INTUITIVE BELIEFS

By

Jawwad Noor

December 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 2216

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Intuitive Beliefs*

Jawwad Noor

December 24, 2019

Abstract

Beliefs are intuitive if they rely on associative memory, which can be described as a network of associations between events. A belief-theoretic characterization of the model is provided, its uniqueness properties are established, and the intersection with the Bayesian model is characterized. The formation of intuitive beliefs is modelled after machine learning, whereby the network is shaped by past experience via minimization of the difference from an objective probability distribution. The model is shown to accommodate correlation misperception, the conjunction fallacy, base-rate neglect/conservatism, etc.

JEL classification: C45, D01, D90

Keywords: Belief formation, Intuition, Associations, Associative memory, Boltzmann machine, Energy-Based Neural Networks, Non-Bayesian updating, Heuristics and Biases, System 1.

1 Introduction

How do people make decisions in complex environments? Echoing the "spontaneous urge to action" that underlies Keynes' notion of animal spirits (Keynes 1936), studies

***Noor is at the Department of Economics, Boston University, 270 Bay State Road, Boston MA 02215. Email: jnoor@bu.edu.** **The author acknowledges valuable comments from Roland Benabou, Yi-Hsuan Lin and Bart Lipman at earlier stages of this project and helpful feedback from seminar audiences at University of Paris 2, Yale, Duke, Hong Kong University, UC-Davis, UCLA, NYU-AD, Erasmus University 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) and the Manchester Economic Theory Workshop. Expert research assistance has been provided by Alonso Ahumada, Kevin Cooke, Enrico De Magistris, Yi-Hsuan Lin and Nick Saponara. 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.**

show that investors often describe gut-feelings as an important factor in decision-making (Salas et al 2010, Hensman and Sadler-Smith 2011, Huang and Pearce 2015, Huang 2018). If agents in complex environments utilize gut feelings alongside deliberative analysis to form judgements and make decisions, then such non-deliberative, intuitive processes may possibly play an important role in observed economic behavior. In this paper we provide a formal theory of intuition.

We build the theory on the classic psychological notion of *associations*. Associations are connections in the mind that exist between mental images of events, whereby the activation of one image activates (or inhibits) the other. A simple illustration is word association, where a given word (say "stop") causes another word (say "red light") to appear in the mind. Stereotyping is another manifestation of associations. An extreme example of an association is a phobia, such as when the sight of a spider is associated with a sense of extreme danger. Philosophers and psychologists have extensively studied how associative connections can be formed through frequency, salience and similarity, and can be strengthened through reinforcement, or weakened through counter-conditioning or through decay over time. A central observation is that the triggering of associations is an *involuntary* process that entails no cognitive effort. In this regard, associative thinking is fundamentally different from deliberative reasoning.

We identify intuition with a *reliance* on associative memory. Specializing to the case of intuitive assessments about the likelihood of events, we model associative memory in terms of a *network of associations* between events. Our model takes inspiration from computer science where networks are designed to acquire associative memory and accomplish tasks such as pattern recognition. For instance, a network takes inputs from pixels on a touch screen and produces a prediction of which letter of the alphabet has been written on it. An important class of networks for such tasks is the class of *energy-based neural networks*, which include the Hopfield network (which produces deterministic predictions) and the Boltzmann machine (a stochastic counterpart which produces a probability distribution over predictions). The *training* of a network involves adjusting the weights in the network until its predictions accord well enough with the data being used to train it.

The primitive of our model is a family of normalized set functions that describe beliefs $p(\mathbf{x}|\mathbf{z})$ over a multi-dimensional event $\mathbf{x} = (\mathbf{x}_1, .., \mathbf{x}_k)$ given information $\mathbf{z} = (\mathbf{z}_1, .., \mathbf{z}_k)$, and view each $\mathbf{x}_i$ and $\mathbf{z}_j$ as a node in a network. Using a function a that describes the associative weight $a(\mathbf{x}_i, \mathbf{x}_j)$ between any pair of nodes $\mathbf{x}_i, \mathbf{x}_j$, we define an *Intuitive Belief with General Updating* representation, which reinterprets and adapts the mathematical structure of the Boltzmann machine to permit non-Bayesian updating, non-additivity, and more. The model allows for a general class of updating rules that nests Bayesian updating. A special case of the model is presented as our theory of *Intuitive Beliefs with Intuitive Updating*.

We explore the properties on beliefs that are necessary and sufficient for the existence of the Intuitive Belief representations, and the additional properties that characterize Intuitive updating. We prove a uniqueness theorem that describes the class of functions a that represents a given Intuitive Belief and show how a can be identified. We also characterize the narrow intersection of Bayesian updating and Intuitive updating.

Next, we study the *formation* of Intuitive Beliefs. Taking inspiration from machine learning, we think of belief formation in terms of the training of the agent's associative network by an objective probability distribution q^* that governs the occurrence of states of the world. We illustrate the formation of Intuitive Beliefs that exhibit Intuitive updating, and show how properties on the objective distribution give rise to some of the biases observed in the psychology literature (Tversky and Kahneman 1974).

The contributions of this paper are as follows:

1. Our model of Intuitive Beliefs with General Updating embodies the main contributions of this paper: the idea of reliance of beliefs on associative memory and the formalization of such reliance in terms of an energy-based network.

2. The special case of the model, Intuitive Beliefs with Intuitive Updating, contributes a novel theory of updating. Here priors and posteriors come from a parsimonious network of associations. The corresponding (Intuitive) updating rule is driven by the prior assessment of statistical dependence between events. The model is shown to accommodate well-known evidence from psychology.

3. In addition to providing formal models of Intuitive Beliefs, this paper provides a theory of belief formation. The weights of the agent's associative network are determined by experience in a manner similar to machine learning. This gives rise to a mapping between data and the prior.

The remainder of this paper is organized as follows. We present our model in Sections 2 and 3, present its exhaustive testable implications in Section 4, explore its uniqueness properties in Section 5 and characterize the intersection with the Bayesian model in Section 6. We discuss the formation of intuitive beliefs in Section 7 and provide an illustration in Section 8. A literature review is given in Section 9 and a conclusion in Section 10. All proofs are contained in appendices.

2 Primitives

2.1 States

The set $\Gamma = \{1, ..., N\}$ of *sources or elements of uncertainty* is a finite subset of $\mathbb{N}_+$ with cardinality $N > 1$ and generic elements $i, j, k, ...$ For each source of uncertainty $i \in \Gamma$, the 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 $x = (x_1, ..., x_N)$.

For instance, suppose there are two assets $\Gamma = \{1, 2\}$, and the elementary state space $\Omega_i = \{h_i, m_i, l_i\}$ for each asset $i \in \Gamma$ consists of a high, medium and low return that may be realized for the asset. The full state space is given by:

$$\Omega = \Omega_1 \times \Omega_2.$$

For instance, "both assets yield high return" is a state of the world given by the full state $h = (h_1, h_2)$.

2.2 Events

An *elementary event* in source i is a subset of Ω_i, generically denoted by $\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i, ..$ To consider an elementary event $\mathbf{x}_i$ is to consider the scenario where the true elementary state lies in $\mathbf{x}_i$ and all the elementary states outside $\mathbf{x}_i$ are ruled out. For each $i \in \Gamma$, fix some algebra of elementary events $\Sigma_i \subset 2^{\Omega_i}$.[1] An *event* in the full state space is a vector $\mathbf{x} = (\mathbf{x}_i, \mathbf{x}_j, ...\mathbf{x}_k)$ of elementary events, and the space of all events is given by:

$$\Sigma = \prod_{i \in \Gamma} \Sigma_i.$$

Note that an event would normally be defined as any subset of Ω rather than as a subset with a product structure. See Section 3.4 for extensions of our model that can accommodate general events.

To illustrate, let $\mathbf{x}_i = \{\text{high, medium}\} \in \Sigma_i$ define positive news about asset i. A full event with positive news about asset 1 and no information about asset 2 is $(\mathbf{x}_1, \Omega_2) \in \Sigma = \Sigma_1 \times \Sigma_2$.

Notation: For any elementary event $\mathbf{x}_i \in \Sigma_i$, we abuse notation and denote the full event $\mathbf{x}_i\Omega_{-i} \in \Sigma$ by

$$\mathbf{x}_i := \mathbf{x}_i\Omega_{-i}.$$

Moreover, for any pair of events $\mathbf{x}, \mathbf{z} \in \Sigma$ we define set inclusion by pair-wise set inclusion:

$$\mathbf{x} \subset \mathbf{z} \Longleftrightarrow \mathbf{x}_i \subset \mathbf{z}_i \text{ for all } i.$$

2.3 Beliefs

In this setting where the state space has a product structure, we define:[2]

[1] The results in this paper do not rely on the algebra structure for Σ_i. They require only that $\Omega_i \in \Sigma_i$ and that Σ_i is closed under intersections.

[2] The terms "belief function", "mass function" and "basic belief assignment" that are used in Dempster and Shafer's theory (Shafer 1976) are not related with our use of the term "belief".

Definition 1 *(Conditional Beliefs) A* belief $p(\cdot|\mathbf{z})$ *over* (Ω, Σ) *that is* conditioned *on some event* $\mathbf{z} \in \Sigma$ *is a set function that assigns* $p(\mathbf{x}|\mathbf{z}) \in [0,1]$ *for each event* $\mathbf{x} \in \Sigma$ *and satisfies:*

(i) $p(\Omega|\mathbf{z}) = 1$,
(ii) $p(\mathbf{x}|\mathbf{z}) = 0$ *if* $\mathbf{x}_i = \phi$ *for some* i,
(iii) $p(\mathbf{x} \cap \mathbf{z}|\mathbf{z}) = p(\mathbf{x}|\mathbf{z})$.

Condition (i) in the definition states that the agent is certain that some state is true. Condition (ii) states that any event that involves an empty event is deemed impossible. Condition (iii) captures the idea that the agent understands that her information $\mathbf{z}$ rules out the possibility of states outside $\mathbf{z}$, and so when evaluating the likelihood of event $\mathbf{x}$ she in fact evaluates the likelihood of $\mathbf{x} \cap \mathbf{z}$.

Consider a two period time line where at time 0 the agent holds *prior* beliefs $p(\cdot|\Omega)$, and at time 1 she learns an event $\mathbf{z}$ and forms a *posterior* $p(\mathbf{x}|\mathbf{z})$. Once a prior is defined, one can define the set of ex-ante possible events by:

$$\Sigma^+ = \{\mathbf{z} \in \Sigma \colon \ p(\mathbf{z}|\Omega) > 0\}.$$

Our primitive consists of a family of belief assessments conditioned on events in Σ^+:

Definition 2 *(Beliefs) The agent's* beliefs *are a family of conditional beliefs*

$$\mathbf{p} = \{p(\cdot|\mathbf{z}) : \mathbf{z} \in \Sigma^+\}.$$

Experimentalists have incentive compatible methods of eliciting beliefs in the experimental laboratory (Schotter and Trevino, 2014). Alternatively beliefs may be derived in an appropriate manner from betting behavior. In light of these observations, beliefs can be treated as in principle observable objects.

For later reference, we define two subclasses of interest. Say that a conditional belief satisfies *monotonicity* if for $\mathbf{x}, \mathbf{y} \in \Sigma$,

$$\mathbf{x} \subset \mathbf{y} \implies p(\mathbf{x}|\mathbf{z}) \leq p(\mathbf{y}|\mathbf{z}).$$

Monotonicity requires the agent to deem larger events as more likely.

Definition 3 *(Monotone Beliefs) Beliefs* $\mathbf{p}$ *are* Monotone *if each conditional belief satisfies monotonicity.*

Monotone beliefs are known in the literature as *non-additive probability, fuzzy measure* and *capacity*, and have been used notably to model beliefs under ambiguity (Schmeidler 1989). In order to formulate corresponding models of decision-making, Choquet integration is used to define expectation with respect to non-additive probability. Expectations can be defined in terms of non-monotone beliefs using signed Choquet integration (Waegenaere and Wakker, 2001)

Definition 4 *(Bayesian Beliefs) Beliefs* $\mathbf{p}$ *are* Bayesian *if they satisfy the Bayesian conditioning formula: for any* $\mathbf{x} \in \Sigma$ *and* $\mathbf{z} \in \Sigma^+$,

$$p(\mathbf{x}|\mathbf{z}) = p^{BU}(\mathbf{x}|\mathbf{z}) := \frac{p(\mathbf{x} \cap \mathbf{z}|\Omega)}{p(\mathbf{z}|\Omega)}.$$

That is, the agent is Bayesian if her posterior is related to the prior as in the standard definition of conditional probability. The term "Bayesian" is usually used for additive probability measures, but we use the term to describe generically non-additive beliefs that satisfy the Bayesian conditioning formula.[3]

3 Intuitive Beliefs

3.1 Associative Network

We begin by defining:

Definition 5 *(Associative Network) An* associative network *is a tuple* (a, b) *that consists of*

(i) an association function a *that maps each* $\mathbf{x}_i, \mathbf{x}_j \in \cup_{k\in\Gamma}\Sigma_k$ *to a symmetric*[4] associative weight $a(\mathbf{x}_i\mathbf{x}_j) \in \mathbb{R} \cup \{-\infty\}$,

(ii) a bias function b *that maps each* $\mathbf{x}_i \in \cup_{k\in\Gamma}\Sigma_k$ *and* $\mathbf{z} \in \Sigma^+$ *to some* bias $b(\mathbf{x}_i|\mathbf{z}) \in \mathbb{R} \cup \{-\infty\}$.

An associative network (a, b) *is* regular *if for all* $\mathbf{x}_i \in \cup_{k\in\Gamma}\Sigma_k$ *and* $\mathbf{z} \in \Sigma^+$, *the network satisfies* $b(\mathbf{x}_i|\mathbf{z}_i) = a(\mathbf{x}_i\mathbf{z}_i)$ *and*

$$b(\mathbf{x}_i|\mathbf{z}) > -\infty \implies a(\mathbf{x}_i\mathbf{z}_j) > -\infty \text{ for all } j \in \Gamma \text{ and } b(\mathbf{x}_i|\Omega) > -\infty.$$

Suppose that the elementary events $\cup_{k\in\Gamma}\Sigma_k$ form the nodes of a fully connected network. The nodes contribute to the "associative energy" in the network, where the energy contributed can be positive or negative and can take on the value $-\infty$. The bias $b(\mathbf{x}_i|\mathbf{z})$ of a node $\mathbf{x}_i$ is the energy it contributes to the network on account of its association with the information $\mathbf{z}$. The associative weight $a(\mathbf{x}_i\mathbf{x}_j)$ for any pair of nodes $\mathbf{x}_i, \mathbf{x}_j$ is the energy contributed by the pair independently of the information. Define the *associative energy* in the nodes $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N)$ given information

[3] It is easily shown that beliefs are Bayesian iff they satisfy *Bayes' Rule*: for any $\mathbf{x} \in \Sigma$ and $\mathbf{z} \in \Sigma^+$, $p(\mathbf{x}|\mathbf{z}) = \begin{cases} p(\mathbf{x}|\Omega) \times \frac{p(\mathbf{z}|\mathbf{x})}{p(\mathbf{z}|\Omega)} & \text{if } p(\mathbf{x}|\Omega) > 0 \\ 0 & \text{otherwise} \end{cases}$. Moreover, Bayesian beliefs must be Monotone since whenever $\mathbf{x} \subset \mathbf{z} \in \Sigma^+$, it must be that $\frac{p(\mathbf{x}|\Omega)}{p(\mathbf{z}|\Omega)} = p(\mathbf{x}|\mathbf{z}) \leq 1$.

[4] That is $a(\mathbf{x}_i\mathbf{x}_j) = a(\mathbf{x}_j\mathbf{x}_i)$ for all $i, j \in \Gamma$ and $\mathbf{x}_i, \mathbf{x}_j \in \cup_{k\in\Gamma}\Sigma_k$.

$\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, ...\mathbf{z}_N)$ by[5]

$$\Lambda(\mathbf{x}|\mathbf{z}) := -[\sum_{i<j} a(\mathbf{x}_i\mathbf{x}_j) + \sum_{i\in\Gamma} b(\mathbf{x}_i|\mathbf{z})].$$

In the simple example where $\Gamma = \{i, j\}$, the associative energy at nodes $\mathbf{x}_i\mathbf{x}_j$ given information $\mathbf{z}_i\mathbf{z}_j$ is:

$$\Lambda(\mathbf{x}_i\mathbf{x}_j|\mathbf{z}_i\mathbf{z}_j) = -[a(\mathbf{x}_i\mathbf{x}_j) + b(\mathbf{x}_i|\mathbf{z}_i\mathbf{z}_j) + b(\mathbf{x}_j|\mathbf{z}_i\mathbf{z}_j)].$$

The associative weights (and the bias) can be interpreted in at least two ways, corresponding to whether we think of associations that manifest consciously in the agent's mind, or subconsciously in the form of gut feeling. If conscious, then $a(\mathbf{x}_i\mathbf{x}_j)$ is a measure of the extent to which a mental image of $\mathbf{x}_i$ encourages or discourages a mental image of $\mathbf{x}_j$. If subconscious, then $a(\mathbf{x}_i\mathbf{x}_j)$ is a measure of the extent to which observing $\mathbf{x}_i$ produces or suppresses the gut feeling that $\mathbf{x}_j$ will occur. In either case, the interpretation is that an associative memory is accessed that produces the mental images or the gut feeling. The bias is interpreted similarly in terms of the associative memory involving an elementary event $\mathbf{x}_i$ and the information $\mathbf{z}$. The interpretation of $a(\mathbf{x}_i\mathbf{x}_j) = -\infty$ is that the elementary event $\mathbf{x}_i$ is associated with the *non-occurrence* of $\mathbf{x}_j$.

Associative weights are presumed to be symmetric and independent of information. One can possibly conceive of asymmetric associations. For instance, when presented with the word "differentiable" one may think of the word "continuous" very readily, but when presented with "continuous" one may not think of "differentiable" as readily. In our model, such asymmetries will be connected with the presentation of information and will be captured by the bias function rather than the association function.

Finally, the regularity property ties together bias across different information. The property states that if the occurrence of $\mathbf{x}_i$ is associated with information $\mathbf{z} = (\mathbf{z}_1, .., \mathbf{z}_N)$, then it must be that the occurrence of $\mathbf{x}_i$ is associated with each $\mathbf{z}_j$, and with the prior information Ω.

3.2 Intuitive Beliefs with General Updating

We think of beliefs $\mathbf{p}$ as intuitive if they can be viewed as arising from an associative network. Our general model is given by:

Definition 6 *(Intuitive Beliefs) Beliefs* $\mathbf{p}$ *are* Intuitive Beliefs with General Updating (IBGU) *if there exists a regular associative network* (a, b) *and a function*

[5]In order to avoid double-counting of these symmetric associations, the sum $\sum_{i<j} a(\mathbf{x}_i\mathbf{x}_j)$ is taken over all i, j where $i < j$. An equivalent formulation that exists in the computer science literature is $\frac{1}{2}\sum_{i\neq j} a(\mathbf{x}_i\mathbf{x}_j)$.

$Z : \Sigma^+ \to \mathbb{R}_{++}$ *such that for any* $(\mathbf{x}, \mathbf{z}) \in \Sigma \times \Sigma^+$ *s.t.* $\mathbf{x} \subset \mathbf{z}$,

$$p(\mathbf{x}|\mathbf{z}) = \frac{1}{Z(\mathbf{z})} \times exp\left[\sum_{i<j} a(\mathbf{x}_i\mathbf{x}_j) + \sum_{i\in\Gamma} b(\mathbf{x}_i|\mathbf{z})\right].$$

The associative network (a, b) *is said to* represent $\mathbf{p}$.

The model asserts that the belief $p(\mathbf{x}|\mathbf{z})$ is determined by exponentiating the (absolute value of the) energy of the associative network when assessing $\mathbf{x}$ given information $\mathbf{z}$, and then scaling it by a factor $Z(\mathbf{z})$ that ensures that the range of p is the unit interval. Since p must satisfy $p(\mathbf{z}|\mathbf{z}) = 1$, the scaling factor $Z(\mathbf{z})$ is determined by the energy of the network when evaluating the likelihood of the information $\mathbf{z}$:

$$Z(\mathbf{z}) = exp[\sum_{i<j} a(\mathbf{z}_i\mathbf{z}_j) + \sum_{i\in\Gamma} b(\mathbf{z}_i|\mathbf{z})] = -\Lambda(\mathbf{z}|\mathbf{z}).$$

Consequently, Z is completely determined by (a, b), which is why we refer to (a, b), rather than (a, b, Z), as a representation for $\mathbf{p}$.

The definition of Intuitive Beliefs is stated only for nested events $\mathbf{x} \subset \mathbf{z}$. By definition, beliefs satisfy $p(\mathbf{x} \cap \mathbf{z}|\mathbf{z}) = p(\mathbf{x}|\mathbf{z})$, and so the representation extends to non-nested events simply by replacing $\mathbf{x}$ with $\mathbf{x} \cap \mathbf{z}$ in the right-hand side expression.[6]

Our model is inspired by the notion of associative memory, by models of energy-based networks, and by the mathematical structure of the Boltzmann-Gibbs distribution - see Section 9.2 for the relationship with the computer science literature.

3.3 Intuitive Beliefs with Intuitive Updating

The IBGU model defines a family of beliefs where the energy between nodes is defined by a single association function and the bias at each node depends on the information in an arbitrary way. Since the bias function is what captures the relationship between the prior and posteriors, the model can accommodate a wide range of updating rules (which, as we shall confirm in Proposition 4, includes Bayesian updating). We use a special case of the model to define our theory of *Intuitive updating.*

Definition 7 *(IBIU) Beliefs* $\mathbf{p}$ *are* Intuitive Beliefs with Intuitive Updating (IBIU) *if they are IBGU with a regular associative network* (a, b) *that satisfies for all* $\mathbf{x}_i \in \cup_{k\in\Gamma}\Sigma_k$ *and* $\mathbf{z} \in \Sigma^+$,

$$b(\mathbf{x}_i|\mathbf{z}) = \sum_{j\in\Gamma} a(\mathbf{x}_i\mathbf{z}_j),$$

[6]In particular, for any $\mathbf{x}, \mathbf{z}$, the energy is given by:

$$\Lambda(\mathbf{x}|\mathbf{z}) := -[\sum_{i<j} a(\mathbf{x}_i \cap \mathbf{z}_i, \mathbf{x}_j \cap \mathbf{z}_j) + \sum_{i\in\Gamma} b(\mathbf{x}_i \cap \mathbf{z}_i|\mathbf{z})].$$

in which case, for any $(\mathbf{x}, \mathbf{z}) \in \Sigma \times \Sigma^+$ *s.t.* $\mathbf{x} \subset \mathbf{z}$*,*

$$p(\mathbf{x}|\mathbf{z}) = \frac{1}{Z(\mathbf{z})} \times exp\left[\sum_{i<j} a(\mathbf{x}_i\mathbf{x}_j) + \sum_{i\in\Gamma}\sum_{j\in\Gamma} a(\mathbf{x}_i\mathbf{z}_j)\right].$$

The association function a *is said to* represent $\mathbf{p}$.

Our theory of Intuitive updating is defined by a bias function that is defined in terms of the association function. Specifically, the bias of node $\mathbf{x}_i$ given information $\mathbf{z}$ is the sum of the associative weights between $\mathbf{x}_i$ and each $\mathbf{z}_j$. While IBGU allows the energy of the network to change with information $\mathbf{z}$ in an arbitrary way, IBIU is a parsimonious model where energy is defined completely by the association function alone. For completeness we note that regularity of (a, b) with Intuitive Updating reduces to a *regularity condition on* a that requires: for all $\mathbf{x}_i \in \Sigma$ and $\mathbf{z} \in \Sigma^+$,

$$\sum_{j\in\Gamma} a(\mathbf{x}_i\mathbf{z}_j) > -\infty \implies \sum_{j\in\Gamma} a(\mathbf{x}_i\Omega_j) > -\infty,$$

that is, for any given $\mathbf{z}$, if $\mathbf{x}_i$ is associated with the occurrence of elementary states in each $\mathbf{z}_j$ then it must be associated with the occurrence of elementary states in each Ω_j.

