equalities (11) are solved by the network $a \in A$ defined by: for any $x_i x_j$ that appears in D,*

$$exp[a(x_i x_j)] = \frac{q(x_i x_j \overline{z}_{-ij})}{[q(x_i \overline{z}_{-i})]^{\frac{N-2}{N-1}} [q(x_j \overline{z}_{-j})]^{\frac{N-2}{N-1}}} q(\overline{z})^{\frac{N-3}{N-1}}.$$

Proof. For $\overline{z} \in D$ the desired equality yields $exp[\sum_{i<j} a(\overline{z}_i \overline{z}_j)] = \frac{q(\overline{z})}{q(\overline{z})} = 1$, which is satisfied by any $a \in A$ due to the normalization $a(\overline{z}_i \overline{z}_j) = 0$ for all i, j.

Since $a(x_i \overline{z}_j) = a(x_i \overline{z}_k)$ for all distinct $i, j, k \in \Gamma$, let us write $a(x_i \overline{z}_j) = \alpha(x_i)$ for all $i, j \in \Gamma$ and $x_i \neq \overline{z}_i$. For $(x_i \overline{z}_{-i}) \in D$ with $x_i \neq \overline{z}_i$ the desired equality yields

$\iff exp[\sum_{i \neq j \in \Gamma} \alpha(x_i) + 0] = \frac{q(x_i \overline{z}_{-i})}{q(\overline{z})}$

$\iff exp[(N-1)\alpha(x_i)] = \frac{q(x_i \overline{z}_{-i})}{q(\overline{z})}$

$\iff exp[\alpha(x_i)] = \left[\frac{q(x_i \overline{z}_{-i})}{q(\overline{z})}\right]^{\frac{1}{N-1}}$.

Finally, for $(x_i x_j \overline{z}_{-ij}) \in D$ with $x_i \neq \overline{z}_i$ and $x_j \neq \overline{z}_j$, we obtain

$exp[a(x_i x_j) + \sum_{i,j \neq k \in \Gamma} \alpha(x_i) + \sum_{i,j \neq k \in \Gamma} \alpha(x_j) + 0] = \frac{q(x_i x_j \overline{z}_{-ij})}{q(\overline{z})}$

$\iff exp[a(x_i x_j)] exp[(N-2)\alpha(x_i)] exp[(N-2)\alpha(x_j)] = \frac{q(x_i x_j \overline{z}_{-ij})}{q(\overline{z})}$

$\iff exp[a(x_i x_j)] \left[\frac{q(x_i \overline{z}_{-i})}{q(\overline{z})}\right]^{\frac{N-2}{N-1}} \left[\frac{q(x_j \overline{z}_{-j})}{q(\overline{z})}\right]^{\frac{N-2}{N-1}} = \frac{q(x_i x_j \overline{z}_{-ij})}{q(\overline{z})}$

$\iff exp[a(x_i x_j)] = \frac{q(x_i x_j \overline{z}_{-ij})}{[q(x_i \overline{z}_{-i})]^{\frac{N-2}{N-1}} [q(x_j \overline{z}_{-j})]^{\frac{N-2}{N-1}}} q(\overline{z})^{\frac{N-3}{N-1}}$. This completes the proof. ∎

The preceding lemma specifies the values of a network $(a, 0)$ for the pairs of elementary states $x_i x_j$ that appear in D. By minimality, all other pairs of elementary states must have value $a(x_i x_j) = -\infty$. This fully defines $a \in A$, yielding a minimal Intuitive Belief p over Ω.

H Proof of Proposition 4

The first two claims are a corollary of Proposition 3. For the final claim, we derive the reduced form for p conditioned on $\Omega \subset \Omega^*$.

