Do Process-1 simulations generate the epistemic feelings that drive Process-2 decision making?

Chris Fields
23 Rue des Lavandières
11160 Caunes Minervois, FRANCE
fieldsres@gmail.com
ORCID: 0000-0002-4812-0744

and

James F. Glazebrook
Department of Mathematics and Computer Science
Eastern Illinois University, 600 Lincoln Ave.
Charleston, IL 61920–3099, USA
jfglazebrook@eiu.edu
Adjunct Faculty
Department of Mathematics
University of Illinois at Urbana–Champaign
Urbana, IL 61801, USA

June 14, 2020

Abstract

We apply previously developed Chu space and Channel Theory methods, focusing on the construction of Cone-Cocone Diagrams (CCCDs), to study the role of epistemic feelings, particularly feelings of confidence, in dual-process models of problem solving. We specifically consider "Bayesian brain" models of probabilistic inference within a global neuronal workspace (GNW) architecture. We develop a formal representation of Process-1 problem solving in which a solution is reached if and only if a CCCD is completed. We show that in this representation, Process-2 problem solving can be represented as multiply-iterated Process-1 problem solving and has the same formal solution conditions. We then model the generation of explicit, reportable subjective probabilities from implicit, experienced confidence as a simulation-based, reverse-engineering process, and show that this process can also be modeled as a CCCD construction.

Keywords: Bayesian inference, Dual-process models, Epistemic feelings, Chu space, Channel Theory, Cone-Cocone Diagram, Problem solving.

1 Introduction

Since the pioneering work of Simon (1967, 1972) and Tversky and Kahneman (1973, 1974), dual process models that distinguish fast, heuristic, and highly-automated ("Process-1") from slow, deliberative, and effortful ("Process-2") reasoning, decision making, and problem solving processes have become commonplace (see Evans (2006, 2010); Kahneman (2011); Evans and Stanovich (2013) for reviews, Moors and De Houwer (2006); Melnikoff and Bargh (2018) for criticism, and Frankish (2010) for a philosophical perspective with extensive historical background). A wealth of information is available on the functional aspects of this distinction, including the speed-accuracy tradeoff (SAT; Wickelgren, 1977; Heitz, 2014), the susceptibility of Process-1 to an array of biases and its consequent ease of manipulation (Kahneman, 2011), and the sensitive dependence of Process-2 on attentional resources, in part to counteract elements of bias endemic to Process-1 (Evans, 2008). There are also salient phenomenological differences between Process-1 and Process-2 cognition. Problem solving using Process-2 is experienced as a *process*: it feels effortful and extended in time, and typically involves a sequence of experienced intermediate steps. Problem solving using Process-1, in contrast, is not experienced as a multistep process. Only the problem to be solved and its solution are experienced. These differences are often summarized by characterizing Process-1 and Process-2 cognition as "unconscious" and "conscious" (Wason and Evans, 1975), "automatic" and "deliberate" (Evans, 2006), or "intuitive" and "rational" (Kahneman, 2011) respectively.

As Moors and De Houwer (2006); Melnikoff and Bargh (2018) point out in their critical reviews, aligning the opposed pairs of phenomenological characteristics often associated with Process-1 and Process-2 does not cleanly separate all of cognition into two distinct, non-overlapping categories. Much "Process-1" cognition is deliberate, for example, in that it is performed intentionally for the purpose of achieving some goal, and "Process-2" cognition can in some cases fail to be rational, or even to meet minimal standards of coherence. In what follows, therefore, we will focus on the kinds of canonical cases that motivate the distinction (Tversky and Kahneman, 1973, 1974). Canonical Process-2 cognition is exemplified by non-expert solutions to difficult mathematical "word" problems. Here problem solving requires multiple, consciously distinguished, individually reportable, sequential steps, conscious recall of step-relevant facts and rules, and conscious judgements about relevance, correctness, and whether progress is being made toward a solution. This kind of conscious, sequential process is what is commonly called "thinking through" a problem. In contrast, canonical Process-1 cognition is exemplified by the construction of simple grammatical sentences in one's native language within a relaxed conversational setting. While sentence production in such a setting is clearly driven by communicative goals, it "happens naturally" in a single step, without conscious recall of grammatical rules or alternative formulations, and without awareness of exactly what will be said until the utterance is actually produced. Sentence production can become Process-2, e.g. when "speaking carefully" to an opponent in a dispute. Here the phenomenology is more similar to solving mathematical problems than to relaxed conversation. We will use the terms "Process-1" and "Process-2" in what follows to refer to such canonical cases.

Positive epistemic feelings, e.g. feelings of familiarity, recognition, agency, ownership, and confidence (see Arango-Muñoz, 2014; Proust, 2015; Schwarz, 2012, for reviews) signal successful completion for both kinds of problem solving, with the "Aha!" of insight (Kounios and Beeman, 2015) as an extreme example. More negative epistemic feelings such as doubt or frustration are typical of the early stages of canonical Process-2 problem solving and may persist until a solution is finally reached. Negative epistemic feelings may also accompany outright failure of either Process-1 or Process-2, for example, when one blurts out an entirely unintended and unfortunate statement in conversation (Process-1) or reaches the end of a difficult mathematical "proof" only to realize that it contains a contradiction (Process-2). Process-2 problem solving also often involves, and in the rational-agent models of classical, normative decision theory mechanistically requires (e.g. Parmigiani and Inoue, 2009), a particular kind of epistemic feeling not encountered in Process-1 problem solving: a feeling of confidence made explicit and at least semi-quantitative as a subjective probability or "degree of belief" (de Finetti, 1974) in a recalled background fact, an intermediate step, or a considered possible solution. In a traditional representational model of problem solving (e.g. Fodor, 1983; Pylyshyn, 1984), such degrees of belief would be explicitly stored, e.g. as quantitative values associated with propositions, for recall when needed by Process-2. Both theoretical and empirical considerations now weigh against such models, suggesting instead that explicit, experienced subjective probabilities are constructed (Sanborn and Chater, 2016; Chater, 2018) or "read out" (Meyniel, Sigman and Mainen, 2015) from an underlying implicit representation. Here we suggest a particular model of this process: that explicit, experienced subjective probabilities are synthesized by running multiple Process-1 simulations of closely-related problems or problem components. The energy and time required to do this, we suggest, contribute to Process 2 being effortful and slow.

We begin in §2 by describing Process-1 problem solving in a general model framework based in category theory (e.g. Adámek, Herrlich and Strecker, 2004; Awodey, 2010) originally developed to provide a representation for bidirectional constraint flow in perceptual categorization (Fields and Glazebrook, 2019a,b). We show how quasihierarchical networks of Bayesian inferences can be represented by a conceptually simple category-theoretic construction, a scale-free information-processing architecture called a *Cone-Cocone Diagram* (CCCD). This representation naturally generalizes from perceptual categorization to Process-1 problem solving in any domain, a generalization consistent both with the view of cognition as essentially perceptual proposed by Chater (2018), and with standard predictive Processing (PP), hierarchical predictive coding (HPC), or "Bayesian Brain" models (see Knill and Pouget, 2004; Friston, 2010; Clark, 2013; Hohwy, 2013; Seth, Suzuki and

Critchley, 2012, among many others). The CCCD representation does not in any way contradict standard PP/HPC models, but rather expresses them in a more general mathematical language that provides an explicit representation of the logical constraints that hold between representations at either the same or different levels of the hierarchy. A CCCD requires, by its construction, the inferential coherence that is enforced in Bayesian models by probability-theoretic constraints (Fields and Glazebrook, 2020), but with the additional features that inferences are logical as well as probabilistic and are intrinsically semantic. The CCCD representation also naturally accommodates contextuality, including "intrinsic" contextuality as discussed by Dzhafarov and Kujala (2017a); Dzhafarov and Kon (2018), by allowing probability distributions to be context-dependent in a well-defined way (Fields and Glazebrook, 2020). A problem is solvable in the CCCD representation if and only if a CCCD modeling the solution process can be constructed.

We proceed in §3 to discuss subjective probabilities, distinguishing their implicit representation by connection strengths or activation values, their expression as feelings of confidence in outcomes of Process-1 problem solving, and their explicit representation as reportable degrees of belief in Process-2 problem solving. We formulate in §4 the question of how highly-distributed sets of implicitly-represented probabilties could be repackaged into and summarized by feelings of confidence in a Process-1 solution, using a generalized global neuronal workspace (GNW) model (Baars and Franklin, 2003; Baars, Franklin and Ramsoy, 2013; Dehaene and Naccache, 2001; Dehaene, Sergent and Changeux, 2003; Dehaene, Charles and King, 2014; Shea and Frith, 2019) that includes both perception and interoception. Here we follow the basic PP/HPC dictum that the brain assimilates a cascade of competing model-based simulations to address "what is this new sensory input like?" (Bar, 2009) while acknowledging that preservation of bodily state is both the fundamental expectation around which active inference is organized and a sensitive indicator, via affective feedback, of the possible consequences for bodily state of projected world states or actions (Barrett and Simmons, 2015; Barrett, 2017; Seth, 2013; Van de Cruys, 2017).

Within the GNW model, feelings of confidence, (epistemic) uncertainty, familiarity, and other kinds of epistemic feelings, are synthetic interoceptions that blend cortical performance-monitoring signals with subcortical valence and reward signals. Interestingly, we are close to the recent hypothesis of Shea and Frith (2019) that workspace representations involve metacognitive parameters such as confidence and uncertainty, including, as we propose here, epistemic feelings. While prediction error itself indicates short-term performance, Van de Cruys (2017) has argued that the rate of change of prediction error is the key correlate of experienced affect, with rapidly improving predictions generating positive and rapidly worsening predictions generating negative experiences. How do signals dependent on temporally-extended processing modulate the uninterrupted bidirectional flow of constraint information through the parallel, effectively domain-specific channels "below" the GNW? How do worsening predictions interrupt fluid Process-1 performance, resulting in slower, stepwise Process-2 performance? Within a GNW picture, an interrupt signal must be strong enough to "ignite" conscious processing (Hohwy, 2013; Barrett, 2017; Whyte, 2019). With this background, we outline our primary hypothesis, that Process-2 cognition is implemented by multiple rounds of Process-1 cognition, and then show in §5 how a model of experienced, explicit subjective probabilities as synthesized from confidence measures naturally follows. If this model is correct, conscious deliberation is effectively a reverse-engineering activity: it is discovering, by the experimental tactic of running multiple simulations with slightly varying initial conditions, what one thinks (cf. the notion of a "representational exchange mechanism" recently proposed by Cushman (2020)). As this reverse-engineering process may be employed to a greater or lesser extent in different contexts, one can expect canonical Process-1 and Process-2 cognition to be two ends of a continuum with a substantial grey area in between, perhaps explaining some of the "misalignments" of phenomenological markers discussed by Melnikoff and Bargh (2018). We close by formulating model predictions and discussing possible experimental approaches (§6).

2 Visual object recognition as a model of problem solving

In Fields and Glazebrook (2019b), we showed how the processes of visual object categorization and individual object recognition can be represented by a mathematical object, a CCCD, that describes a quasi-hierarchical network of inferential constraints. Here we first outline the basic intuitions behind the CCCD formalism. Next we summarize, from Fields and Glazebrook (2019a), the mathematical tools needed (§2.2) and use these tools to construct a CCCD (§2.3). We then describe a CCCD in intuitive terms, using visual object recognition as an example (§2.4). Readers uninterested in the technical details can skip directly to this intuitive example. All "canonical" Process-1 problem solving, we then claim (§2.5), can be represented using the same structure, i.e. a CCCD, used for visual object recognition.

2.1 The basic intuitions

Suppose Alice is examining a map of London. Some large-scale pattern of activity spanning Alice's occipital, temporal, and parietal lobes represents the map "to Alice" as a conscious visual percept. At the same time, the map is a conventionalized representation of London. Because the semantic relation instantiated in Alice's neurocognitive system and the semantic relation instantiated by the social conventions of cartography are connected by an object, the map that Alice is examining, Alice's visual experience when examining the map provides information not just about the map, but also about London.

Combining semantic information ("meanings") of different kinds or from different sources to produce "higher-level" representations is a fundamental and ubiquitous function of cognition. How can it be described in a way that captures the logic of such combinations, but is independent of both the algorithms or other procedures used and how these are implemented? Barwise and Seligman (1997) employed mathematical tools from category theory, the most general available mathematical language (see e.g. Adámek, Herrlich and Strecker, 2004; Awodey, 2010, for accessible introductions), to develop a way of describing the relations between the semantics of different representations without having to explicitly model the structures or implementations of those representations. The "language" used to do this is, like category theory itself, diagrammatic. The relations between

Alice's neurocognitive activity, her map, and London can be diagrammed as:

$$\begin{tikzcd}
& & \text{Combined representation} & & \\
& \text{Visual representation} \arrow[ur, "\phi"] & & \text{Conventional representation} \arrow[ul, "\psi"'] & \\
\text{Alice's brain} \arrow[ur, "f_1"] & & \text{The map} \arrow[ul, "f_2"'] \arrow[ur, "g_1"] & & \text{London} \arrow[ul, "g_2"']
\end{tikzcd}$$

The arrows in this diagram represent a bottom-up flow of logical constraints on the meanings of representations. The activity pattern in Alice's brain involves visual-processing areas, hence her representation of the map is constrained (arrow f_1) to be visual. The shapes depicted on the map constrain (f_2) the "shape" that Alice imaginatively experiences. The shapes depicted on the map likewise constrain (g_1) the geographical locations that it can represent while remaining consistent with the conventions of cartography; the "shape" of London itself similarly constrains (g_2) its set of possible cartographically-accurate maps. The semantics of Alice's visual representation and the semantics of the conventional cartographic representation similarly constrain (ϕ and ψ) the combined representation to be simultaneously visual, cartographic, and of London. This kind of relationship between constraints from different sources is called a "distributed system" in the language of Barwise and Seligman (1997); we provide a formal definition and generalize the above diagram in (2.5) below.