It is worth noting that our two models differ in how they use the weights $a(\mathbf{x}_i\mathbf{z}_i)$ involving two elementary events in the same source. These weights play an almost negligible role in the IBGU model: they do not appear in the representation but only in the definition of regularity of the network, where it matters only whether $a(\mathbf{x}_i\mathbf{z}_i)$ equals $-\infty$ or not. In contrast, the exact value of $a(\mathbf{x}_i\mathbf{z}_i)$ is a direct input into the value of the bias in the IBIU model.

3.4 Extensions to General Events

Our definition of an "event" excludes events like "one asset has a high return and the other has a low return", which specify (h_1, l_2) and (l_1, h_2) as the only possible states and do not correspond to a product of elementary events.[7] A more standard space of events consists of all subsets of Ω,

$$\Sigma^g := 2^{\prod_{i\in\Gamma}\Omega_i}$$

whereas we consider $\Sigma = \prod_{i\in\Gamma} 2^{\Omega_i} \subset \Sigma^g$. We adhere to Σ in this paper for the simplicity that it provides in studying the basic structure of the models (for instance the

[7] The usual definition of additivity does not apply either, since the union of two events in Σ may not belong to Σ. Additivity in our set up means ***Source Additivity:*** $p(\mathbf{x}_i \cup \mathbf{x}'_i, \mathbf{x}_{-i}|\mathbf{z}) = p(\mathbf{x}_i\mathbf{x}_{-i}|\mathbf{z}) + p(\mathbf{x}'_i\mathbf{x}_{-i}|\mathbf{z})$ for all $\mathbf{x}, \mathbf{x}', \mathbf{z} \in \Sigma$ and $i \in \Gamma$ s.t. $\mathbf{x}_i \cap \mathbf{x}'_i = \phi$.

calculation of $Z(\mathbf{z})$ becomes very simple). Nevertheless, for the sake of completeness, we show how Intuitive Belief style models can be formulated for general events Σ^g.

Take any $\mathbf{x}, \mathbf{z} \in \Sigma^g$. Consider the energy $\Lambda(x|z)$ that describes the association between states $x \in \mathbf{x}$ and $z \in \mathbf{z}$ where $\mathbf{x} \subset \mathbf{z}$,

$$\Lambda(x|z) = -\left[\sum_{i<j} a(x_i x_j) + \sum_{i \in \Gamma} \sum_{j \in \Gamma} a(x_i z_j)\right].$$

Then an *additive* Intuitive Belief can be defined by

$$p(\mathbf{x}|\mathbf{z}) = \frac{1}{Z(\mathbf{z})} \sum_{x \in \mathbf{x}} exp\left[-\sum_{z \in \mathbf{z}} \Lambda(x|z)\right],$$

and a *log-additive* counterpart can be defined by

$$p(\mathbf{x}|\mathbf{z}) = \frac{1}{Z(\mathbf{z})} exp\left[-\sum_{x \in \mathbf{x}} \sum_{z \in \mathbf{z}} \Lambda(x|z)\right].$$

Such models have more structure than the ones we consider in this paper. An analysis of these models is left for future research.

4 Characterizations

What structure does the model place on how the agent assesses statistical dependence between sources? How does the agent's posterior belief deviate from the Bayesian update? Is there some distinctive structure in the beliefs that lends itself to interpretation in terms of associative networks? We explore and organize the key testable implications of the model to answer these questions. Our main finding is that, in our models, beliefs over arbitrary events are, in a sense, constructed from beliefs over simpler events.

4.1 Intuitive Beliefs with General Updating

Recall that a belief p is said to exhibit *statistical independence* if it can be expressed as the product of its marginals, that is, $p(\mathbf{x}|\mathbf{z}) = \prod_{j \in \Gamma} p(\mathbf{x}_j|\mathbf{z})$, for all events $\mathbf{x}$ given information $\mathbf{z}$. This inspires the following measure of *statistical dependence* $SD(\mathbf{x}|\mathbf{z})$ *exhibited by* $\mathbf{x}$ *given information* $\mathbf{z}$:

$$SD(\mathbf{x}|\mathbf{z}) := \frac{p(\mathbf{x}|\mathbf{z})}{\prod_{j \in \Gamma} p(\mathbf{x}_j|\mathbf{z})}.$$

If $SD(\mathbf{x}|\mathbf{z}) \geq 1$ (resp. $SD(\mathbf{x}|\mathbf{z}) \leq 1$) then the elementary events in $\mathbf{x}$ are *positively associated* (resp. *negatively associated*).

Our first condition below states that the statistical dependence $SD(\mathbf{x}|\mathbf{z})$ among elementary events $\mathbf{x}_1, ..., \mathbf{x}_N$ is built in terms of the statistical dependence $SD(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})$ of all possible *pairs*.[8] More specifically, it is a product.

Definition 8 *(ASD) Beliefs* $\mathbf{p}$ *satisfy* Aggregated Statistical Dependence (ASD) *if for any* $\mathbf{x} \in \Sigma$ *and* $\mathbf{z} \in \Sigma^+$ *s.t.* $p(\mathbf{x}_i|\mathbf{z}) > 0$ *for all* $i \in \Gamma$*,*

$$\frac{p(\mathbf{x}|\mathbf{z})}{\prod_{i\in\Gamma} p(\mathbf{x}_i|\mathbf{z})} = \prod_{i<j} \frac{p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z})}.$$

Observe that ASD places no restriction on pairwise statistical dependence $\frac{p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z})}$. The content lies only in the claim the agent's evaluation of statistical dependence $\frac{p(\mathbf{x}|\mathbf{z})}{\prod_{i\in\Gamma} p(\mathbf{x}_i|\mathbf{z})}$ is rooted in pairwise statistical dependence.

The next condition places structure on deviations from Bayesian updating. It requires that the deviation from Bayesian updating is rooted in how each marginal belief deviates from its corresponding Bayesian update.

Definition 9 *(AUB) Beliefs* $\mathbf{p}$ *satisfy* Aggregated Updating Bias (AUB) *if for any* $\mathbf{x}, \mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x}_i\mathbf{z}_{-i} \in \Sigma^+$ *for all* i*,*

$$\frac{p(\mathbf{x}|\mathbf{z})}{p^{BU}(\mathbf{x}|\mathbf{z})} = \prod_{i\in\Gamma} \frac{p(\mathbf{x}_i|\mathbf{z})}{p^{BU}(\mathbf{x}_i|\mathbf{z})}.$$

Since the Bayesian update is defined purely in terms of the prior, AUB in fact specifies a relationship between the posterior, the posterior marginals and the prior. This connects beliefs across different information.

It is noteworthy that AUB has no bite when considering posterior marginals $p(\mathbf{x}_i|\mathbf{z})$ (in which case AUB reduces to the tautology $\frac{p(\mathbf{x}_i|\mathbf{z})}{p^{BU}(\mathbf{x}_i|\mathbf{z})} = \frac{p(\mathbf{x}_i|\mathbf{z})}{p^{BU}(\mathbf{x}_i|\mathbf{z})}$). This implies that AUB places no restriction on how marginal beliefs are updated, just as ASD places no restriction on pairwise statistical dependence.

A final condition connects the posteriors with the prior in a minimal way.

Definition 10 *(Regularity) Beliefs* $\mathbf{p}$ *satisfy* Regularity *if for any* $\mathbf{x} \in \Sigma$ *and* $\mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x} \subset \mathbf{z}$*,*

$$p(\mathbf{x}|\mathbf{z}) > 0 \iff p(\mathbf{x}|\Omega) > 0 \text{ and } p(\mathbf{x}_i|\mathbf{z}) > 0 \text{ for all } i,j \in \Gamma.$$

[8]Pairwise statistical dependence $\frac{p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z})}$ between elementary events is reminiscent of the notion of pointwise mutual information $\frac{p(XY)}{p(X)p(Y)}$, which is a measure of the association between a pair of random variables X, Y used in information theory. We thank Fabio Maccheroni for pointing out this similarity.

Regularity states that $\mathbf{x}$ is possible under information $\mathbf{z}$ if and only if $\mathbf{x}$ is ex-ante possible, and each of its marginals $\mathbf{x}_i$ is possible under information $\mathbf{z}$. This property is satisfied by Bayesian beliefs.[9] Bayesian beliefs also satisfy:

Definition 11 *(Information Regularity) Beliefs* $\mathbf{p}$ *satisfy* Information Regularity *if for any* $i \in \Gamma$, $\mathbf{x}_i \in \Sigma$ *and* $\mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x}_i \subset \mathbf{z}_i$,

$$p(\mathbf{x}_i|\mathbf{z}) > 0 \implies p(\mathbf{x}_i|\mathbf{z}_i) > 0.$$

That is, if $\mathbf{x}_i$ is possible given $\mathbf{z}$, it is possible also under less information, specifically, when she is only informed of $\mathbf{z}_i$.

The first result in this paper states that the above properties constitute the exhaustive testable implications of the model for $\mathbf{p}$.[10]

Theorem 1 *Beliefs* $\mathbf{p}$ *are Intuitive Beliefs with General Updating if and only if they satisfy Aggregated Statistical Dependence, Aggregated Updating Bias, Regularity and Information Regularity.*

The main take-away from this result is that, in the IBGU model, *beliefs over complex events are built from beliefs over simple events*. In the Aggregate Statistical Dependence property, beliefs over arbitrary events are constructed from beliefs over events with up to 2 dimensions, that is, events of the form $\mathbf{x}_i$ and $\mathbf{x}_i\mathbf{x}_j$. In the Aggregated Updating Bias property, the updating bias in the posterior is constructed from the updating bias exhibited by each marginal posterior. These properties reflect the binariness of associations that define the network. We view this as the observable manifestation of reliance of beliefs on associative memory.

It is worth rearranging ASD and observing that IBGU beliefs have a reduced form:

$$p(\mathbf{x}|\mathbf{z}) = [\prod_{i\in\Gamma} p(\mathbf{x}_i|\mathbf{z})] \times \prod_{i<j} \frac{p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z})}.$$

That is, these beliefs are an adjustment of a hypothetical statistically independent belief $\prod_{i\in\Gamma} p(\mathbf{x}_i|\mathbf{z})$ by using pairwise statistical dependence. It is also worth noting that ASD imposes a restriction on marginals $p(\mathbf{x}_i|\mathbf{z})$ and $p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})$, namely, that their product in the above expression must be a number less than 1. This restriction is satisfied if beliefs satisfy Monotonicity (Proposition 3). More generally, however, and with applications in mind, one may imagine generalizing the model by introducing a parameter that relieves the marginals from having to be restricted in any way.

[9] In the "if" part, $p(\mathbf{x}|\Omega) > 0$ is sufficient to guarantee that $p(\mathbf{x}|\mathbf{z}) = \frac{p(\mathbf{x}|\Omega)}{p(\mathbf{z}|\Omega)} > 0$ when $\mathbf{z} \in \Sigma^+$. For the "only if" part, first note that $p(\mathbf{x}|\mathbf{z}) = \frac{p(\mathbf{x}|\Omega)}{p(\mathbf{z}|\Omega)} > 0$ clearly implies $p(\mathbf{x}|\Omega) > 0$. To see that $p(\mathbf{x}_i|\mathbf{z}) > 0$, it suffices to note that Bayesian beliefs must satisfy the Monotonicity property that $[\mathbf{y} \subset \mathbf{z} \in \Sigma^+ \implies p(\mathbf{y}|\Omega) \leq p(\mathbf{z}|\Omega)]$, since $\frac{p(\mathbf{y}|\Omega)}{p(\mathbf{z}|\Omega)} = p(\mathbf{y}|\mathbf{z}) \leq 1$.

[10] The proof reveals that we can in fact weaken ASD so that its statement holds only for the prior. Then AUB ensures that the statement holds also for the posteriors.

4.2 Intuitive Beliefs with Intuitive Updating

The IBGU conditions do not place any restriction on the agent's updating of marginal beliefs. Consider the following concrete restriction:

Definition 12 *(IU) Beliefs* $\mathbf{p}$ *satisfy* Intuitive Updating (IU) *if for any distinct* $i, j \in \Gamma$ *and* $\mathbf{x}_i\mathbf{z}_j \in \Sigma^+$,

$$\frac{p(\mathbf{x}_i|\mathbf{z}_j)}{p^{BU}(\mathbf{x}_i|\mathbf{z}_j)} = \frac{p(\mathbf{x}_i\mathbf{z}_j|\Omega)}{p(\mathbf{x}_i|\Omega)p(\mathbf{z}_j|\Omega)}.$$

Intuitive Updating requires that updating bias must be intimately connected with the agent's *prior* evaluation of *statistical dependence.* Evidently, the agent's update is unbiased only if the sources are deemed to be unconnected (in the sense that the joint belief equals the product of marginals). If the agent perceives positive (resp. negative) statistical dependence, the update will overshoot (resp. undershoot) the Bayesian update. Note that since statistical dependence is only defined across distinct sources, there is no restriction on how the agent updates any elementary event $\mathbf{x}_i$ in source i with respect to information $\mathbf{z}_i$ on source i.

The additive separability between the associations in the representation finds further expression (beyond that in ASD and AUB) in the following property.

Definition 13 *(IS) Beliefs* $\mathbf{p}$ *satisfy* Information Separability (IS) *if for any distinct* $i, j \in \Gamma$, $\mathbf{x}_i, \mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x}_i \subset \mathbf{z}$ *and* $p(\mathbf{x}_i|\Omega_j\mathbf{z}_{-j}) > 0$,

$$\frac{p(\mathbf{x}_i|\mathbf{z}_j\mathbf{z}_{-j})}{p(\mathbf{x}_i|\Omega_j\mathbf{z}_{-j})} = \frac{p(\mathbf{x}_i|\mathbf{z}_j)}{p(\mathbf{x}_i|\Omega)}.$$

Information Separability makes several claims. First, if $p(\mathbf{x}_i|\mathbf{z}_j\mathbf{z}_{-j}) > 0$ then $p(\mathbf{x}_i|\Omega_j\mathbf{z}_{-j}) > 0$, that is, if $\mathbf{x}_i$ is possible given $\mathbf{z}$ then it must be possible given $\mathbf{z}_{-j}$. This suggests that the possibility of $\mathbf{x}_i$ is evaluated on the basis of each separate elementary event $\mathbf{z}_j$, and not on the full vector of elementary events in the information. Information Separability strengths this conclusion further by requiring that the ratio $\frac{p(\mathbf{x}_i|\mathbf{z}_j\mathbf{z}_{-j})}{p(\mathbf{x}_i|\Omega_j\mathbf{z}_{-j})}$ is not just positive, but also independent of $\mathbf{z}_{-j}$.

Finally, consider a strengthening of Information Regularity:

Definition 14 *(Strong IR) Beliefs* $\mathbf{p}$ *satisfy* Strong Information Regularity (Strong IR) *if for any* $i, j \in \Gamma$, $\mathbf{x}_i \in \Sigma$ *and* $\mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x}_i \subset \mathbf{z}_i$,

$$p(\mathbf{x}_i|\mathbf{z}) > 0 \implies p(\mathbf{x}_i|\Omega_j\mathbf{z}_{-j}) > 0.$$

Now we can state:

Theorem 2 *Suppose beliefs* $\mathbf{p}$ *are Intuitive Beliefs with General Updating. Then* $\mathbf{p}$ *are Intuitive Beliefs with Intuitive Updating if and only if they satisfy Intuitive Updating, Information Separability and Strong Information Regularity.*

The theorem reveals that the main content of the IBIU model is that the posteriors are connected with the prior through the latter's assessment of statistical dependence.

5 Uniqueness and Identification

5.1 Uniqueness Theorem

Next we present a characterization of the uniqueness properties of the representation, which will allow us to define a very useful normalization of the representation.

There is a source of non-uniqueness in the representation when it comes to elementary events for which $p(\mathbf{x}_i|\Omega) = 0$.[11] To sidestep this non-uniqueness, we restrict attention to representations that satisfy the property that for any $\mathbf{x}_i \in \cup_{l\in\Gamma}\Sigma_l$ and $i \neq j \in \Gamma$,

$$a(\mathbf{x}_i\Omega_j) = -\infty \text{ or } b(\mathbf{x}_i|\Omega) = -\infty \implies a(\mathbf{x}_i\mathbf{z}_k) = -\infty \text{ for all } \mathbf{z}_k \in \cup_{l\in\Gamma}\Sigma_l. \tag{1}$$

This allows us to state our uniqueness result as:

Theorem 3 *Consider two regular IBGU representations* (a, b) *and* (α, β) *that satisfy (1). Then* (a, b) *and* (α, β) *represent the same* $\mathbf{p}$ *if and only if there exist real-valued functions* $(\mathbf{x}, i, j) \mapsto \gamma(\mathbf{x}_i\Omega_j)$ *and* $(\mathbf{z}, i) \mapsto \psi(\mathbf{z}_i|\mathbf{z})$ *such that for any distinct* $i, j \in \Gamma$, $\mathbf{x} \in \Sigma$ *and* $\mathbf{z} \in \Sigma^+$ *for which* $\mathbf{x} \subset \mathbf{z}$,

$$a(\mathbf{x}_i\mathbf{z}_j) = \alpha(\mathbf{x}_i\mathbf{z}_j) + [\gamma(\mathbf{x}_i\Omega_j) + \gamma(\mathbf{z}_j\Omega_i) - \gamma(\Omega_i\Omega_j)],$$

and

$$b(\mathbf{x}_i|\mathbf{z}) = \beta(\mathbf{x}_i|\mathbf{z}) + \psi(\mathbf{z}_i|\mathbf{z}) - \sum_{i\neq j\in\Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)] .$$

The result reveals how, starting with some representation (α, β), we can find any other representation (a, b). Specifically, we can transform (α, β) with *any* real-valued functions $\gamma(\mathbf{x}_i\Omega_j)$ and $\psi(\mathbf{z}_i|\mathbf{z})$ in the manner described in the result in order to obtain a new representation (a, b).

It is noteworthy that there is no restriction on $a(\mathbf{x}_i\mathbf{z}_i)$, since this parameter does not feature in the IBGU representation. It does however feature in the IBIU representation.

Theorem 4 *Consider two regular IBIU representations* a *and* α *that satisfy (1). Then* a *and* α *represent the same* $\mathbf{p}$ *if and only if*

(i) there exists a real-valued function $(\mathbf{x}, i, j) \mapsto \gamma(\mathbf{x}_i\Omega_j)$ *satisfying, for each* $i \in \Gamma$ *and* $\mathbf{x}_i \in \Sigma^+$,

$$\gamma(\mathbf{x}_i\Omega_i) = \gamma(\Omega_i\Omega_i) - 2\sum_{i\neq j\in\Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\Omega_i\Omega_j)] ,$$

[11] The marginal belief $p(\mathbf{x}_i|\Omega) = 0$ arises if and only if $\sum_j a(\mathbf{x}_i\Omega_j) = -\infty$ or $b(\mathbf{x}_i|\Omega) = -\infty$, and in particular the value of any $a(\mathbf{x}_i\Omega_j) > -\infty$ is immaterial for the representation. Furthermore, by Regularity, $p(\mathbf{x}_i|\Omega) = 0$ implies $p(\mathbf{x}_i|\mathbf{z}) = 0$ and similarly the value of any $a(\mathbf{x}_i\mathbf{z}_j) > -\infty$ is immaterial for the representation.

(ii) there exists a function $(i, j, \mathbf{z}_i) \mapsto \lambda(\mathbf{z}_i, j) \in \mathbb{R}$ *satisfying* $\lambda(\mathbf{z}_i, j) = 0$ *if* $j \neq i$, *and*

(iii) for any $i, j \in \Gamma$, $\mathbf{x} \in \Sigma$ *and* $\mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x} \subset \mathbf{z}$,

$$a(\mathbf{x}_i\mathbf{z}_j) = \alpha(\mathbf{x}_i\mathbf{z}_j) + [\gamma(\mathbf{x}_i\Omega_j) + \gamma(\mathbf{z}_j\Omega_i) - \gamma(\Omega_i\Omega_j)] + \lambda(\mathbf{z}_i, j).$$

Condition (i) requires a restriction on the γ function. Specifically, it relates the value $\gamma(\mathbf{x}_i\Omega_i) - \gamma(\Omega_i\Omega_i)$ for source i with the values $\gamma(\mathbf{x}_i\Omega_j) - \gamma(\Omega_i\Omega_j)$ for distinct i, j. When $i \neq j$, condition (ii) requires $\lambda(\mathbf{z}_i, j) = 0$, and so in this case, condition (iii) is exactly the condition on $a(\mathbf{x}_i\mathbf{z}_j)$ and $\alpha(\mathbf{x}_i\mathbf{z}_j)$ in Theorem 3. When $i = j$, then we have an extra degree of freedom in that, for any given $\mathbf{z}_i$ and any $\mathbf{x}_i \subset \mathbf{z}_i$, we can transform $\alpha(\mathbf{x}_i\mathbf{z}_i)$ with some constant $\lambda(\mathbf{z}_i, i)$.

We use the uniqueness theorem to specify a particular normalized IBGU representation.

Corollary 1 *For any representation* (α, β) *of an IBGU* $\mathbf{p}$ *there exists another representation a that is* normalized *in that it satisfies (1) and the following properties:*

(a) $a(\mathbf{x}_i\Omega_j) = 0$ *for any* $\mathbf{x} \in \Sigma^+$ *and* distinct $i, j \in \Gamma$,

(b) $a(\mathbf{x}_i\mathbf{x}_i) = 0$ *for any* $\mathbf{x} \in \Sigma^+$ *and* $i \in \Gamma$,

(c) $b(\mathbf{z}_i|\mathbf{z}) = 0$ *for any* $\mathbf{z} \in \Sigma^+$ *and* $i \in \Gamma$.

For any representation α *of an IBIU* $\mathbf{p}$ *there exists another representation a that is* normalized *in that it satisfies (1) and properties (a) and (b).*

The normalization sets to zero the weight $a(\mathbf{x}_i\Omega_j)$ between *distinct* sources i, j, and also any weight $a(\mathbf{x}_i\mathbf{x}_i)$ between an elementary event and itself. The normalization does not restrict the value of $a(\mathbf{x}_i\Omega_i)$ when $\mathbf{x}_i \neq \Omega_i$.

5.2 Identification Results

In general, the network is revealed through beliefs in the following way:

Proposition 1 *For IBGU* $\mathbf{p}$, *the following statements hold:*

(i) For any $\mathbf{x} \in \Sigma$ *and* $\mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x} \subset \mathbf{z}$ *and any distinct* $i, j \in \Gamma$ *s.t.* $p(\mathbf{x}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z}) > 0$,

$$\frac{p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z})} = exp\left[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{x}_j) + a(\mathbf{z}_i\mathbf{z}_j)\right]$$

(ii) For any $i \in \Gamma$ *and* $\mathbf{x}_i\mathbf{z}_{-i}, \mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x}_i \subset \mathbf{z}_i$,

$$\frac{p(\mathbf{x}_i|\mathbf{z})}{BU(\mathbf{x}_i|\mathbf{z})} = exp[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) - b(\mathbf{x}_i|\Omega) + b(\mathbf{z}_i|\Omega)].$$

The main content of the Proposition is that pairwise statistical dependence reflects associative weights, and updating bias reflects associative bias.

A sharp identification result obtains for associative weights in the normalized models.

Proposition 2 *For IBGU* $\mathbf{p}$ *with* ***normalized*** (a, b),

$$\frac{p(\mathbf{x}_i\mathbf{x}_j|\Omega)}{p(\mathbf{x}_i|\Omega)p(\mathbf{x}_j|\Omega)} = exp\left[a(\mathbf{x}_i\mathbf{x}_j)\right]$$

for any distinct $i, j \in \Gamma$ *and* $\mathbf{x}_i\mathbf{x}_j \in \Sigma$ *s.t.* $p(\mathbf{x}_i|\Omega)p(\mathbf{x}_j|\Omega) > 0$, *and*

$$p(\mathbf{x}_i|\mathbf{z}_i) = exp[a(\mathbf{x}_i\mathbf{z}_i)]$$

for any $i \in \Gamma$ *and* $\mathbf{x}_i \in \Sigma$ *and* $\mathbf{z}_i \in \Sigma^+$ *s.t.* $\mathbf{x}_i \subset \mathbf{z}_i$. *Finally,*

$$\frac{p(\mathbf{x}_i|\mathbf{z})p^{BU}(\mathbf{x}_i|\mathbf{z}_i)}{p^{BU}(\mathbf{x}_i|\mathbf{z})} = exp[b(\mathbf{x}_i|\mathbf{z})]$$

for any $i \in \Gamma$ *and* $\mathbf{x}_i, \mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x}_i\mathbf{z}_{-i} \in \Sigma^+$.