Lemma 17 *For any $x \in \Omega_{D^*} \cap \Omega$, the belief $p(x)$ can be written as:*

$$p(x) = \frac{1}{W} \times \left[\prod_{i<j} \frac{q(x_i x_j)}{q(x_i) q(x_j)}\right] \times \left[\prod_{i \in I} q(x_i)\right].$$

Proof. For any $(x_i x_j) \in \Omega_{ij}$, note that our above expression $exp[a(x_i x_j)] = \frac{q^*(x_i x_j z^*_{-i})}{q^*(x_i z^*_{-i})^{\frac{N-2}{N-1}} q(x_j z^*_{-j})^{\frac{N-2}{N-1}}} q(z^*)^{\frac{(N-2)}{N}}$ is similar (up to a scalar multiple) of the expression $exp[a_{normalized}(x_i x_j)] = p(\overline{z})^{\frac{N-3}{N-1}} \frac{p(x_i x_j \overline{z}_{-ij})}{[p(x_i \overline{z}_{-i}) p(x_j \overline{z}_{-j})]^{\frac{N-2}{N-1}}}$ established in Lemma 8. Arguing exactly as in Lemma 9 while using only $x \in \Omega_{D^*} \cap \Omega$ yields a reduced form:

$$p(x) = \frac{1}{Z} \times \left[\prod_{i<j} \frac{q^*(x_i x_j z^*_{-ij}) q^*(z^*)}{q^*(x_i z^*_{-i}) q^*(x_j z^*_{-j})} \right] \times \left[\prod_{i \in \Gamma} \frac{q^*(x_i z^*_{-j})}{q^*(z^*)} \right]$$

with an appropriately defined Z (which absorbs the terms that appear due to the noted scalar multiple). Inserting the expressions for q

$$p(x) = \frac{1}{Z} \times \left[\prod_{i<j} \frac{\sigma^2(1-\sigma)^{N-2} q(x_i x_j) \times (1-\sigma)^N}{\sigma(1-\sigma)^{N-1} q(x_i) \times \sigma(1-\sigma)^{N-1} q(x_j)} \right] \times \left[\prod_{i \in I} \frac{\sigma(1-\sigma)^{N-1} q(x_i)}{(1-\sigma)^N} \right]$$

$$= \frac{1}{W} \times \left[\prod_{i<j} \frac{q(x_i x_j)}{q(x_i) q(x_j)} \right] \times \left[\prod_{i \in I} q(x_i) \right],$$

where W absorbs the terms involving σ. This completes the proof. ■

I Appendix: Small Probabilities

We present a general result that subsumes the observations in Section 5.2. Fix any $\phi \neq E \subset \Omega$. Say that elementary state x_i *appears in* E if there is a state $z \in \Omega$ such that $x_i z_{-i} \in E$. By definition, for each $z \in E$, the elementary state z_i appears in E for all N dimensions $i \in \Gamma$. Consider all states constructed using $x_i x_j$ that jointly appear in E, in the sense that there is a state $z \in \Omega$ s.t. $x_i x_j z_{-ij} \in E$:

$$\Omega_E = \{x \in \Omega : \forall i, j \in \Gamma, \exists z \in \Omega \text{ s.t. } x_i x_j z_{-ij} \in E\}.$$

Clearly, $E \subset \Omega_E$. Consider next those states $x \notin E$ for which x_i appears in E for only $N-1$ dimensions:

$$F = \{x \in \Omega : x_i \text{ appears in } E \text{ in only } N-1 \text{ dimensions } i \in \Gamma\}.$$

To illustrate, if $E = \{z\}$ then F consists of z and all states x that deviate from z in exactly one dimension. If every elementary state appears in E then $F = \Omega$. Finally, consider all remaining states: $G = \Omega \backslash (\Omega_E \cup F)$.

Consider a full support distribution q. For each $\alpha \in (0,1)$, consider Intuitive Beliefs p_α that are trained by q_α given by: for all $x \in \Omega$,

$$q_\alpha(x) = \alpha q(x) + (1-\alpha) q(x|E).$$

Proposition 7 *If Intuitive Beliefs p_α are trained by q_α for each α then the following properties hold:*

(a) $0 < lim_{\alpha \to 0} \frac{p_\alpha(\Omega_E)}{p_\alpha(F)} = \infty$.