The flow of semantic constraints in map reading, or in distributed systems generally is not, however, only bottom-up. If Alice gets "turned around" while navigating London, she may realize by seeing the actual layout of some part of the city that she has read the map incorrectly. She may see that the map is wrong or out of date. In either case, the constraints in the above diagram also flow downward, from Alice's comparison of the map to London to semantic features of her visual representation, the map, or perhaps even London itself. Alice's ability to navigate using a crude sketch on a bar napkin, for example, loosens somewhat the usual cartographic conventions, at least for Alice. As we will see in §2.3 and then use in §2.4, the CCCD formalism is based on recognition of the bidirectionality of semantic constraint flow in distributed systems. First, however, it is useful to lay out the basic conceptual tools with which a CCCD is defined.

2.2 Basics of Channel Theory

Category Theory is a general mathematical language for describing objects and relations (Adámek, Herrlich and Strecker, 2004; Awodey, 2010). Here we give a short outline of the basic ideas. A *category* $\mathfrak{C}$ comprises a set of objects and a set of arrows (i.e. directed relations, or morphisms) between objects, satisfying two requirements: 1) arrows compose associatively, i.e. for objects A, B, C, D, if $f : A \to B$, $g : B \to C$, and $h : C \to D$, then $hgf : A \to D$, and 2) each object has an identity arrow $\mathrm{id}_A : A \to A$. Readers familiar with mathematically paired concepts such as sets with functions, sets with relations, vector spaces with linear mappings, groups with group homomorphisms, topological spaces with continuous mappings, as the respective objects and arrows, will see these as familiar examples of categories. One of the simplest categories is the category of Chu spaces, any object in which is defined as follows:

Definition 2.1. *A (dyadic or two-valued) Chu space* C consists of a triple $(C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$ where C_{o} is a set of *objects*, C_{a} is a set of *attributes*, along with a *satisfaction relation* (or *evaluation*) $\Vdash_{\mathsf{C}} \subseteq C_{\mathsf{o}} \times C_{\mathsf{a}} \longrightarrow \mathsf{K}$, given a set K (with no structure assumed).

With $\mathsf{K} = \{0, 1\}$, take a set (of objects) $X = \{a, b, c\}$. This can be represented as the Chu space

$\Vdash$								
a	0	1	0	1	0	1	0	1
b	0	0	1	1	0	0	1	1
c	0	0	0	0	1	1	1	1

By associating objects with attributes, Chu spaces provide a natural model of the process of categorizing objects by their attributes; the categorization process implements the satisfaction relation $\Vdash$. They easily generalize to multi-valued satisfaction relations, e.g. relations satisfied with some probability. The arrows in the category of Chu spaces relate one categorization process to another:

Definition 2.2. A *Chu transform* of a Chu space $\mathsf{C} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$ to a Chu space $\mathsf{D} = (D_{\mathsf{o}}, \Vdash_{\mathsf{D}}, D_{\mathsf{a}})$ is a pair of functions $(f_{\mathsf{a}}, f_{\mathsf{o}})$ with $f_{\mathsf{o}} : C_{\mathsf{o}} \longrightarrow D_{\mathsf{o}}$, and $f_{\mathsf{a}} : D_{\mathsf{a}} \longrightarrow C_{\mathsf{a}}$, such that for any $x \in C_{\mathsf{o}}$, and $y \in D_{\mathsf{a}}$, we have $f_{\mathsf{o}}(x) \Vdash_{\mathsf{D}} y$, if and only if $x \Vdash_{\mathsf{C}} f_{\mathsf{a}}(y)$.

These transforms constrain both the objects and attributes of the Chu spaces to which they relate, forcing the two categorization processes to "line up" in the way one would intuitively expect.

Chu spaces were originally developed within category theory itself (Barr, 1979, 1991). They are more general than topological spaces, and have been extensively applied in theoretical computer science (Pratt, 1995, 1999a,b), physics (Abramsky, 2012; Fields and Glazebrook, 2020; Gratus and Porter, 2006) as well as elsewhere in mathematics, mainly because of the considerable scope for the choice of the relation $\Vdash$, and hence the structure afforded by the corresponding Chu space. These include probabilistic (in particular, conditional) relations (Allwein, Moskowitz and Chang, 2004; Nhuy and Van Quang, 2001), spatial observations (Gratus and Porter, 2006), and "degree to which it belongs" fuzzy-type relations (Papadopoulos and Syropoulos, 2000); numerous examples are discussed in Fields and Glazebrook (2019a). Significantly, all theories of 'relational structures' (with and without topological structure) can be modeled by Chu spaces (Pratt, 1997); such 'structural' models include those pertaining to analogy and metaphor (e.g. Brown and Porter, 2006; Fields, 2011, 2013; Gentner, 1983; Old and Priss, 2001).

For the present purposes, the most important application of Chu spaces is in modeling semantic (Dretske, 1981) or "pragmatic" (Roederer, 2010) information flow and inference, which are presented here in the guise of Chu spaces known as *Channel Theory* (Barwise and Seligman, 1997). The fundamental concept in this case is the idea of a "Classification" relating "Tokens" to the "Types" that encompass them:

Definition 2.3. A *Classification* (sometimes called a "classifier") $\mathcal{A} = \langle \mathrm{Tok}(\mathcal{A}), \mathrm{Typ}(\mathcal{A}), \Vdash_{\mathcal{A}} \rangle$ consists of a set $\mathrm{Tok}(\mathcal{A})$ consisting of the *tokens of* $\mathcal{A}$, a set $\mathrm{Typ}(\mathcal{A})$ consisting of the *types of* $\mathcal{A}$, and a classification relation

$$\Vdash_{\mathcal{A}} \subseteq \mathrm{Tok}(\mathcal{A}) \times \mathrm{Typ}(\mathcal{A}), \tag{2.1}$$

that classifies tokens to types.

Example 2.1. Following Barwise and Seligman (1997, Ex. 2.2, p.28), a first order language L is a classification, where Tok(L) consists of a set M of certain mathematical/logical structures, and Typ(L) consists of sentences in L, and $M \Vdash \varphi$, if and only if φ is true in the token M. The type set of a token M is the set of all sentences of L true in M, called the *theory* of M (see Barwise and Seligman (1997, Ch.9) for the development of formal details pertaining to the latter concept).

In Channel Theory, Chu transforms are formulated as "infomorphisms" mapping one classification to another (Barwise and Seligman, 1997):

Definition 2.4. Given two classifications $\mathcal{A} = \langle \mathrm{Tok}(\mathcal{A}), \mathrm{Typ}(\mathcal{A}), \Vdash_{\mathcal{A}} \rangle$ and $\mathcal{B} = \langle \mathrm{Tok}(\mathcal{B}), \mathrm{Typ}(\mathcal{B}), \Vdash_{\mathcal{B}} \rangle$, an *infomorphism* $f : \mathcal{A} \rightleftarrows \mathcal{B}$, is a pair of contravariant maps

i) $\overrightarrow{f} : \mathrm{Typ}(\mathcal{A}) \longrightarrow \mathrm{Typ}(\mathcal{B})$

ii) $\overleftarrow{f} : \mathrm{Tok}(\mathcal{B}) \longrightarrow \mathrm{Tok}(\mathcal{A})$

such that for all $b \in \mathrm{Tok}(\mathcal{B})$, and for all $a \in \mathrm{Typ}(\mathcal{A})$, we have

$$\overleftarrow{f}(b) \Vdash_{\mathcal{A}} a, \text{ if and only if } b \Vdash_{\mathcal{B}} \overrightarrow{f}(a). \tag{2.2}$$

This last condition may be schematically represented by:

$$\begin{tikzcd}
\mathrm{Typ}(\mathcal{A}) \arrow[r, "\overrightarrow{f}"] \arrow[d, dash, "\Vdash_{\mathcal{A}}"'] & \mathrm{Typ}(\mathcal{B}) \arrow[d, dash, "\Vdash_{\mathcal{B}}"] \\
\mathrm{Tok}(\mathcal{A}) & \mathrm{Tok}(\mathcal{B}) \arrow[l, "\overleftarrow{f}"']
\end{tikzcd} \tag{2.3}$$

In this definition, types play the role of Chu-space objects while tokens play the role of attributes, an example of the duality between objects and arrows characteristic of category theory in general. The satisfaction relations $\Vdash_{\mathcal{A}}$ and $\Vdash_{\mathcal{B}}$ are explicitly regarded as enforcing semantic, not merely syntactic or set-theoretic constraints, rendering both classifications and infomorphisms intrinsically semantic notions. The intuition is that an infomorphism transmits the information from one classification to another, so that, e.g. "b is type B" can encode or represent the information "a is type A". Note that "information" here does not simply specify a quantity of bits, as typical of Shannon information (Cover and Thomas, 2006), but it is rather the set of logical constraints imposed by Definition 2.4. Thus it may be viewed as "pragmatic information" as proposed in Roederer (2010).

Example 2.2. We exemplify these latter concepts with a straightforward example. Consider the classification $\mathbf{M} = \langle Messages, Contents, \Vdash_{\mathbf{M}} \rangle$ where Messages are classified by their Contents (Allwein, Moskowitz and Chang, 2004).
Suppose we have another such classification $\mathbf{M}' = \langle Messages', Contents', \Vdash_{\mathbf{M}'} \rangle$. An infomorphism $f : \mathbf{M} \longrightarrow \mathbf{M}'$ may represent a function decoding messages from $\mathbf{M}'$ to messages in $\mathbf{M}$, so that whatever can be noted about the translation, may be mapped into something noted in the original message. That is, $m^f \Vdash_{\mathbf{M}} C \Leftrightarrow m \Vdash_{\mathbf{M}'} C^f$.

This idea of an infomorphism as a mapping between classifications provides the basic building block for constructing multi-level, quasi-hierarchical classification systems. Like the connections between "processing layers" in brains, infomorphisms are intrinsically bidirectional. Channel Theory provides a *flow of reasoning* thanks to the allied concepts of *local logic* and *logic infomorphism* (Barwise and Seligman, 1997, Ch.12) in which infomorphisms become manifestly bidirectional maps between sets of logical relations.* This structure is occasionally represented by concept lattices, as in e.g. for reference ontologies (Kalfoglou and Schorlemmer, 2003), or for metaphor (Old and Priss, 2001). We refer the interested reader to Barwise and Seligman (1997, Ch.12) for the complete formal details (see also Fields and Glazebrook (2019a,b) with examples), but for the sake of a present workable explanation, let us say that a classification can be extended to a local logic by specifying a subset (possibly a singleton) of tokens satisfying the types of a (regular) theory (such as exemplified in Example 2.1) that specifies the logical aspects of some situation. Accordingly, an infomorphism is extended to a logic infomorphism which preserves this additional (logical) structure. In fact, any classification admits its own local logic (Barwise and Seligman, 1997, 9.1). Significantly, the flow of information through a network of such logic infomorphisms can naturally be interpretable as "inference" in the usual sense. As the satisfaction relation $\Vdash$ can be considered time- and context-dependent, these inferential processes can be regarded as having similar dependencies. Putting it

*The role of an infomorphism contrasts with the superficially-similar treatment of Ehresmann and Vanbremeersch (2007); Ehresmann and Gomez-Ramirez (2015), which basically implements maps between neurons or co-activated functional assemblies of neurons (christened "cat-neurons").

another way, if component classifiers represent causal "if – then" relations, the inferences implemented are likewise causal. We can also interpret classifications themselves as probabilistic (cf. Barwise (1997)), and to see this we introduce an important ingredient of a local logic, namely the *sequent* which we define now:

Definition 2.5. A *sequent* $M \Vdash_{\mathcal{A}} N$ holding of a classification $\mathcal{A}$ is a pair of subsets M, N of $Typ(\mathcal{A})$ such that $\forall\ x \in Tok(\mathcal{A})$, $x \models_{\mathcal{A}} M \Rightarrow x \Vdash_{\mathcal{A}} N$.

We observe that a sequent encodes a semantic, e.g. causal constraint that in information flow functions effectively as a logical gate. As pointed out in Allwein (2004); Allwein, Yang and Harrison (2011) (see also Fields and Glazebrook (2019a)), the sequent can be relaxed by requiring only that if $x \Vdash_{\mathcal{A}} M$, there is some probability $P(N|M)$ such that $x \Vdash_{\mathcal{A}} N$. This is essentially how a conditional probability interprets the logical implication "$\Rightarrow$" (Adams, 1998), leading to a representation of probabilistic inference, including Bayesian inference as applied in Fields and Glazebrook (2019b, 2020). We will use this interpretation in §2.5 where the information flow itself generates probability distributions (an earlier discussion of inference along similar lines was proposed in McClelland (1998)).

2.3 Constructing a CCCD

As the first step in constructing a CCCD, we define a finite *information channel* **Chan** as a finite indexed family $\{f_i : \mathcal{A}_i \rightleftarrows \mathbf{C}\}_{i \in \mathcal{I}}$ of infomorphisms having a common codomain $\mathbf{C}$, called the *core of the channel* **Chan**:

$$\begin{tikzcd} & \mathbf{C} & \\ \mathcal{A}_1 \arrow[ur, "f_1"] & \mathcal{A}_2 \arrow[u, "f_2"] & \ldots\ \mathcal{A}_k \arrow[ul, "f_k"'] \end{tikzcd} \tag{2.4}$$

The core $\mathbf{C}$, which itself is a classification, functions as a carrier of information flow between the f_i, and hence between the component classifications $\mathcal{A}_i$. Intuitively, the channel can be viewed as a "wire" connecting two agents (i.e. classifiers) to a "blackboard" or other shared memory via which they can exchange information. As the shared memory $\mathbf{C}$ is itself a classifier, it admits a structure regulating how information is written to and read from it; for example, it may be a "smart blackboard" that incorporates a function such as a multi-language translation (Fields and Glazebrook, 2019a). This idea of a channel core as a shared memory is developed in the setting of interacting channels forming a *distributed system* (Barwise and Seligman, 1997), for example:

$$
\begin{tikzcd}
 & & \mathbf{C} & & \\
 & \mathbf{B}_1 \arrow[ur, "\phi"] & & \mathbf{B}_2 \arrow[ul, "\psi"'] & \\
\mathcal{A}_1 \arrow[ur, "f_1"] & & \mathcal{A}_2 \arrow[ul, "f_2"'] \arrow[ur, "f_3"] & & \mathcal{A}_3 \arrow[ul, "f_4"']
\end{tikzcd}
\tag{2.5}
$$

which generalizes the relation between a map reader (here $\mathcal{A}_1$), a map ($\mathcal{A}_2$), and the territory being mapped ($\mathcal{A}_3$) discussed in §2.1. This idea of distributing semantic relations across multiple systems underlies much of the following constructions.