Therefore associative weights between elementary events in *distinct* sources correspond to pairwise statistical dependence $\frac{p(\mathbf{x}_i\mathbf{x}_j|\Omega)}{p(\mathbf{x}_i|\Omega)p(\mathbf{x}_j|\Omega)}$, and the weights within a source are reflected in the source-specific posterior $p(\mathbf{x}_i|\mathbf{z}_i)$. This identifies a for both IBGU and IBIU. For IBGU, in addition, we can identify b through the marginal posteriors adjusted by bayesian posteriors.

6 Bayesian Intuitive Beliefs

In this section we characterize subclasses of Intuitive Beliefs that satisfy standard properties.

6.1 Monotonicity

We first consider Monotonicity.

Proposition 3 *The following are equivalent for Intuitive Beliefs* $\mathbf{p}$ *with General Updating: for any* $\mathbf{z} \in \Sigma^+$,

(i) $p(\textbf{.}|\mathbf{z})$ *satisfies Monotonicity.*

(ii) For any $\mathbf{x}, \mathbf{y} \in \Sigma$ *s.t.* $p(\mathbf{x}|\mathbf{z}) > 0$, *if* $\mathbf{x}_i \subset \mathbf{y}_i$ *and* $\mathbf{y} = \mathbf{y}_i\mathbf{x}_{-i}$, *then*

$$p(\mathbf{x}_i|\mathbf{z}) \times \prod_{i \neq j \in \Gamma \ s.t. \ \mathbf{x}_j \neq \mathbf{z}_j} \frac{p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})} \leq p(\mathbf{y}_i|\mathbf{z}) \times \prod_{i \neq j \in \Gamma \ s.t. \ \mathbf{x}_j \neq \mathbf{z}_j} \frac{p(\mathbf{y}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{y}_i|\mathbf{z})}.$$

(iii) The associative network (a, b) *satisfies: for any* $\mathbf{x}, \mathbf{y}_i \in \Sigma$ *s.t.* $p(\mathbf{x}|\mathbf{z}) > 0$,

$$\mathbf{x}_i \subset \mathbf{y}_i \implies \sum_{i \neq j} a(\mathbf{x}_i \mathbf{x}_j) + b(\mathbf{x}_i|\mathbf{z}) \leq \sum_{i \neq j} a(\mathbf{y}_i \mathbf{x}_j) + b(\mathbf{y}_i|\mathbf{z}).$$

When $\mathbf{x} = \mathbf{x}_i \mathbf{x}_j \mathbf{z}_{-ij}$ is effectively 2-dimensional event given information $\mathbf{z}$, condition (ii) reduces to $p(\mathbf{x}_i|\mathbf{z}) \frac{p(\mathbf{x}_i \mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})} \leq p(\mathbf{y}_i|\mathbf{z}) \frac{p(\mathbf{y}_i \mathbf{x}_j|\mathbf{z})}{p(\mathbf{y}_i|\mathbf{z})}$ and so $p(\mathbf{x}_i \mathbf{x}_j|\mathbf{z}) \leq p(\mathbf{y}_i \mathbf{x}_j|\mathbf{z})$. Therefore the condition is a strengthening of the requirement of Monotonicity for two-dimensional events. The corresponding restriction on the representation is that the sum $\sum_{i \neq j} a(\mathbf{x}_i \mathbf{x}_j) + b(\mathbf{x}_i|\mathbf{z})$ must be increasing in $\mathbf{x}_i$ with respect to set-inclusion.

The corresponding result for Intuitive Beliefs is the same (except that the bias b would be written with the additional structure). The additional structure connects all the conditional beliefs, but the above proposition explores the monotonicity property of a given conditional belief.

6.2 Updating

In this section we characterize the class of Bayesian Intuitive Beliefs.

Proposition 4 *The following are equivalent for Intuitive Beliefs* $\mathbf{p}$ *with General Updating:*

(i) $\mathbf{p}$ *is Bayesian.*

(ii) For all $\mathbf{x}_i, \mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x}_i \subset \mathbf{z}_i$,

$$p(\mathbf{x}_i|\mathbf{z}) = p^{BU}(\mathbf{x}_i|\mathbf{z}) > 0.$$

(iii) In any representation (a, b), *for all* $\mathbf{x}_i, \mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x}_i \subset \mathbf{z}_i$ *and* $p(\mathbf{x}_i|\mathbf{z}) > 0$,

$$b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) = b(\mathbf{x}_i|\Omega) - b(\mathbf{z}_i|\Omega).$$

It follows readily from Aggregated Updating Bias that if the marginal beliefs are Bayesian then beliefs must be Bayesian. In terms of the representation, as one might expect, Bayesian updating for IBGU limits how the bias changes with information. In particular, Bayesian updating can be modelled through information-independent bias: $b(\mathbf{x}_i|\mathbf{z}) = b(\mathbf{x}_i|\Omega)$ for all $\mathbf{x}_i, \mathbf{z}$.

With more structure on the bias, the restriction on beliefs is much more sharp:

Proposition 5 *The following are equivalent for Intuitive Beliefs* $\mathbf{p}$ *with Intuitive Updating:*

(i) $\mathbf{p}$ *is Bayesian.*

(ii) For any $\mathbf{x}, \mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x} \subset \mathbf{z}$ *and* $p(\mathbf{x}|\mathbf{z}) > 0$,

$$p(\mathbf{x}|\mathbf{z}) = \frac{\prod_{i \in \Gamma} p(\mathbf{x}_i|\Omega)}{\prod_{i \in \Gamma} p(\mathbf{z}_i|\Omega)}$$

(iii) In any normalized representation a, for any distinct $i, j \in \Gamma$ *and events* $\mathbf{x}_i, \mathbf{z}_i, \mathbf{x}_i\mathbf{z}_j \in \Sigma^+$ *s.t.* $\mathbf{x}_i \subset \mathbf{z}_i$,

$$a(\mathbf{x}_i\mathbf{z}_j) = 0 \text{ and } a(\mathbf{x}_i\mathbf{z}_i) - a(\mathbf{z}_i\mathbf{z}_i) = a(\mathbf{x}_i\Omega_i) - a(\mathbf{z}_i\Omega_i).$$

The proposition reveals that the intersection of Intuitive updating and Bayesian updating is narrow, requiring that the conditional beliefs $p(\mathbf{x}|\mathbf{z})$ must satisfy statistical independence for all $\mathbf{x} \in \Sigma^+$, which, unsurprisingly, requires that the associative weight between elementary events across different sources must be 0. Observe that the statistical independence property is not global, in that it is assumed to hold only for events $\mathbf{x} \in \Sigma^+$ with strictly positive prior likelihood. This permits the possibility that all marginals have a positive likelihood but the joint is assigned zero likelihood, a property not consistent with global statistical independence.

The proposition suggests that Intuitive updating can mimic Bayesian reasoning, but only if the uncertainty is perceived to have a very simple structure.

7 Formation of Intuitive Beliefs

Having formulated a model of intuitive assessments in terms of associative networks, we proceed to hypothesize how the parameters of the network might be determined. We model belief formation on the idea that associations between events are built from the agent's experience in a manner similar to machine learning.

7.1 Training

Suppose that states of the world occur according to an *objective* probability distribution:

$$q^*(\cdot|\Omega) \text{ over } (\Omega, \Sigma),$$

and let $\mathbf{q}^*$ denote the corresponding family of Bayesian beliefs. Fix some class $\mathbf{P}$ of beliefs, which at this point do not have to be limited to Intuitive Beliefs.

We consider two notions of training. The first defines training in terms of training the prior.[12]

Definition 15 *(Training) Prior beliefs* $p(\centerdot|\Omega)$ *are* trained *by* $q^*(\centerdot|\Omega)$ *if they solve:*

$$min_{\mathbf{p}\in\mathbf{P}, p(\centerdot|\Omega)\in\mathbf{p}} L((\mathbf{x}, q^*(\mathbf{x}|\Omega), p(\mathbf{x}|\Omega))_{x\in\Sigma}),$$

for some real-valued function L.

[12] In machine learning, Bayesian probability distributions are presumed and the prior is trained using a loss function given by Kullback-Leibler divergence.

To interpret, imagine that the agent, consciously or unconsciously, is exposed to different events according to $q^*(\mathbf{.}|\Omega)$, which her intuitive process picks up. If some event $\mathbf{x} = (\mathbf{x}_1, .., \mathbf{x}_n)$ happens very frequently, the repeated experience strengthens the associations defined by all the elementary events $\mathbf{x}_1, .., \mathbf{x}_n$ and universal elementary events $\Omega_1, .., \Omega_n$. Similarly, if an event occurs very infrequently, then the corresponding associations will be weak. The function L, which can be referred to as a *loss function* following the computer science literature, captures some notion of distance. In computer science, computational complexity plays a large role in the computer scientist's choice of loss function. In the context of beliefs, the modeller may similarly choose the loss function for analytical convenience, but may also have other considerations. For instance, an analyst may hypothesize that the intuitive process trains only with low-dimensional events on account of limited ability to absorb complex (high-dimensional) information.

The content of Definition 15 is that *only prior beliefs are trained by the data.* A natural question concerns whether training the prior leads to a unique belief. The answer is clearly negative in the absence of more structure, and it depends on what class $\mathbf{P}$ of beliefs is considered. If, for instance, the class $\mathbf{P}$ consists of Bayesian beliefs then the posteriors are uniquely determined by the prior, but the solution of the minimization problem may not produce a unique prior. For the IBIU class, even if there is a unique solution to the minimization problem, the prior does not fully identify the posterior: training determines the associative weights $a(\mathbf{x}_i\Omega_j)$, $a(\mathbf{x}_j\Omega_i)$ and $a(\mathbf{x}_i\mathbf{x}_j)$ for $i \neq j$ for all elementary events, but does not constrain the association $\overline{a}(\mathbf{x}_i\mathbf{z}_i)$ as these terms never appear in the prior.

The non-uniqueness of beliefs arising from training has economic content in that it provides a reason for why people may hold heterogenous beliefs when faced with the same data. But it is also of interest to explore directions that might reduce the heterogeneity, and possibly give rise to common priors in a population. We consider two possible directions:

A simple one is to adopt a refinement of the solution to the training problem. A natural hypothesis is that the agent's beliefs are deliberative (that is, Bayesian) when assessing marginal beliefs about $\mathbf{x}_i$ given information $\mathbf{z}_i$ in the *same* source i. That is, she is able to deliberate effectively when assessing marginal beliefs in light of the simplest information but relies on intuitive assessments to evaluate more complex events given more complex information:

Definition 16 *(Semi-Deliberative) Beliefs are* semi-deliberative *if they are Bayesian with respect to any single source: for any $\mathbf{x}_i \in \Sigma$ and $\mathbf{z}_i \in \Sigma^+$,*

$$p(\mathbf{x}_i|\mathbf{z}_i) = \frac{p(\mathbf{x}_i|\Omega)}{p(\mathbf{z}_i|\Omega)}.$$

An alternative route is to strengthen the notion of training. Imagine that the agent experiences not just prior objective probabilities, but also extensively experiences the

(Bayesian) objective posterior probabilities conditional on information. The network is then shaped by entire family of objective distributions.

Definition 17 *(Strong Training) Beliefs* $\mathbf{p}$ *are* strongly trained *by* $\mathbf{q}^*$ *if they solve*

$$min_{\mathbf{p}\in\mathbf{P}} max_{\mathbf{z}\in\Sigma^+} L((\mathbf{x}, q^*(\mathbf{x}|\mathbf{z}), p(\mathbf{x}|\mathbf{z}))_{x\in\Sigma}).$$

One might imagine that in most contexts people are exposed to prior objective probabilities more extensively than posterior objective probabilities. However agents with the most extensive experience in a market – experts – are more likely to be strongly trained.

8 Illustration

Suppose that there are only two dimensions, $\Gamma = \{1, 2\}$, and that the loss function L is the supnorm:

$$L(q^*(\centerdot|\Omega), p(\centerdot|\Omega)) = sup_{x\in\Sigma}|q^*(\mathbf{x}|\Omega) - p(\mathbf{x}|\Omega)|.$$

Suppose also that the agent's beliefs have the structure of Intuitive Beliefs with Intuitive Updating.

Our results in this illustration are presented in terms of the objective prior statistical dependence between pairs of elementary events:

$$SD^*(\mathbf{x}_i\mathbf{x}_j|\Omega) := \frac{q^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}{q^*(\mathbf{x}_i|\Omega)q^*(\mathbf{x}_j|\Omega)}.$$

8.1 Trained Intuitive Beliefs

In general, training may not give rise to an IBIU that is semi-deliberative in the sense of Definition 16. However, such beliefs can arise under some restriction on the objective distribution:

Proposition 6 *Suppose that for all* $\mathbf{x}, \mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x} \subset \mathbf{z}$,

$$\frac{q^*(\mathbf{z}_i|\Omega)}{q^*(\mathbf{x}_i|\Omega)} \leq \left[\frac{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}\right]^{1.5}$$

Then there exists a unique semi-deliberative Intuitive Belief with Intuitive Updating that is trained by q^**. Moreover:*

(i) The prior is perfectly trained in the sense that

$$p(\centerdot|\Omega) = q^*(\centerdot|\Omega).$$

(ii) The posterior is given by

$$p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z}) = [q^*(\mathbf{x}_i|\mathbf{z}_i)q^*(\mathbf{x}_j|\mathbf{z}_j)]\left[\frac{SD^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]\left[\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right].$$

The agent's prior belief is therefore a standard probability distribution. We study her posterior reaction to information.

8.2 Correlation Perception

To what extent does the agent's assessment of statistical dependence remain accurate after observing information?

Proposition 7 *For any* $\mathbf{x}, \mathbf{z} \in \mathbf{\Sigma}^+$,

$$\frac{p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z})} = \frac{q^*(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{q^*(\mathbf{x}_i|\mathbf{z})q^*(\mathbf{x}_j|\mathbf{z})}.$$

In the two-dimensional setup, the agent's perception of statistical dependence remains accurate. Although her posterior beliefs deviate from the objective posterior distribution, the deviation of $p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})$ and the marginals $p(\mathbf{x}_i|\mathbf{z})$, $p(\mathbf{x}_j|\mathbf{z})$ are such that they cancel out in the ratio $\frac{p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z})}$.

More generally, however, when the state space has higher dimensionality, the agent's perception of correlation may not be correct even at the prior stage.[13] One can construct examples where q^* exhibits statistical independence for 2-dimensional events $\mathbf{x}_i\mathbf{x}_j$ but not for events with higher dimensions,[14] but the trained beliefs satisfy statistical independence for all events. This is because Intuitive Beliefs are unable to perceive higher dimensional dependence due to the simple binary nature of associative connections.

8.3 Non-Bayesian Updating

In the experimental literature (see Tversky and Kahneman, 1974), overshooting (resp. undershooting) relative to the Bayesian update corresponds to *base rate neglect* (resp. *conservatism*). Our model produces one or the other depending on the strength of associations involved after information is presented.

Proposition 8 *For any* $\mathbf{x}, \mathbf{z} \in \Sigma^+$,

$$p(\mathbf{x}|\mathbf{z}) \geq p^{BU}(\mathbf{x}|\mathbf{z}) \iff SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega) \times SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega) \geq SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)^2.$$

The proposition tells us that the agent over-updates if and only if the event $\mathbf{x} = (\mathbf{x}_i, \mathbf{x}_j)$ being evaluated is very strongly associated with the information $\mathbf{z} = (\mathbf{z}_i, \mathbf{z}_j)$, where $SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)^2$ determines the threshold. The association between $\mathbf{x}_i$ and $\mathbf{z}_i$ does not factor in because we have assumed that beliefs are semi-deliberative.

[13] See Ellis and Piccione for recent decision-theoretic work on correlation misperception.

[14] For example, if there are three assets and each can take value H or L, and if $q(HLH|\Omega) = q(LHH|\Omega) = q(HHL|\Omega) = q(LLL|\Omega) = \frac{1}{4}$ then q satisfies pairwise statistical independence but fails statistical independence.

8.4 Non-Monotonicity

The experimental literature has documented the existence of violations of monotonicity of beliefs (Tversky and Kahneman, 1974). For instance, in a well-known experiment, based on a description of some hypothetical person named "Linda", subjects found it less likely that Linda was a bank teller than a bank teller that is active in the feminist movement. Since a standard property of probabilities is that the conjunction of two events can only be less likely than either of the events, this finding is referred to as the *conjunction fallacy*. We see that:

Proposition 9 *For any* $\mathbf{y}_i\mathbf{x}_j, \mathbf{x}_i\mathbf{x}_j, \mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x}_i \subset \mathbf{y}_i$*,*

$$p(\mathbf{y}_i\mathbf{x}_j|\mathbf{z}) \geq p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z}) \iff q^*(\mathbf{y}_i\mathbf{x}_j|\Omega) \geq q^*(\mathbf{x}_i\mathbf{x}_j|\Omega) \times \frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{y}_i\mathbf{z}_j|\Omega)}.$$

By monotonicity of objective probabilities, $q^*(\mathbf{y}_i\mathbf{x}_j|\Omega) \geq q^*(\mathbf{x}_i\mathbf{x}_j|\Omega)$. But if $\mathbf{x}_i$ is very strongly associated with $\mathbf{z}_j$ (relative to how strongly $\mathbf{y}_i$ is associated with $\mathbf{z}_j$), then $\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{y}_i\mathbf{z}_j|\Omega)} > 1$ and the proposition reveals that information $\mathbf{z}_j$ may give rise to $p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z}) > p(\mathbf{y}_i\mathbf{x}_j|\mathbf{z})$, which illustrates the conjunction fallacy.

8.5 Other

The *disjunction fallacy* occurs when the union of two events is deemed less likely than each constituent event, which points to a violation of monotonicity, as does the conjunction fallacy, and can be accommodated similarly. To accommodate the *gambler's fallacy* or the *hot-hand effect*, one can incorporate the idea in the literature that beliefs about the realization of, say, the next coin toss, depend on how beliefs about some underlying parameter evolve with each coin toss (Rabin and Vayanos 2010).

9 Related Literature

9.1 Psychology

The philosophical school of Associationism, which dates back to the writings of Locke and Hume and served as the foundation of behavioral psychology in the early 20th century, recognized the creation of associations as the most basic function of the mind. The Associationists sought to reduce all mental life to associations, an endeavor that survived until the cognitive revolution in the mid 20th century.

Associations and memory have been modelled as networks in cognitive psychology, specifically 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 computer science (see Section 9.2).

Though not based on mathematical models of associative memory, Morewedge and Kahneman (2010) posit that intuitive judgements are made through automatic, non-deliberative "System 1" processing, of which associative memory is a part.[15] The Heuristics and Biases program of Kahneman and Tversky posits that people's beliefs are heuristic-based intuitive judgements (Tversky and Kahneman 1974). The three heuristics discussed in this literature are Availability, Representativeness and Anchoring and Adjustment heuristics, and our model relates to the first two (see Section 8). According to the Availability heuristic, people assess likelihoods in terms of how many salient examples come to their minds. According to the Representativeness heuristic, people update their likelihoods based on the representativeness of the information for the proposition being evaluated, which are driven by similarity considerations. Salience and similarity play a role in determining the strength of associations, and in our model prior and posterior likelihoods are determined by the strength of associations.

9.2 Computer Science

Developments in neuroscience in the 20th century inspired the study of artificial neural networks in computer science. Hopfield (1982) shows how associative memory can be modelled using an energy-based network. To outline the so-called *Hopfield network*, consider a fully connected network with nodes, $\Gamma = \{1, ..., N\}$. Each node i can either be "on" or "off", thereby taking a value x_i in the binary set $\Omega_i = \{1, 0\}$ (the values 1 and -1 are also often used). A state $x = (x_1, ..., x_n)$ is a vector of configurations of activation of these nodes. For instance, suppose there is a screen (that can display only black and white images) where each pixel is a node that can be on or off.

In the Hopfield network, at each state, the nodes can update their activation, giving rise to a new state and leading to a dynamic evolution of the state. Let the weight between two nodes be given by some real number $a(i, j)$. Fix a threshold θ_i for each node. The updating rule requires that if the state of the system is $x = (x_1, ..., x_n)$, then each node i is updated

$$x'_i = \begin{cases} 1 & \text{if } \sum_j a(i,j)x_j \geq \theta_i \\ 0 & \text{otherwise.} \end{cases}$$

The nodes are updated in a (pre-determined or random) sequence. If the network is well-behaved, repeated updating leads the Hopfield network to converge to some steady state. The set of steady states can be characterized in terms of the "energy" of the network. The energy at state $x = (x_1, ..., x_n)$ is defined as the (conventionally

[15] Rubinstein (2016) categorizes actions by experimental subjects as instinctive or contemplative based on their response time.

negative) quantity:

$$\Lambda(x_1, ..., x_n) = -[\sum_{i<j} a(i,j)x_i x_j + \sum_i b(i)x_i].$$

The connection between the two nodes i, j contributes $a(i,j)x_i x_j$ to the energy (so that the contribution is zero unless both are on) and furthermore, each node i has a "bias" $b(i)x_i$ that indicates the energy it contributes to the system even if all other nodes are off. The set of steady states of the Hopfield network is the set of local minima of Λ.

To illustrate how the Hopfield network is trained, suppose that we wish to train it to recognize the letter "A". Suppose the ideal letter corresponds to state (1,1,1,1), but to allow for variation in people's handwriting, we wish to associate any state for which $x_1 = x_2 = 1$ with that letter. In order to do this, the values of nodes 1 and 2 are "clamped" to $x_1 = x_2 = 1$, and the remaining nodes are allowed to update according to the updating rule above until the system reaches a steady state. If the steady state is not (1,1,1,1), then the *weights* are updated by some pre-specified rule (such as gradient descent) and the network is run again. The process stops if the network correctly settles to (1,1,1,1) – in which case the Hopfield network has successfully acquired the desired associative memory – but the stop condition may also specify a limit on how many cycles of training the network has gone through.

The existence of local minima can be a shortcoming (since revising weights by rules such as gradient descent may not take the network towards its global minima), and there are several solutions available. The Boltzmann machine avoids the issue of local minima by adapting the Hopfield network's updating rule for nodes (Hinton and Sejnowski 1983, Ackley, Hinton and Sejnowski 1985). Specifically, borrowing from models in physics, the Boltzmann machine requires nodes to be updated probabilistically as a logistic function of the energy of the network. Repeated updating causes the probability distribution over states to converge to an equilibrium distribution (referred to as its "thermal equilibrium"), which takes the form of the so-called Boltzmann-Gibbs distribution:[16]

$$p(x_1, ..., x_n) = \frac{1}{Z} exp[-\Lambda(x_1, ..., x_n)]. \tag{2}$$

The Boltzmann machine is mainly used to represent (that is, capture the properties of) the data. Thus, weights are adjusted until the equilibrium probability distribution is sufficiently close (with respect to Kullback-Leibler divergence) to the distribution in the data being used to train the network. Then, for instance, the machine can provide a maximum likelihood estimate of the letter written on a screen.

Our models of Intuitive Beliefs reinterpret and adapt the formalism that defines the Boltzmann machine. We reinterpret its probability distribution over states as the

[16]The scaling factor is called the *partition function* in computer science and is conventionally denoted by the letter Z.

agent's beliefs over states. States are now mental images in the agent's mind or her gut-feelings, and the energy of the network at a state is a measure of the extent to which the agent's thoughts are populated by a particular mental image, or the extent to which her gut feeling is driven by stored associative memory. Circumventing the process by which thermal equilibrium comes about, we assume the agent's beliefs have already settled for the network, and therefore takes the Boltzmann-Gibbs form.

In our model, each elementary event is a node, which can be on or off – the belief about an event $\mathbf{x}$ is the likelihood assigned to the configuration where only the nodes in $\mathbf{x_1}, .., \mathbf{x}_n$ are on. Therefore, a state in the Boltzmann machine corresponds to an event in our model.

The Boltzmann machine is understood to represent a Bayesian system of probabilities, and so for any event the Bayesian posterior may be calculated. For any general event $\mathbf{x} \subset \prod_i \Omega_i$, the Bayesian posterior is given by

$$p(x_1, ..., x_n|\mathbf{x}) = \frac{exp[-\Lambda(x_1, ..., x_n)]}{\sum_{(x'_1,...,x'_n)\in(\mathbf{x}_1,...,\mathbf{x}_n)} exp[-\Lambda(x'_1, ..., x'_n)]}.$$

Our IBGU model makes a critical departure here. We suppose that the prior and posteriors each have a Boltzmann-Gibbs form (2) that share the same associative weights, but that the bias is an arbitrary function of the information. Exploiting the fact that the bias now involves pairs of elementary events and full events, our IBIU model provides a parsimonious special case where bias is computed with associative weights alone.