(b) If $G \neq \phi$ then $lim_{\alpha \to 0} \frac{p_\alpha(E \cup F)}{p_\alpha(G)} < \infty$.

Proof. If x_i does not appear in E, then for any x_j,

$$q_\alpha(x_i x_j) = \alpha q(x_i x_j) \text{ and } q_\alpha(x_i) = \alpha q(x_i).$$

If z_i and z_j appear in E, then

$$q_\alpha(z_i z_j) = \alpha q(z_i z_j) + (1-\alpha) q(z_i z_j|E), \text{ and } q_\alpha(z_i) = \alpha q(z_i) + (1-\alpha) q(z_i|E).$$

We use these marginals to compute objective PMI and their limit as $\alpha \to 0$: If neither x_i nor x_j appear in E, and we have

$$\frac{q_\alpha(x_i x_j)}{q_\alpha(x_i) q_\alpha(x_j)} = \frac{\alpha q(x_i x_j)}{\alpha q(x_i) \alpha q(x_j)} = \frac{1}{\alpha} \frac{q(x_i x_j)}{q(x_i) q(x_j)} \to \infty.$$

If z_i and z_j appear in E then

$$\frac{q_\alpha(z_i z_j)}{q_\alpha(z_i) q_\alpha(z_j)} = \frac{\alpha q(z_i z_j) + (1-\alpha) q(z_i z_j|E)}{[\alpha q(z_i) + (1-\alpha) q(z_i|E)][\alpha q(z_{2j}) + (1-\alpha) q(z_j|E)]} \to \frac{q(z_i z_j|E)}{q(z_i|E) q(z_j|E)} \in (0, \infty)$$

Finally, if z_i appears in E and x_j does not, then

$$\frac{q_\alpha(z_i x_j)}{q_\alpha(z_i) q_\alpha(x_j)} = \frac{\alpha q(z_i x_j)}{[\alpha q(z_i) + (1-\alpha) q(z_i|E)] \alpha q(x_j)} = \frac{q(z_i x_j)}{[\alpha q(z_i) + (1-\alpha) q(z_i|E)] q(x_j)} \to \frac{q(z_i x_j)}{q(z_i|E) q(x_j)} \in (0, \infty).$$

By definition of a q_α-trained Intuitive Belief p_α, there exists a scalar Z_α such that for all events $A \subset \Omega$:

$$p_\alpha(A) = \frac{1}{Z_\alpha} \sum_{x \in A} \left[\prod_{i<j} \frac{q_\alpha(x_i x_j)}{q_\alpha(x_i) q_\alpha(x_j)} \right] \times \prod_{i \in \Gamma} q_\alpha(x_i).$$

For each $x \in \Omega_E$ and each i, j, it is trivially the case that x_i, x_j and $x_i x_j$ appear in E. Therefore, both the PMI and marginals converge to a strictly positive finite limit, and in particular, the limit of $Z_\alpha p_\alpha(\Omega_E)$ as $\alpha \to 0$ must be strictly positive and finite:

$$lim_{\alpha \to 0} Z_\alpha p_\alpha(\Omega_E) = lim_{\alpha \to 0} \sum_{x \in \Omega_E} \left[\prod_{i<j} \frac{q_\alpha(x_i x_j)}{q_\alpha(x_i) q_\alpha(x_j)} \right] \times \prod_{i \in \Gamma} q_\alpha(x_i) \in (0, \infty).$$

On the other hand, $lim_{\alpha \to 0} Z_\alpha p_\alpha(F) = 0$: for each $x \in F$, although all PMI terms converge to a strictly positive limit, there is an elementary state x_i that does not appear in E and so $q_\alpha(x_i) \to 0$.

