Brain and Cognition 112 (2017) 25–38

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Partial information decomposition as a unified approach to the specification of neural goal functions

Michael Wibral [a,*], Viola Priesemann [b], Jim W. Kay [c], Joseph T. Lizier [d], William A. Phillips [e]

[a] MEG Unit, Brain Imaging Center, Goethe University, Heinrich Hoffmann Straße 10, 60528 Frankfurt am Main, Germany
[b] Department of Non-linear Dynamics, Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany
[c] Department of Statistics, University of Glasgow, Glasgow G12 8QQ, UK
[d] Complex Systems Research Group, School of Civil Engineering, The University of Sydney, NSW, Australia
[e] School of Natural Sciences, University of Stirling, Stirling, UK

ARTICLE INFO

Article history:
Received 2 June 2015
Revised 15 September 2015
Accepted 16 September 2015
Available online 21 October 2015

Keywords:
Information theory
Unique information
Shared information
Synergy
Redundancy
Predictive coding
Neural coding
Coherent infomax
Neural goal function

ABSTRACT

In many neural systems anatomical motifs are present repeatedly, but despite their structural similarity they can serve very different tasks. A prime example for such a motif is the canonical microcircuit of six-layered neo-cortex, which is repeated across cortical areas, and is involved in a number of different tasks (e.g. sensory, cognitive, or motor tasks). This observation has spawned interest in finding a common underlying principle, a 'goal function', of information processing implemented in this structure. By definition such a goal function, if universal, cannot be cast in processing-domain specific language (e.g. 'edge filtering', 'working memory'). Thus, to formulate such a principle, we have to use a domain-independent framework. Information theory offers such a framework. However, while the classical framework of information theory focuses on the relation between one input and one output (Shannon's mutual information), we argue that neural information processing crucially depends on the combination of *multiple* inputs to create the output of a processor. To account for this, we use a very recent extension of Shannon Information theory, called partial information decomposition (PID). PID allows to quantify the information that several inputs provide individually (unique information), redundantly (shared information) or only jointly (synergistic information) about the output. First, we review the framework of PID. Then we apply it to reevaluate and analyze several earlier proposals of information theoretic neural goal functions (predictive coding, infomax and coherent infomax, efficient coding). We find that PID allows to compare these goal functions in a common framework, and also provides a versatile approach to design new goal functions from first principles. Building on this, we design and analyze a novel goal function, called 'coding with synergy', which builds on combining external input and prior knowledge in a synergistic manner. We suggest that this novel goal function may be highly useful in neural information processing.

1. Introduction

In many neural systems anatomical and physiological motifs are present repeatedly in the service of a variety of different functions. A prime example is the canonical cortical microcircuit that is found in many different regions of the six-layered mammalian neocortex. These different regions serve various sensory, cognitive, and motor functions, but how can a common circuit be used for such a variety of different purposes? This issue has spawned interest in finding a common abstract framework within which the relevant information processing functions can be specified.

Several solutions for such an abstract framework have been proposed previously, among them approaches that still use semantics to a certain extent (predictive coding with its initial focus on sensory perception), teleological ones that prescribe a goal based on statistical physics of the organism and its environment (free energy principle) and information theoretic ones that focus on local operations on information (Coherent Infomax). While these are all encouraging developments, they also beg the question of how to compare these approaches, and how many more possibilities of defining new approaches of this kind exist. Ideally, an abstract framework that would comprise these approaches as specific cases would be desirable. This article suggests a possible starting point for the development of such a unifying framework.

* Corresponding author.
E-mail address: wibral@em.uni-frankfurt.de (M. Wibral).

http://dx.doi.org/10.1016/j.bandc.2015.09.004
0278-2626/

By definition this framework cannot be cast in processing-domain specific language, such as 'edge-filtering' or 'face perception, or 'visual working memory, for example, but must avoid any use of semantics beyond describing the elementary operations that information processing is composed of.[1] A framework that has these properties is information theory. In fact, information theory is often criticized exactly for its lack of semantics, i.e. for ignoring the *meaning* of the information that is processed in a system. As we will demonstrate here, this apparent shortcoming can be a strength when trying to provide a unified description of the goals of neural information processing. Moreover, by identifying separate component processes of information processing, information theory provides a meta-semantics that serves to better understand what neural systems do at an abstract level (for more details see Wibral, Lizier, & Priesemann, 2015). Last, information theory is based on evaluating probabilities of events and thereby closely related to the concepts and hypotheses of probabilistic inference that are at the heart of predictive coding theory (Clark, 2013; Hohwy, 2013; Lee & Mumford, 2003; Rao & Ballard, 1999). Thus information theory is naturally linked to the domain-general semantics of this and related theories.