9.3 Economics

The literature on non-Bayesian beliefs in economics generalizes the Bayesian model in order to accommodate and study particular non-Bayesian features. For instance, Rabin (2002) models the law of small numbers, Rabin and Vayanos (2010) model the gambler's and hot-hand fallacies, Benjamin et al (2019) model base-rate neglect, Gennaioli and Shleifer (2010) and Zhao (2017) model the Representativeness heuristic, Ortoleva (2012) models an agent who changes her prior upon receiving ex ante unlikely information, and so on. This paper differs in that it does not set out to model a specific non-Bayesian feature. However in providing a model of intuition, it provides a framework that can accommodate non-Bayesian features. A framework for non-Bayesian updating is provided in Epstein et al (2008) but there the driving force is a temptation to retroactively change the prior.

There are few psychology-inspired models of memory in economics. Memory is modelled as a *set* of past cases or experiences in Gilboa and Schmeidler (1995) (in the context of choice under subjective uncertainty) and Bordalo et al (2019) (in the context of choice under certainty). Past experiences are explicitly considered by the agent in the former, and automatically come to the agent's mind in the latter, and in both cases some notion of similarity determines the extent to which something is considered or recalled. In our model, the agent's past experiences are encoded in her associative network, and notions such as similarity are subsumed in the more abstract notion of association. As motivation for their seminal work on case-based decision theory, Gilboa and Schmeidler (1995) cite Hume, an early Associationist philosopher, who stated that "[f]rom causes which appear *similar* we expect similar effects" (Hume 1748).

In the narrow literature on belief formation, Gilboa and Schmeidler (2003) write a case-based model of an agent whose likelihood assessment about a proposition is a similarity-weighted sum of support for that proposition offered by a set of past cases that the agent has access to. Spiegler (2016) uses Bayesian networks to describe an agent who uses data in conjunction with her causal theory of the world in order to form beliefs. Using a directed acyclic graph (DAG), a Bayesian network graphically represents the conditional independence between different variables, and uses the chain rule to aggregate conditional probabilities to form a prior. The agent draws objective conditional probabilities from some objective distribution q, and constructs her prior by the chain rule.[17] Posterior beliefs are calculated using Bayes' Rule.

Similar to Spiegler (2016), we draw from the computer science literature, but we do so by adapting energy-based networks, which involve fully connected networks with symmetric weights and share little with Bayesian networks. Similar to Gilboa and Schmeidler (2003) and Spiegler (2016), we are interested in the mapping from data to beliefs. However, our interest is in the formation of intuitive beliefs whereas these papers are more readily interpreted in terms of beliefs arrived at through deliberative reasoning. For a concrete illustration of the conceptual difference, observe that associations can be built if the agent is faced with *copies* of the data. For instance, a news story that is aired continuously on television can form an association without one even being conscious of it, thereby affecting beliefs, while deliberative beliefs would treat the news as one data point.

10 Summary and Future Directions

While decision-makers in complex environments may seek to reduce the complexity of the problem in various ways,[18] this paper is motivated by the idea that people often rely on their intuitive assessment of the problem. To enable an exploration of this idea, we construct a formal theory of intuition. We identify intuition with reliance on associative memory. In order to model associative memory, we take inspiration from energy-based network models. We formulate two nested models of Intuitive beliefs: one with general updating, and one with Intuitive updating.

[17]For instance, the graph $x_2 \to x_1 \leftarrow x_3$ represents the belief that x_1 is conditionally dependent on x_2 and x_3, which are conditionally independent of each other and x_1. The chain rule yields $p(x_1x_2x_3) = q(x_1|x_2x_3)q(x_2)q(x_3)$.

[18]For instance, see Jehiel (2005) for a model where the agent creates analogy classes.

The characterization of our model reveals the structure we might expect from beliefs that are based on memories of simple associations: beliefs about complex events are built entirely from beliefs over simple events. The formation of beliefs is modelled as the creation of associative memory from experience. The special case of the model that exhibits Intuitive updating is shown to accommodate findings from the psychology literature on intuitive judgement. The model also provides an explanation for why people may fail to recognize correlation in the data.

Beyond exploring possible applications of the model to understand economic phenomena, or normatively exploring the idea that intuition should be regarded as a noisy but informative signal, there are several directions for theoretical research, such as:

- We saw that in the IBIU model, Bayesian beliefs can exist only in a very narrow class. One way to think about this result is that limitations in how associative memory is encoded (such as the binary structure of the network in the case of IBIU) are a source of deviation from Bayesian updating. To study this idea further, one could construct associative networks using hypergraphs, where connections exist between sets of nodes rather than pairs of nodes.[19]

- It is well-known that after observing an infinite sequence of signals, Bayesian beliefs eventually learn the true state. What can Intuitive Beliefs learn? It is important to distinguish between updating beliefs that arise from a given network (as in the models in this paper) versus an update of the associative network due to information, which subsequently changes the agent's beliefs.

- The goal of our characterization theorems was to gain some immediate insight into the structure of an unfamiliar model. A natural direction for future research is to explore whether the model can be derived instead from appealing axioms on behavior. The primitive may consist of ordinal likelihood relations that represent betting preferences, or could be a preference defined on a standard choice domain consisting of state-contingent consequences. Signal structures would give more power than the kind of partitional information structure we presupposed.

- The formation of beliefs lends itself to a host of hypotheses. If an agent holds an asset, she may be more sensitive to parts of the data that speak to the possibility of making large gains or losses, and this focus may bias how her associative memory is trained by the available data. In particular, the agent's utility function may also enter a theory that links associative networks with data. Another possible hypothesis to consider is the role of repetition in forming associations. To distinguish the occurrence of a data point from the frequency with which it is encountered, the notion of data could be enriched by using multi-sets.

[19] We thank Ryota Iijima for this observation.

A Appendix: Basic Expressions

Without being explicit about it, we will sometimes use the following trivial equalities when manipulating the representation:

Lemma 1 *For any function* $(\mathbf{x}, i, j) \mapsto a(\mathbf{x}_i\mathbf{x}_j)$,
(i) $\sum_{i,j\in\Gamma:i<j} a(\mathbf{x}_i\mathbf{x}_j) = \frac{1}{2}\sum_{i,j\in\Gamma:i\neq j} a(\mathbf{x}_i\mathbf{x}_j)$,
(ii) $\frac{1}{2}\sum_{i\in\Gamma}\sum_{i\neq j\in\Gamma} a(\mathbf{x}_i\mathbf{z}_j) = \frac{1}{2}\sum_{i\in\Gamma}\sum_{i\neq j\in\Gamma} a(\mathbf{x}_j\mathbf{z}_i)$.

Some expressions that we will use often:

Lemma 2 *For any IBGU, and any* $\mathbf{x} \in \Sigma$ *and* $\mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x} \subset \mathbf{z}$,
(Ai) $p(\mathbf{x}|\mathbf{z}) = exp[\sum_{i<j}[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i\in\Gamma}[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]]$.
(Aii) $p(\mathbf{x}_i|\mathbf{z}) = exp[\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + [b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]]$.
(Aiii)

$$p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z}) =$$

$$exp[[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i,j\neq k}[a(\mathbf{x}_i\mathbf{z}_k) - a(\mathbf{z}_i\mathbf{z}_k)] + \sum_{i,j\neq k}[a(\mathbf{x}_j\mathbf{z}_k) - a(\mathbf{z}_j\mathbf{z}_k)]$$

$$+[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})] + [b(\mathbf{x}_j|\mathbf{z}) - b(\mathbf{z}_j|\mathbf{z})]$$

For any IBIU, and any $\mathbf{x} \in \Sigma$ *and* $\mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x} \subset \mathbf{z}$,
(Bi) $p(\mathbf{x}|\mathbf{z}) = exp[\sum_{i<j}[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i\in\Gamma}\sum_{j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)]]$.
(Bii) $p(\mathbf{x}_i|\mathbf{z}) = exp[2\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + [a(\mathbf{x}_i\mathbf{z}_i) - a(\mathbf{z}_i\mathbf{z}_i)]]$.

Proof. We prove only some of these claims.

(Ai) Since $p(\mathbf{z}|\mathbf{z}) = 1$, we can compute that $Z(\mathbf{z}) = exp[\sum_{i<j} a(\mathbf{z}_i\mathbf{z}_j) + \sum_{i\in\Gamma} b(\mathbf{z}_i|\mathbf{z})]$. Insert this into the model to get the expression.

(Aii) Compute that
$p(\mathbf{x}_i|\mathbf{z})$
$= exp[\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + [b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]]$
$+ \sum_{i\neq k\in\Gamma}[b(\mathbf{z}_k|\mathbf{z}) - b(\mathbf{z}_k|\mathbf{z})]]$
$= exp[\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + [b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]]$.

(Aiii) Compute that
$p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})$
$= exp[[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i,j\neq k}[a(\mathbf{x}_i\mathbf{z}_k) - a(\mathbf{z}_i\mathbf{z}_k)] + \sum_{i,j\neq k}[a(\mathbf{x}_j\mathbf{z}_k) - a(\mathbf{z}_j\mathbf{z}_k)]$
$+[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})] + [b(\mathbf{x}_j|\mathbf{z}) - b(\mathbf{z}_j|\mathbf{z})]$
$+ \sum_{i,j\neq l\in\Gamma}[b(\mathbf{z}_l|\mathbf{z}) - b(\mathbf{z}_l|\mathbf{z})]$
where the last term equals 0.

(Bii) Compute that
$p(\mathbf{x}_i|\mathbf{z})$
$= exp[\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + [b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]]$.
$= exp[\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)]]$

$= exp[\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + [a(\mathbf{x}_i\mathbf{z}_i) - a(\mathbf{z}_i\mathbf{z}_i)]]$
$= exp[2\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + [a(\mathbf{x}_i\mathbf{z}_i) - a(\mathbf{z}_i\mathbf{z}_i)]]$.
■

Lemma 3 *For IBGU, any* $\mathbf{z} \in \Sigma^+$ *and distinct* $i, j \in \Gamma$,

$$a(\mathbf{z}_i\mathbf{z}_j), b(\mathbf{z}_i|\mathbf{z}) > -\infty.$$

For IBIU, in addition, $a(\mathbf{z}_i\mathbf{z}_i) > -\infty$.

Proof. For any $\mathbf{z} \in \Sigma^+$, since $p(\mathbf{z}|\mathbf{z}) = 1$, we obtain $Z(\mathbf{z}) = exp[\sum_{i<j} a(\mathbf{z}_i\mathbf{z}_j) + \sum_{i\in\Gamma} b(\mathbf{z}_i|\mathbf{z})]$and $Z(\mathbf{z}) = exp[\sum_{i<j} a(\mathbf{z}_i\mathbf{z}_j) + \sum_{i\in\Gamma}\sum_{j\in\Gamma} a(\mathbf{z}_i\mathbf{z}_j)]$ for the two models. Since Z is strictly positive by definition, it follows that all the terms on the RHS are strictly greater than $-\infty$. ■

Lemma 4 *For IBGU,*

(Ai) $p(\mathbf{x}|\mathbf{z}) > 0 \iff a(\mathbf{x}_i\mathbf{x}_j), b(\mathbf{x}_i|\mathbf{z}) > -\infty$ *for all distinct* i, j.

(Aii) $p(\mathbf{x}_i|\mathbf{z}) > 0 \iff a(\mathbf{x}_i\mathbf{z}_j)$ *for all* $j \in \Gamma$ *and* $b(\mathbf{x}_i|\mathbf{z}) > -\infty$.

(Aiii) $p(\mathbf{x}_i|\mathbf{z}_i) > 0 \iff a(\mathbf{x}_i\Omega_j)$ *for all* $i \neq j \in \Gamma$ *and* $a(\mathbf{x}_i|\mathbf{z}_i) > -\infty$.

For IBIU,

(Bi) $p(\mathbf{x}|\mathbf{z}) > 0 \iff a(\mathbf{x}_i\mathbf{x}_j) > -\infty$ *for all distinct* i, j *and* $a(\mathbf{x}_i\mathbf{z}_j) > -\infty$ *for all* i, j.

(Bii) $p(\mathbf{x}_i|\mathbf{z}) > 0 \iff a(\mathbf{x}_i\mathbf{z}_j) > -\infty$ *for all* j.

Proof. (Ai): First establish the result for $\mathbf{z} = \Omega$. Since Lemma 3 yields $a(\Omega_i\Omega_j), b(\Omega_i|\Omega) > -\infty$ for all distinct i, j, then: $p(\mathbf{x}|\Omega) > 0$ is equivalent to

$$exp[\sum_{i<j}[a(\mathbf{x}_i\mathbf{x}_j) - a(\Omega_i\Omega_j)] + \sum_{i\in\Gamma}[b(\mathbf{x}_i|\Omega) - b(\Omega_i|\Omega)]] > 0$$

which is equivalent to [$a(\mathbf{x}_i\mathbf{x}_j) > -\infty$ for all distinct i, j and $b(\mathbf{x}_i|\Omega) > -\infty$ for all i], as desired.

More generally, take any $\mathbf{z} \in \Sigma^+$. By the preceding, $a(\mathbf{z}_i\mathbf{z}_j) > -\infty$ for all distinct i, j. Moreover, $b(\mathbf{z}_i|\mathbf{z}) > -\infty$ for any i by Lemma 3. Given Lemma 2, then: $p(\mathbf{x}|\mathbf{z}) > 0$ is equivalent to

$$exp[\sum_{i<j}[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i\in\Gamma}[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]] > 0$$

which is equivalent to the statement that $a(\mathbf{x}_i\mathbf{x}_j) > -\infty$ for all distinct i, j and $b(\mathbf{x}_i|\mathbf{z}) > -\infty$ for all i.

(Aii): Given that $p(\mathbf{x}_i|\mathbf{z}) = exp[\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + [b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]]$, argue as in (Ai) to establish that [$p(\mathbf{x}_i|\mathbf{z}) > 0$ is equivalent to $a(\mathbf{x}_i\mathbf{z}_j), b(\mathbf{x}_i|\mathbf{z}) > -\infty$ for all distinct i, j]. But, by regularity of the network, $b(\mathbf{x}_i|\mathbf{z}) > -\infty$ implies

$a(\mathbf{x}_i\mathbf{z}_i) > -\infty$. Therefore $p(\mathbf{x}_i|\mathbf{z}) > 0$ implies $a(\mathbf{x}_i\mathbf{z}_j) > -\infty$ for all distinct $j \in \Gamma$ and $b(\mathbf{x}_i|\mathbf{z}) > -\infty$. Conversely, these conditions subsume the noted conditions for $p(\mathbf{x}_i|\mathbf{z}) > 0$, and therefore the converse holds.

(Aiii): Apply (Aii) and the property $b(\mathbf{x}_i|\mathbf{z}_i) = a(\mathbf{x}_i|\mathbf{z}_i)$ that defines a regular network.

(Bi) and (Bii): Use the definition $b(\mathbf{x}_i|\mathbf{z}) = \sum_j a(\mathbf{x}_i\mathbf{z}_j)$ to note that $b(\mathbf{x}_i|\mathbf{z}) > -\infty$ is equivalent to $a(\mathbf{x}_i\mathbf{z}_j) > -\infty$ for all i, j. Apply this to (Ai) and (Aii). ■

B Proof of Proposition 1

Lemma 5 *For IBGU and any $\mathbf{x} \in \Sigma$ and $\mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$, s.t. $p(\mathbf{x}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z}) > 0$*

$$p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z}) = p(\mathbf{x}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z}) \times exp\left[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{x}_j) + a(\mathbf{z}_i\mathbf{z}_j)\right].$$

Proof. By Lemma 4, $p(\mathbf{x}_i|\mathbf{z}) > 0$ implies $a(\mathbf{x}_i\mathbf{z}_j) > -\infty$ for all j and similarly $p(\mathbf{x}_j|\mathbf{z}) > 0$ implies $a(\mathbf{x}_j\mathbf{z}_i) > -\infty$ for all i. We use this fact, along with the expression for marginals (Lemma 2) in the following observations:

$p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})$

$= exp[\Big[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j) + \sum_{i,j\neq k\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_k) - a(\mathbf{z}_i\mathbf{z}_k)] + \sum_{i,j\neq k\in\Gamma}[a(\mathbf{z}_k\mathbf{x}_j) - a(\mathbf{z}_k\mathbf{z}_j)]\Big]$
$+b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) + b(\mathbf{x}_j|\mathbf{z}) - b(\mathbf{z}_j|\mathbf{z})]$

$= exp[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j) - a(\mathbf{x}_i\mathbf{z}_j) + a(\mathbf{z}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{x}_j) + a(\mathbf{z}_i\mathbf{z}_j)$
$+ \sum_{i\neq k\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_k) - a(\mathbf{z}_i\mathbf{z}_k)] + \sum_{j\neq k\in\Gamma}[a(\mathbf{z}_k\mathbf{x}_j) - a(\mathbf{z}_k\mathbf{z}_j)]$
$+ b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) + b(\mathbf{x}_j|\mathbf{z}) - b(\mathbf{z}_j|\mathbf{z})]$ (where we added and subtracted $a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j) + a(\mathbf{z}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)$)

$= exp[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{x}_j) + a(\mathbf{z}_i\mathbf{z}_j)$
$+ \Big[\sum_{i\neq k\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_k) - a(\mathbf{z}_i\mathbf{z}_k)] + b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})\Big]$
$+ \Big[\sum_{j\neq k\in\Gamma}[a(\mathbf{z}_k\mathbf{x}_j) - a(\mathbf{z}_k\mathbf{z}_j)] + b(\mathbf{x}_j|\mathbf{z}) - b(\mathbf{z}_j|\mathbf{z})]\Big]$

$= p(\mathbf{x}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z}) \times exp\left[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{x}_j) + a(\mathbf{z}_i\mathbf{z}_j)\right]$ (by Lemma 2). This completes the argument. ■

Define the bayesian update of the prior by $p^{BU}(\mathbf{x}|\mathbf{z}) := \frac{p(\mathbf{x}\cap\mathbf{z}|\Omega)}{p(\mathbf{z}|\Omega)}$.

Lemma 6 *(a) For IBGU and any $\mathbf{x}, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$,*

$$p^{BU}(\mathbf{x}|\mathbf{z}) = exp[\sum_{i<j}[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i\in\Gamma}[b(\mathbf{x}_i|\Omega) - b(\mathbf{z}_i|\Omega)]]$$

(b) For IBGU and any $\mathbf{x}, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$,

$$\frac{p(\mathbf{x}|\mathbf{z})}{p^{BU}(\mathbf{x}|\mathbf{z})} = exp[\sum_{i\in\Gamma}[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) - [b(\mathbf{x}_i|\Omega) - b(\mathbf{z}_i|\Omega)]].$$

Moreover, if $\mathbf{x}_i\mathbf{z}_{-i} \in \Sigma^+$,

$$\frac{p(\mathbf{x}_i|\mathbf{z})}{p^{BU}(\mathbf{x}_i|\mathbf{z})} = exp[[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) - [b(\mathbf{x}_i|\Omega) - b(\mathbf{z}_i|\Omega)]],$$

Proof. It is immediate from the expression for prior beliefs that

$$p^{BU}(\mathbf{x}|\mathbf{z}) = \frac{p(\mathbf{x}|\Omega)}{p(\mathbf{z}|\Omega)} = exp[\sum_{i<j}[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i\in\Gamma}[b(\mathbf{x}_i|\Omega) - b(\mathbf{z}_i|\Omega)]]$$

and

$$p(\mathbf{x}|\mathbf{z}) = exp[\sum_{i<j}[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i\in\Gamma}[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]],$$

$$= p^{BU}(\mathbf{x}|\mathbf{z}) \times exp[\sum_{i\in\Gamma}[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) - [b(\mathbf{x}_i|\Omega) - b(\mathbf{z}_i|\Omega)]].$$

The claims in the lemma follow from this. ■

C Appendix: Proof of Theorem 1 (Necessity)

C.1 IBGU implies Regularity

Proof. Observe that
$p(\mathbf{x}|\mathbf{z}) > 0$
$\iff a(\mathbf{x}_i\mathbf{x}_j) > -\infty$ for all distinct i, j and $b(\mathbf{x}_i|\mathbf{z}) > -\infty$ for all i (Lemma 2)
$\iff a(\mathbf{x}_i\mathbf{x}_j)a(\mathbf{x}_i\mathbf{z}_j) > -\infty$ for all distinct i, j, $a(\mathbf{x}_i\mathbf{z}_i) > -\infty$ for all i and $b(\mathbf{x}_i|\mathbf{z}), b(\mathbf{x}_i|\Omega) > -\infty$ for all i, j (by the regularity of the associative network)
$\iff [a(\mathbf{x}_i\mathbf{x}_j) > -\infty$ for all distinct i, j and $b(\mathbf{x}_i|\Omega) > -\infty$ for all $i]$
and $[a(\mathbf{x}_i\mathbf{z}_j) > -\infty$ for all i, j and $b(\mathbf{x}_i|\mathbf{z}) > -\infty$ for all $i]$ (rearranging terms)
$\iff p(\mathbf{x}|\Omega) > 0$ and $p(\mathbf{x}_i|\mathbf{z}) > 0$ for all i (Lemma 2). ■

C.2 IBGU implies Information Regularity

Proof. Observe that $p(\mathbf{x}_i|\mathbf{z}) > 0$
$\implies a(\mathbf{x}_i\mathbf{z}_j) > -\infty$ for all j and $b(\mathbf{x}_i|\mathbf{z}) > -\infty$ (Lemma 2 (Aii))
$\implies a(\mathbf{x}_i\mathbf{z}_i) > -\infty$ and $b(\mathbf{x}_i|\Omega) > -\infty$ (by regularity of the network)
$\implies a(\mathbf{x}_i\mathbf{z}_i) > -\infty$ and $a(\mathbf{x}_i\Omega_j) > -\infty$ for all $j \neq i$ (by regularity of the network)
$\implies p(\mathbf{x}_i|\mathbf{z}_i) > 0$ (Lemma 2 (Aiii)), as desired. ■

C.3 IBGU implies ASD

Proof. WLOG suppose $\mathbf{x} \subset \mathbf{z}$. Given Lemmas 1 and 2, recall that

$$p(\mathbf{x}|\mathbf{z}) = exp[\sum_{i<j}[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i\in\Gamma}[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]]$$

$$= exp[\frac{1}{2}\sum_{i\in\Gamma}\sum_{i\neq j}[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i\in\Gamma}[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]]$$

and

$$p(\mathbf{x}_i|\mathbf{z}) = exp[\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})].$$

Observe that

$$\prod_{i\in\Gamma}p(\mathbf{x}_i|\mathbf{z}) = exp[\sum_{i\in\Gamma}\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i\in\Gamma}[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]].$$

So

$$\frac{p(\mathbf{x}|\mathbf{z})}{\prod_{i\in\Gamma}p(\mathbf{x}_i|\mathbf{z})} = exp[\frac{1}{2}\sum_{i\in\Gamma}\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] - \sum_{i\in\Gamma}\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)]]$$

$$= exp[\frac{1}{2}\sum_{i\in\Gamma}\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] - \frac{1}{2}\sum_{i\in\Gamma}\sum_{i\neq j\in\Gamma}[2a(\mathbf{x}_i\mathbf{z}_j) - 2a(\mathbf{z}_i\mathbf{z}_j)]$$

$$= exp[\frac{1}{2}\sum_{i\in\Gamma}\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] - \frac{1}{2}\sum_{i\in\Gamma}\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) + a(\mathbf{z}_i\mathbf{x}_j) - 2a(\mathbf{z}_i\mathbf{z}_j)]]$$

$$= exp[\frac{1}{2}\sum_{i\in\Gamma}\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{x}_j) + a(\mathbf{z}_i\mathbf{z}_j)]]$$

$$= exp[\sum_{i<j}[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{x}_j) + a(\mathbf{z}_i\mathbf{z}_j)]]$$

$$= \prod_{i<j} exp\left[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{x}_j) + a(\mathbf{z}_i\mathbf{z}_j)\right].$$

Invoking Lemma 5 we obtain:

$$\frac{p(\mathbf{x}|\mathbf{z})}{\prod_{i\in\Gamma}p(\mathbf{x}_i|\mathbf{z})} = \prod_{i<j}\frac{p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z})},$$

thereby establishing the desired result. ■

C.4 IBGU implies AUB

Proof. By Lemma 6,

$$\frac{p(\mathbf{x}|\mathbf{z})}{p^{BU}(\mathbf{x}|\mathbf{z})} = exp[\sum_{i\in\Gamma}[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) - [b(\mathbf{x}_i|\Omega) - b(\mathbf{z}_i|\Omega)]],$$

and

$$\frac{p(\mathbf{x}_i|\mathbf{z})}{p^{BU}(\mathbf{x}_i|\mathbf{z})} = exp[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) - b(\mathbf{x}_i|\Omega) + b(\mathbf{z}_i|\Omega)].$$

Put together, we obtain

$$p(\mathbf{x}|\mathbf{z}) = p^{BU}(\mathbf{x}|\mathbf{z}) \prod_{i\in\Gamma} \frac{p(\mathbf{x}_i|\mathbf{z})}{p^{BU}(\mathbf{x}_i|\mathbf{z})},$$

which rearranges to yield the desired result. ■

D Appendix: Proof of Theorem 1 (Sufficiency)

Define the bias function b as follows:

For any $i \in \Gamma$ and $\mathbf{x}_i \in \Sigma$ define $b(\mathbf{x}_i|\Omega)$ by the condition:

$$p(\mathbf{x}_i|\Omega) = exp[b(\mathbf{x}_i|\Omega)] \tag{3}$$

For any $(\mathbf{z}, i)$ s.t. $\mathbf{z} \in \Sigma^+$, define

$$b(\mathbf{z}_i|\mathbf{z}) = 0. \tag{4}$$

For any $\mathbf{x} \in \Sigma$, $\mathbf{z} \in \Sigma^+$ and $i \in \Gamma$ s.t. $\mathbf{x}_i \subset \mathbf{z}_i$, if $p(\mathbf{x}_i\mathbf{z}_{-i}|\Omega) = 0$ then let

$$b(\mathbf{x}_i|\mathbf{z}) = -\infty. \tag{5}$$

For any $\mathbf{x} \in \Sigma$, $\mathbf{z} \in \Sigma^+$ and $i \in \Gamma$ s.t. $\mathbf{x}_i \subset \mathbf{z}_i$, if $p(\mathbf{x}_i\mathbf{z}_{-i}|\Omega) > 0$ then define $b(\mathbf{x}_i|\mathbf{z})$ by

$$\frac{p(\mathbf{x}_i|\mathbf{z})}{p^{BU}(\mathbf{x}_i|\mathbf{z})} = exp[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) - b(\mathbf{x}_i|\Omega) + b(\mathbf{z}_i|\Omega)]. \tag{6}$$

Next we define an association function a as follows.