In the sense of maximally abstract, while preserving the mathematical structure of interest, the most general channel on a finite set of classifiers corresponds to the category-theoretic notion of a (finite) *cocone* (the prefix "co-" indicating dual, in this case of a cone), with the core $\mathbf{C}'$ the colimit of all possible upward-going structure-preserving maps from the classifiers $\mathcal{A}_i$ (Awodey, 2010). Such a colimit core, provided it exists, can be regarded as "containing" or "binding" in its structure as a classifier, all of the information that is common to the component classifications $\mathcal{A}_i$ (Fields and Glazebrook, 2019a, provides a detailed construction and examples). The cocone must commute, i.e. the rightward arrows $\mathcal{A}_i \to \mathcal{A}_j$ between the component classifications must be such that $f_i = f_j g_{ij}$ for all i, j, where g_{ij} can be any composition of arrows $\mathcal{A}_i \to \cdots \to \mathcal{A}_j$. This commutativity requirement makes explicit the role of $\mathbf{C}'$ as a "wire" or shared memory for the component classifiers (Barwise and Seligman, 1997; Allwein, Yang and Harrison, 2011). It assures inferential coherence by assuring joint activation of all of the classifiers covered by the cocone core

$\mathbf{C}'$ and hence joint, parallel use of all possible inferential paths through the cocone diagram (CCD):

$$\begin{tikzcd}
 & \mathbf{C}' & \\
\mathcal{A}_1 \arrow[ur, "f_1"] \arrow[r, "g_{12}"'] & \mathcal{A}_2 \arrow[u, "f_2"'] \arrow[r, "g_{23}"'] & \dots\ \mathcal{A}_k \arrow[ul, "f_k"']
\end{tikzcd} \tag{2.6}$$

A commuting finite cone of infomorphisms is the dual construction, in which all the arrows are reversed. In this case the core of the (dual) channel is the limit of all possible downward-going structure-preserving maps to the classifiers $\mathcal{A}_i$.

We can now define the central idea of a finite, commuting *Cone-Cocone Diagram* (CCCD) as comprising both a cone and a cocone on a single finite set of classifications $\mathcal{A}_i$.

$$\begin{tikzcd}
 & \mathbf{C}' & \\
\mathcal{A}_1 \arrow[ur, "f_1"] \arrow[r, shift left, "g_{21}"] & \mathcal{A}_2 \arrow[u, "f_2"'] \arrow[l, shift left, "g_{12}"] \arrow[r, shift left, "g_{32}"] & \dots\ \mathcal{A}_k \arrow[ul, "f_k"'] \arrow[l, shift left, "g_{23}"] \\
 & \mathbf{D}' \arrow[ul, "h_1"] \arrow[u, "h_2"] \arrow[ur, "h_k"'] &
\end{tikzcd} \tag{2.7}$$

It is natural to interpret this diagram as depicting a flow of constraint information, represented by the component classifications, from D' through a set of component classifiers to C'; as noted above, this "information flow" is inferential in a natural sense. Commutativity here requires that any path from D' to C', including any number of lateral maps between component classifiers, yields the same result; this assures that all available inferential paths are brought to bear on the "input" encoded by D', and hence assures inferential coherence. The bidirectional maps $\mathcal{A}_i \leftrightarrow \mathcal{A}_j$ between component classifiers are naturally interpreted as encoding mutual, lateral constraints on the behavior of the component classifiers; these are logical constraints imposed by the component classifiers on each other. The maps f_i and h_j can be arbitrarily finitely expanded by inserting intermediate "layers" of additional classifications, e.g. $f_i : \mathcal{A}_i \to C' \Rightarrow f_i = f_{ib}f_{ia} : \mathcal{A}_i \to \mathcal{B}_i \to C'$ for some intermediate classification $\mathcal{B}_i$; hence a CCCD can have an arbitrary finite number of layers of classifiers. Below we will explain how this descriptive mechanism can be applied, and then employ it to represent

Process-1 problem solving.

2.4 Using CCCDs to model object recognition

Here we elaborate on the intuitions behind the construction of the CCCD as introduced in the previous section, and demonstrate its utility in the context of visual object identification. The most basic and most salient feature of the construction is that it provides a means of capturing a simultaneous upward and downward flow of semantic constraints implemented by the dynamics of the object-recognition process. This already makes clear the structural similarity between CCCDs and recurrent neural network models. The formal relationship between a cocone diagram and a general feedforward neural network has been developed in detail (Kikuchi et al., 2003); reversing the arrows, and superposing yields a recurrent network.

A system capable of both object history construction and its dual, a form of category learning, necessitating construction and maintenance of a "single-entity" category, was presented in Fields and Glazebrook (2019b). Here we applied CCCDs to model the process of visual object categorization, and individual recognition (i.e. the categorization to a single-entity category) from the level of instantaneous spatially-mapped static features and motion vectors, through timestamped object files, to temporal sequences of categorized object tokens embedded in episodic memories (see Treisman, 2006; Flombaum, Scholl and Santos, 2008; Zimmer and Ecker, 2010; Fields, 2011, 2012, 2016, for relevant background). The basic intuition and methodology we briefly recall as follows. Each layer in this process corresponds to a network of mutually-constraining classifiers, with constraint information collected over progressively longer clock times as the level of abstraction increases. The model can be summarized as stating that each detectable feature (technically, the classification criteria that detect each detectable feature) is the core of some cone of classifiers, and every object token (technically, the classification criteria that define each object token) is the core of some cocone of classifiers. The central idea is that of category-theoretic duality: every component classifier at every level serves as both a token that is classified by progressively more abstract types moving up the hierarchy, and as a type that classifies tokens at levels below it. Consistent with the global duality of a CCCD, these "upward-going" roles can also be reversed, viewing every component classifier at every level as both a token that demonstrates the mutual consistency of the types above it, and as a type that is shown to be consistent with some, but never all, of the other types at its level by at least one token below it. A classifier that identifies the static feature [black], for example, demonstrates consistency along the color dimension it defines between all black objects; similarly the classifier that identifies instances of the abstract category [cat] demonstrates consistency along the dimensions defined by its constituent properties between all object tokens representing cats. We can reasonably expect that such a general scale-free architecture can be fruitfully applied to other sensory modalities, both functionally and neuro-architecturally (as for instance in Hochstein and Ahissar (2002)).

An object categorization or object recognition problem is solvable, in this CCCD representation, if and only if a (colimit) cocone over the relevant subset of classifiers exists; the existence of such a cocone assures, in turn, that this subset of classifiers is fully bidirectionally linked, and hence can be coherently activated (Fig. 1). Solutions are dense if almost all subsets of classifiers have cocones and hence correspond to identifiable categories or objects; they are sparse if relatively few subsets have cocones. Perceptual systems tend to be sparse; most collections of arbitrarily-selected features do not correspond to identifiable objects. Abstract systems, e.g. the binary representation of numbers, may be dense. Fields and Glazebrook (2019b) also show how to extend this model to CCCD representations of mereological relations, thus capturing the semantics of containment for structured, multi-component objects and events. These models make use of the association of both induced local logics (Barwise and Seligman, 1997; Seligman, 2009) and finite-element geometries (Cordier and Porter, 1989; Gratus and Porter, 2006) with networks of classifications (see Fields and Glazebrook, 2019a, for details). Overall, the CCCD design is meaningful, not only for encompassing distributed systems, but also for a wide range of types of two-way parallel processing; for instance, with $\longrightarrow$ representing "associative", and $\longleftarrow$ representing "analytic" relative to a suitable choice of $\Vdash$ (e.g. Baars and Franklin, 2003; Sloman, 1996).

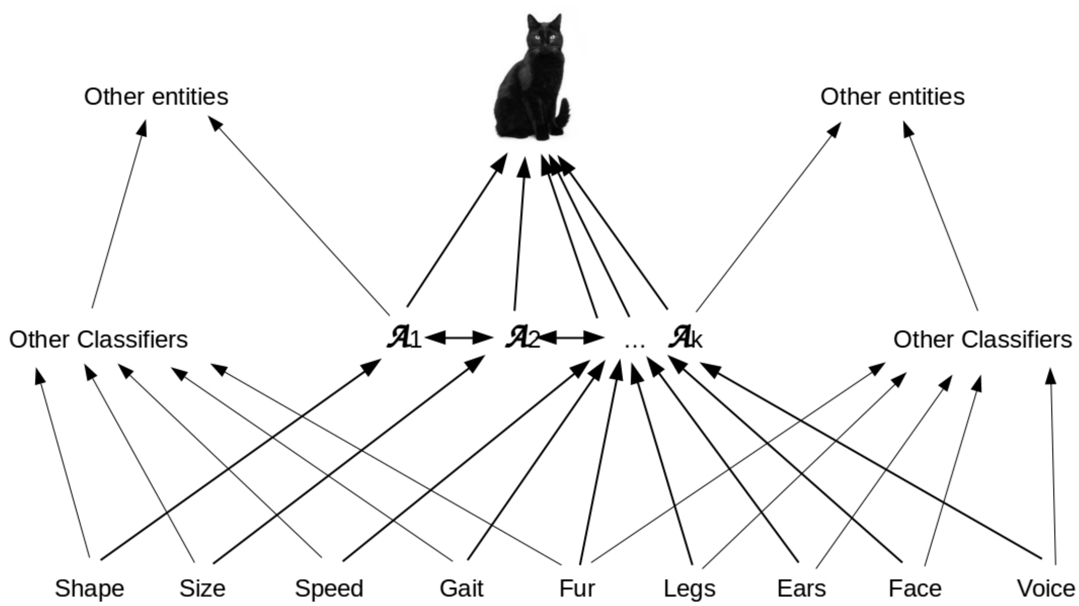

Fig. 1: An object recognition problem is solved when a cocone above the relevant subset of classifiers exists. Lower-level features (i.e. cores of cones) typically map to many classifiers, which in turn may map to distinct categories or individual objects (i.e. cores of cocones). Individual objects are treated as single-member categories as in Fields (2012).

The downward flow of constraints in a CCCD provides a natural model of attention or, in Bayesian terms, precision adjustment (Fields and Glazebrook, 2019b). Modeling anything beyond binary attentional control requires a non-trivial notion of the "weight" of an arrow in a CCCD, a notion most naturally modeled by a probability structure (i.e. by requiring all weights in or out of a node to sum to unity). Such structures can be represented as classifications, with discrete probability spaces corresponding to finite classifications (Allwein, 2004; Allwein, Moskowitz and Chang, 2004); again see Fields and Glazebrook (2019a) for details. We address the general issue of probability in a Bayesian framework in the next section (§3).

This classifier-based approach can be contrasted with the pattern-based approach of Ehresmann and Vanbremeersch (2007); Ehresmann and Gomez-Ramirez (2015), which makes similar use of the cocone construction to represent upward-going pattern abstraction and hence categorization. [†] If we identify patterns in such an abstracted sense with types as defined by Barwise and Seligman (1997), and take the colimit to represent a binding agent in a hierarchical structure, then a classifier in the current sense can be considered to represent the (logical, inferential) process of identifying a pattern. As all types serve as tokens one level lower in the inferential hierarchy, a classifier is in this case a pattern transformation, i.e. a mapping from one "cat-neuron" to another. With these identifications, each "object" in Fig. 1 (e.g. the "shape" feature) is implemented by a network of neural connections, and hence can be identified with the effective mapping between functionally-definable groups of conditionally-coactive neurons; the arrows in Fig. 1 are implemented by the dynamic process of transferring activation from one "layer" of functional connectivity to another.

2.5 From object recognition to Process-1 problem solving

It is commonplace to view problem solving as a kind of perception: one is faced with a problem and, perhaps after some deliberative (i.e. Process-2) thought, "sees" an answer. Intuitive (i.e. Process-1), along with insightful problem solving are generally described in perceptual terms (Kounios and Beeman, 2015). Chater (2018) takes this a step further, proposing that all problem solving is essentially perceptual.

This common view of problem solving is consistent with, and indeed entailed by, the "Bayesian Brain" hypothesis that the brain is a Bayesian optimizer, or at least a "Bayesian satisficer", the goal of which is to maximize the predictability of its environment (Knill and Pouget, 2004; Friston, 2010; Clark, 2013; Hohwy, 2013). With this hypothesis, the organizational structure and algorithms (in this case, Bayesian inference) of problem solving are domain-independent, although both the inputs (posterior probability distributions) and the knowledge (prior probability distributions) to

[†] A "pattern" in this latter approach models an activation pattern (a "cat-neuron" in the terminology of Ehresmann and Vanbremeersch, 2007) over a subset of conditionally jointly-activated neurons or, more technically, a class of functionally-equivalent (in the context of the overall categorization system) jointly-activated subsets of neurons. Such a "cat-neuron" is the colimit, in the construction of Ehresmann and Vanbremeersch (2007), of all such (classes of) synchronous assemblies of neurons that are activated by the same input, the colimit representing a kind of "binding agent" in a hierarchical structure.

which they are compared are domain-specific, and the processing is at least partially modularized. Substantial evidence now supports this view, though particular models, such as HPC, may differ in details (e.g. Brown et al., 2013; Friston, 2010; Maloney and Zhang, 2010; Parr, Rees and Friston, 2018; Spratling, 2016, 2017; Badcock et al., 2019, see Bar (2009); Grossberg (2013) for prediction-based models that are not Bayesian). We assume a broadly Bayesian, predictive-coding model of problem solving here, deferring details beyond those covered previously (Fields and Glazebrook, 2019b, 2020) to future work. We claim, in this case, that the bidirectional flow of constraints in problem solving can, in general, be represented by a CCCD. The conditions for problem solution are the same as for object recognition: a cocone must exist over the subset of classifiers coherently activated by the input. As described in §2.3 above, coherence is assured by the commutativity of the CCCD.