Based on the domain-generality of information theory several variants of information theoretic goal functions for neural networks have been proposed. The optimization of these abstract goal functions on artificial neural networks leads to the emergence of properties also found in biological neural systems – this can be considered an amazing success of the information theoretic approach given that we still know very little about general cortical algorithms. This success raises hopes for finding unifying principles in the flood of phenomena discovered in experimental neuroscience. Examples of successful, information-theoretically defined goal functions are Linsker's infomax (Linsker, 1988) – producing receptive fields and orientation columns similar to those observed in primary visual cortex V1 (Bell & Sejnowski, 1997), recurrent infomax – producing neural avalanches, and an organization to synfire-chain like behavior (Tanaka, Kaneko, & Aoyagi, 2009), and coherent infomax (Phillips, Kay, & Smyth, 1995). The goal function of coherent infomax is to find coherent information between two streams of inputs from different sources, one conceptualized as sensory input, the other as internal contextual information. As coherent infomax requires the precomputation of an integrated receptive field input as well as an integrated contextual input to be computable efficiently (and thereby, in a biologically plausible way), the theory predicted the recent discovery of two distinct sites of neural integration in neocortical pyramidal cells (Larkum, 2013). For details see the contribution of Phillips to this special issue. We will revisit some of these goal functions below and demonstrate how they fit in the larger abstract framework aiming at a unified description that is presented here.

Apart from the desire for a unified description of the common goals of repeated anatomical motifs, there is a second argument in favor of using an abstract framework. This argument is based on the fact that a large part of neural communication relies on axonal transmission of action potentials and on their transformation into post-synaptic potentials by the receiving synapse. Thus, for neurons, there is only one currency of information. This fact has been convincingly demonstrated by the successful rewiring of sensory organs to alternative cortical areas that gave rise to functioning, sense-specific perception (see for example the cross-wiring, cross-modal training experiments in von Melchner, Pallas, & Sur, 2000). In sum, neurons only see the semantics inherent in the train of incoming action potentials, not the semantics imposed by the experimenter. Therefore, a neurocentric framework describing information processing must be necessarily abstract. From this perspective information theory is again a natural choice.

Classic Shannon information theory, however, mostly deals with the transmission of information through a communication channel with one input and one output variable. In a neural setting this would amount to asking how much information present at the soma of one cell reaches the soma of another cell across the connecting axons, synapses and dendrites, or how much information is passed from one circuit to another. Information processing, however, comprises more operations on information than just its transfer. A long tradition dating back all the way to Turing has identified the elementary operations of information as information transfer, active storage, and modification. Correspondingly, measures of information transfer have been extended to cover more complex cases than Shannon's channels, incorporating directed and dynamic couplings (Schreiber, 2000) and multivariate interactions (Lizier, Prokopenko, & Zomaya, 2008), and also measures of active information storage have been introduced (Lizier, Prokopenko, & Zomaya, 2012). Information modification, seemingly comprising of subfunctions such as *de novo* creation and fusion of information, however, has been difficult to define (Lizier, Flecker, & Williams, 2013).

One reason for extending our view of information processing to more complicated cases is that even the most simple function from Boolean logic that any other logic function can be composed of (NAND, see for example Jaynes, 2003, chap. 1) uses two distinct input variables and one output. While such a logic function could be described as a channel between the two inputs and the outputs, this does not do justice to the way the two inputs interact with each other. What is needed instead is an extension of classic information theory to three way systems, describing how much information in the output of this Boolean function, or any other three-way processor of information, comes uniquely from one input, uniquely from the other input, how much they share about the output, and how much output information can only be obtained from evaluating both inputs jointly.

These questions can be answered using an extension of information theory called partial information decomposition (PID) (Bertschinger, Rauh, Olbrich, Jost, & Ay, 2014; Griffith & Koch, 2014; Harder, Salge, & Polani, 2013; Williams & Beer, 2010).

This article will introduce PID and show how to use it to specify a generic goal function for neural information processing. This generic goal function can then be adapted to represent previously defined neural information processing goals such as infomax, coherent infomax and predictive coding. This representation of previous neural goal functions in just one generic framework is highly useful to understand their differences and commonalities. Apart from a reevaluation of existing neural goal functions, the generic neural goal function introduced here also serves to define novel goals not investigated before.

The remainder of the text will first introduce partial information decomposition, and then demonstrate its use to decompose the total output information of a neural processor. From this decomposition we derive a generic neural goal function "G", and then express existing neural goal functions as specific parameterizations of G. We will then discuss how the use of G simplifies the comparison of these previous goal functions and how it helps to develop new ones.

2. Partial information decomposition

In this section we will describe the framework of partial information decomposition (PID) to the extent that is necessary to

[1] To be truly generic, the framework should also avoid to resort too strongly to semantics in terms of "survival of the organism" as even that maybe not desirable for each and every individual organism in certain species. This is because "programmed death" will allow a more rapid turnover of generations and thereby more rapid evolutionary adaptation.