For any i, j and any $\mathbf{x}_i \notin \Sigma^+$, let

$$a(\mathbf{z}_j\mathbf{x}_i) = a(\mathbf{x}_i\mathbf{z}_j) = -\infty \text{ for all } \mathbf{z}_j. \tag{7}$$

For any $\mathbf{x} \in \Sigma$ and distinct $i, j \in \Gamma$ s.t. $\mathbf{x}_i, \mathbf{x}_j \in \Sigma^+$, define $a(\mathbf{x}_i\mathbf{x}_j)$ uniquely by:

$$\frac{p(\mathbf{x}_i\mathbf{x}_j|\Omega)}{p(\mathbf{x}_i|\Omega)p(\mathbf{x}_j|\Omega)} = exp\left[a(\mathbf{x}_i\mathbf{x}_j)\right]. \tag{8}$$

Finally, for any $i \in \Gamma$, $\mathbf{x}_i \in \Sigma^+$ and $\mathbf{z}_i \in \Sigma^+$ s.t. $\mathbf{x}_i \subset \mathbf{z}_i$, define

$$a(\mathbf{x}_i\mathbf{z}_i) := b(\mathbf{x}_i|\mathbf{z}_i). \tag{9}$$

We note in passing that as defined, a satisfies the following properties that define a normalized representation (see Corollary 1).

Lemma 7 *The functions (a, b) satisfy: for any distinct i, j and any $\mathbf{x}_i, \mathbf{z} \in \Sigma^+$,*

$$a(\mathbf{x}_i\Omega_j) = a(\Omega_i\Omega_j) = a(\mathbf{x}_i\mathbf{x}_i) = 0, \tag{10}$$

Proof. The claim that $a(\mathbf{x}_i\Omega_j) = a(\Omega_i\Omega_j) = 0$ follows immediately from (8), and the remaining claim follows from (4). ■

Given the definition of the functions (a, b), the definition (9) and the properties (Ai) and (Biii) below establishe that these functions constitute a regular associative network.

Lemma 8 *The functions (a, b) satisfy*
(Ai) $a(\mathbf{x}_i\mathbf{x}_j) = a(\mathbf{x}_j\mathbf{x}_i)$,
(Aii) for any $\mathbf{x}_i\mathbf{z}_j$, $[a(\mathbf{x}_i\mathbf{z}_j) > -\infty \implies a(\mathbf{x}_i\Omega_j) > -\infty]$
(Aiii) for any $\mathbf{x}_i \in \Sigma$, $a(\mathbf{x}_i\Omega_i) = b(\mathbf{x}_i|\Omega)$.
(Bi) $\mathbf{x}_i \in \Sigma^+ \iff b(\mathbf{x}_i|\Omega) > -\infty$.
(Bii) $\mathbf{z} \in \Sigma^+$ and $p(\mathbf{x}_i|\mathbf{z}) = 0 \implies b(\mathbf{x}_i|\mathbf{z}) = -\infty$.
(Biii) for any $\mathbf{z} \in \Sigma^+$ and $i \in \Gamma$,

$$b(\mathbf{x}_i|\mathbf{z}) > -\infty \implies a(\mathbf{x}_i\mathbf{z}_j) > -\infty \textit{ for all } j \textit{ and } b(\mathbf{x}_i|\Omega) > -\infty.$$

Proof.
(Ai) Follows from (8) and (7).

(Aii) We show that the contrapositive of (Aii) is satisfied. There are two cases to consider, one where $i = j$ and one where i, j are distinct. If $a(\mathbf{x}_i\Omega_i) = -\infty$ then by definition, $p(\mathbf{x}_i|\Omega) = 0$, that is, $\mathbf{x}_i \notin \Sigma^+$ and so by (7), $a(\mathbf{x}_i\mathbf{z}_i) = -\infty$ for all $\mathbf{z}_i$. This establishes the contrapositive of (Aii) for $i = j$. Next consider $a(\mathbf{x}_i\Omega_j) = -\infty$ for any distinct i, j. Since, by (10), $a(\mathbf{x}_i\Omega_j) = 0$ for all $\mathbf{x} \in \Sigma^+$ and distinct i, j, it follows that $\mathbf{x}_i \notin \Sigma^+$. Therefore by (7), $a(\mathbf{x}_i\mathbf{z}_j) = -\infty$ for all $\mathbf{z}_j$. This establishes the contrapositive of (Aii) for disjoint i, j. Thus (Aii) is established.

(Aiii) This follows from the equivalence of the definitions in (9) and (3).

(Bi) Follows from (3).

(Bii) Suppose $p(\mathbf{x}_i|\mathbf{z}) = 0$ and consider two cases. If $p(\mathbf{x}_i\mathbf{z}_{-i}|\Omega) = 0$ then $b(\mathbf{x}_i|\mathbf{z}) = -\infty$ by definition. On the other hand, if $p(\mathbf{x}_i\mathbf{z}_{-i}|\Omega) > 0$ then $\mathbf{x}_i \in \Sigma^+$ by Regularity, and in turn $b(\mathbf{x}_i|\Omega) > -\infty$ by (Bi). Moreover, given $\mathbf{z} \in \Sigma^+$ and the definition of a above, we have $b(\mathbf{z}_i|\Omega), b(\mathbf{z}_i|\mathbf{z}) > -\infty$. Consequently, by (6) and $b(\mathbf{x}_i|\Omega), b(\mathbf{z}_i|\Omega), b(\mathbf{z}_i|\mathbf{z}) > -\infty$, it follows that $p(\mathbf{x}_i|\mathbf{z}) = 0$ implies $b(\mathbf{x}_i|\mathbf{z}) = -\infty$, thereby establishing (Bii).

(Biii) Suppose $b(\mathbf{x}_i|\mathbf{z}) > -\infty$. We see that
$b(\mathbf{x}_i|\mathbf{z}) > -\infty$
$\implies p(\mathbf{x}_i|\mathbf{z}) > 0$ by (Bii)
$\implies p(\mathbf{x}_i\mathbf{z}_{-i}|\Omega) > 0$ by Regularity
$\implies p(\mathbf{x}_i\mathbf{z}_{-i}|\Omega) > 0$ and $p(\mathbf{x}_i|\Omega), p(\mathbf{z}_j|\Omega) > 0$ for all $i \neq j \in \Gamma$ by Regularity
$\implies p(\mathbf{x}_i\mathbf{z}_j|\mathbf{z}) > 0$ for all $i \neq j \in \Gamma$ by ASD
$\implies a(\mathbf{x}_i\mathbf{z}_j) > -\infty$ for all $i \neq j \in \Gamma$ by (8), which establishes part of the desired assertion. Since $p(\mathbf{x}_i|\Omega) > 0$, we also obtain the assertion $b(\mathbf{x}_i|\Omega) > -\infty$ by (3). Finally, since we have already seen that $p(\mathbf{x}_i|\mathbf{z}) > 0$ and $p(\mathbf{x}_i\mathbf{z}_{-i}|\Omega) > 0$, invoke Information Regularity to obtain $p(\mathbf{x}_i|\mathbf{z}_i) > 0$, and in turn use (9) and (6) to obtain $a(\mathbf{x}_i\mathbf{z}_i) > -\infty$. This completes the proof. ■

Lemma 9 *The functions* (a, b) *are a IBGU representation for* $\mathbf{p}$*: for all* $\mathbf{x} \in \Sigma$ *and* $\mathbf{z} \in \Sigma^+$*,*

$$p(\mathbf{x}|\mathbf{z}) = exp[\sum_{i<j} [a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{j\in\Gamma} [b(\mathbf{x}_j|\mathbf{z}) - b(\mathbf{z}_j|\mathbf{z})]].$$

Proof. We prove this lemma in a sequence of steps.

Step 1. Show that the IBGU representation holds for any $\mathbf{x} \in \Sigma$ and $\mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$ and $p(\mathbf{x}|\mathbf{z}) = 0$.

Consider any $\mathbf{x} \in \Sigma$ and $\mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$ and $p(\mathbf{x}|\mathbf{z}) = 0$. By Regularity, either $p(\mathbf{x}_i|\mathbf{z}) = 0$ for some i, or [$p(\mathbf{x}_i|\mathbf{z}) > 0$ for all i but $p(\mathbf{x}|\Omega) = 0$]. In the former case, (Bii) (in Lemma 8) implies $b(\mathbf{x}_i|\mathbf{z}) = -\infty$, and in the latter case, ASD implies $p(\mathbf{x}_i\mathbf{x}_j|\Omega) = 0$ for some distinct i, j, and so by (8), it must be that $a(\mathbf{x}_i\mathbf{x}_j) = -\infty$. That is, either $b(\mathbf{x}_i|\mathbf{z}) = -\infty$ for some i or $a(\mathbf{x}_i\mathbf{x}_j) = -\infty$ for some distinct i, j, and so we see that

$$p(\mathbf{x}|\mathbf{z}) = 0 = exp[\sum_{i<j} [a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{j\in\Gamma} [b(\mathbf{x}_j|\mathbf{z}) - b(\mathbf{z}_j|\mathbf{z})]],$$

establishing that the model represents $\mathbf{p}$ on events assigned a zero likelihood.

Step 2. Show that the IBGU representation holds for any $\mathbf{x} \in \Sigma$ s.t. $p(\mathbf{x}|\Omega) > 0$.

Take $\mathbf{x} \in \Sigma$ s.t. $p(\mathbf{x}|\Omega) > 0$. By Regularity, $\mathbf{x}_i \in \Sigma^+$ for all i and so (8) comes into play. By ASD, (8), (3) and (4),

$$\begin{aligned} p(\mathbf{x}|\Omega) &= \left[\prod_{i<j} \frac{p(\mathbf{x}_i\mathbf{x}_j|\Omega)}{p(\mathbf{x}_i|\Omega)p(\mathbf{x}_j|\Omega)}\right] \times \prod_{i\in\Gamma} p(\mathbf{x}_i|\Omega) \\ &= exp[\sum_{i<j} a(\mathbf{x}_i\mathbf{x}_j) + \sum_{i\in\Gamma} [b(\mathbf{x}_i|\Omega) - b(\Omega_i|\Omega)]]. \end{aligned}$$

Since (10) yields $a(\Omega_i\Omega_j) = 0$, this can be written in the desired form:

$$= exp[\left[\sum_{i<j} a(\mathbf{x}_i\mathbf{x}_j) - a(\Omega_i\Omega_j)\right] + \sum_{i\in\Gamma} [b(\mathbf{x}_i|\Omega) - b(\Omega_i|\Omega)]].$$

Step 3. Show that the IBGU representation holds for any $\mathbf{x} \in \Sigma$ and $\mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$ and $p(\mathbf{x}|\mathbf{z}) > 0$.

Take any $\mathbf{x} \in \Sigma$ and $\mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$ and $p(\mathbf{x}|\mathbf{z}) > 0$. By Regularity $p(\mathbf{x}|\Omega) > 0$. For any i, Regularity also implies $p(\mathbf{x}_i|\mathbf{z}) > 0$, that is $p(\mathbf{x}_i\mathbf{z}_{-i}|\mathbf{z}) > 0$, and yet another application of Regularity yields $p(\mathbf{x}_i\mathbf{z}_{-i}|\Omega) > 0$. Therefore we can bring AUB into play. Use the result of Step 2 to compute that

$$p^{BU}(\mathbf{x}|\mathbf{z}) := \frac{p(\mathbf{x}|\Omega)}{p(\mathbf{z}|\Omega)} = exp[\left[\sum_{i<j} a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)\right] + \sum_{j\in\Gamma}[b(\mathbf{x}_j|\Omega) - b(\mathbf{z}_j|\Omega)]].$$

Use this together with AUB and (6) to determine that

$$p(\mathbf{x}|\mathbf{z}) = p^{BU}(\mathbf{x}|\mathbf{z}) \prod_{i\in\Gamma} \frac{p(\mathbf{x}_i|\mathbf{z})}{p^{BU}(\mathbf{x}_i|\mathbf{z})}$$

$$= exp[\left[\sum_{i<j} a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)\right] + \sum_{j\in\Gamma}[b(\mathbf{x}_j|\Omega) - b(\mathbf{z}_j|\Omega)]]$$
$$+ \sum_{j\in\Gamma}[b(\mathbf{x}_j|\mathbf{z}) - b(\mathbf{z}_j|\mathbf{z})] - \sum_{j\in\Gamma}[b(\mathbf{x}_j|\Omega) - b(\mathbf{z}_j|\Omega)]]$$
$$= exp[\left[\sum_{i<j} a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)\right] + \sum_{j\in\Gamma}[b(\mathbf{x}_j|\mathbf{z}) - b(\mathbf{z}_j|\mathbf{z})]],$$

as desired. ■

E Appendix: Proof for Theorem 2

E.1 Necessity of IS and Strong IR

Proof. Take a normalized IBIU representation a. We first make a few observations.

Take $\mathbf{x}_i, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x}_i \subset \mathbf{z}_i$. By Lemma 2,

$$p(\mathbf{x}_i|\mathbf{z}) = exp[2 \sum_{i\neq j\in\Gamma} [a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + [a(\mathbf{x}_i\mathbf{z}_i) - a(\mathbf{z}_i\mathbf{z}_i)]].$$

Since $\mathbf{x}_i, \mathbf{z} \in \Sigma^+$ (in which case, by Regularity, $\mathbf{x}_i, \mathbf{z}_j \in \Sigma^+$ for all j) we can invoke Proposition 2 to obtain

$$p(\mathbf{x}_i|\mathbf{z}) = \left[\prod_{i\neq j\in\Gamma} \frac{SD(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]^2 p(\mathbf{x}_i|\mathbf{z}_i).$$

For any $i \neq k \in \Gamma$, write

$$p(\mathbf{x}_i|\mathbf{z}) = \left[\frac{SD(\mathbf{x}_i\mathbf{z}_k|\Omega)}{SD(\mathbf{z}_i\mathbf{z}_k|\Omega)}\right]^2 \left[\prod_{i,k\neq j\in\Gamma} \frac{SD(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]^2 p(\mathbf{x}_i|\mathbf{z}_i).$$

Since $SD(\mathbf{x}_i\Omega_k|\Omega) = SD(\mathbf{z}_i\Omega_k|\Omega) = 1$, we see that

$$p(\mathbf{x}_i|\Omega_k\mathbf{z}_{-k}) = \left[\frac{SD(\mathbf{x}_i\Omega_k|\Omega)}{SD(\mathbf{z}_i\Omega_k|\Omega)}\right]^2 \left[\prod_{i,k\neq j\in\Gamma} \frac{SD(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]^2 p(\mathbf{x}_i|\mathbf{z}_i)$$

$$= \left[\prod_{i,k\neq j\in\Gamma} \frac{SD(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]^2 p(\mathbf{x}_i|\mathbf{z}_i).$$

Similarly, when $\mathbf{z}_i$ is replaced with Ω_i then

$$p(\mathbf{x}_i|\Omega_i\mathbf{z}_{-i}) = \left[\frac{SD(\mathbf{x}_i\mathbf{z}_k|\Omega)}{SD(\Omega_i\mathbf{z}_k|\Omega)}\right]^2 \left[\prod_{i,k\neq j\in\Gamma} \frac{SD(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD(\Omega_i\mathbf{z}_j|\Omega)}\right]^2 p(\mathbf{x}_i|\Omega_i)$$

$$= [SD(\mathbf{x}_i\mathbf{z}_k|\Omega)]^2 \left[\prod_{i,k\neq j\in\Gamma} SD(\mathbf{x}_i\mathbf{z}_j|\Omega)\right]^2 p(\mathbf{x}_i|\Omega_i).$$

To establish Strong IR, suppose $p(\mathbf{x}_i|\mathbf{z}) > 0$. Then by Regularity, $\mathbf{x}_i \in \Sigma^+$, and so we can invoke the above expression for $p(\mathbf{x}_i|\mathbf{z})$ to deduce that all its components are strictly positive. It follows that, for any $i \neq k \in \Gamma$, all the components of the expression for $p(\mathbf{x}_i|\Omega_k\mathbf{z}_{-k})$ are strictly positive too, and consequently, $p(\mathbf{x}_i|\Omega_k\mathbf{z}_{-k}) > 0$. Similarly, the squared terms in the expression for $p(\mathbf{x}_i|\Omega_i\mathbf{z}_{-i})$ are positive, and by two applications of Regularity, $p(\mathbf{x}_i|\mathbf{z}) > 0$ implies $p(\mathbf{x}_i|\Omega_i) > 0$. Consequently $p(\mathbf{x}_i|\Omega_i\mathbf{z}_{-i}) > 0$ and Strong IR is established.

To establish IS, observe that for any $i \neq k \in \Gamma$ s.t. $p(\mathbf{x}_i|\Omega_k\mathbf{z}_{-k}) > 0$, Regularity implies $\mathbf{x}_i \in \Sigma^+$ and so by the preceding expressions,

$$\frac{p(\mathbf{x}_i|\mathbf{z}_k\mathbf{z}_{-k})}{p(\mathbf{x}_i|\Omega_k\mathbf{z}_{-k})} = \frac{\left[\frac{SD(\mathbf{x}_i\mathbf{z}_k|\Omega)}{SD(\mathbf{z}_i\mathbf{z}_k|\Omega)}\right]^2}{\left[\frac{SD(\mathbf{x}_i\Omega_k|\Omega)}{SD(\mathbf{z}_i\Omega_k|\Omega)}\right]^2} = \left[\frac{SD(\mathbf{x}_i\mathbf{z}_k|\Omega)}{SD(\mathbf{z}_i\mathbf{z}_k|\Omega)}\right]^2 = \frac{p(\mathbf{x}_i|\mathbf{z}_k)}{p(\mathbf{x}_i|\Omega_k)},$$

as desired. ■

E.2 Necessity of IU

Proof. Suppose $i \neq j$. Take any $\mathbf{x}_i\mathbf{z}_j \in \Sigma^+$. By Regularity $\mathbf{x}_i, \mathbf{z}_j \in \Sigma^+$. Then by Lemma 6 ,

$$\frac{p(\mathbf{x}_i|\mathbf{z}_j)}{p^{BU}(\mathbf{x}_i|\mathbf{z}_j)} = exp[[b(\mathbf{x}_i|\mathbf{z}_j) - b(\Omega_i|\mathbf{z}_j) - [b(\mathbf{x}_i|\Omega) - b(\Omega_i|\Omega)]],$$

$$= exp[a(\mathbf{x}_i\mathbf{z}_j) - a(\Omega_i\mathbf{z}_j) - a(\mathbf{x}_i\Omega_j) + a(\Omega_i\Omega_j) + \sum_{k\neq j} a(\mathbf{x}_i\Omega_k) - a(\Omega_i\Omega_k) - a(\mathbf{x}_i\Omega_k) + a(\Omega_i\Omega_k)]$$

$$= exp[a(\mathbf{x}_i\mathbf{z}_j) - a(\Omega_i\mathbf{z}_j) - a(\mathbf{x}_i\Omega_j) + a(\Omega_i\Omega_j)]$$

and by Lemma 5,

$$\frac{p(\mathbf{x}_i\mathbf{z}_j|\Omega)}{p(\mathbf{x}_i|\Omega)p(\mathbf{z}_j|\Omega)} = exp\left[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{x}_i\Omega_j) - a(\Omega_i\mathbf{z}_j) + a(\Omega_i\Omega_j)\right].$$

Conclude that $\frac{p(\mathbf{x}_i|\mathbf{z}_j)}{p^{BU}(\mathbf{x}_i|\mathbf{z}_j)} = \frac{p(\mathbf{x}_i\mathbf{z}_j|\Omega)}{p(\mathbf{x}_i|\Omega)p(\mathbf{z}_j|\Omega)}$. ■

E.3 Proof of Sufficiency

It will be useful to first note that

Lemma 10 *Information Separability and Strong Information Regularity imply that for any* $\mathbf{x}_i \in \Sigma$ *and* $\mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x}_i \subset \mathbf{z}_i$*,*

$$p(\mathbf{x}_i|\mathbf{z}) > 0 \iff p(\mathbf{x}_i|\mathbf{z}_j) > 0 \text{ for all } j \in \Gamma.$$

Proof. Take any $\mathbf{x}_i \in \Sigma$ and $\mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x}_i \subset \mathbf{z}_i$. If $p(\mathbf{x}_i|\mathbf{z}) > 0$ then, by Strong IR, $p(\mathbf{x}_i|\Omega_j\mathbf{z}_{-j}) > 0$ for any j. Repeatedly applying this condition yields $p(\mathbf{x}_i|\mathbf{z}_j) > 0$ for all $j \in \Gamma$. Conversely, suppose $p(\mathbf{x}_i|\mathbf{z}_j) > 0$ for all $j \in \Gamma$. Then by IS, $p(\mathbf{x}_i|\mathbf{z}_i) > 0$ implies

$$\frac{p(\mathbf{x}_i|\mathbf{z}_j\mathbf{z}_i)}{p(\mathbf{x}_i|\mathbf{z}_i)} = \frac{p(\mathbf{x}_i|\mathbf{z}_j)}{p(\mathbf{x}_i|\Omega)},$$

and since $\frac{p(\mathbf{x}_i|\mathbf{z}_j)}{p(\mathbf{x}_i|\Omega)} > 0$ it follows that $p(\mathbf{x}_i|\mathbf{z}_j\mathbf{z}_i) > 0$. This argument can be repeated using $p(\mathbf{x}_i|\mathbf{z}_j\mathbf{z}_i) > 0$ and $p(\mathbf{x}_i|\mathbf{z}_k) > 0$ to obtain $p(\mathbf{x}_i|\mathbf{z}_k\mathbf{z}_j\mathbf{z}_i) > 0$. Continue in this fashion to establish $p(\mathbf{x}_i|\mathbf{z}) > 0$. ■

By hypothesis, $\mathbf{p}$ admits an IBGU representation. Consider the regular (a, b) tuple defined in the proof of Theorem 1. Then there exists an IBIU representation if (a, b) are such that:

Lemma 11 *For any* $\mathbf{x}_i \in \Sigma$ *and* $\mathbf{z} \in \Sigma^+$ *s.t.* $\mathbf{x}_i \subset \mathbf{z}_i$*,*

$$b(\mathbf{x}_i|\mathbf{z}) = \sum_{j\in\Gamma} a(\mathbf{x}_i\mathbf{z}_j). \tag{11}$$

Proof. Consider any $\mathbf{z} \in \Sigma^+$, $i \in \Gamma$ and $\mathbf{x}_i \in \Sigma$ s.t. $\mathbf{x}_i \subset \mathbf{z}_i$. If $a(\mathbf{x}_i\mathbf{z}_j) = -\infty$ for some j, then $b(\mathbf{x}_i|\mathbf{z}) = -\infty$ holds by regularity of (a, b), thereby satisfying (11). So consider the case that $a(\mathbf{x}_i\mathbf{z}_j) > -\infty$ for all $j \in \Gamma$. Observe that by (7) it must be that $\mathbf{x}_i \in \Sigma^+$

We first show that $p(\mathbf{x}_i|\mathbf{z}) > 0$ and $p(\mathbf{x}_i|\mathbf{z}_j) > 0$ for all $j \in \Gamma$: By (9) and (3), $a(\mathbf{x}_i\mathbf{z}_i) > -\infty$ implies $p(\mathbf{x}_i|\mathbf{z}_i) > 0$. Moreover, given $\mathbf{x}_i \in \Sigma^+$ and (8), for any $j \neq i$, $a(\mathbf{x}_i\mathbf{z}_j) > -\infty$ implies $\frac{p(\mathbf{x}_i\mathbf{z}_j|\Omega)}{p(\mathbf{x}_i|\Omega)p(\mathbf{z}_j|\Omega)} > 0$, which by IU implies $p(\mathbf{x}_i|\mathbf{z}_j) > 0$, as desired. Therefore, $p(\mathbf{x}_i|\mathbf{z}_j) > 0$ for all $j \in \Gamma$, and lemma 10 then implies $p(\mathbf{x}_i|\mathbf{z}) > 0$, as desired.