Viewing Bayesian inference generally as implemented by a CCCD subtly alters its standard interpretation. As noted earlier, a cocone by definition represents what is common to the classifiers beneath it: this is a consequence of the commutativity requirement. The classification at the "top" of a CCCD – e.g. the cat classifier in Fig. 1 – can, therefore, be regarded as encoding a *consensus* among the classifiers below it. Consistent with the usual competitive interpretation of Bayesian inference, this classifier produces a single "best" answer given its input – a cat, or not – but the satisfaction relation it implements is effectively a consensus satisfaction relation (Fields and Glazebrook, 2020). As its input is, in Bayesian terms, the posterior distribution at its level in the hierarchy, this interpretative point can equally well be made by considering the posterior distribution itself to be a consensus distribution resulting from the domain-specific contributions of the contributing lower-level classifiers. "Consensus" implies a loss of information about potentially-conflicting details, i.e. a coarse-graining. This loss of information is complemented by the gain of semantic information, e.g. about function or expected behavior, that is only encoded at higher levels of the categorization hierarchy (Fields and Glazebrook, 2019b).

This Bayesian picture of problem solving as generalized perception applies straightforwardly, however, only to Process-1 problem solving. There is, in particular, nothing about either the process of classification or the structure of a CCCD to suggest the discontinuities, backtracking, or searches for additional relevant information typical of Process-2 problem solving. Process-2 problem solving may, nonetheless, be subjectively Bayesian, though typically it is not (Kahneman, 2011). Bayesian or not, it typically involves *experienced* subjective probabilities, whether prior, posterior, or both. These are encoded by epistemic feelings, often described in terms of "degrees of belief" or "degrees of assent" or "willingness to bet" on an outcome (however see Eriksson and Hájek (2007) for a discussion of the problems that arise when one tries to *define* subjective probabilities in these terms). Process-2 problem solving typically also involves feelings of familiarity, recognition, agency, ownership, confidence, doubt, and frustration as well as experienced beliefs, goals, questions, intermediate steps, and emotions. As with more traditional emotions (Barrett and Simmons, 2015; Barrett, 2017; Seth, Suzuki and Critchley, 2012; Van de Cruys, 2017), there is every reason to consider such feelings as generated by a predictive-coding architecture. They are implemented, at least in GNW-type models (e.g. Baars and Franklin, 2003; Baars, Franklin and Ramsoy, 2013; Dehaene and Naccache, 2001; Dehaene, Charles and King, 2014; Shanahan, 2012; Wallace, 2005), by large-scale activation patterns spanning much of the cortex (particularly anterior insular, anterior cingulate, and medial prefrontal cortex) and extending into the midbrain (particularly basal and parathalamic "reward" nuclei and amygdala; see McCall and Franklin (2013) for an overview of programmable architecture for workspace predictive coding). As Shea and Frith (2019) emphasize, all such epistemic feelings are *about* cognitive states; hence GNW models that include them are intrinsically metacognitive. Critically, however, such models are not required by this to represent the "higher-order" metacognitions – "I think I believe X" – sometimes encountered in Process-2 cognition. Both problem solutions and their associated epistemic feelings are co-represented as "global broadcasts" within the GNW, as discussed in more detail below (§4.1). From an implementation perspective, the GNW comprises salience, default-mode (DMN), and executive control (ECN) networks (Hagmann et al., 2008; Menon and Uddin, 2010; Sporns and Honey, 2006; Uddin, 2015; van den Heuvel and Sporns, 2013) with the highest degree of mutual connection comprising the connectome "rich club" (Sporns, 2013) or the workspace connective core (Shanahan, 2012).

How such large-scale, episodic processes, with their relatively long durations and rich experiential content relate to the uninterrupted bidirectional flow of constraint information through modularized, effectively domain-specific channels postulated by either CCCD or hierarchical Bayesian network models is the main question posed in this paper. To begin to address it, we focus first on the epistemic feeling without which deliberative problem solving would be hard to define, the feeling of subjective probability.

3 Subjective probabilities, implicit and explicit

From the 'perceptron' model (Rosenblatt, 1961) to current deep learning systems (LeCun, Bengio and Hinton, 2015), neuromorphic computers (Schuman et al., 2017), and computational models of processing pathways in brains (e.g. Chadhuri et al., 2015; Bezaire et al., 2016), relatively simple processors linked by weighted connections have been used to model computation in networks of neurons or abstracted neuron-like elements. Models of neural circuits as Bayesian predictive-coding systems are no exception (Bastos et al., 2012; Kanai et al., 2015; Shipp, Adams and Friston, 2013; Spratling, 2016, 2017). It is natural to interpret the connection weights in such models as encoding probabilities of information flow from one model component to another; as discussed in §2.3, this interpretation is also natural for CCCDs as networks of infomorphisms (Allwein, 2004; Allwein, Moskowitz and Chang, 2004). The encoding of probabilities by connection weights is implicit or procedural (Johnson-Laird, 1977); as Sanborn and Chater (2016) point out, an explicit or declarative encoding of the probability distributions and calculations assumed by typical Bayesian models is computationally infeasible in a network architecture. Such implicitly-encoded probabilities are "subjective" in a very natural sense: they represent the computing system's representation of the strength of association between the sending element and the receiving element, and hence the strength of association between whatever these processing elements represent in the context of the computation. As the values of these connection weights are typically set by learning – including in those cases where measured values from biological neurons are used – they reflect the developmental and experiential history of the system they characterize, and are thus unique to it. Here it is relevant that, in the context of prediction error, Barrett (2017) proposes that all new learning is concept learning, since the brain condenses a redundancy of firing patterns into more efficient, more cost-effective multi-modal summaries, stored within limbic cortices to construct lower dimensional such summaries geared to abstraction.

Experienced subjective probabilities, from an inner-speech statement that the probability of a fair coin flip landing heads is 50% to a "gut feeling" that a proposed risk is too high, are however explicit representations corresponding, in GNW models, to (components of) GNW activation patterns. Experiences of the former kind are commonly called "conceptual", while the latter are "nonconceptual" (Arango-Muñoz, 2014), but both are declarative in the sense of being explicitly experienced. While experienced subjective probabilities *per se* have been somewhat neglected within the literature on epistemic feelings, experiences of confidence in a belief or an answer are broadly acknowledged as epistemic feelings (Arango-Muñoz, 2014; Proust, 2015; Schwarz, 2012) and a substantial literature addresses their implementation (e.g. Koriat and Levy-Radot, 1999; Bach and Dolan, 2012; Koriat, 2012; Meyniel, Sigman and Mainen, 2015; Paz et al., 2016; Navajas et al., 2017). Meyniel, Sigman and Mainen (2015) explicitly identify felt confidence with subjective probability, with the acknowledgement that feelings of confidence typically admit only qualitative, not quantitative, gradation. While experiments that demand a quantitative expression of subjective probability – e.g. betting experiments – receive one, this quantitative statement is the *output* of a problem solving process (even if it is only a memory-access process), and is accompanied by qualitative felt confidence. Except in the simplest cases, the human experience of subjective probability is rough and ready (Kahneman, 2011); the simulation mechanism proposed below §5) provides an explanation of why this should be so.

Why would cognition be organized this way, with subjective probabilities that are mostly implicit and not experienced, encoded by weights on network connections as expected on Bayesian Brain models, but sometimes explicit and experienced, either as felt qualitative confidence, or, as conceptualized and quantitated in a modality such as inner speech? While experienced subjective probabilities, in the form of feeling of confidence, accompany solutions reached by Process-1 problem solving, in Process-2 problem solving such feelings appear to play a guiding or directing role, with explicit computations using explicitly represented, quantitative subjective probabilities as called for by normative decision theory as an extreme example. The question of why subjective probabilities are experienced can, therefore, be considered a special case of the question of why conscious, deliberative cognition, i.e. Process-2 exists: if *only* Process-1 existed, they would seem unnecessary. Yet many animals – not just humans – appear to experience feelings of confidence; SAT experiments suggest such feelings extend to insects (Heitz, 2014). This question of why such feelings exist is hence at bottom a question about evolution, and its most compelling answer is flexibility (see, e.g. Dennett, 2017, for a recent attempt to provide a fully worked-out version of this answer)[‡]. A flexible system can devote more resources to searching either the environment or memory for information, i.e. exhibit SAT, when the expected consequences of an error are serious. What is the connection between devoting resources to search – particularly of memory, where waiting for a changing environment to reveal more information is not an issue – and experienced subjective probabilities? We suggest here that this search process is effectively a process of reverse engineering, with the aim of discovering the probabilities embedded in the Bayesian network. To make this suggestion precise, we first consider fast, Process-1 cognition in the context of a GNW-type architecture that includes interoception as well as perception, and show how such a model naturally supports epistemic feelings as synthetic interoceptions. We then ask how such feelings could play a guiding role in Process-2 cognition.

4 Epistemic feelings as synthetic interoceptions

4.1 A GNW model of Process-1

While GNW models are often presented and criticized as models of consciousness *per se* (see e.g. the recent exchange between Dehaene, Lau and Kouider, 2017, and Carter et al. (2018)), they are, strictly speaking, models of information access, and hence of "access consciousness" (Block, 1995). Even when epistemic feelings are recognized as metacognitive (Shea and Frith, 2019), it is access to

[‡]Dennett (2017) characterizes Process-1 as relatively inflexible "competence without comprehension" that is available, to varying degrees, to all organisms. Process-2 is, from an evolutionary point of view, more recent and possibly unique to human cognition (see also Evans (2010); Frankish (2010)). The paradigmatic descriptions in, e.g. Kahneman (2011) can reasonably be tied to cultural theories of 'duality', as in the anthropological setting of Nisbett (2003) who compares the traditional modes of Western analytic/rational-based thinking to those of the Asian intuitive/holistic thinking. The warnings of Henrich, Heine and Norenzayan (2010) are clearly of relevance here.

such feelings that is granted by the GNW. The process of "ignition" that grants conscious access to percepts, memories, emotions, epistemic feelings, or other contents has been suggested, from a PP/HPC point of view, to depend on the magnitudes of prediction errors crossing some context-dependent threshold (Hohwy, 2013; Barrett, 2017; Whyte, 2019). We do not adopt any particular account of this process here, but only assume that "ignition" occurs by some means or other.

Early GNW models largely focused on how information from either perception or memory is made available for the control of intentional behavior, including the verbal reporting of conscious contents (e.g. Baars and Franklin, 2003; Dehaene and Naccache, 2001). In these models, the GNW comprises a highly-connected prefrontal to parietal network (Dehaene and Naccache, 2001; Dehaene, Sergent and Changeux, 2003) that serves effectively as a shared working memory accessed by function-specific modules, admission into which ("ignition") requires both sufficient amplification from the source module and sufficient feedback from the GNW itself (for recent reviews, see Dehaene, Charles and King (2014); Mashour et al. (2020); for criticism, see Koch et al. (2016)). With increased understanding of the role of subcortical structures and interoception of bodily state in regulating salience, and the representation of the self, largely via insular and ventrolateral frontal cortex (e.g. Craig, 2009, 2010; Menon and Uddin, 2010; Uddin, 2015), and of the contribution of insula – cingulate – frontal connectivity to executive control (e.g. Dajani and Uddin, 2015; Dosenbach et al., 2008), the scope of GNW models has increased to include these effects (Baars, Franklin and Ramsoy, 2013; Dehaene, Charles and King, 2014). In this broadened conception, the GNW can be seen as making the current states of the "world model" derived from perception and memory, and the (internal) "body model" derived from interoception mutually accessible, and hence as enabling regulatory actions on the body to be coordinated with behavioral actions on the world (Fig. 2). As emphasized by Barrett (2017), such coupling of world and body models is essential to the maintenance of allostasis, and can be viewed as the primary "self modeling" function of the DMN.[§] This joint representation promotes the options available for active inference – altering

[§]Barrett (2017) also emphasizes that such allostasis-maintenance functions must be universal across animals. The DMN is best known as the primary locus of self-referential rumination (e.g. Buckner, Andrews-Hanna and Schacter, 2008; Qin and Northoff, 2011) and hence much of Process-2 thought, functions that are presumably human-specific. Evidence that individual differences in DMN connectivity correlate with genetic differences observable in pedigree studies (Glahn et al., 2010) suggests significant recent evolutionary change. Whether ancient and recent functions of the DMN can be teased apart at the architectural level remains to be seen.

beliefs (prior probabilities), altering sensory input (posterior probabilities), or altering the relative importance of the two (Bayesian precision) (Friston, 2010; Friston et al., 2015) – to organism-scale options of altering the self through regulation, the world through behavior, or what is most salient through attention shifts. Attention, itself, here corresponds to inferring precision via sensory signals; optimizing this precision accounts for attention within perceptual inference in which the registering, or non-registering, of simultaneous stimuli evokes a biased competition towards attaining to a Bayesian optimal outcome (Feldman and Friston, 2010). In the overall perspective, these kinds of integrative, organism-scale models have recently been explored in a hierarchical Bayesian framework by Badcock et al. (2019), although not in explicit GNW terms.

