Cognition 179 (2018) 56–70

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Basic functional trade-offs in cognition: An integrative framework

Marco Del Giudice[a,*], Bernard J. Crespi[b]

[a] Department of Psychology, University of New Mexico, United States
[b] Department of Biological Sciences, Simon Fraser University, Canada

ARTICLE INFO

Keywords:
Design
Efficiency
Flexibility
Performance
Robustness
Trade-offs

ABSTRACT

Trade-offs between advantageous but conflicting properties (e.g., speed vs. accuracy) are ubiquitous in cognition, but the relevant literature is conceptually fragmented, scattered across disciplines, and has not been organized in a coherent framework. This paper takes an initial step toward a general theory of cognitive trade-offs by examining four key properties of goal-directed systems: performance, efficiency, robustness, and flexibility. These properties define a number of basic functional trade-offs that can be used to map the abstract "design space" of natural and artificial cognitive systems. Basic functional trade-offs provide a shared vocabulary to describe a variety of specific trade-offs including speed vs. accuracy, generalist vs. specialist, exploration vs. exploitation, and many others. By linking specific features of cognitive functioning to general properties such as robustness and efficiency, it becomes possible to harness some powerful insights from systems engineering and systems biology to suggest useful generalizations, point to under-explored but potentially important trade-offs, and prompt novel hypotheses and connections between disparate areas of research.

1. Introduction

Trade-offs—balances between separately advantageous but conflicting traits—are fundamental aspects of all goal-directed systems, whether they are artificial machines or biological mechanisms designed through evolution by natural selection. Trade-offs are also ubiquitous in cognitive systems. Enhanced computational performance does not come for free; the same is true of other desirable properties such as speed, flexibility, or the ability to withstand damage. Crucially, improving a system on one front will typically worsen it in other ways. For example, the speed of decisions can be increased by sacrificing their accuracy (Heitz, 2014), and more flexible learning algorithms also tend to be more computationally demanding (Daw & Dayan, 2014). The design of cognitive systems is thus shaped by constraints, compromises, and opposing priorities that can be understood only in relation to the underlying trade-offs.

Cognitive trade-offs have been addressed in many disciplines, from neuroscience and psychology to behavioral ecology and computer science. Unfortunately, the relevant literature remains scattered, limited in scope, and conceptually fragmented. Different research traditions tend to focus on different trade-offs, largely ignore each other's contribution, and often employ different terms for similar or overlapping constructs. To the best of our knowledge, there have been no attempts to organize this literature within a coherent framework. Here we take an initial step in this direction by offering an integrative overview of what we label *basic functional trade-offs*: a set of highly general trade-offs that apply to all natural or artificial systems designed to perform a function, including cognitive systems whose function can be described as manipulation of information (Piccinini & Scarantino, 2011; more on this in Section 2).

Basic functional trade-offs are defined by four key properties of goal-directed systems: *performance*, *efficiency*, *robustness*, and *flexibility* (Fig. 1). Together, these properties map the abstract "design space" of any natural or artificial system endowed with a function; when they are applied to cognitive systems (as we do here), they provide a shared vocabulary to describe a variety of specific characteristics such as speed, accuracy, reliability, memory use, and so on. By linking specific features of cognitive functioning to general properties such as robustness and efficiency, it becomes possible to harness some powerful insights from systems engineering and systems biology, two related disciplines that explicitly investigate the design of complex functional mechanisms (Alderson & Doyle, 2010; Doyle & Csete, 2011; Kitano, 2004, 2007).

We have identified the four properties in Fig. 1 as basic after surveying an extensive literature on trade-offs in biology and engineering, as detailed in the remainder of this paper. We could not find other examples of properties that were both universal (i.e., would apply to all functional systems) and similarly general (i.e., were not already encompassed by the basic ones). This assertion does not mean that the classification we propose is fully exhaustive or that it cannot be

* Corresponding author at: Department of Psychology, University of New Mexico. Logan Hall, 2001 Redondo Dr. NE, Albuquerque, NM 87131, United States.
E-mail address: marcodg@unm.edu (M. Del Giudice).

https://doi.org/10.1016/j.cognition.2018.06.008
Received 8 January 2018; Received in revised form 5 June 2018; Accepted 11 June 2018

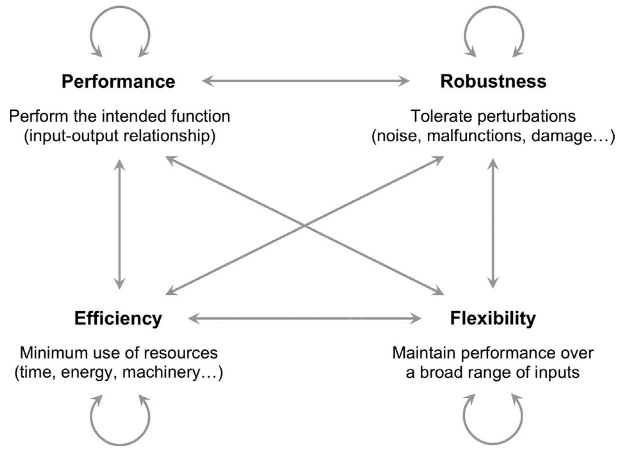

Fig. 1. A map of basic functional trade-offs. Performance, efficiency, robustness, and flexibility are the key properties of all functional systems, including natural and artificial cognitive systems. Straight arrows represent trade-offs between properties; curved arrows represent trade-offs between different aspects of the same property.