Next observe that since, by hypothesis, $\mathbf{p}$ admits an IBGU representation, and since $b(\mathbf{z}_i|\mathbf{z}) = 0$ (by (4)), for any $j \neq i$ we can write

$$p(\mathbf{x}_i|\mathbf{z}) = exp\left[\sum_{i \neq k \in \Gamma} [a(\mathbf{x}_i\mathbf{z}_k) - a(\mathbf{z}_i\mathbf{z}_k)] + [b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]\right]$$

$$= exp\left[\sum_{i,j \neq k \in \Gamma} [a(\mathbf{x}_i\mathbf{z}_k) - a(\mathbf{z}_i\mathbf{z}_k)]\right] \times exp\left[[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)]\right] + b(\mathbf{x}_i|\mathbf{z})\right].$$

Then

$$\frac{p(\mathbf{x}_i|\Omega_j\mathbf{z}_{-j})}{p(\mathbf{x}_i|\mathbf{z}_j\mathbf{z}_{-j})} = \frac{exp\left[[a(\mathbf{x}_i\Omega_j) - a(\mathbf{z}_i\Omega_j)] + b(\mathbf{x}_i|\Omega_j\mathbf{z}_{-j})\right]}{exp\left[[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + b(\mathbf{x}_i|\mathbf{z}_j\mathbf{z}_{-j})\right]}$$

$$= \frac{exp\left[a(\mathbf{x}_i\Omega_j) - a(\mathbf{z}_i\Omega_j)\right]}{exp\left[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)\right]} exp[b(\mathbf{x}_i|\Omega_j\mathbf{z}_{-j}) - b(\mathbf{x}_i|\mathbf{z}_j\mathbf{z}_{-j})].$$

We use this expression below.

Recall that we are in a case where $i \in \Gamma$, $\mathbf{x}_i, \mathbf{z} \in \Sigma^+$ and $p(\mathbf{x}_i|\mathbf{z}) > 0$. Respectively by Strong IR and by hypothesis, $p(\mathbf{x}_i|\Omega_j\mathbf{z}_{-j}) > 0$ and $p(\mathbf{x}_i|\mathbf{z}_j\Omega_{-j}) > 0$. By IS, for any $j \neq i$ we can therefore write $\frac{p(\mathbf{x}_i|\Omega_j\mathbf{z}_{-j})}{p(\mathbf{x}_i|\mathbf{z}_j\mathbf{z}_{-j})} = \frac{p(\mathbf{x}_i|\Omega_j\Omega_{-j})}{p(\mathbf{x}_i|\mathbf{z}_j\Omega_{-j})}$. Applying the above expression to these ratios yields a key equation:

$$b(\mathbf{x}_i|\Omega_j\mathbf{z}_{-j}) - b(\mathbf{x}_i|\mathbf{z}_j\mathbf{z}_{-j}) = b(\mathbf{x}_i|\Omega_j\Omega_{-j}) - b(\mathbf{x}_i|\mathbf{z}_j\Omega_{-j}).$$

We also observe that by IU (and Lemmas 5 and 6),

$$exp[b(\mathbf{x}_i|\mathbf{z}_j) - b(\Omega_i|\mathbf{z}_j) - b(\mathbf{x}_i|\Omega) + b(\Omega_i|\Omega)] = \frac{p(\mathbf{x}_i|\mathbf{z}_j)}{p^{BU}(\mathbf{x}_i|\mathbf{z}_j)}$$

$$= \frac{p(\mathbf{x}_i\mathbf{z}_j|\Omega)}{p(\mathbf{x}_i|\Omega)p(\mathbf{z}_j|\Omega)} = exp\left[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{x}_i\Omega_j) - a(\Omega_i\mathbf{z}_j) + a(\Omega_i\Omega_j)\right]$$

which, by (10), yields a second key equation:

$$b(\mathbf{x}_i|\mathbf{z}_j) - b(\mathbf{x}_i|\Omega) = a(\mathbf{x}_i\mathbf{z}_j).$$

Using the two key equations, we obtain for any $j, k \neq i$,

$b(\mathbf{x}_i|\mathbf{z}_j\mathbf{z}_{-j}) = b(\mathbf{x}_i|\mathbf{z}_j) - b(\mathbf{x}_i|\Omega) + b(\mathbf{x}_i|\Omega_j\mathbf{z}_{-j})$
$= a(\mathbf{x}_i\mathbf{z}_j) + b(\mathbf{x}_i|\Omega_j\mathbf{z}_{-j})$
$= a(\mathbf{x}_i\mathbf{z}_j) + a(\mathbf{x}_i\mathbf{z}_k) + b(\mathbf{x}_i|\Omega_j\Omega_k\mathbf{z}_{-jk})$
$\vdots$
$= \sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j)] + b(\mathbf{x}_i|\mathbf{z}_i\Omega_{-i})$
$= \sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j)] + a(\mathbf{x}_i\mathbf{z}_i)$ by (9)
$= \sum_{j\in\Gamma} a(\mathbf{x}_i\mathbf{z}_j)$, as desired. ■

F Appendix: Proof of Theorem 3

Proof. Consider two associative networks (a, b) and (α, β) that satisfy (1). Suppose that (a, b) and (α, β) represent the same IBGU $\mathbf{p}$. For any $i, j \in \Gamma$ and any $\mathbf{x}_i \in \Sigma^+$ define $\gamma(\mathbf{x}_i, \Omega_j)$ by

$$\gamma(\mathbf{x}_i\Omega_j) = a(\mathbf{x}_i\Omega_j) - \alpha(\mathbf{x}_i\Omega_j).$$

Since $\mathbf{x}_i \in \Sigma^+$, by Lemma 4 $\gamma(\mathbf{x}_i\Omega_j)$ is a real number for each distinct i, j and by symmetry of a, α, it is also true that $\gamma(\mathbf{x}_i\Omega_j) = a(\Omega_j\mathbf{x}_i) - \alpha(\Omega_j\mathbf{x}_i)$. If $\mathbf{x}_i \notin \Sigma^+$ then let $\gamma(\mathbf{x}_i\Omega_j)$ be an arbitrary real number.

For any $i \in \Gamma$ and $\mathbf{x}_i, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x}_i \subset \mathbf{z}_i$ and $p(\mathbf{x}_i|\mathbf{z}) > 0$ define $\psi(\mathbf{x}_i|\mathbf{z})$ by

$$\psi(\mathbf{x}_i|\mathbf{z}) = b(\mathbf{x}_i|\mathbf{z}) - \beta(\mathbf{x}_i|\mathbf{z}).$$

Given $p(\mathbf{x}_i|\mathbf{z}) > 0$, Lemma 4 implies that $\psi(\mathbf{x}_i|\mathbf{z})$ must be real-valued.

We proceed in sequence of steps.

Step 1: Show that for any distinct $i, j \in \Gamma$, $\mathbf{x}_i \in \Sigma$ and $\mathbf{z}_j \in \Sigma^+$,

$$a(\mathbf{x}_i\mathbf{z}_j) = \alpha(\mathbf{x}_i\mathbf{z}_j) + [\gamma(\mathbf{x}_i\Omega_j) + \gamma(\mathbf{z}_j\Omega_i) - \gamma(\Omega_i\Omega_j)].$$

Take any distinct $i, j \in \Gamma$, $\mathbf{x}_i \in \Sigma$ and $\mathbf{z}_j \in \Sigma^+$. First suppose $\mathbf{x}_i \notin \Sigma^+$. By Lemma 4, $a(\mathbf{x}_i\Omega_j) = -\infty$ or $b(\mathbf{x}_i|\Omega) = -\infty$. By (1), $a(\mathbf{x}_i\mathbf{z}_j) = \alpha(\mathbf{x}_i\mathbf{z}_j) = -\infty$ for all $\mathbf{z}_j \in \Sigma_j$. Since $\gamma > -\infty$ by definition, the desired equality holds trivially.

Suppose next that $\mathbf{x}_i \in \Sigma^+$. Given Lemma 5 and the definition of γ, we see that:
$exp\,[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{x}_i\Omega_j) - a(\Omega_i\mathbf{z}_j) + a(\Omega_i\Omega_j)]$
$= \frac{p(\mathbf{x}_i\mathbf{z}_j|\Omega)}{p(\mathbf{x}_i|\Omega)p(\mathbf{z}_j|\Omega)}$
$= exp\,[\alpha(\mathbf{x}_i\mathbf{z}_j) - \alpha(\mathbf{x}_i\Omega_j) - \alpha(\mathbf{z}_j\Omega_i) + \alpha(\Omega_i\Omega_j)]$
$= exp\,[\alpha(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{x}_i\Omega_j) + \gamma(\mathbf{x}_i\Omega_j) - a(\Omega_i\mathbf{z}_j) + \gamma(\mathbf{z}_j\Omega_i) + a(\Omega_i\Omega_j) - \gamma(\Omega_i\Omega_j)]$,that is,

$$a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{x}_i\Omega_j) - a(\Omega_i\mathbf{z}_j) + a(\Omega_i\Omega_j)$$
$$= \alpha(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{x}_i\Omega_j) + \gamma(\mathbf{x}_i\Omega_j) - a(\Omega_i\mathbf{z}_j) + \gamma(\mathbf{z}_j\Omega_i) + a(\Omega_i\Omega_j) - \gamma(\Omega_i\Omega_j).$$

Since $p(\mathbf{x}_i|\Omega)p(\mathbf{z}_j|\Omega) > 0$, Lemmas 3 and 4 imply $a(\mathbf{x}_i\Omega_j), a(\Omega_i\mathbf{z}_j), a(\Omega_i\Omega_j) > -\infty$ for all i, j. If $p(\mathbf{x}_i\mathbf{z}_j|\Omega) > 0$ then $a(\mathbf{x}_i\mathbf{z}_j), \alpha(\mathbf{x}_i\mathbf{z}_j) > -\infty$ and we obtain the desired equality from the preceding.[20] If $p(\mathbf{x}_i\mathbf{z}_j|\Omega) = 0$ then $a(\mathbf{x}_i\mathbf{z}_j), \alpha(\mathbf{x}_i\mathbf{z}_j) = -\infty$ and the desired equality holds anyway. This completes the proof of the step.

Step 2: Show that for each $i \in \Gamma$ and $\mathbf{x}_i, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x}_i \subset \mathbf{z}_i$ and $p(\mathbf{x}_i|\mathbf{z}) > 0$,

$$\sum_{i \neq j \in \Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)] + \psi(\mathbf{x}_i|\mathbf{z}) - \psi(\mathbf{z}_i|\mathbf{z}) = 0$$

Take any $i \in \Gamma$ and $\mathbf{x}_i, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x}_i \subset \mathbf{z}_i$ and $p(\mathbf{x}_i|\mathbf{z}) > 0$. By Lemma 2, the definition of ψ and step 1,

$$\begin{aligned}
&exp[\textstyle\sum_{i\neq j} [a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + [b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]] \\
&= p(\mathbf{x}_i|\mathbf{z}) \\
&= exp[\textstyle\sum_{i\neq j} [\alpha(\mathbf{x}_i\mathbf{z}_j) - \alpha(\mathbf{z}_i\mathbf{z}_j)] + [\beta(\mathbf{x}_i|\mathbf{z}) - \beta(\mathbf{z}_i|\mathbf{z})]] \\
&= exp[\textstyle\sum_{i\neq j} [a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j) + [\gamma(\mathbf{x}_i\Omega_j) + \gamma(\mathbf{z}_j\Omega_i)] - [\gamma(\mathbf{z}_i\Omega_j) + \gamma(\mathbf{z}_j\Omega_i)]] \\
&\quad + [b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) + \psi(\mathbf{x}_i|\mathbf{z}) - \psi(\mathbf{z}_i|\mathbf{z})]] \\
&= exp[\textstyle\sum_{i\neq j} [a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j) + \gamma(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)]] \\
&\quad + [b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) + \psi(\mathbf{x}_i|\mathbf{z}) - \psi(\mathbf{z}_i|\mathbf{z})]] \\
&= exp[\textstyle\sum_{i\neq j} [a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + [b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})] \\
&\quad + \textstyle\sum_{i\neq j} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)] + \psi(\mathbf{x}_i|\mathbf{z}) - \psi(\mathbf{z}_i|\mathbf{z})].
\end{aligned}$$

Since $p(\mathbf{x}_i|\mathbf{z}) > 0$, all the terms in the first an last expression are real-valued. The assertion follows.

Step 3: Show that for each $i \in \Gamma$, $\mathbf{x}_i \in \Sigma$ and $\mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x}_i \subset \mathbf{z}_i$,

$$b(\mathbf{x}_i|\mathbf{z}) = \beta(\mathbf{x}_i|\mathbf{z}) + \psi(\mathbf{z}_i|\mathbf{z}) - \sum_{i \neq j \in \Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)]\,.$$

First consider the case $\mathbf{x}_i\mathbf{z}_{-i} \notin \Sigma^+$. Then by Regularity $\mathbf{x}_i \notin \Sigma^+$ (since by Regularity $\mathbf{z}_j \in \Sigma^+$ for any $\mathbf{z} \in \Sigma^+$), which by Lemma 4, implies that either $a(\mathbf{x}_i\Omega_j) = -\infty$ for some j or $b(\mathbf{x}_i|\Omega) = -\infty$. In the former case, the regularity of the network implies $b(\mathbf{x}_i|\Omega) = -\infty$, as well. But by regularity of the network, $b(\mathbf{x}_i|\Omega) = -\infty$ in turn implies $b(\mathbf{x}_i|\mathbf{z}) = -\infty$. By the same argument, $\beta(\mathbf{x}_i|\mathbf{z}) = -\infty$ as well. Consequently the desired equality in the statement of the step is trivially satisfied, given that $b(\mathbf{z}_i|\mathbf{z}), \beta(\mathbf{z}_i|\mathbf{z}) > -\infty$ on account of $\mathbf{z} \in \Sigma^+$, and since $\psi > -\infty$.

Consider next the case that $\mathbf{x}_i\mathbf{z}_{-i} \in \Sigma^+$. Invoke Lemma 6 and the definition of ψ to see that

$$\begin{aligned}
&exp[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) - b(\mathbf{x}_i|\Omega) + b(\mathbf{z}_i|\Omega)] \\
&= \tfrac{p(\mathbf{x}_i|\mathbf{z})}{p^{BU}(\mathbf{x}_i|\mathbf{z})} \\
&= exp[\beta(\mathbf{x}_i|\mathbf{z}) - \beta(\mathbf{z}_i|\mathbf{z}) - \beta(\mathbf{x}_i|\Omega) + \beta(\mathbf{z}_i|\Omega)] \\
&= exp[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) + \psi(\mathbf{x}_i|\mathbf{z}) - \psi(\mathbf{z}_i|\mathbf{z}) - b(\mathbf{x}_i|\Omega) + b(\mathbf{z}_i|\Omega) - \psi(\mathbf{x}_i|\Omega) + \psi(\mathbf{z}_i|\Omega)].
\end{aligned}$$

[20]Note that if $\mathbf{x}_i = \Omega_i$ then the desired equality holds trivially by definition of γ. Similarly if $\mathbf{z}_j = \Omega_j$.

Since $\mathbf{z} \in \Sigma^+$ and $p(\mathbf{x}_i|\mathbf{z}) > 0$, all the terms are real-valued, and so form the equalities we obtain

$$\psi(\mathbf{x}_i|\mathbf{z}) - \psi(\mathbf{z}_i|\mathbf{z}) = \psi(\mathbf{x}_i|\Omega) - \psi(\mathbf{z}_i|\Omega). \tag{12}$$

Moreover, by definition of ψ,

$$b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) = \beta(\mathbf{x}_i|\mathbf{z}) - \beta(\mathbf{x}_i|\mathbf{z}) + [\psi(\mathbf{x}_i|\mathbf{z}) - \psi(\mathbf{z}_i|\mathbf{z})].$$

Combine this with (12) and step 2 to obtain

$$b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) = \beta(\mathbf{x}_i|\mathbf{z}) - \beta(\mathbf{z}_i|\mathbf{z}) - \sum_{i \neq j \in \Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)].$$

Finally, use $\psi(\mathbf{z}_i|\mathbf{z}) = b(\mathbf{z}_i|\mathbf{z}) - \beta(\mathbf{z}_i|\mathbf{z})$ to establish step 3.

Step 4: Establish the converse.

Suppose that p is represented by (a, b), and suppose that (α, β) is an associative network that satisfies the conditions in the theorem. We show that p must be represented by (α, β) as well.

By Lemma 2, for any $\mathbf{x} \subset \mathbf{z} \in \Sigma^+$,

$$p(\mathbf{x}|\mathbf{z}) = exp[\sum_{i<j} [a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i \in \Gamma} [b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})].$$

By Lemma 4, $a(\mathbf{z}_i\mathbf{z}_j), b(\mathbf{z}_i|\mathbf{z}) > -\infty$ for any $\mathbf{z} \in \Sigma^+$ and i, j. By the conditions, we also have $\alpha(\mathbf{z}_i\mathbf{z}_j), \beta(\mathbf{z}_i|\mathbf{z}) > -\infty$ for any i, j.

First suppose that $p(\mathbf{x}|\mathbf{z}) = 0$. By Lemma 4, $a(\mathbf{x}_i\mathbf{x}_j) = -\infty$ for some distinct i, j or $b(\mathbf{x}_i|\mathbf{z}) = -\infty$ for some i, and by the conditions relating (α, β) with (a, b), it follows that either $\alpha(\mathbf{x}_i\mathbf{x}_j) = -\infty$ for some distinct i, j or $\beta(\mathbf{x}_i|\mathbf{z}) = -\infty$ for some i. Thenwe obtain the desired conclusion that

$$p(\mathbf{x}|\mathbf{z}) = 0 = exp[\sum_{i<j} [\alpha(\mathbf{x}_i\mathbf{x}_j) - \alpha(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i \in \Gamma} [\beta(\mathbf{x}_i|\mathbf{z}) - \beta(\mathbf{z}_i|\mathbf{z})].$$

Next suppose $p(\mathbf{x}|\mathbf{z}) > 0$. Then Lemma 4 implies $a(\mathbf{x}_i\mathbf{x}_j), b(\mathbf{x}_i|\mathbf{z}) > -\infty$ for all i, j. Moreover, by Regularity, $\mathbf{x}_i \in \Sigma^+$ for all i. Then by the conditions relating (a, b) and (α, β) we obtain

$$\begin{aligned} p(\mathbf{x}|\mathbf{z}) &= exp[\textstyle\sum_{i<j} [a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] \\ &\quad + \textstyle\sum_{i \in \Gamma} [b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]] \\ &= exp[\textstyle\sum_{i<j} [\alpha(\mathbf{x}_i\mathbf{x}_j) - \alpha(\mathbf{z}_i\mathbf{z}_j) + [\gamma(\mathbf{x}_i\Omega_j) + \gamma(\mathbf{x}_j\Omega_i)] - [\gamma(\mathbf{z}_i\Omega_j) + \gamma(\mathbf{z}_j\Omega_i)]] \\ &\quad + \textstyle\sum_{i \in \Gamma} \left[\beta(\mathbf{x}_i|\mathbf{z}) - \beta(\mathbf{z}_i|\mathbf{z}) - \sum_{i \neq j \in \Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)]\right]] \\ &= exp[\textstyle\sum_{i<j} [\alpha(\mathbf{x}_i\mathbf{x}_j) - \alpha(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i \in \Gamma} [\beta(\mathbf{x}_i|\mathbf{z}) - \beta(\mathbf{z}_i|\mathbf{z})] + K(\mathbf{x}|\mathbf{z})] \end{aligned}$$

where

$$K(\mathbf{x}|\mathbf{z}) = \sum_{i<j} [[\gamma(\mathbf{x}_i\Omega_j) + \gamma(\mathbf{x}_j\Omega_i)] - [\gamma(\mathbf{z}_i\Omega_j) + \gamma(\mathbf{z}_j\Omega_i)]] - \sum_{i \in \Gamma} \sum_{i \neq j \in \Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)].$$

The proof is complete once we show that $K(\mathbf{x}|\mathbf{z}) = 0$. To this end, observe simply that the first summation term equals the last, since

$$\sum_{i<j} [[\gamma(\mathbf{x}_i\Omega_j) + \gamma(\mathbf{x}_j\Omega_i)] - [\gamma(\mathbf{z}_i\Omega_j) + \gamma(\mathbf{z}_j\Omega_i)]]$$

$$= \sum_{i<j}[\gamma(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)] + \sum_{i<j}[\gamma(\mathbf{x}_j\Omega_i)] - \gamma(\mathbf{z}_j\Omega_i)]$$

$$= \frac{1}{2}\sum_i \sum_{i\neq j\in\Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)] + \frac{1}{2}\sum_i \sum_{i\neq j\in\Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)]$$

$$= \sum_{i\in\Gamma} \sum_{i\neq j\in\Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)],$$

as desired. ■

G Appendix: Proof of Theorem 4

Proof. Consider two associative networks (a, b) and (α, β) that satisfy (1). Suppose (a, b) and (α, β) represent the same IBIU. As in the proof of Theorem 3, define $\gamma(\mathbf{x}_i\Omega_j)$ by

$$\gamma(\mathbf{x}_i\Omega_j) = a(\mathbf{x}_i\Omega_j) - \alpha(\mathbf{x}_i\Omega_j).$$

for any $\mathbf{x}_i \in \Sigma^+$, and let $\gamma(\mathbf{x}_i\Omega_j)$ be an arbitrary real number for any $\mathbf{x}_i \notin \Sigma^+$. We proceed in steps.

Step 1: Show that for any distinct $i, j \in \Gamma$, $\mathbf{x}_i \in \Sigma$ and $\mathbf{z}_j \in \Sigma^+$,

$$a(\mathbf{x}_i\mathbf{z}_j) = \alpha(\mathbf{x}_i\mathbf{z}_j) + [\gamma(\mathbf{x}_i\Omega_j) + \gamma(\mathbf{z}_j\Omega_i) - \gamma(\Omega_i\Omega_j)].$$

This follows from Theorem 3.