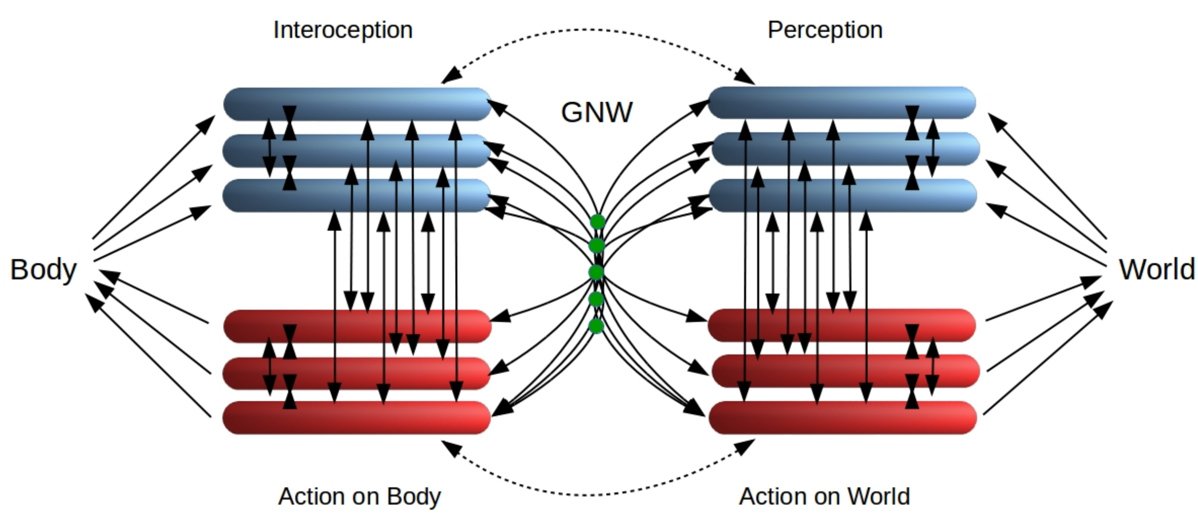

Fig. 2: Simplified cartoon of a GNW multi-parallel distributed architecture incorporating both perception of the external world and interoception of bodily state, and both action of the external world (i.e. overt behavior) and regulation of bodily state (e.g. regulation of hormonal signals and blood pressure). Bidirectional vertical arrows indi-

cate non-GNW modulatory connections between input and output processing modules on either the perception or interoception side; green dots represent GNW hubs. Upper and lower dashed arcs connecting perception and action systems are lateral (i.e. cross-modulatory) connections induced by requiring the GNW nodes to be cores of cocones spanning both world- and body-directed systems. These arcs transfer the information needed to generate epistemic feelings as synthetic interoceptions blending cortical with subcortical information.

Following the logic of §2.5, it is natural to interpret the central GNW nodes in Fig. 2 as access-providing processes instead of states or connections, and hence to interpret Fig. 2 itself as an extended CCCD with the central GNW nodes as cores of cocones. Doing so requires, to maintain commutativity across all classifiers in the CCCD, the existence of lateral, cross-modulatory connections between the world- and body-directed systems. This interpretation reconceptualizes the GNW from an arena for competition between the outputs of independent modules, to a mechanism for generating a consensus combined body-world model that maximizes predictive power.

The GNW model in Fig. 2, like all such models, explicitly assumes that the agent it models is conscious (up to salience) of the *outputs* of its perception and interoception systems, and of the *inputs* to its body-directed and world-directed action systems; in standard models these are identical. It is, again like all such models, consistent with the agent having inner, imaginative experience such as inner speech or imagery, i.e. with activation of perception or interoception systems by top-down inputs (Edelman, Gally and Baars, 2011; Kosslyn, Ganis and Thompson, 2001; Fields, 2002). It supports complex, perceptually guided behaviors such as visual navigation or language use, including reports from memory, and admits representations from different domain-specific systems to interact and subsequently be co-processed. That metacognitive parameters are representable in the workspace is evidenced from explicit reports of confidence, how the latter relate to cognitive load, and programmable, automatic error-detection (Shea and Frith, 2019).

Epistemic feelings are not only allowed by this model, but are required as consequences of the lateral connections between world- and body-directed systems (Barrett and Simmons, 2015; Barrett, 2017; Seth, Suzuki and Critchley, 2012, cf. Shea and Frith (2019)). Both rapid detection of (particularly severe) prediction errors by the frontolimbic threat detection system (Shechner and Bar-Haim, 2016) and fast autonomic feedback triggered by midbrain-directed cingular-cortex conflict signals (Sequeira et al., 2009; Critchley, Eccles and Garfinkel, 2013) can be expected to couple vascular, visceral, and other bodily sensations to cortical representations of prediction error or its rate of change. Hence we regard epistemic feelings as "synthetic interoceptions" that combine midbrain-mediated bodily interoception, e.g. via amygdala or reward system activation, with information from the cortex that reflects the flow and rate of perceptual and memory processing (Winkielman et al., 2003). Feelings of perceptual familiarity, for example, blend information from the perirhinal cortex that reports categorization fluency (Eichenbaum, Yonelinas and Ranganath, 2007) with introceptively-sourced information, with systematic failures leading to misidentification syndromes (Feinberg and Roane, 2005). Feelings of confidence and doubt blend information from insular and cingulate cortex that reports processing fluency, or in Bayesian terms, rate of reduction of prediction error (Brown et al., 2013; Joffily and Coricelli, 2013; Pliushch, 2015; Van de Cruys, 2017), and the presence or absence of cognitive conflict (Craig, 2009) with introceptively-sourced information, a process dramatically revealed during insular cortex seizures (Gschwind and Picard, 2016). Feelings of ownership of experience and "reality" blend information from both anterior prefrontal and insular cortex (Craig, 2010; Simons, Garrison and Johnson, 2017) with introceptively-sourced information, producing varying feelings of unease when source attribution is uncertain, and pathology, e.g. schizophrenia (Simons, Garrison and Johnson, 2017) or Cotard's syndrome (Debruyne et al., 2009), when this uncertainty is severe or chronic. Sampling models reproduce evidence from psychophysics experiments for short-term choice confidence (Koriat, 2012; Paz et al., 2016), consistent with the suggestions of Sanborn and Chater (2016) that "prior" probabilities are not fixed *a priori* and that Bayesian "inference" in the brain is implemented mainly by Bayesian sampling, and of Bar (2009) that low-resolution information drives the initial stages of processing, with higher-resolution information brought to bear as SAT requirements demand. Substantial individual differences in the types of processing-fluency information contributing to feelings of confidence have also been reported (Navajas et al., 2017). We can speculate that the generation of other epistemic feelings uses similar sampling methods, that such feelings are often low-resolution and hence qualitatively indistinct, and that there are substantial individual differences in what is sampled and experienced in given circumstances; all these remain to be tested.

Epistemic feelings generated as synthetic interoceptions are "metacognitive" in the sense employed by Fleming and Daw (2017); they report an assessment of the performance of a separate system, here the perception component of the GNW. The are not, however, yet metacognitive in the sense of mechanistically regulating Process-2 problem solving. The GNW model in Fig. 2 does not provide any specific mechanism for awareness of subjective probabilities as explicit, quantitative conceptual representations, e.g. their formulation as quantitative probability reports. Nor does it provide any mechanism for intentionally interrupting a problem-solving process in order to search for more information or "think about" the problem. [¶] It provides, in other words, a model of Process-1 problem solving, included arbitrarily-complex problem-solving abilities that expertise has rendered automatic (Bargh and Ferguson, 2000), but it does not provide any specific mechanism for Process-2 problem solving. The model supports the experiencing of epistemic feelings in the course of Process-1 problem solving, but does not provide a mechanism to support their role in Process 2. In this respect, it is "flat" in the sense of Chater (2018).

Process-2 problem solving is induced under a wide variety of conditions, including Process-1 unsolvability (e.g. multi-digit arithmetic), conflicting Process-1 solutions (e.g. conflicting perceptions of a Necker cube, or conflicting political intuitions), or motivational settings that bias SAT toward accuracy (e.g. competitive exams). What kind of metacognitive process could effect this process-switching behavior? How does this process couple with the "flat" GNW of Fig. 2? What role do epistemic feelings play in triggering Process-2 induction? Assuming that Process-1 has failed, which kinds of problems become solvable by inducing Process 2, and which kinds remain unsolvable?

[¶] In this respect, Arango-Muñoz (2014) suggests that some epistemic feelings are non-conceptual experiences. An individual need not have to realize specific concepts of certainty or uncertainty in order to have epistemic feelings, such as *feeling* certain or uncertain about something (e.g. on a dull day an individual may grab at an umbrella before leaving home while thinking about details of her conference paper). In a similar spirit, Proust (2015) proposes epistemic feelings to be both intentional and directional, and thus seen as types of mental affordances that induce a cognitive response, and then determine how to implement it.

4.2 Process-2 as iterated Process-1

The CCCD formalism introduced in §2.3, (2.6), (2.7), allows us to formulate this question precisely, and places a strong constraint on its answer. As noted in §2.5, a problem is solvable by Process-1 only if the cocone encoding its solution exists. Hence we can represent the first two Process-1 failure conditions above as in Fig. 3: in either case, problem solving failure occurs because the cocone structures required to link, and hence establish consensus between, the classifiers activated by the input do not exist. The missing structures include, critically, the lateral maps between classifiers that assure commutativity. Identifying the required cocone cores with GNW nodes as in Fig. 2, these lateral maps must not only connect classifiers within the perception and interoception systems individually, but also establish lateral links between these systems. Process-1 failures are, in these cases, partial solutions within the perception system that do not sufficiently activate the interoception system to generate the "confidence" signals required for cocone completion. Competition within Process-1 due to a lack of consensus as in Fig. 3b produces instability at the GNW level, e.g. conflicting opinions or the visual instability of a Necker cube. Even if Process-1 is successful, insufficient lateral activation of the interoception system leads to failure at the GNW level that is experienced as a lack of confidence. Motivational settings such as competitive exams or high-stakes bets, place the required confidence bar high, hence strong lateral activation is required to complete and hence accept a Process-1 solution.

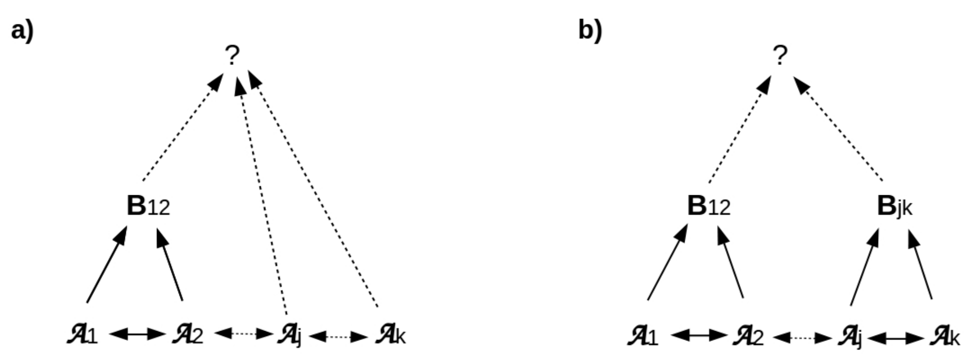

Fig. 3: Failure modes for Process-1 problem solving. a) Failure to complete a cocone (indicated by '?') over a set $\mathcal{A}_1 ... \mathcal{A}_k$ of activated classifiers. Dashed arrows indicate missing maps; note the absence of the lateral maps between classifiers that are required for commutativity. b) Failure to complete a cocone spanning two partial solutions.

If reaching a solution is completing a cocone that spans both perception and interoception systems, then the key to problem solving is activating the right collection of classifiers within each system. This is the task of the ECN (see e.g. Menon and Uddin (2010)). The tools available are behavioral or regulatory change and attention switching, or in Bayesian terms, modification of either the posterior distributions or the precision assignments. Modifying prior probability distributions – in a CCCD, the satisfaction relations of component classifiers – is a time-consuming learning process; however, the prior probability distributions that are deployed in response to an input can be varied by either behavior / regulation or attention switching. Critically, these ECN tools apply not just to world-directed behavior and the perception system, but also to body-directed regulation and the interoception system. The broad correlation between low stress, positive affect, and problem-solving performance (e.g. Arnsten, 2009; Kahneman, 2011; Kounios and Beeman, 2015) makes sense in this context. If prediction errors remain sufficiently small that ECN activity remains below the threshold of consciousness, in particular, a "flow" state in which performance is both fluid and enjoyable results Csikzentmihalyi (2014).

The effect of either behavior / regulation or attention switching is to re-initiate Process-1 with either a new input or a new collection of classifiers, starting a new "cognitive cycle" in the terminology of Chater (2018). This is effectively a feedback process, as shown in Fig. 4a: partial or intermediate results obtained in one round are available to be attended to and hence fed back as inputs in subsequent rounds, and may or may not be valuable. The information fed back includes the confidence values obtained on earlier rounds (Folke et al., 2016). Success yields a Process-1 solution, i.e. a completed cocone over the activated classifiers.