This establishes the first desired claim since

$$lim_{\alpha \to 0} \frac{p_\alpha(\Omega_E)}{p_\alpha(F)} = lim_{\alpha \to 0} \frac{Z_\alpha p_\alpha(\Omega_E)}{Z_\alpha p_\alpha(F)} = \frac{lim_{\alpha \to 0} Z_\alpha p_\alpha(\Omega_E)}{lim_{\alpha \to 0} Z_\alpha p_\alpha(F)} = \infty.$$

Next suppose $G \neq \phi$. We verify that the limit $lim_{\alpha \to 0} Z_\alpha p_\alpha(G)$ must be strictly positive. Take any $x^* \in G$ such that $p(x^*) > 0$ (which exists since $G \neq \phi$ and p_α has full support). Then

$$\begin{aligned} & lim_{\alpha \to 0} Z_\alpha p_\alpha(G) \\ &= lim_{\alpha \to 0} \sum_{x \in G} \left[\prod_{i<j} \frac{q_\alpha(x_i x_j)}{q_\alpha(x_i) q_\alpha(x_j)} \right] \times \prod_{i \in \Gamma} q_\alpha(x_i) \\ &\geq lim_{\alpha \to 0} \left[\prod_{i<j} \frac{1}{\alpha} \frac{q(x_i^* x_j^*)}{q(x_i^*) q(x_j^*)} \right] \times \prod_{i \in \Gamma} \alpha q(x_i^*) \\ &= lim_{\alpha \to 0} \left[\frac{\prod_i \alpha}{\prod_{i<j} \alpha} \right] \left[\prod_{i<j} \frac{q(x_i^* x_j^*)}{q(x_i^*) q(x_j^*)} \right] \times \prod_{i \in \Gamma} q(x_i^*) \end{aligned}$$

$$= lim_{\alpha\to 0}\left[\frac{1}{\alpha^{N(N-3)}}\right] Zp(x^*) > 0.$$

where $\frac{\prod_i \alpha}{\prod_{i<j}\alpha} = \frac{1}{\alpha^{N(N-3)}}$ since the product $\prod_{i<j}$ involves $\frac{N!}{2!(N-2)!} = \frac{N(N-1)}{2}$ combinations, so that $\frac{\prod_i \alpha}{\prod_{i<j}\alpha} = \frac{1}{\alpha^{\frac{N(N-1)}{2}-N}} = \frac{1}{\alpha^{N(N-3)}}$. Therefore, we conclude that

$$lim_{\alpha\to 0}\frac{p_\alpha(\Omega_E)}{p_\alpha(G)} = lim_{\alpha\to 0}\frac{Z_\alpha p_\alpha(\Omega_E)}{Z_\alpha p_\alpha(G)} < \infty.$$

In fact, when $N > 3$ we see from the above expressions that $lim_{\alpha\to 0} Z_\alpha p_\alpha(G) \geq lim_{\alpha\to 0}\left[\frac{1}{\alpha^{N(N-3)}}\right] Zp(x^*) = \infty$, and so $lim_{\alpha\to 0}\frac{p_\alpha(\Omega_E)}{p_\alpha(G)} = lim_{\alpha\to 0}\frac{Z_\alpha p_\alpha(\Omega_E)}{Z_\alpha p_\alpha(G)} = 0$. ■

As a corollary we observe a pathological possibility that arises when the complexity is high in the sense that $N > 3$.