Step 2: Show that for each $i \in \Gamma$ and $\mathbf{x}_i \in \Sigma^+$,

$$\gamma(\mathbf{x}_i\Omega_i) - \gamma(\Omega_i\Omega_i) = -2\sum_{i\neq j\in\Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\Omega_i\Omega_j)].$$

Take any $i \in \Gamma$ and $\mathbf{x}_i \in \Sigma^+$. Observe that by Lemma 2 and the definition of γ,

$exp[2\sum_{i\neq j\in\Gamma} [a(\mathbf{x}_i\Omega_j) - a(\Omega_i\Omega_j)] + a(\mathbf{x}_i\Omega_i) - a(\Omega_i\Omega_i)]$
$= p(\mathbf{x}_i|\Omega)$
$= exp[2\sum_{i\neq j\in\Gamma} [\alpha(\mathbf{x}_i\Omega_j) - \alpha(\Omega_i\Omega_j)] + \alpha(\mathbf{x}_i\Omega_i) - \alpha(\Omega_i\Omega_i)]$
$= exp[2\sum_{i\neq j\in\Gamma} [a(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{x}_i\Omega_j) - a(\Omega_i\Omega_j) + \gamma(\Omega_i\Omega_j)]$
$+a(\mathbf{x}_i\Omega_i) - \gamma(\mathbf{x}_i\Omega_i) - a(\Omega_i\Omega_i) + \gamma(\Omega_i\Omega_i)].$

By Lemma 4, $[a(\mathbf{x}_i\Omega_j) > -\infty$ for all $j]$ for any $\mathbf{x}_i \in \Sigma^+$. Therefore

$$2\sum_{i\neq j\in\Gamma} [-\gamma(\mathbf{x}_i\Omega_j) + \gamma(\Omega_i\Omega_j)] - \gamma(\mathbf{x}_i\Omega_i) + \gamma(\Omega_i\Omega_i) = 0,$$

which establishes the step.

Step 3: Show that there exists a function $(i, \mathbf{z}_i) \mapsto \lambda(\mathbf{z}_i) \in \mathbb{R}$ s.t. for each $i \in \Gamma$, $\mathbf{x}_i \in \Sigma$ and $\mathbf{z}_i \in \Sigma^+$,

$$a(\mathbf{x}_i\mathbf{z}_i) = \alpha(\mathbf{x}_i\mathbf{z}_i) + [\gamma(\mathbf{x}_i\Omega_i) + \gamma(\mathbf{z}_i\Omega_i) - \gamma(\Omega_i\Omega_i)] + \lambda(\mathbf{z}_i)$$

Define a function by:

$$\lambda(\mathbf{z}_i) := -\gamma(\mathbf{z}_i\Omega_i) - \sum_{i\neq j\in\Gamma} [\gamma(\Omega_i\Omega_j)] + \psi(\mathbf{z}_i|\mathbf{z}_i),$$

which is real-valued since γ, ψ are real-valued. Consider the following cases.

If $\mathbf{x}_i \notin \Sigma^+$ then $a(\mathbf{x}_i\Omega_j) = -\infty$ for some j and by assumption (1) it follows that $a(\mathbf{x}_i\mathbf{z}_i) = -\infty$ for any $\mathbf{z}_i \in \Sigma^+$. Similarly, $\alpha(\mathbf{x}_i\mathbf{z}_i) = -\infty$ for any $\mathbf{z}_i \in \Sigma^+$. Consequently, the desired equality is satisfied.

Next suppose $\mathbf{x}_i \in \Sigma^+$. By definition of IBIU, $b(\mathbf{x}_i|\mathbf{z}_i) = a(\mathbf{x}_i\mathbf{z}_i) + \sum_{i\neq j\in\Gamma} a(\mathbf{x}_i\Omega_j)$ and similarly $\beta(\mathbf{x}_i|\mathbf{z}_i) = \alpha(\mathbf{x}_i\mathbf{z}_i) + \sum_{i\neq j\in\Gamma} \alpha(\mathbf{x}_i\Omega_j)$. Applying these and Step 1 in what follows, we observe that by Theorem 3,

$$b(\mathbf{x}_i|\mathbf{z}_i) = \beta(\mathbf{x}_i|\mathbf{z}_i) + \psi(\mathbf{z}_i|\mathbf{z}_i) - \sum_{i\neq j\in\Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\Omega_i\Omega_j)]$$

$$\begin{aligned}
\implies & a(\mathbf{x}_i\mathbf{z}_i) + \textstyle\sum_{i\neq j\in\Gamma} a(\mathbf{x}_i\Omega_j) \\
& = \alpha(\mathbf{x}_i\mathbf{z}_i) + \textstyle\sum_{i\neq j\in\Gamma} \alpha(\mathbf{x}_i\Omega_j) + \psi(\mathbf{z}_i|\mathbf{z}_i) - \sum_{i\neq j\in\Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\Omega_i\Omega_j)] \\
\implies & a(\mathbf{x}_i\mathbf{z}_i) = \alpha(\mathbf{x}_i\mathbf{z}_i) - \textstyle\sum_{i\neq j\in\Gamma}[a(\mathbf{x}_i\Omega_j) - \alpha(\mathbf{x}_i\Omega_j)] \\
& +\psi(\mathbf{z}_i|\mathbf{z}_i) - \textstyle\sum_{i\neq j\in\Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\Omega_i\Omega_j)] \\
\implies & a(\mathbf{x}_i\mathbf{z}_i) = \alpha(\mathbf{x}_i\mathbf{z}_i) - \textstyle\sum_{i\neq j\in\Gamma} \gamma(\mathbf{x}_i\Omega_j) \\
& +\psi(\mathbf{z}_i|\mathbf{z}_i) - \textstyle\sum_{i\neq j\in\Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\Omega_i\Omega_j)] \\
\implies & a(\mathbf{x}_i\mathbf{z}_i) = \alpha(\mathbf{x}_i\mathbf{z}_i) - \textstyle\sum_{i\neq j\in\Gamma} \gamma(\Omega_i\Omega_j) \\
& +\psi(\mathbf{z}_i|\mathbf{z}_i) - 2\textstyle\sum_{i\neq j\in\Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\Omega_i\Omega_j)] \\
\implies & a(\mathbf{x}_i\mathbf{z}_i) = \alpha(\mathbf{x}_i\mathbf{z}_i) - \textstyle\sum_{i\neq j\in\Gamma} \gamma(\Omega_i\Omega_j) \\
& +\psi(\mathbf{z}_i|\mathbf{z}_i) + \gamma(\mathbf{x}_i\Omega_i) - \gamma(\Omega_i\Omega_i) \text{ (by step 2).}
\end{aligned}$$

Apply the definition $\lambda(\mathbf{z}_i)$ given at the start of the proof of this step to establish the result.

The step is established by defining $\lambda(\mathbf{z}_i)$ by

$$\lambda(\mathbf{z}_i) := -\gamma(\mathbf{z}_i\Omega_i) - \sum_{i\neq j\in\Gamma} [\gamma(\Omega_i\Omega_j)] + \psi(\mathbf{z}_i|\mathbf{z}_i)$$

and noting that it must be real valued since γ, ψ are real-valued.

Step 4: Establish the converse.

Suppose that p is represented by (a, b), and suppose that (α, β) is an associative network that satisfies the conditions in the statement of the Theorem. We show that p must be represented by (α, β) as well.

By Theorem 3, it suffices to show that there exists a function $\psi(\mathbf{z}_i|\mathbf{z})$ such that

$$b(\mathbf{x}_i|\mathbf{z}) = \beta(\mathbf{x}_i|\mathbf{z}) + \psi(\mathbf{z}_i|\mathbf{z}) - \sum_{i\neq j\in\Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)]\,. \tag{13}$$

So observe that

$b(\mathbf{x}_i|\mathbf{z}) = a(\mathbf{x}_i\mathbf{z}_i) + \sum_{i\neq j\in\Gamma} a(\mathbf{x}_i\mathbf{z}_j)$
$= \alpha(\mathbf{x}_i\mathbf{z}_i) + \gamma(\mathbf{x}_i\Omega_i) + \gamma(\mathbf{z}_i\Omega_i) - \gamma(\Omega_i\Omega_i)] + \lambda(\mathbf{z}_i)$
$+ \sum_{i\neq j\in\Gamma}[\alpha(\mathbf{x}_i\mathbf{z}_j) + [\gamma(\mathbf{x}_i\Omega_j) + \gamma(\mathbf{z}_j\Omega_i) - \gamma(\Omega_i\Omega_j)]] + 0$
$= \sum_{j\in\Gamma} \alpha(\mathbf{x}_i\mathbf{z}_j) + \sum_{i\neq j\in\Gamma}[\gamma(\mathbf{x}_i\Omega_j) - \gamma(\Omega_i\Omega_j)] + [\gamma(\mathbf{x}_i\Omega_i) - \gamma(\Omega_i\Omega_i)]$
$+\lambda(\mathbf{z}_i) + \sum_{j\in\Gamma} \gamma(\mathbf{z}_j\Omega_i)$

By the condition

$$\gamma(\mathbf{x}_i\Omega_i) - \gamma(\Omega_i\Omega_i) = -2\sum_{i\neq j\in\Gamma} [\gamma(\mathbf{x}_i\Omega_j) - \gamma(\Omega_i\Omega_j)]\,,$$

and $\beta(\mathbf{x}_i|\mathbf{z}) = \sum_{j\in\Gamma} \alpha(\mathbf{x}_i\mathbf{z}_j)$, the preceding yields

$b(\mathbf{x}_i|\mathbf{z}) = \beta(\mathbf{x}_i|\mathbf{z}) - \sum_{i\neq j\in\Gamma}[\gamma(\mathbf{x}_i\Omega_j) - \gamma(\Omega_i\Omega_j)]$
$+\lambda(\mathbf{z}_i) + \sum_{j\in\Gamma} \gamma(\mathbf{z}_j\Omega_i)$
$= \beta(\mathbf{x}_i|\mathbf{z}) - \sum_{i\neq j\in\Gamma}[\gamma(\mathbf{x}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)]$
$+ \sum_{i\neq j\in\Gamma}[\gamma(\Omega_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)] + \lambda(\mathbf{z}_i) + \sum_{j\in\Gamma} \gamma(\mathbf{z}_j\Omega_i)$.

Define

$$\psi(\mathbf{z}_i|\mathbf{z}) := \sum_{i\neq j\in\Gamma} [\gamma(\Omega_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)] + \lambda(\mathbf{z}_i) + \sum_{j\in\Gamma} \gamma(\mathbf{z}_j\Omega_i)$$

and this yields (13), and completes the proof. ■

H Appendix: Proof of Corollary 1

Proof. Given any representation (α, β), we define a normalized representation by taking for any $i, j \in \Gamma$ and $\mathbf{x}_i \in \Sigma^+$ s.t. $\alpha(\mathbf{x}_i\Omega_j), \alpha(\mathbf{x}_i\mathbf{x}_i) > -\infty$

$$\gamma(\mathbf{x}_i\Omega_j) = \begin{cases} -\alpha(\mathbf{x}_i\Omega_j) & \text{if } i \neq j \text{ or } \mathbf{x}_i = \Omega_i \\ \frac{-\alpha(\mathbf{x}_i\mathbf{x}_i)+\gamma(\Omega_i\Omega_i)}{2} & \text{if } i = j \text{ and } \mathbf{x}_i \neq \Omega_i \end{cases}$$

If $\alpha(\mathbf{x}_i\Omega_j) = -\infty$ then by assumption (1), it must be that $\alpha(\mathbf{x}_i\mathbf{x}_i) = -\infty$ and in fact beliefs satisfy $\mathbf{x}_i \notin \Sigma^+$. The new representation must also then satisfy $a(\mathbf{x}_i\Omega_j) = a(\mathbf{x}_i\mathbf{x}_i) = -\infty$.

An arbitrary $\gamma(\mathbf{x}_i\Omega_i)$ for $\mathbf{x}_i \neq \Omega_i$ can be taken for the IBGU model, but for the IBIU model this is determined by the condition on γ specified in the Theorem 4.

For the IBGU model, let

$$\psi(\mathbf{z}_i|\mathbf{z}) = -\beta(\mathbf{z}_i|\mathbf{z})$$

for all $\mathbf{z}_i$, and for the IBIU model let $\lambda(\mathbf{z}_i, j) = 0$ for all $\mathbf{z}_i, j$.

Take any $\mathbf{x}_i \in \Sigma^+$. For both the IBGU and IBGU model, it is readily computed that for any distinct i, j or $\mathbf{x}_i = \Omega_i$,

$$a(\mathbf{x}_i\Omega_j) = \alpha(\mathbf{x}_i\Omega_j) + [\gamma(\mathbf{x}_i\Omega_j) + \gamma(\Omega_i\Omega_j) - \gamma(\Omega_i\Omega_j)] = \alpha(\mathbf{x}_i\Omega_j) - \alpha(\mathbf{x}_i\Omega_j) = 0.$$

In the IBIU model, for any i and $\Omega \neq \mathbf{x} \in \Sigma^+$,

$$a(\mathbf{x}_i\mathbf{x}_i) = \alpha(\mathbf{x}_i\mathbf{x}_i) + [\gamma(\mathbf{x}_i\Omega_i) + \gamma(\mathbf{x}_i\Omega_i) - \gamma(\Omega_i\Omega_i)]$$

$$= \alpha(\mathbf{x}_i\mathbf{x}_i) + 2\frac{-\alpha(\mathbf{x}_i\mathbf{x}_i) + \gamma(\Omega_i\Omega_i)}{2} - \gamma(\Omega_i\Omega_i) = 0.$$

We show shortly that this $a(\mathbf{x}_i\mathbf{x}_i) = 0$ is also a normalization for IBGU.

In the IBGU model, for any i and $\mathbf{z} \in \Sigma^+$,

$$b(\mathbf{z}_i|\mathbf{z}) = \beta(\mathbf{z}_i|\mathbf{z}) + \psi(\mathbf{z}_i|\mathbf{z}) - \sum_{i \neq j \in \Gamma} [\gamma(\mathbf{z}_i\Omega_j) - \gamma(\mathbf{z}_i\Omega_j)] = 0.$$

Moreover, by the regularity property of the network, $a(\mathbf{x}_i\mathbf{x}_i) = b(\mathbf{x}_i|\mathbf{x}_i) = 0$.

Invoke Theorem 3 and 4 to establish that the new network represents $\mathbf{p}$. ■

I Appendix: Proof of Proposition 2

Proof. Lemmas 2 and 5 show that

- $p(\mathbf{x}_i\mathbf{x}_j|\Omega) = p(\mathbf{x}_i|\Omega)p(\mathbf{x}_j|\Omega) \times exp\left[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{x}_i\Omega_j) - a(\Omega_i\mathbf{x}_j) + a(\Omega_i\Omega_j)\right],$
- $p(\mathbf{x}_i|\mathbf{z}_i) = exp[2\sum_{i \neq j \in \Gamma}[a(\mathbf{x}_i\Omega_j) - a(\mathbf{z}_i\Omega_j)] + [b(\mathbf{x}_i|\mathbf{z}_i) - b(\mathbf{z}_i|\mathbf{z}_i)]]$.

Simply apply the normalizations to these expressions, and the property of regularity of the network that requires $a(\mathbf{x}_i\mathbf{z}_i) = b(\mathbf{x}_i|\mathbf{z}_i)$ to obtain the identification result for a. To identify b, first note that from the expression for $p(\mathbf{x}_i|\mathbf{z}_i)$ we can derive $p(\mathbf{x}_i|\Omega) = exp[b(\mathbf{x}_i|\Omega)]$. By Lemma 6, whenever $\mathbf{x}_i\mathbf{z}_{-i} \in \Sigma^+$,

$$\frac{p(\mathbf{x}_i|\mathbf{z})}{p^{BU}(\mathbf{x}_i|\mathbf{z})} = exp[[b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) - [b(\mathbf{x}_i|\Omega) - b(\mathbf{z}_i|\Omega)]]$$

$$= \frac{1}{p^{BU}(\mathbf{x}_i|\mathbf{z}_i)} exp[b(\mathbf{x}_i|\mathbf{z})]$$

and so $exp[b(\mathbf{x}_i|\mathbf{z})] = \frac{p(\mathbf{x}_i|\mathbf{z})p^{BU}(\mathbf{x}_i|\mathbf{z}_i)}{p^{BU}(\mathbf{x}_i|\mathbf{z})}$. ■

J Proof of Proposition 3

We establish the Proposition in the following lemmas.

Lemma 12 *For IBGU* $\mathbf{p}$*, the following are equivalent for any* $\mathbf{z} \in \Sigma^+$*:*
(a) $\mathbf{p}$ *satisfies Monotonicity.*
(b) for any $\mathbf{x}, \mathbf{y} \in \Sigma$ *s.t.* $p(\mathbf{x}|\mathbf{z}) > 0$, $\mathbf{y} = \mathbf{y}_i\mathbf{x}_{-i}$,

$$\mathbf{x}_i \subset \mathbf{y}_i \implies p(\mathbf{x}|\mathbf{z}) \leq p(\mathbf{y}_i\mathbf{x}_{-i}|\mathbf{z}),$$

(c) for any $\mathbf{x}, \mathbf{y} \in \Sigma$ *s.t.* $p(\mathbf{x}|\mathbf{z}) > 0$, $\mathbf{y} = \mathbf{y}_i\mathbf{x}_{-i}$,

$$\mathbf{x}_i \subset \mathbf{y}_i \implies$$

$$p(\mathbf{x}_i|\mathbf{z}) \times \prod_{i \neq j \in \Gamma} \frac{p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})} \leq p(\mathbf{y}_i|\mathbf{z}) \times \prod_{i \neq j \in \Gamma} \frac{p(\mathbf{y}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{y}_i|\mathbf{z})}.$$

Proof. Condition (b) trivially follows from (a). To show the converse, suppose condition (b) holds. If $p(\mathbf{x}|\mathbf{z}) = 0$ then trivially $p(\mathbf{x}|\mathbf{z}) = 0 \leq p(\mathbf{y}_i\mathbf{x}_{-i}|\mathbf{z})$ for each i. Consequently, condition (b) can be replaced by the stronger condition that for any $\mathbf{x}, \mathbf{y} \in \Sigma$ s.t. $\mathbf{y} = \mathbf{y}_i\mathbf{x}_{-i}$,

$$\mathbf{x}_i \subset \mathbf{y}_i \implies p(\mathbf{x}|\mathbf{z}) \leq p(\mathbf{y}_i\mathbf{x}_{-i}|\mathbf{z}).$$

Apply this inductively to establish that for any $\mathbf{x}, \mathbf{w} \in \Sigma$ s.t. $\mathbf{x}_i \subset \mathbf{w}_i$ for all i, we must have $p(\mathbf{x}|\mathbf{z}) \leq p(\mathbf{w}|\mathbf{z})$, thereby establishing condition (a).

Turn to the equivalence of conditions (b) and (c). Suppose condition (b) holds, and take any $\mathbf{x}, \mathbf{y} \in \Sigma$ and $\mathbf{z} \in \Sigma^+$ s.t. $p(\mathbf{x}|\mathbf{z}) > 0$ and $\mathbf{y} = \mathbf{y}_i\mathbf{x}_{-i}$. By condition (b), $p(\mathbf{x}|\mathbf{z}) > 0$ implies $p(\mathbf{y}_i\mathbf{x}_{-i}|\mathbf{z}) > 0$. By Regularity, $p(\mathbf{x}|\mathbf{z})p(\mathbf{y}_i\mathbf{x}_{-i}|\mathbf{z}) > 0$ implies $p(\mathbf{x}_j|\mathbf{z}) > 0$ for all j and $p(\mathbf{y}_i|\mathbf{z}) > 0$. Condition (b) implies immediately that $p(\mathbf{x}_i|\mathbf{z}) \leq p(\mathbf{y}_i|\mathbf{z})$ which establishes the desired inequality for the case that $\mathbf{x}_{-i} = \mathbf{z}_{-i}$. More generally, observe that if $\mathbf{x}_i \subset \mathbf{y}_i$, then by condition (b), $p(\mathbf{x}|\mathbf{z}) \leq p(\mathbf{y}_i\mathbf{x}_{-i}|\mathbf{z})$. But then by ASD,

$$p(\mathbf{x}|\mathbf{z}) \leq p(\mathbf{y}_i\mathbf{x}_{-i}|\mathbf{z})$$

$$\iff p(\mathbf{x}_i|\mathbf{z}) \times \prod_{i \neq j \in \Gamma} p(\mathbf{x}_j|\mathbf{z}) \times \prod_{i \neq j} \frac{p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z})} \times \prod_{i \neq k < j \neq i} \frac{p(\mathbf{x}_k\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_k|\Omega)p(\mathbf{x}_j|\mathbf{z})}$$

$$\leq p(\mathbf{y}_i|\mathbf{z}) \times \prod_{i \neq j \in \Gamma} p(\mathbf{x}_j|\mathbf{z}) \times \prod_{i \neq j} \frac{p(\mathbf{y}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{y}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z})} \times \prod_{i \neq k < j \neq i} \frac{p(\mathbf{x}_k\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_k|\mathbf{z})p(\mathbf{x}_j|\mathbf{z})}$$

$$\iff p(\mathbf{x}_i|\mathbf{z}) \times \prod_{i \neq j} \frac{p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})} \leq p(\mathbf{y}_i|\mathbf{z}) \times \prod_{i \neq j} \frac{p(\mathbf{y}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{y}_i|\mathbf{z})}. \tag{14}$$

This shows that condition (b) implies condition (c).