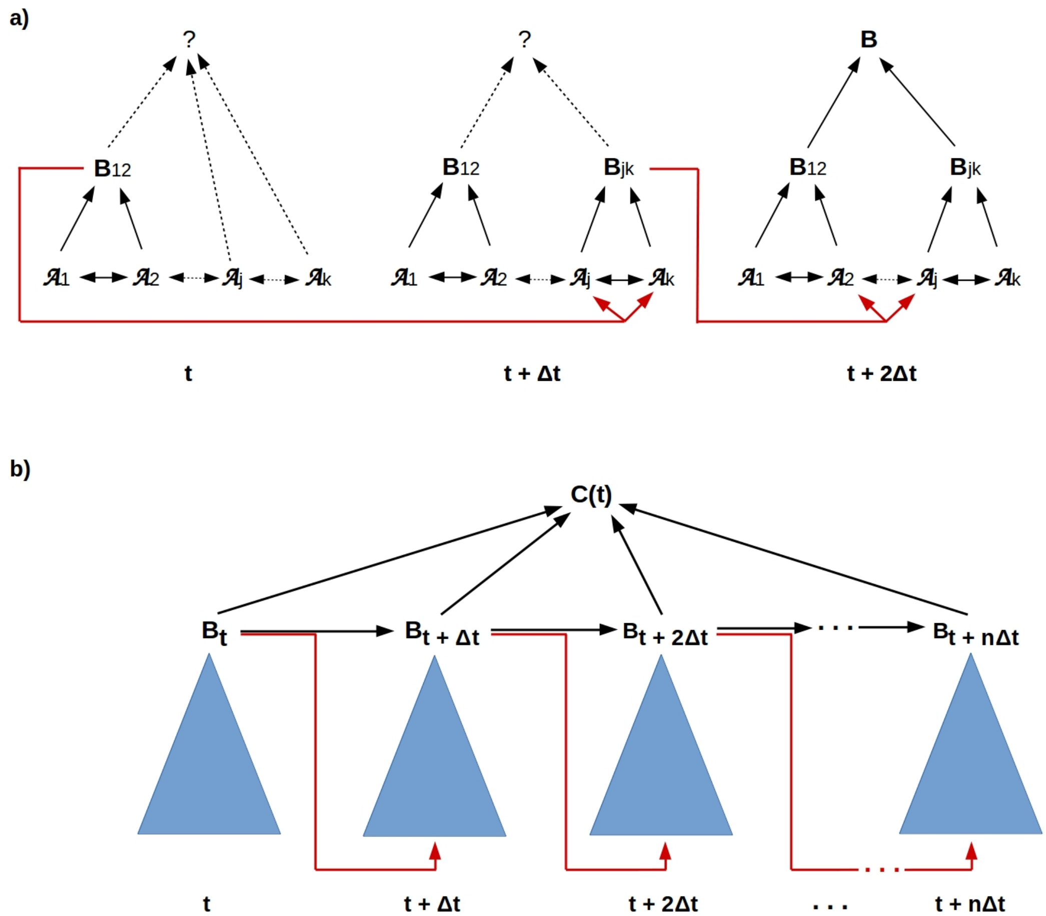

Fig. 4: Problem solving as an iterative process. a) Feeding partial solution information back (red arrows) into a Process-1 CCCD, either by behaving in the world (or regulating the body) and detecting the results by perception (or interoception), or by switching attention, until a solution is reached. Here the world (or body) serves as a short-term memory for partial results. In a Channel Theory sense, the depiction is effectively a sequence of distributed systems through time. b) Employing a time-indexed classifier as a short-term memory allows partial solutions to a complex problem to be represented as a sequence of "steps" toward an overall solution. Blue triangles represent the upper

cocones of CCCDs, which may or may not include the same classifier layers; red arrows represent feedback via either behavior/regulation or attention switching. Over a suitable time frame (as $\mathbf{t}$ increases), $\mathbf{C}(\mathbf{t})$ approaches a colimit; in other words, a "binding agent".

In task environments requiring speed, i.e. task environments in which task-relevant input features are changing within the 200 - 300+ ms required for a single Process-1 cycle (Sergent, Baillet and Dehaene, 2005; Schendan and Maher, 2009; Zmigrod and Hommel, 2010), the world (and / or body) serves as an effective 1-step short-term memory; hence feedback can, indeed must given the time constraints, be passed through the world / body. As the time available for problem solving increases to minutes, hours or even days, however, such external feedback is insufficient; even if the environment is modified by a process such as taking notes, additional internal memory resources must be committed to recording the fact that such actions were taken. This additional internal resource must, in addition, be able to sequence and link into a causal chain the Process-1 cycles that are deployed. These recording, sequencing, and linking functions are all provided if it is assumed that the internal memory resource has the form of a temporally-indexed cocone (Fig. 4b). The core of this cocone is a classifier, one that solves a multistep problem over extended time by iterating Process-1 cycles.

Process-2 problem solving in a CCCD-based model is, therefore, iterated Process-1 problem solving. It succeeds, just as Process-1 succeeds, when it constructs a complete cocone; the cocone core in this case represents the solution as a satisfaction relation. When faced with a problem that cannot be solved by Process-1 and given sufficient time, the goal of the ECN is to construct such a cocone by a process of active inference. Each step in the process is accompanied by synthetic interoceptions as experienced as feelings of doubt, confidence, being on the right or wrong track (see e.g. the subjective reports compiled by Shaw, 1999; Amabile et al., 2005). The interoceptive component of the solution state is experienced as knowledge, belief, assent, confidence, or other positive outcome markers, perhaps including the "Aha!" of long-delayed insight.

This model is consistent with Chater's view of insight as resulting from "trying it again", or "improvising as we go along" in a different attentional context (Chater, 2018). It is inconsistent with any model that assumes a specific "rational" alternative to Process-1 problem solving. Process-2 cognition is, on this model, not intrinsically better than Process-1 cognition, just slower. Its benefit derives from its ability to work on subproblems sequentially, feeding back partial results at each stage as inputs. This benefit comes at minimal evolutionary cost: only a single additional layer, with the same architecture and solution conditions, needs to be added to the underlying Process-1 system.

As noted by Evans (2008) and Frankish (2010), and emphasized by dual-process critics such as Melnikoff and Bargh (2018), sharp distinctions between the features typically characterizing Process-1 and Process-2 cognition and reasoning may not be fully justified. For instance, Evans (2008) suggests a possibility of Process-1 combined with a blend of Process-1 and Process-2, where the need for such a blend is tied to a limited capacity working memory. Barrett, Tugade and Engle (2004) have similarly considered that complex tasks may be sequentially implemented by domain-specific Process-1 operations, but under the control of a domain-general Process-2 mechanism for allocating attention. Such interleaving of Process-1 and Process-2 strategies, driven by SAT and working-memory availability, is understandable in the framework proposed here.

5 Reverse-engineering subjective probabilities

We are now in a position to address the question raised and deferred in §3: what enables people to explicitly estimate their subjective probabilities for events, particularly in cases in which a convenient, memory-retrievable formula for calculating them is not available? How, in particular, are people capable of performing Bayesian reasoning, even rough-and-ready Bayesian reasoning, outside of simple situations like card games in which probabilities can be remembered or recalculated explicitly? It is well known that people do not do this very well (Kahneman, 2011). At question is how they can do it at all. How does anyone know what their "degree of belief" in any given proposition is? If asked, what are they doing when trying to come up with an answer? Why is answering such a question typically a slow, deliberative process?

The model of Process-2 problem solving as iterated Process-1 problem solving outlined above suggests a natural answer to this question. Except in cases of rote learning, explicit subjective probabilities – degrees of belief – are not encoded anywhere, and hence are not available to be retrieved from memory. They must instead be inferred, with the feelings of confidence associated with Process-1 solutions as the only available evidence. Such feelings are, however, both implicit and qualitative. Making them explicit and precise – as reportable, numerical, subjective probabilities – requires, we suggest, a reverse engineering process in which multiple Process-1 simulations are used to generate a range of qualitative responses that can be summarized quantitatively. Such simulations can be implemented by repeating a Process-1 solution while injecting noise into its input; such noise injection is widely used to broaden search processes in simulated annealing (Kirkpatrick, Gelatt and Vecchi, 1983) and related techniques. While it is implausible that the executive system can generate true noise, high-frequency modulation of precision assignments during problem solving would provide a noise-like input. "Standards" of high- and low-confidence are also required; inputs such as "1 + 1 = 2" or "I exist" might provide the former, and "1 = 0" or "black is white" the latter. The mechanism for summarizing confidence estimates across Process-1 replicates needs only to be capable of averaging and computing distance of the average from the standards; a quantitative classifier is suitable for this task. We suggest that summary confidence estimates in the form of explicit subjective probabilities are, when the task demands it, provided as input to Process-2 problem solving as shown in Fig. 5: macroscopic time intervals Δt – at least 10s of seconds – are employed to summarize the confidence results from multiple, noise-injected Process-1 replicates of the original Process-1 solution. This is effectively a reverse engineering process in which the subject estimates, by running simulations, her degrees of belief in a partial solution.

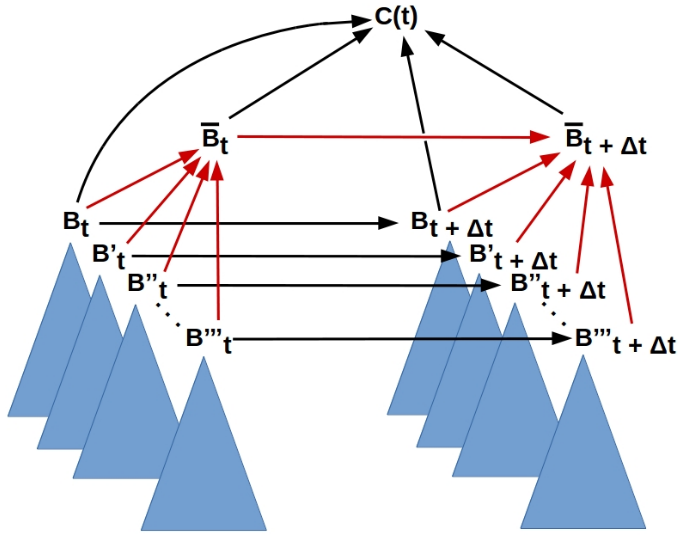

Fig. 5: Summarizing the confidence components of parallel Process-1 solutions to yield an explicit, reportable representation of confidence as a subjective probability. A cocone (vertical red arrows) assembles the parallel solutions for a macroscopic interval Δt into a single representation as a quantitative classifier. These are linked through time (horizontal red arrow) by the temporally-indexed cocone with the diachronic classifier $\mathbf{C(t)}$ as in Fig. 4b. All other maps are suppressed for clarity.

This mechanism requires no new architectural components to handle explicit subjective probabilities; it merely adds a "layer" of cocone-based processing to Fig. 4. It completely removes, moreover, the need for such probabilities to be stored in memory. If this mechanism operates in humans, it is highly unlikely that reportable subjective probabilities will remain constant across problem-solving contexts for all but the simplest, most easily-calculated examples.

6 Predictions

The model advanced here rests on two assumptions: 1) that the architecture implementing problem solving has the formal structure of a CCCD; and 2) that the GNW combines perception of and action on the world with interoception and regulation of the body. The former is reasonable mathematically (Fields and Glazebrook, 2019a,b), while the latter is well-supported experimentally (Baars, Franklin and Ramsoy, 2013; Dehaene, Sergent and Changeux, 2003; Dehaene, Charles and King, 2014; Mashour et al., 2020). It nonetheless generates a number of predictions, some of which make contact with existing experimental results.

- Process-2 problem solving that does not employ explicit subjective probabilities is mechanistically simpler than problem solving that does. Hence it is expected to be faster, besides evolving and developing at earlier stages. As the notion of a probability must be explicitly learned, it is not clear that this can be tested without circularity.

- Reportable subjective probabilities are neither stored in or retrieved from memory, but synthesized "on the fly". Hence they can be expected to change with implicit learning, general experience, and question context. Probability judgements are known to be sensitive to context and framing (Kahneman, 2011). To what degree they can be generated by the simulation-based mechanism outlined here is deferred to later work. On a deeper level, it is not clear, as Sanborn and Chater (2016) point out, that any of the prior probabilities required for Bayesian inference, even if this is viewed only as satisficing, are well-defined at the implementation level. To the extent that human behavior displays intrinsic contextuality (Dzhafarov and Kujala, 2017a; Dzhafarov and Kon, 2018), real-time switching between different sets of "prior probabilities" may be required. We have shown that the CCCD formalism allows such switching (Fields and Glazebrook, 2020), but how such a mechanism can be incorporated into a GNW framework remains to be determined.

- "Noising" by the ECN is required for generation and use of explicit, numerical probabilities. Such a mechanism would be expected to disrupt normal cognition if dysregulated, consistent with theoretical characterizations of autism as a disorder of Bayesian precision assignments (Lawson, Rees and Friston, 2014; Van de Cruys et al., 2014).

- Epistemic feelings are, in general, synthetic interoceptions as opposed to being purely metacognitive. Hence such feelings should be detectable and reportable in resting states in which not only are there no external task demands, but internal, imaginative mind-wandering and rumination are also turned off. The "witness" or "as-such" states sometimes achieved by experienced meditators (Josipovic, 2019) may be evidence for such activity. Affective feelings are intermediaries between implicit/automatic (antecedent Process-1) and explicit controlled (consequential Process-2) thinking, as underlying a putative duality between metacognitive feelings and metacognitive judgements (Koriat and Levy-Radot, 1999). Likewise, epistemic feelings are intermediaries in judgements of learning through applying non-analytic heuristics (Koriat and Levy-Radot, 1999)(cf. Arango-Muñoz (2014)).

- If explicit subjective probabilities are not stored in memory, they cannot be components of an explicit self model. This is consistent with the "self" being a real-time construct (Craig, 2010) and the "propositional attitude" model of belief (Fodor, 1983) being fundamentally incorrect, as argued by Chater (2018).

- The CCCD formalism deployed here is equivalent to the standard hierarchical Bayesian formalism in situations in which probabilities are non-contextual (in the sense of Dzhafarov and Kujala, 2017a; Dzhafarov and Kon, 2018) and otherwise well-behaved. However, it generalizes the hierarchical Bayesian formalism in domains in which probabilities are not well behaved, and in particular, provides a mechanism for maintaining inferential coherence in the face of contextuality (Fields and Glazebrook, 2020). While this is at present a theoretical result, not an empirical prediction, it is a significant issue for the applicability of the theory.

We are currently pursuing a formal analysis of the last of these questions, focussing in particular on the issue of contextuality.

7 Conclusion

Damásio (1994) re-instated emotion as a component of cognition 25 years ago. Since that time, the role of affect in problem solving and decision making has become increasingly well-understood. Dual-process models have, however, largely relegated this role to Process-1, while Process-2 has remained, at least in the popular imagination, the domain of the idealized rational agent of classical decision theory.