Corollary 1 *Suppose $G \neq \phi$ and $N > 3$ then $lim_{\alpha\to 0}\frac{p_\alpha(\Omega_E)}{p_\alpha(G)} = 0$.*

That is, although at $\alpha = 0$ Intuitive Beliefs will regard G as impossible, for α close to 0 the possibility of states in G is so salient that the agent regards it as almost certain. This is an artifact of the way we defined q_α as a linear combination. For any $x \in G$ and any i, j, the marginals $q_\alpha(x_i)$, $q_\alpha(x_j)$ and $q_\alpha(x_i x_j)$ go to 0 uniformly, leading to an infinitely high PMI in the limit: $\frac{q_\alpha(x_i x_j)}{q_\alpha(x_i)q_\alpha(x_j)} = \frac{\alpha q(x_i x_j)}{\alpha q(x_i)\alpha q(x_j)} = \frac{1}{\alpha}\frac{q(x_i x_j)}{q(x_i)q(x_j)} \to \infty$.

References

[1] Ackley, D., G. Hinton and T. Sejnowski (1985): "A Learning Algorithm for Boltzmann Machines", *Cognitive Science* 9(1), pp. 147–169.

[2] Anderson, J. R. (1983): "A Spreading Activation Theory of Memory", *Journal of Verbal Learning and Verbal Behavior* 22(3), pp. 261-295.

[3] Argenziano, R. and I. Gilboa (2019): "Second-Order Induction in Prediction Problems", *Proceedings of the National Academy of Sciences* 116 (21): pp. 10323–10328.

[4] Benjamin, D. (2019): "Errors in Probabilistic Reasoning and Judgment Biases." Chapter for the *Handbook of Behavioral Economics* (edited by Doug Bernheim, Stefano DellaVigna, and David Laibson). Elsevier Press.

[5] Benjamin, D., A. Bodoh-Creed, and M. Rabin (2019): "Base-Rate Neglect: Foundations and Implications", working paper.

[6] Benjamin, D., M. Rabin and C. Raymond (2016): "A Model of Non-Belief in the Law of Large Numbers", *Journal of the European Economic Association* 14(2), pp 515-544.

[7] Betch, T. (2007): "The Nature of Intuition and Its Neglect in Research on Judgment and Decision Making", in E. Plessner, C. Betsch and T. Betsch (Eds.), *Intuition in Judgement and Decision Making*, Psychology Press; 1st edition.

[8] Billot, A., I. Gilboa, D. Samet, and D. Schmeidler (2005): "Probabilities as Similarity-Weighted Frequencies", *Econometrica* 73(4), pp. 1125-1136.

[9] Bodoh-Creed, A. (2020): "Mood, Memory and the Evaluation of Asset Prices", *Review of Finance* 24(1), pp. 227–262.

[10] Bordalo, P., N. Gennaioli, and A. Shleifer (2019): "Memory, Attention and Choice", working paper.

[11] Brunnerimeier, M. (2009), "Deciphering the Liquidity and Credit Crunch 2007- 2008", *Journal of Economic Perspectives* 23, pp. 77–100.

[12] Collins, A. and E. Loftus (1975): "A Spreading-Activation Theory of Semantic Processing", *Psychological Review* 82(6), pp. 407–428.

[13] Ellis, A. and M. Piccione (2017): "Correlation Misperception in Choice", *American Economic Review* 107(4), pp. 1264-92.

[14] Enke, B. and F. Zimmerman (2013): "Correlation Neglect in Belief Formation", *Review of Economic Studies* 86(1), pp. 313–332.

[15] Eyster, E. and G. Weizsäcker (2016): "Correlation Neglect in Financial Decision Making: Lab Evidence." Mimeo.

[16] Gennaioli, N. and A. Shleifer (2010): "What Comes to Mind", *Quarterly Journal of Economics* 125 (4), pp. 1399-1433.

[17] Gilboa, I. and D. Schmeidler (2001): *A Theory of Case-Based Decisions*, Cambridge University Press.

[18] Gilboa, I. and D. Schmeidler (2003): "Inductive Inference: An Axiomatic Approach", *Econometrica* 71(1), pp. 1-26.

[19] Hensman, A. and E. Sadler-Smith (2011): "Intuitive Decision-Making in Banking and Finance", *European Management Journal* 29, pp. 51–66.