Conversely, assume condition (c) and take any $\mathbf{x}, \mathbf{y} \in \Sigma$ and $\mathbf{z} \in \Sigma^+$ s.t. $\mathbf{y} = \mathbf{y}_i \mathbf{x}_{-i}$ and $\mathbf{x}_i \subset \mathbf{y}_i$. For the case where $\mathbf{x}_{-i} = \mathbf{z}_{-i}$, condition (c) yields $0 < p(\mathbf{x}_i|\mathbf{z}) \leq p(\mathbf{y}_i|\mathbf{z})$. More generally, condition (c) and this implication $p(\mathbf{y}_i|\mathbf{z}) > 0$ yields the inequality in (14), which we have already seen is equivalent to $p(\mathbf{x}|\mathbf{z}) \leq p(\mathbf{y}_i \mathbf{x}_{-i}|\mathbf{z})$, as desired. ■

Lemma 13 *For IBGU* $\mathbf{p}$*, the following are equivalent:*

(i) $\mathbf{p}$ *satisfies Monotonicity.*

(ii) For any $\mathbf{x}, \mathbf{y}_i \in \Sigma$ *and* $\mathbf{z} \in \Sigma^+$ *s.t.* $p(\mathbf{x}|\mathbf{z}) > 0$ *and* $\mathbf{x}_i \subset \mathbf{y}_{-i}$*,*

$$\mathbf{x}_i \subset \mathbf{y}_i \implies \sum_{i \neq j} a(\mathbf{x}_i \mathbf{x}_j) + b(\mathbf{x}_i|\mathbf{z}) \leq \sum_{i \neq j} a(\mathbf{y}_i \mathbf{x}_j) + b(\mathbf{y}_i|\mathbf{z}).$$

Proof. Lemmas 2 and 5 imply

$$p(\mathbf{x}_i|\mathbf{z}) = exp[\sum_{i \neq j \in \Gamma} [a(\mathbf{x}_i \mathbf{z}_j) - a(\mathbf{z}_i \mathbf{z}_j)] + [b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]]$$

$$\prod_{i \neq j \in \Gamma} \frac{p(\mathbf{x}_i \mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z}) p(\mathbf{x}_j|\mathbf{z})} = exp\left[\sum_{i \neq j \in \Gamma} [a(\mathbf{x}_i \mathbf{x}_j) - a(\mathbf{z}_i \mathbf{x}_j)] - \sum_{i \neq j \in \Gamma} [a(\mathbf{x}_i \mathbf{z}_j) - a(\mathbf{z}_i \mathbf{z}_j)]\right]$$

and therefore we obtain

$$p(\mathbf{x}_i|\mathbf{z}) \times \prod_{i \neq j \in \Gamma} \frac{p(\mathbf{x}_i \mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z}) p(\mathbf{x}_j|\mathbf{z})}$$

$$= exp[\sum_{i \neq j \in \Gamma} [a(\mathbf{x}_i \mathbf{x}_j) - a(\mathbf{z}_i \mathbf{x}_j)] + [b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z})]]$$

For any $\mathbf{x}, \mathbf{y} \in \Sigma$ and $\mathbf{z} \in \Sigma^+$ s.t. $\mathbf{y} = \mathbf{y}_i \mathbf{x}_{-i}$ and $p(\mathbf{x}|\mathbf{z}), p(\mathbf{y}|\mathbf{z}) > 0$, Regularity implies $p(\mathbf{x}_i|\mathbf{z}), p(\mathbf{y}_i|\mathbf{z}) > 0$. By Lemma 12, Monotonicity is equivalent to the condition stated there. Observe that

$$p(\mathbf{x}_i|\mathbf{z}) \times \prod_{i \neq j} \frac{p(\mathbf{x}_i \mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})} \leq p(\mathbf{y}_i|\mathbf{z}) \times \prod_{i \neq j} \frac{p(\mathbf{y}_i \mathbf{x}_j|\mathbf{z})}{p(\mathbf{y}_i|\mathbf{z})}$$

$$\iff p(\mathbf{x}_i|\mathbf{z}) \times [\prod_{i \neq j} \frac{p(\mathbf{x}_i \mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})}] \times [\prod_{i \neq j} \frac{1}{p(\mathbf{x}_j|\mathbf{z})}] \leq p(\mathbf{y}_i|\mathbf{z}) \times [\prod_{i \neq j} \frac{p(\mathbf{y}_i \mathbf{x}_j|\mathbf{z})}{p(\mathbf{y}_i|\mathbf{z})}] \times [\prod_{i \neq j} \frac{1}{p(\mathbf{x}_j|\mathbf{z})}]$$

$$\iff p(\mathbf{x}_i|\mathbf{z}) \times \prod_{i \neq j} \frac{p(\mathbf{x}_i \mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z}) p(\mathbf{x}_j|\mathbf{z})} \leq p(\mathbf{y}_i|\mathbf{z}) \times \prod_{i \neq j} \frac{p(\mathbf{y}_i \mathbf{x}_j|\mathbf{z})}{p(\mathbf{y}_i|\mathbf{z}) p(\mathbf{x}_j|\mathbf{z})}$$

$$\iff \sum_{i \neq j} a(\mathbf{x}_i \mathbf{x}_j) + b(\mathbf{x}_i|\mathbf{z}) \leq \sum_{i \neq j} a(\mathbf{y}_i \mathbf{x}_j) + b(\mathbf{y}_i|\mathbf{z}).$$

This equivalence implies the desired equivalence between (i) and (ii). ■

K Appendix: Proof of Proposition 4

Proof. By Regularity and the definition of the Bayesian update, $p(\mathbf{x}|\Omega) = 0 \implies p(\mathbf{x}|\mathbf{z}) = 0 = p^{BU}(\mathbf{x}|\mathbf{z})$. That is, Bayesian conditioning is satisfied whenever $p(\mathbf{x}|\Omega) = 0$. Therefore we only need to establish conditions for which beliefs are Bayesian for any $\mathbf{x}, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$.

First we show that Bayesian updating for any $\mathbf{x}, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$ is equivalent to condition (ii) in the statement of the Proposition. The implication " $\implies$ " is trivial. For the converse, take any $\mathbf{x}, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$. Then $\mathbf{x}_i \subset \mathbf{z}_i$ for all i, and by Regularity, $\mathbf{x}_i \in \Sigma^+$. That is, the hypothesis for condition (ii) is satisfied for each i. Consequently, by condition (ii), $p(\mathbf{x}_i|\mathbf{z}) = p^{BU}(\mathbf{x}_i|\mathbf{z})$ for each i and moreover, $p(\mathbf{x}_i|\mathbf{z}) > 0$ for each i. By Regularity, $p(\mathbf{x}_i|\mathbf{z}) > 0$ implies $\mathbf{x}_i\mathbf{z}_{-i} \in \Sigma^+$ and so we can invoke AUB to obtain $p(\mathbf{x}|\mathbf{z}) = p^{BU}(\mathbf{x}|\mathbf{z})$, as desired.

Next we show that Bayesian updating is equivalent to condition (iii). We proceed by first establishing:

Claim: For any $\mathbf{x}, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$ and $p(\mathbf{x}|\mathbf{z}) > 0$, the model satisfies $p(\mathbf{x}|\mathbf{z}) = p^{BU}(\mathbf{x}|\mathbf{z})$ if and only if

$$\sum_{j\in\Gamma}[b(\mathbf{x}_j|\mathbf{z}) - b(\mathbf{z}_j|\mathbf{z})] = \sum_{j\in\Gamma}[b(\mathbf{x}_j|\Omega) - b(\mathbf{z}_j|\Omega)]]. \tag{15}$$

To show this, taking any $\mathbf{x}, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$ and $p(\mathbf{x}|\mathbf{z}) > 0$, and recall that the Bayesian update is given by

$$p^{BU}(\mathbf{x}|\mathbf{z}) = \frac{p(\mathbf{x}|\Omega)}{p(\mathbf{z}|\Omega)} = exp[\left[\sum_{i<j} a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)\right] + \sum_{j\in\Gamma}[b(\mathbf{x}_j|\Omega) - b(\mathbf{z}_j|\Omega)]],$$

while the actual update is

$$p(\mathbf{x}|\mathbf{z}) = exp[\left[\sum_{i<j} a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)\right] + \sum_{j\in\Gamma}[b(\mathbf{x}_j|\mathbf{z}) - b(\mathbf{z}_j|\mathbf{z})]].$$

Since $p(\mathbf{x}|\Omega), p(\mathbf{x}|\mathbf{z}) > 0$, all the terms in the above expressions are real-valued. The claim follows.

Now we prove the lemma. If beliefs are Bayesian, then for any $\mathbf{x}_i, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x}_i \subset \mathbf{z}_i$ and $p(\mathbf{x}_i|\mathbf{z}) > 0$ we have $p(\mathbf{x}_i|\mathbf{z}) = p^{BU}(\mathbf{x}_i|\mathbf{z})$, and then condition (15) clearly implies the desired condition, $b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) = b(\mathbf{x}_i|\Omega) - b(\mathbf{z}_i|\Omega)$. Conversely, suppose that this desired condition holds. We show that (15) must hold: Take any $\mathbf{x}, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$ and $p(\mathbf{x}|\mathbf{z}) > 0$. By Regularity, $p(\mathbf{x}_i|\mathbf{z}) > 0$ for any $i \in \Gamma$. Consequently we can invoke $b(\mathbf{x}_i|\mathbf{z}) - b(\mathbf{z}_i|\mathbf{z}) = b(\mathbf{x}_i|\Omega) - b(\mathbf{z}_i|\Omega)$ for all i and conclude that (15) must hold, as desired. ■

L Appendix: Proof of Proposition 5

Proof. By Regularity and the definition of the Bayesian update, $p(\mathbf{x}|\Omega) = 0 \implies p(\mathbf{x}|\mathbf{z}) = 0 = p^{BU}(\mathbf{x}|\mathbf{z})$. That is, Bayesian conditioning is satisfied whenever $p(\mathbf{x}|\Omega) = 0$. Therefore we need to establish conditions for which beliefs are Bayesian for any $\mathbf{x}$ s.t. $p(\mathbf{x}|\Omega) > 0$.

Step 1. Show that beliefs are Bayesian iff for all $\mathbf{x}_i, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x}_i \subset \mathbf{z}_i$ and $p(\mathbf{x}_i|\mathbf{z}) > 0$,

$$\sum_{j\in\Gamma} a(\mathbf{x}_i\mathbf{z}_j) - \sum_{j\in\Gamma} a(\mathbf{z}_i\mathbf{z}_j) = \sum_{j\in\Gamma} a(\mathbf{x}_i\Omega_j) - \sum_{j\in\Gamma} a(\mathbf{z}_i\Omega_j). \tag{16}$$

This is just an application of Proposition 4 to IBIU.

Step 2. Show that, for a normalized representation, condition (16) holds iff
(i) $a(\mathbf{x}_i\mathbf{z}_j) = 0$ if $\mathbf{x}_i\mathbf{z}_j \in \Sigma^+$ s.t. $j \neq i$, and
(ii) $a(\mathbf{x}_i\mathbf{z}_i) - a(\mathbf{z}_i\mathbf{z}_i) = a(\mathbf{x}_i\Omega_i) - a(\mathbf{z}_i\Omega_i)$ if $\mathbf{x}_i, \mathbf{z}_i \in \Sigma^+$ s.t. $\mathbf{x}_i \subset \mathbf{z}_i$.

Assume the two conditions and take any $\mathbf{x}, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$ and $p(\mathbf{x}|\mathbf{z}) > 0$. By Regularity $p(\mathbf{x}|\mathbf{z}) > 0$ implies $p(\mathbf{x}_i|\mathbf{z}) > 0$ for any i. By Lemma 10, it follows that $p(\mathbf{x}_i|\mathbf{z}_j) = p(\mathbf{x}_i\mathbf{z}_j|\mathbf{z}_j) > 0$ and in turn by Regularity $\mathbf{x}_i\mathbf{z}_j \in \Sigma^+$. Since $\mathbf{x}, \mathbf{z} \in \Sigma^+$, Regularity also implies that $\mathbf{x}_i, \mathbf{z}_i \in \Sigma^+$ for each i. Consequently we can invoke the two conditions to obtain (16).

Conversely, assume (16). Consider information $\mathbf{z}_j \in \Sigma^+$ and event $\mathbf{x}_i\mathbf{z}_j \in \Sigma^+$ s.t. $j \neq i$. Since (by (16) and Step 1) beliefs are Bayesian, it must be that $p(\mathbf{x}_i|\mathbf{z}_j) > 0$. Apply (16) to obtain the first desired condition:

$a(\mathbf{x}_i\Omega_i) + a(\mathbf{x}_i\mathbf{z}_j) - a(\Omega_i\Omega_i) - a(\Omega_i\mathbf{z}_j)$
$= a(\mathbf{x}_i\Omega_i) + a(\mathbf{x}_i\Omega_j) - a(\Omega_i\Omega_i) - a(\Omega_i\Omega_j)$, and so

$$a(\mathbf{x}_i\mathbf{z}_j) = 0.$$

Next consider $\mathbf{x}_i, \mathbf{z}_i \in \Sigma^+$ s.t. $\mathbf{x}_i \subset \mathbf{z}_i$. Since beliefs are Bayesian, $p(\mathbf{x}_i|\mathbf{z}_i) > 0$. Applying (16), and first desired condition that we just established, we see that the second desired condition, $a(\mathbf{x}_i\mathbf{z}_i) - a(\mathbf{z}_i\mathbf{z}_i) = a(\mathbf{x}_i\Omega_i) - a(\mathbf{z}_i\Omega_i)$, holds, as desired.

Step 3. Show that beliefs are Bayesian iff for any $\mathbf{x}, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$ and $p(\mathbf{x}|\mathbf{z}) > 0$,

$$p(\mathbf{x}|\mathbf{z}) = \prod_{i\in\Gamma} p(\mathbf{x}_i|\mathbf{z}) \text{ and } p(\mathbf{x}_i|\mathbf{z}) = p^{BU}(\mathbf{x}_i|\mathbf{z})$$

First suppose that IBIU beliefs are Bayesian and take any $\mathbf{x}, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$ and $p(\mathbf{x}|\mathbf{z}) > 0$. Take any distinct i, j. Then $p(\mathbf{x}_i|\mathbf{z}) > 0$ (by Regularity) and $p(\mathbf{x}_i|\mathbf{z}_j) > 0$ (by Lemma 10, and in turn, by another application of Regularity, $\mathbf{x}_i\mathbf{z}_j \in \Sigma^+$. Therefore we can invoke the first condition in step 2 to obtain $a(\mathbf{x}_i\mathbf{z}_j) = 0 = a(\mathbf{x}_i\mathbf{x}_j)$ for any distinct $i, j \in \Gamma$. This implies that

$$p(\mathbf{x}|\mathbf{z}) = exp[\sum_{i<j}[a(\mathbf{x}_i\mathbf{x}_j) - a(\mathbf{z}_i\mathbf{z}_j)] + \sum_{i\in\Gamma}\sum_{j\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_j) - a(\mathbf{z}_i\mathbf{z}_j)]]$$
$$= exp[\sum_{i\in\Gamma}[a(\mathbf{x}_i\mathbf{z}_i) - a(\mathbf{z}_i\mathbf{z}_i)]]$$

and also
$$p(\mathbf{x}_i|\mathbf{z}) = exp[a(\mathbf{x}_i\mathbf{z}_i) - a(\mathbf{z}_i\mathbf{z}_i)].$$

Given this, and recalling that beliefs are Bayesian, conclude that for any $\mathbf{x}, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$ and $p(\mathbf{x}|\mathbf{z}) > 0$,

$$p(\mathbf{x}|\mathbf{z}) = \prod_{i\in\Gamma} p(\mathbf{x}_i|\mathbf{z}),$$

and
$$p(\mathbf{x}_i|\mathbf{z}) = p(\mathbf{x}_i|\mathbf{z}_i) = p^{BU}(\mathbf{x}_i|\mathbf{z}_i) = \frac{p(\mathbf{x}_i|\Omega)}{p(\mathbf{z}_i|\Omega)},$$

as desired.

Conversely, note that for any $\mathbf{x}, \mathbf{z} \in \Sigma^+$ s.t. $\mathbf{x} \subset \mathbf{z}$ and $p(\mathbf{x}|\mathbf{z}) > 0$,

$$p(\mathbf{x}_i|\mathbf{z}) = \frac{p(\mathbf{x}_i|\Omega)}{p(\mathbf{z}_i|\Omega)} = \frac{p(\mathbf{x}_i|\Omega)}{p(\mathbf{z}_i|\Omega)} \times \prod_{i\neq j} \frac{p(\mathbf{z}_j|\Omega)}{p(\mathbf{z}_j|\Omega)} = \frac{p(\mathbf{x}_i\mathbf{z}_{-i}|\Omega)}{p(\mathbf{z}|\Omega)} = p^{BU}(\mathbf{x}_i|\mathbf{z})$$

and so by AUB, $p(\mathbf{x}|\mathbf{z}) = p^{BU}(\mathbf{x}|\mathbf{z})$ (we can invoke AUB since by Regularity $p(\mathbf{x}|\mathbf{z}) > 0$ implies $p(\mathbf{x}_i|\mathbf{z}) > 0$ and in turn $p(\mathbf{x}_i\mathbf{z}_{-i}|\Omega) > 0$). On the other hand if $\mathbf{x} \notin \Sigma^+$ then Regularity and the definition of the Bayesian update imply $p(\mathbf{x}|\mathbf{z}) = 0 = p^{BU}(\mathbf{x}|\mathbf{z})$. This completes the characterization of Bayesian Intuitive Beliefs. ■

M Appendix: Proof of Proposition 6

Proof. Consider a normalized a that satisfies:

$$exp\,[a(\mathbf{x}_i\mathbf{x}_j)] = SD^*(\mathbf{x}_i\mathbf{x}_j|\Omega) := \frac{q^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}{q^*(\mathbf{x}_i|\Omega)q^*(\mathbf{x}_j|\Omega)},$$

$$exp\,[a(\mathbf{x}_i\Omega_i)] = q^*(\mathbf{x}_i|\Omega),$$

$$\text{and } exp\,[a(\mathbf{x}_i\mathbf{z}_i)] = q^*(\mathbf{x}_i|\mathbf{z}_i).$$

This association function defines a prior that satisfies:

$$p(\mathbf{x}_i\mathbf{x}_j|\Omega) = [q^*(\mathbf{x}_i|\Omega)q^*(\mathbf{x}_j|\Omega)] \times \left[\frac{q^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}{q^*(\mathbf{x}_i|\Omega)q^*(\mathbf{x}_j|\Omega)}\right] = q^*(\mathbf{x}_i\mathbf{x}_j|\Omega).$$

That is, the training problem can be solved perfectly (in that $p(\cdot|\Omega) = q^*(\cdot|\Omega)$) with this association function. The posterior is given by:

$$p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z}) = [q^*(\mathbf{x}_i|\mathbf{z}_i)q^*(\mathbf{x}_j|\mathbf{z}_j)]\left[\frac{SD^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]\left[\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right].$$

However, we need to check if this is well-defined in the sense that $p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z}) \leq 1$.

Observe that

$$p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z}) \leq 1$$

$$\iff [q^*(\mathbf{x}_i|\mathbf{z}_i)q^*(\mathbf{x}_j|\mathbf{z}_j)]\left[\frac{SD^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]\left[\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right] \leq 1$$

$$\iff \left[\frac{q^*(\mathbf{x}_i|\Omega_i)q^*(\mathbf{x}_j|\Omega_j)}{q^*(\mathbf{z}_i|\Omega_i)q^*(\mathbf{z}_j|\Omega_j)}\right]\left[\frac{SD^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]\left[\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right] \leq 1$$

$$\iff \left[\frac{q^*(\mathbf{x}_i|\Omega_i)q^*(\mathbf{x}_j|\Omega_j)}{q^*(\mathbf{z}_i|\Omega_i)q^*(\mathbf{z}_j|\Omega_j)}\right]\left[\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right] \leq \frac{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}$$

$$\iff \left[\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right] \leq \frac{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}. \tag{17}$$

We show that the last statement is implied by the condition on q^* given in the proposition. Note that the condition can be written as:

$$\frac{q^*(\mathbf{z}_i|\Omega)}{q^*(\mathbf{x}_i|\Omega)} \leq \left[\frac{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}\right]^{1.5} \iff \frac{q^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{q^*(\mathbf{z}_i|\Omega)}{q^*(\mathbf{x}_i|\Omega)} \leq \left[\frac{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}\right]^{0.5}$$

$$\iff \frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)} \leq \left[\frac{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}\right]^{0.5}$$

By the preceding and by Monotonicity of q^*,

$$\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)} \leq \left[\frac{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}\right]^{0.5}\left[\frac{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}\right]^{0.5}$$

$$\leq \left[\frac{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}\right]^{0.5}\left[\frac{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}\right]^{0.5} = \frac{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{x}_i\mathbf{x}_j|\Omega)},$$

thereby satisfying (17). This confirms $p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z}) \leq 1$ and completes the proof. ■

N Appendix: Proof of Proposition 7

Proof. The posterior marginal is given by

$p(\mathbf{x}_i|\mathbf{z}) = p(\mathbf{x}_i\mathbf{z}_j|\mathbf{z})$

$= [q^*(\mathbf{x}_i|\mathbf{z}_i)q^*(\mathbf{z}_j|\mathbf{z}_j)]\left[\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]\left[\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]$

$= [q^*(\mathbf{x}_i|\mathbf{z}_i)]\left[\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]^2$, and similarly $p(\mathbf{x}_j|\mathbf{z}) = [q^*(\mathbf{x}_j|\mathbf{z}_j)]\left[\frac{SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]^2$

Then

$$\frac{p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{p(\mathbf{x}_i|\mathbf{z})p(\mathbf{x}_j|\mathbf{z})} = \left[\frac{SD^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]\left[\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]^{-1}$$

$$= SD^*(\mathbf{x}_i\mathbf{x}_j|\Omega)\left[\frac{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}\right]$$

$$= \left[\frac{q^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}{q^*(\mathbf{x}_i|\Omega)q^*(\mathbf{x}_j|\Omega)}\right]\frac{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{z}_i|\Omega)q^*(\mathbf{z}_j|\Omega)}\frac{q^*(\mathbf{x}_i|\Omega)q^*(\mathbf{z}_j|\Omega)}{q^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}\frac{q^*(\mathbf{z}_i|\Omega)q^*(\mathbf{x}_j|\Omega)}{q^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}$$

$$= \frac{q^*(\mathbf{x}_i\mathbf{x}_j|\Omega)q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{x}_i\mathbf{z}_j|\Omega)q^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}$$

$$= \frac{q^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{x}_i\mathbf{z}_j|\Omega)q^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}$$

$= \frac{q^*(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})}{q^*(\mathbf{x}_i|\mathbf{z})q^*(\mathbf{x}_j|\mathbf{z})}$, where we have used the definition of Bayesian beliefs in the last step. ■

O Appendix: Proof of Proposition 8

Proof. Determine that

$$p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})$$

$$= [q^*(\mathbf{x}_i|\mathbf{z}_i)q^*(\mathbf{x}_j|\mathbf{z}_j)]\left[\frac{SD^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]\left[\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]$$

$$= \left[\frac{q^*(\mathbf{x}_i|\Omega)}{q^*(\mathbf{z}_i|\Omega)}\frac{q^*(\mathbf{x}_j|\Omega)}{q^*(\mathbf{z}_j|\Omega)}\right]\left[\frac{q^*(\mathbf{x}_i\mathbf{x}_j|\Omega)q^*(\mathbf{z}_i|\Omega)q^*(\mathbf{z}_j|\Omega)}{q^*(\mathbf{x}_i|\Omega)q^*(\mathbf{x}_j|\Omega)q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]\left[\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]$$

$$= \frac{q^*(\mathbf{x}_i\mathbf{x}_j|\Omega)}{q^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\left[\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]$$

$= p^{BU}(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})\left[\frac{SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\frac{SD^*(\mathbf{z}_i\mathbf{x}_j|\Omega)}{SD^*(\mathbf{z}_i\mathbf{z}_j|\Omega)}\right]$,and the assertion follows. ■

P Appendix: Proof of Proposition 9

Proof. Observe that

$$p(\mathbf{y}_i\mathbf{x}_j|\mathbf{z}) \geq p(\mathbf{x}_i\mathbf{x}_j|\mathbf{z})$$

$$\iff q^*(\mathbf{y}_i|\mathbf{z}_i)SD^*(\mathbf{y}_i\mathbf{x}_j|\Omega)SD^*(\mathbf{y}_i\mathbf{z}_j|\Omega) \geq q^*(\mathbf{x}_i|\mathbf{z}_i)SD^*(\mathbf{x}_i\mathbf{x}_j|\Omega)SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)$$

$$\iff \frac{q^*(\mathbf{y}_i|\Omega)}{q^*(\mathbf{z}_i|\Omega)}SD^*(\mathbf{y}_i\mathbf{x}_j|\Omega)SD^*(\mathbf{y}_i\mathbf{z}_j|\Omega) \geq \frac{q^*(\mathbf{x}_i|\Omega)}{q^*(\mathbf{z}_i|\Omega)}SD^*(\mathbf{x}_i\mathbf{x}_j|\Omega)SD^*(\mathbf{x}_i\mathbf{z}_j|\Omega)$$

$$\iff q^*(\mathbf{y}_i\mathbf{x}_j|\Omega)\frac{q^*(\mathbf{y}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{y}_i|\Omega)} \geq q^*(\mathbf{x}_i\mathbf{x}_j|\Omega)\frac{q^*(\mathbf{x}_i\mathbf{z}_j|\Omega)}{q^*(\mathbf{x}_i|\Omega)},$$

which yields the desired expression. ■

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

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

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

[6] Epstein, L., J. Noor and A. Sandroni (2008): "Non-Bayesian Updating: A theoretical framework", *Theoretical Economics* 3, pp. 193–229.

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

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

[9] Gilboa, I. and D. Schmeidler (1995): "Case-Based Decision Theory". *Quarterly Journal of Economics* 110(3), pp. 605–639.

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

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

[12] 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.

[13] 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.

[14] 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.

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

[16] Hume, D. (1748). *An Inquiry Concerning Human Understanding.* Oxford Clarendon Press, second edition 1966.

[17] Jehiel, P (2005): "Analogy-Based Expectation Equilibrium", *Journal of Economic Theory* 123, pp. 81–104.

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

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

[20] Ortoleva, P. (2012): “Modeling the Change of Paradigm: Non-Bayesian Reaction to Unexpected News”, *American Economic Review* 102(6), pp. 2410-2436.

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

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

[23] Rubinstein, A. (2016): “A Typology of Players: Between Instinctive and Contemplative”, *Quarterly Journal of Economics*, pp. 859–890.

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

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

[26] Schmeidler, D. (1989): “Subjective Probability and Expected Utility without Additivity”, *Econometrica* 57(3), pp. 571-587.

[27] Shafer, G. (1976). *A Mathematical Theory of Evidence.* Princeton University Press.

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

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

[30] Waegenaere A. and P. Wakker (2001): “Nonmonotonic Choquet Integrals”, *Journal of Mathematical Economics* 36, pp. 45–60.

[31] Zhao, C. (2017): “Pseudo-Bayesian Updating”, mimeo.