Here we have employed some general category-theoretic tools, particularly the CCCD construction of distributed information flow, together with current GNW models, to argue that Process-2 problem solving is multiply-iterated Process-1 problem solving driven, to a significant extent, by synthetic interoceptions experienced as qualitative confidence. Process-2 is invoked, in this model, when the rate of change, or delay between changes, in task-relevant input, including internally-generated imaginations, is too slow for either the world, or for body states to serve as effective short-term memories. Quantitative confidence in the form of the degrees of belief of classical decision theory can be reproduced in this model by adding a "layer" to a Process-2 CCCD that summarizes the results of multiple Process-1 simulations with slightly different inputs or precision assignments. The CCCD construction provides a built-in mechanism – diagram commutativity – that enforces inferential coherence. Adapting and extending this mechanism in order to deal with task environments exhibiting true contextuality (Dzhafarov and Kon, 2018) is an interesting question relating to the risk of inconsistency in decision-making, which we defer to future work.

Finally, we remark that we have inadvertently, and somewhat surprisingly, connected with two particular recent advances. Firstly, in understanding the *modus operandi* of the GNW with respect to working memory and metacognition (Shea and Frith, 2019). Secondly, the transformations of Process-1 to a metacognitive Process-2 can be viewed as a "representational exchange mechanism", a multi-parallel processing of information based upon the peculiarities and the frequent irrationality of human decision-making (Cushman, 2020): a "useful fiction" since perfunctory rationalization imputes a sense of reason upon a possible conundrum when deciding ("fiction"), and "useful" because the process supports improvising, and improving upon future reasoning (as exemplified in

Cushman (2020); cf. Chater (2018)).

Acknowledgements

We wish to thank two anonymous referees for their constructive criticism and suggestions which helped to improve the overall presentation of this paper. We also thank the Handling Editor for further assessment and useful comments.

Compliance with Ethical Standards

The authors declare that they have no conflicts of interest. The work reported here received no external funding. No experiments involving either animals or humans were performed by either of the authors as part of this work.

References

Abramsky, S. (2012). Big toy models: Representing physical systems as Chu spaces. *Synthese* 186, 697–718.

Adámek, J., Herrlich, H. and Strecker, G. E. (2004). *Abstract and Concrete Categories: The Joy of Cats.* New York: Wiley. Available at http://katmat.math.uni-bremen.de/acc (Accessed May 26, 2019)

Adams, E. W. (1998). *A Primer of Probabilistic Logic.* Chicago: University of Chicago Press.

Alain, C., Arnott, S. R., Hevenor, S., Graham, S. and Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proc. Natl. Acad. Sci. USA* 98, 12301–12306.

Allwein, G. (2004). A qualititative framework for Shannon Information theories. In 'NSPW '04 Proceedings of 2004 Workshop on New Security Paradigms' pp. 23-31, Nova Scotia, Canada, September 20-23, 2004. ACM New York NY, 2004.

Allwein, G., Yang, Y. and Harrison, W. L. (2011). Qualitative decision theory via Channel Theory. *Logic and Logical Philosophy* 20, 1–30.

Allwein, G., Moskowitz, I. S. and Chang, L.-W. (2004). A new framework for Shannon information theory. Technical Report A801024 Naval Research Laboratory, 17pp. Washington DC.

Amabile, T. M., Barsade, S. G., Mueller, J. S., and Staw, B. M. (2005). Affect and creativity at work. *Admin. Sci. Quart.* 50, 367–403.

Arango-Muñoz, S. (2014). The nature of epistemic feelings. *Philos. Psych.* 27(2), 193–211.

Arnsten, A. F. T. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nat. Rev. Neurosci.* 10, 410–422.

Awodey, S. (2010). *Category Theory.* Oxford Logic Guides, 62. Oxford, UK: Oxford University Press.

Baars, B. J. and Franklin, S. (2003). How conscious experience and working memory interact. *Trends Cogn. Sci.* 7, 166–172.

Baars, B. J., Franklin, S. and Ramsoy, T. Z. (2013). Global workspace dynamics: Cortical "binding and propagation" enables conscious contents. *Front. Psychol.* 4, 200.

Bach, D. R. and Dolan, R. J. (2012). Knowing how much you dont know: A neural organization of uncertainty estimates. *Nat. Rev. Neurosci.* 13, 572–586.

Badcock, P. B., Friston, K. J., Ramstead, M. J. D., Ploeger, A. and Hohwy, J. (2019). The hierarchically mechanistic mind: An evolutionary systems theory of the human brain, cognition, and behavior. *Cogn. Affect. Neurosci.* in press, https://doi.org/10.3758/s13415-019-00721-3

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364(1521), 1235–1243.

Bargh, J. A. and Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychol. Bull.* 126, 925–945.

Barr, M. (1979). *-Autonomous categories, with an appendix by Po Hsiang Chu. *Lecture Notes in Mathematics* 752. Berlin: Springer.

Barr, M. (1991). *-Autonomous categories and linear logic. *Mathematical Structures in Computer Science* 1, 159–178.

Barrett L. F., Tugade, M. M. and Engle, R. W. (2004). Individual differences in working memeory capacity and dual-process theories of the mind. *Psychol. Bull.* 130(4), 553–573.

Barrett, L. F. and Simmons, W. K. (2015) Interoceptive predictions in the brain. Nat. Rev. Neurosci. 16(7): 419–429.

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience* 12(1), 1–23. doi:10.1093/scan/nsw154

Barwise, J. and Seligman, J. (1997). *Information Flow: The Logic of Distributed Systems.* Cambridge Tracts in Theoretical Computer Science, 44. Cambridge, UK: Cambridge University Press.

Barwise, J. (1997). Information and Impossibilities. *Notre Dame Journal of Formal Logic* 38(4), 488–515.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P. and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711.

Bezaire, M. J., Raikov, I., Burk, K., Vyas, D. and Soltesz, I. (2016). Interneuronal mechanisms of hippocampal theta oscillations in a full-scale model of the rodent CA1 circuit. *eLife* 5, e18566.

Blake, R. and Shiffrar, M. (2007). Perception of human motion. *Ann. Rev. Psychol.* 58, 47–73.

Block, N. (1995). On a confusion about a function of consciousness. *Behav. Brain Sci.* 18, 227–247.

Brown, R. and Porter, T. (2006). Category theory: an abstract setting for analogy and comparison. In: *What is Category Theory?* pp. 257–274. Advanced Studies in Mathematics and Logic. Polimetrica Publisher, Italy.

Brown, H., Adams, R. A., Parees, I., Edwards, M. and Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing* 14(4), 411–427.

Buckner, R., Andrews-Hanna, J. and Schacter, D. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Ann. New York Acad. Sci.* 1124, 1–38.

Carter, O., Hohwy, J., van Boxtel, J., Lamme, V., Block, N., Koch, C. and Tsuchiya, N. (2018). Conscious Machines: Defining questions. *Science* 2018, 400.

Chadhuri, R., Knoblauch, K., Gariel, M.-A., Kennedy, H. and Wang, X.-J. (2015). A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron* 88, 419-431.

Chater, N. (2018). *The Mind is Flat. The remarkable shallowness of the improvising brain.* Yale Univ. Press, New Haven and London.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204.

Cordier, J.-M. and Porter, T. (1989) *Shape theory: Categorical Methods of Approximation.* Ellis Horwood, Hemel Hempstead UK.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory.* Wiley, New York.

Craig, A. D. (2009). How do you feel – now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70.

Craig, A. D. (2010). The sentient self. *Brain Struct. Funct.* 214, 563–577.

Critchley, H. D., Eccles, J. and Garfinkel, S. N. (2013). Interaction between cognition, emotion, and the autonomic nervous system. In: R.M. Buijs and D.F. Swaab (Eds.) *Handbook of Clinical Neurology, Vol. 117 (3rd series) Autonomic Nervous System.* Amsterdam: Elsevier (pp. 59–77).

Csikzentmihalyi, M. (2014). *Flow and the Foundations of Positive Psychology.* Dordrecht: Springer.

Cushman, F. (2020). Rationalization is rational. *Behav. Brain Sci.* 43, e28. https://doi.org/10.1017/S0140525X19001730

Dajani, D. R. and Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends Neurosci.* 38, 571–578.

Damásio, A. (1994). *Descartes' Error: Emotion, Reason and the Human Brain.* Putnam, New York.

de Finetti, B. (1974). *Theory of Probability*, Vol. I. New York: Wiley.

Debruyne, H., Portzky, M., Van den Eynde, F. and Audenaert, K. (2000). Cotard's syndrome: A review. *Curr. Psychiatry Rep.* 11, 197–202.

Dehaene, S. and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition, 79*, 1–37.

Dehaene, S. Sergent, C. and Changeux, J.-P. (2003). A neuronal network modeling subjective reports and objective physiological data during conscious perception. *Proc. Natl. Acad. Sci. USA* 100(14), 8520–8525.

Dehaene, S., Charles, L., King, J.-R. and Marti, S. (2014). Toward a computational theory of conscious processing. *Curr. Opin. Neurobiol.* 25, 76–84.

Dehaene, S., Lau, H. and Kouider, S. (2017). What is consciousness, and could machines have it? *Science* 358, 486–492.

Dennett, D. C. (2017). *From Bacteria to Bach and Back: The Evolution of Minds.* London: Penguin.

Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L. and Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends Cogn. Sci.* 12, 99–105.

Dretske, F. (1981). *Knowledge and the Flow of Information.* Cambridge MA: MIT Press.

Dzharfarov, E. N. and Kon, M. (2018). On universality of classical probability with contextually labeled random varaibles. *J. of Math. Psych.* 85, 17-24.

Dzhafarov, E. N. and Kujala, J. V. (2017a). Contextuality-by-Default 2.0: Systems with binary random variables. In: J. A. Barros, B. Coecke and E. Pothos (eds.) *Lecture Notes in Computer Science* 10106, Springer, Berlin, 16–32.

Edelman, G. M., Gally, J. A., and Baars, B. J. (2011). Biology of consciousness, *Frontiers in Psychology* 73(1), 43–52. doi: 10.3389/fpsyg.2011.00004

Ehresmann, A. C. and Vanbremeersch, J.-P. (2007). *Memory Evolutive Systems; Hierarchy, Emergence, Cognition (Studies in Multidisciplinarity).* New York: Elsevier.

Ehresmann, A. C. and Gomez-Ramirez, J. (2015). Conciliating neuroscience and phenomenology via category theory. *Prog. Biophys. Mol. Biol.* 119, 347–359.

Eichenbaum, H., Yonelinas, A. R. and Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Ann. Rev. Neurosci.* 30, 123–152.

Eriksson, L. and Hájek, A. (2007). What are degrees of belief? *Studia Logica* 86, 183–213.

Evans, J. St. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin and Review* 13(3), 378–395.

Evans, J. St. B. T. (2008). Dual processing accounts of reasoning, judgement and social cognition. *Annu. Rev. Psychol.* 59, 255–278.

Evans, J. St. B. T. (2010). Intuition and reasoning: A dual process perspective. *Psychological Inquiry* 21(4) (Special Issue on Intuition (October-December 2010)), 313–326.

Evans, J. St. B. T. and Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspect. Psychol. Sci.* 8, 223–241.

Feinberg, T. E. and Roane, D. M. (2005). Delusional misidentification. *Psychiatr. Clin. N. Am.* 28, 665–683.

Feldman, H. and Friston, K. (2010). Attention, uncertainty, and free energy. *Front. Hum. Neurosci.* 4: 215 doi:10.3389/fnhum.2010.00215

Fields, C. (2002). Why do we talk to ourselves? *J. Expt. Theor. Artif. Intell.* 14, 255–272.

Fields, C. (2011). Trajectory recognition as the basis for object individuation: A functional model of object file instantiation and object token encoding. *Front. Psychol.: Percept. Sci.* 2, 49.

Fields, C. (2012). The very same thing: extending the object token concept to incorporate causal constraints on individual identity. *Adv. Cogn. Psychol.* 8, 234–247.

Fields, C. (2013). Metaphorical motion in mathematical reasoning: further evidence for pre-motor implementation of structure mapping in abstract domains. *Cogn. Process.* 14, 217–229.

Fields, C. (2016). Visual re-identification of individual objects: A core problem for organisms and AI. *Cogn. Process.* 17, 1–13.

Fields, C. (2017). Object Permanence. In: Shackelford, T. K. and Weekes-Shackelford, V. A. (Eds.), *Encyclopedia of Evolutionary Psychological Science.* New York: Springer, (Chap. 2373).

Fields, C. and Glazebrook, J. F. (2019). A mosaic of Chu spaces and Channel Theory I: Category-theoretic concepts and tools. *J. Expt. Theor. Artif. Intell.* 31(2), 177-213. doi:10.1080/0952813X.2018.1544282

Fields, C. and Glazebrook, J. F. (2019). A mosaic of Chu spaces and Channel Theory II: Applications to object identification and mereological complexity. *J. Exper. Theor. Artif. Intell.* 31(2), 237–265. doi: 10.1080/0952813X.2018.1544285

Fields, C. and Glazebrook, J. F. (2020). Representing measurement as a thermodynamic symmetry breaking. *Symmetry* 12, 810. doi:10.3390/sym12050810

Firestone, C. and Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behav. Brain Sci.* 39, e229.

Fleming, S. M. and Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psych. Rev.* 124, 91–114.

Flombaum, J. I., Scholl, B. J. and Santos, L. R. (2008). Spatiotemporal priority as a fundamental principle of object persistence. In Hood, B. and Santos, L. (Eds.), *The Origins of Object Knowledge.* New York: Oxford University Press (pp. 135–164).

Fodor, J. A. (1983). *The Modularity of Mind.* Cambridge, MA: MIT/Bradford.

Folke, T., Jacobsen, C., Fleming, S. M. and De Martino, B. (2016). Explicit representation of confidence informs future value-based decisions. *Nat. Hum. Behav.* 1, 0002.

Frankish, K. (2010). Dual-process and dual system theories of reasoning. *Philosophy Compass* 5(10), 914–926. doi: 10.1111/j.1747-9991.2010.00330.x

Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138.

Friston, K. J. and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Phil. Trans. R. Soc. London B* 364, 1211–1221.

Friston, K. J., Kilner, J. and Harrison, L. (2006). A free energy principle for the brain. *J. Physiology – Paris* 100, 70–87.