[20] Hinton, G. and T. Sejnowski (1983): "Optimal Perceptual Inference", *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, Washington DC, pp. 448-453.

[21] Hopfield, J. J. (1982): "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", *Proceedings of the National Academy of Sciences* 79 (8), pp. 2554–2558.

[22] Huang, L. and J. Pearce (2015): "Managing the Unknowable: The Effectiveness of Early-stage Investor Gut Feel in Entrepreneurial Investment Decisions", *Administrative Science Quarterly* 60(4), pp. 634-670.

[23] Huang, L. (2018): "The Role of Investor Gut Feel in Managing Complexity and Extreme Risk", *Academy of Management Journal* 61(5), pp. 1821–1847.

[24] Kahana, M. (2020): "Computational Models of Memory Search", *Annual Review of Psychology* 71, pp. 107–38.

[25] Kahneman, D. and A. Tversky (1972): "Subjective Probability: A Judgement of Representativeness", *Cognitive Psychology* 3(3), pp. 430-454.

[26] Keynes, J. (1936). *The General Theory of Employment, Interest and Money.* London. Macmillan.

[27] Krajbich, I. and A. Rangel (2011): "Multialternative Drift-Diffusion Model Predicts the Relationship Between Visual Fixations and Choice in Value-Based Decisions", *Proceedings of the National Academy of Sciences* 108 (33), pp. 13852-13857.

[28] Malmendier, U. and S. Nagel (2016): "Learning from Inflation Experiences", *Quarterly Journal of Economics* 131, pp. 53-87.

[29] Moore, D. and P. Healy (2008). “The Trouble with Overconfidence”, *Psychological Review* 115(2), pp. 502–517.

[30] Morewedge, C. and D. Kahneman (2010): “Associative Processes in Intuitive Judgment”, *Trends in Cognitive Science* 14(10), pp. 435–440.

[31] Mullainathan, S. (2002): “A Memory-Based Model of Bounded Rationality”, *Quarterly Journal of Economics* 117(3), pp. 735-774.

[32] Noor, J. (2019): “Intuitive Beliefs”, mimeo.

[33] Rabin, M. (2002), “Inference by Believers in the Law of Small Numbers”, *Quarterly Journal of Economics* 117, pp. 775–816.

[34] Rabin, M. and D. Vayanos (2010): “The Gambler’s and Hot-Hand Fallacies: Theory and Applications”, *Review of Economic Studies* 77, pp. 730–778.

[35] Russo, J. E., and L. D. Rosen (1975): “An Eye Fixation Analysis of Multi-Alternative Choice”, *Memory and Cognition* 3, pp. 267-276.

[36] Salas, E., M. Rosen, D. Diaz-Granados (2010): “Expertise-Based Intuition and Decision Making in Organizations”, *Journal of Management* 36 (4), pp. 941–973.

[37] Savage, L. (1954): *The Foundations of Statistics*. John Wiley & Sons, Inc.

[38] Schotter, A. and I. Trevino (2014): “Belief Elicitation in the Laboratory”, *Annual Review of Economics* 6, pp. 103-128.

[39] Shefrin, H., & Statman, M. (1985). “The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence”, *The Journal of Finance* 40(3), pp. 777-790.

[40] Simon, H. (1995): “Explaining the Ineffable: AI on the Topics of Intuition, Insight and Inspiration”. *Proceedings of the International Joint Conferences on Artificial Intelligence*-95, pp 939–948.

[41] Spiegler, R. (2016): “Bayesian Networks and Boundedly Rational Expectations", *Quarterly Journal of Economics* 131, pp. 1243-1290.

[42] Tversky, A. and D. Kahneman (1974), “Judgement under Uncertainty: Heuristics and biases”, *Science* 185, pp. 1124–1131.

[43] Wasserman, E. and R. Miller (1997): “What’s Elementary About Associative Learning?” *Annual Review of Psychology* 48, pp. 573-607.