Friston, K. J., Rigoli, F., Ognibene, D., Mathys, C., FitzGerald, T. and Pezzulo, G. (2015). Active inference and epistemic value. *Cogn. Neurosci.* 6, 187–214.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7(2), 155–170, 1983.

Glahn D.C., Winkler, A.M., Kochunov, P., Almasy, L., Duggirala, R., Carless, M.A., Curran, J.C., Olvera, R.L., Laird, A.R., Smith S.M., Beckmann, C.F., Fox, P.T. and Blangero, J. (2010). Genetic control over the resting brain. *Proc. Natl. Acad. Sci. U S A* 10(7), 1223–1228.

Gratus, J. and Porter, T. (2006) A spatial view of information. *Theoretical Computer Science* 365, 206–215.

Grossberg, S. (2013). Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks* 37, 1–47.

Gschwind, M. and Picard, F. (2016). Ecstatic epileptic seizures: A glimpse into the multiple roles of the insula. *Front. Behav. Neurosci.* 10, 21.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J. and Sporns, O. (2008). Mapping the structural core of the human cerebral cortex. PLoS Biology 6(7), e159.

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Front. Neurosci.* 8, 150.

Henrich, J., Heine, S. J. and Norenzayan, A. (2010). The weirdest people in the world. *Behav. Brain Sci.* 33, 61–83.

Hochstein, S. and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36, 791–804.

Hohwy, J. (2013). The Predictive Mind. Oxford University Press, Oxford, UK.

Johnson-Laird, P. N. (1977). Procedural semantics. *Cognition* 5, 189–214.

Josipovic, J. (2019). Nondual awareness: Consciousness-as-such as non-representational reflexivity. *Prog. Brain Res.* 244, 273–298.

Joffily, M. and Coricelli, G. (2013). Emotional valence and the Free-energy principle. *PLoS Comput. Biol.* 9(6), e1003094.

Kahneman, D. and Frederick, S. (2002). Representativeness revisited: attribute substitution in intuitive judgement. In *Heuristics and Biases: The Psychology of Inituitive Judgement* (Gilovich, T., Griffin, D. and Kahneman, D. eds.), pp. 49-81. Cambridge Univ. Press, Cambridge, UK.

Kahneman, D. (2011). *Thinking, Fast and Slow.* Penguin, London.

Kahneman, D., Triesman, A. and Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cogn. Psychol.* 24, 175–219.

Kalfoglou, Y. and Schorlemmer, M. (2003). IF-Map: An ontology-mapping method based on information-flow theory. *J. Data Semantics I* (Lecture Notes in Computer Science). Berlin: Springer (pp. 107–127).

Kanai, R., Komura, Y., Shipp, S. and Friston, K. (2015). Cerebral hierarchies: Predictive processing, precision and the pulvinar. *Phil. Trans. R. Soc. B* 370, 20140169.

Keifer, M. and Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex* 7, 805–825.

Kikuchi, M., Nagasaka, I., Toyoda, S. and Kitamura, S. (2003). A mathematical model of interactions in artifact environment. *Proceedings of SICE Annual Conference 2003* (pp. 2085–2090). Fukui, Japan: Fukui University.

Kirkpatrick, S., Gelatt Jr, C. D. and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science* 220, 67–680.

Knill, D. C. and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neuroscience* 27, 712–719.

Koch, C., Massimini, M., Boly, M and Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. *Nature Reviews Neuroscience* 17, 307–320.

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psych. Rev.* 119(1), 80-113.

Koriat, A. and Levy-Radot, R. (1999) Processes underlying metacognitive judgements: Information-based and experience-based monitoring of one's own knowledge. In (Chaiken, S. and Trope, Y. eds.) *Dual-process theories in social psychology* (pp. 483–502). Guilford Press, London-New York.

Kosslyn, S. M., Ganis, G. and Thompson, W. L. (2001). Neural foundations of imagery. *Nat. Rev. Neurosci.* 2, 635–642.

Kounios, J. and Beeman, M. (2015). *The Eureka Factor: Creative Insights and the Brain.* London: Windmill.

Lawson, R. P., Rees, G. and Friston, K. J. (2014). An aberrant precision account of autism. *Front. Human Neurosci.* 8, 302.

LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444.

Mahon, B. Z. and Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Ann. Rev. Psychol.* 60, 27–51.

Maloney, L. T. and Zhang, H. (2010). Decision-theoretic models of visual perception and action. *Vision Res.* 50, 2362–2374.

Mashour, G. A., Roelfsema, P., Changeux, J.-P. and Dehaene, S. (2020). Conscious processing and the Global Neuronal Workspace hypothesis. *Neuron* 105, 776–798.

McCall, R., and Franklin, S. (2013). Cortical learning algorithms with predictive coding for a systems-level cognitive architecture. In *Second Annual Conference on Advances in Cognitive Systems* Poster Collection: 149–166.

McClelland, J. L. (1998). Connectionist models and Bayesian inference: In M. Oakford and N. Chater (Eds.) *Rational Models of Cognition.* Oxford: Oxford University Press (pp. 21–53).

Melnikoff, D. E. and Bargh, J. A. (2018). The mythical number two. *Trends Cogn. Sci.* 22, 280–293.

Menon, V. and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Struct Funct.* 214(5-6), 655–667.

Meyniel, F., Sigman, M. and Mainen, Z. F. (2015). Confidence as Bayesian probability: From neural origins to behavior. *Neuron* 88, 78-92.

Moors, A. and De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychol. Bull.* 132, 297–326.

Navajas, J., Hindocha, C., Foda, H., Keramati, M., Latham, P. E. and Bahrami, B. (2017). The idiosyncratic nature of confidence. *Nat. Hum. Behav.* 1, 810–818.

Nhuy, N. and Van Quang, N. (2001). Chu spaces and conditional probabilities. *Acta Math. Vietnamica* 26(2), 241–248.

Nisbett, R. (2003). *The Geography of Thought: How Asians and Westerners think differently ... and why*, Free Press, New York.

Old, L. J. and Priss, U. (2001). Metaphor and information flow. In 'Proccedings of the 12th Midwest Artificial Intelligence and Cognitive Science Conference', pp. 99–104.

Papadopoulos, B. K. and Syropoulos, A. (2000). Fuzzy sets and fuzzy relational structures as Chu spaces. *Int. Journal of Uncertainty, Fuzziness and Knowledge Based Systems* 8(4), 471–479.

Parmigiani, G. and Inoue, L. Y. T. (2009). *Decision Theory: Principles and Approaches*. Chichester, UK: Wiley.

Parr, T., Rees, G. and Fiston, K. J. (2018). Computational neuropsychology and Bayesian inference. *Frontiers in Human Neuroscience* 12 Article 61. 14pp.

Paz, L., Insabato, A., Zylberberg, A., Deco, G. and Sigman, M. (2016). Confidence through consensus: A neural mechanism for uncertainty monitoring. *Sci. Rep.* 6, 21830.

Pliushch, I. (2015). The extension of the indicator-function of feelings (A commentary on Joëlle Proust). In Metzinger, T. and Windt, J. M. (Eds.) *Open MIND* 31(C). Frankfurt am Main: MIND Group, 12pp. doi:10.15502/9783958570399

Pratt, V. (1995). Chu spaces and their interpretation as concurrent objects. *Lect. Notes in Comp. Sci.* 1000, 392–405.

Pratt, V. (1997). Types as Processes, via Chu spaces. Invited paper. Proceedings 'Express'97: Fourth Workshop on Expressiveness in Concurrency'. Santa Margherita, Italy. September 1997. *Electronic Notes in Theoretical Computer Science* 7 21pp. http://www.elsevier.nl/locate/entcs/volume7.html

Pratt, V. (1999). Chu spaces. *School on Category Theory and Applications (Coimbra 1999)*, Vol. 21 of *Textos Mat. Sér. B*, University of Coimbra, Coimbra. (pp. 39–100).

Pratt, V. (1999). Chu spaces from the representational viewpoint. *Ann. Pure Appl. Logic* 96, 319–333.

Proust, J. (2015). The representational structure of feelings. In Metzinger, T. and Windt, J. M. (Eds.) *Open MIND* 31(T). Frankfurt am Main: MIND Group, 25pp. doi:10.15502/9783958570047

Pylyshyn, Z. (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.

Qin, P. and Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *NeuroImage* 57, 1221–1233.

Roederer, J. (2010). *Information and its Role in Nature.* Heidelberg: Springer.

Rosenblatt, F. (1961). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Spartan Books, Washington DC.

Sanborn, A. N. and Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Science* 20(12), 883–893.

Sathian, K., Lacey, S., Stilla, R., Gibson, G. O., Deshpanda, G., Hu, X., LaConte, S. and Glielmi, C. (2011). Dual pathways for haptic and visual perception of spatial and texture information. *Neuroimage* 57, 462–475.

Schendan, H. E. and Maher, S. M. (2009). Object knowledge during entry-level categorization is activated and modified by implicit memory after 200 ms. *NeuroImage* 44, 1423–1438.

Schuman, C. D., Potok, T. E., Patton, R. M., Birdwell, D., Dean, M. E., Rose, G. S. and Plank, J. S. (2017). A survey of neuromorphic computing and neural networks in hardware. Preprint arXiv:1705.06963v1 [cs.NE].

Schwarz, N. (2012). Feelings-as-information theory. In (Van Lange, P. A. M., Kruglanski, A. W., and Higgins, E. T., eds.) *Handbook of Theories of Social Psychology* (pp. 289–308). Sage Publications Ltd., Thousand Oaks CA.

Seligman, J. (2009). Channels: From logic to probability. In: Sommaruga, G. (Ed.) *Formal Theories of Information.* (Lecture Notes in Computer Science, 5363) Berlin: Springer (pp. 193–233).

Sequeira, H., Hot, P., Silvert, L. and Delaplanque, S. (2009). Electrical autonomic correlates of emotion. *Int. J. Psychophysiology* 71, 50–56.

Sergent, C., Baillet, S. and Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nat. Neurosci.* 8, 1391–1400.

Seth, A. K., Suzuki, K, and Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology* 2 Article 395, 16 pp. doi: 10.3389/fpsyg.2011.00395

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Science* 17 (11): 565–573.

Shanahan. M. (2012). The brain's connective core and its role in animal cognition, *Phil. Trans. Royal Soc. B* 367, 2704–2714.

Shaw, M. P. (1999). On the role of affect in scientific discovery. In S. W. Russ (Ed.) *Affect, Creative Experience and Psychological Adjustment.* London: Taylor & Francis (pp. 147–162).

Shea, N. and Frith, C. D. (2019). The Global Workspace needs metacognition. *Trends in Cognitive Science* 23(7), 560–571.

Shechner, T. and Bar-Haim, Y. (2016). Threat monitoring and attention-bias modification in anxiety and stress-related disorders. *Current Directions in Psychological Science* 25, 431–437.

Shipp, S., Adams, R. A. and Friston, K. J. (2013). Reflections on agranular architecture: Predictive coding in the motor cortex. *Trends Neurosci.* 36, 706–716.

Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psych. Rev.* 74, 29–39.

Simon, H. A. (1972). Theories of bounded rationality. In: McGuire, C. B. and Radner, R. (Eds.) *Decision and Organization.* Amsterdam: North Holland.

Simons, J. S., Garrison, J. R. and Johnson, M. K. (2017). Brain mechanisms of reality monitoring. *Trends Cogn. Sci.* 21, 462–473.

Sloman, R. (1996). The empirical case for two systems of reasoning. *Psych. Bull.* 119(1), 3–22.

23, 162–171.

Sporns, O. and Honey, C. J. (2006). Small worlds inside big brains. *Proc. Natl. Acad. Sci. USA, 111*, 15220–15225.

Spratling, M. W. (2016). Predictive coding as a model of cognition. *Cogn. Proc.* 17, 279–305.

Spratling, M. W. (2017). A hierarchical predictive coding model of object recognition in natural images. *Cogn. Comput.* 9, 151-167.

Treisman, A. (2006). Object tokens, binding and visual memory. In Zimmer, H. D., Mecklinger, A. and Lindenberger, U. (Eds.) *Handbook of Binding and Memory: Perspectives from Cognitive Neuroscience*. Oxford, UK: Oxford University Press (pp. 315–338).

Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 4, 207–232.

Tversky, A. and Kahneman, D. (1974). Judgement under uncertainty: heuristics and biases. *Science* 185 (Issue 4157), 1124–1131.

Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nat. Rev. Neurosci.* 16, 55–61.

van den Heuvel, M. P. and Sporns, O. (2013). Network hubs in the human brain. *Trends in Cognitive Science* 17(12), 683–695.

Van de Cruys, S. (2017). Affective value in the predictive mind. In Metzinger, T. and Wiese, W.(Eds.) *Philosophy and Predictive Processing*. Frankfurt am Main: MIND Group, 21pp. doi:10.15502/9783958573253

Van de Cruys, S., Evers, K., Van der Hallen, R. ... Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychol. Rev.* 121, 649–675.

Wallace, R. (2005) *Consciousness: A Mathematical Treatment of the Global Neuronal Workspace*. Springer, New York.

Wason, P.C. and Evans, J. St. B . T. (1975). Dual processes in reasoning. *Cognition* 3(2), 141–154.

Whyte, C. J. (2019). Integrating the global neuronal workspace into the framework of predictive processing: Towards a working hypothesis. *Consiousness and Cognition* 73, 102763.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information-processing dynamics. *Acta Psychol.* 41, 67-85.

Winkielman, P., Schwarz, N., Fazendeiro, T. and Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluation judgement. In (Musch, J. and Klauer, K. C., eds.) *The Psychology of evaluation: Affective processes in cognition and emotion* (pp. 189–217). Erlbaum, Mahwah, NJ.

Zimmer, H. D. and Ecker, U. K. D. (2010). Remembering perceptual features unequally bound in object and episodic tokens: Neural mechanisms and their electrophysiological correlates. *Neurosci. Biobehav. Rev.* 34, 1066–1079.

Zmigrod, S. and Hommel, B (2010). Temporal dynamics of unimodal and multimodal feature binding. *Atten. Percept. Psychophys.* 72, 142–152.