extended in principle, and we encourage its growth and elaboration. As we discuss in the following sections, even the distinctions between basic properties are not absolute, and admit a degree of conceptual overlap—for example, in particular cases it can be hard to differentiate sharply between robustness and flexibility, or between robustness and performance. While it is important to acknowledge and discuss those cases, the functional properties that we describe have a broad range of application and considerable heuristic power. Their value lies in their ability to integrate many particular examples within a common frame of reference, suggest useful generalizations, and prompt novel hypotheses and connections across scientific domains.

In this paper, we examine the four properties shown in Fig. 1 and discuss the trade-offs that arise between competing properties, as well as between different aspects of each (e.g., trade-offs between multiple aspects of robustness), with a focus on cognitive and neural systems. We also consider the implications of simultaneous trade-offs among more than two properties (e.g., three-way trade-offs between performance, robustness, and efficiency). The framework we present brings together many specific trade-offs that have been investigated in the literature (summarized in Table 1), points to some potentially important trade-offs that have received comparatively little attention so far, and offers a toolkit for clarifying some counterintuitive phenomena such as "less-is-more" effects in the performance of simple cognitive heuristics (Gigerenzer & Brighton, 2009). We conclude by considering possible ways to apply and extend the framework. From a psychological perspective, a better understanding of trade-offs may illuminate typical human cognitive variation as well as mental disorders, some of which appear to involve extremes or dysregulation in the balances between competing cognitive functions (e.g., Baron-Cohen, 2009; Crespi & Go, 2015).

2. Performance

The performance of a system is usually defined as its ability to produce an intended result (or some other roughly equivalent formulation). The concept of performance is meaningless without explicit or implicit reference to *function*, the idea that the system has an identifiable purpose, goal, or rationale. In turn, function implies *design*—in order to fulfill a purpose, a system needs to be structured in an organized, non-random fashion. When the term "design" is employed in this broad sense it does not require the existence of a conscious designer: indeed, the crucial insight of Darwinian biology is that design and function can arise from the blind, undirected, and impersonal process of natural selection (Alderson & Doyle, 2010; Dennett, 2009; Sterling & Laughlin, 2015).

Table 1
Summary of the trade-offs discussed in the text.

Basic trade-offs	Main examples discussed in the text
Efficiency vs. performance	- Speed-accuracy trade-offs - Exploration-exploitation trade-offs - Efficiency trade-offs in neural design
Efficiency vs. robustness	- Robustness-resource trade-offs - Proactive vs. reactive control
Efficiency vs. flexibility	- Generalist-specialist trade-offs - Model-based vs. model-free learning - Fast and frugal heuristics
Performance vs. robustness	- Bias-variance trade-offs - Pessimistic strategies
Performance vs. flexibility	- Generalist-specialist trade-offs
Robustness vs. flexibility	- Stability-flexibility dilemma - Proactive vs. reactive control - Fast and frugal heuristics
Aspects of efficiency	- Space-time trade-offs
Aspects of robustness	- Robustness-fragility trade-offs

In biological systems, goals can exist on an objective level even if they are not represented consciously (or at all) within the system. When bacteria move toward higher concentrations of glucose by chemotaxis, their behavior is regulated by a system of feedback control that alternates straight line swimming and random tumbling. The objective goal of this behavior is to move bacteria toward glucose, even if bacteria themselves have no representation of it—and, interestingly, do not even possess a representation of the direction in which they are swimming (Bechhoefer, 2005). Such real but unrepresented goals are ubiquitous in biological systems; in Daniel Dennett's terminology, they can be described as "free-floating rationales" produced by blind selection (Dennett, 2009). The difference between free-floating rationales and deliberate, fully represented goals (such as those of a human designer) is best understood as a gradient, which is climbed by evolutionary processes through the gradual accumulation of functional specialization and cognitive complexity. For the purpose of this paper, we make no distinction between different types of goals, and the concepts of design and function apply to natural and artificial systems alike.

2.1. Performance in cognitive systems

Broadly defined, a cognitive system is an information-processing mechanism that computes mappings between inputs and outputs (Lewis, Howes, & Singh, 2014). Input-output mappings can be extremely complex; as well, outputs can take many possible forms, including commands to physical effectors (e.g., muscles or motors) but also representations that are used as inputs to other systems (e. g., information transfer between different brain regions). Note that we employ both "computation" and "information" in a broad sense, to include non-algorithmic and non-digital types of computation as well as various types of information (e.g., Shannon vs. semantic information; see Piccinini & Scarantino, 2011). Thus, our working definition of a cognitive system includes both natural and artificial instances of information-processing mechanisms. Some proponents of dynamical approaches to cognition (most notably van Gelder, 1998) have argued that cognitive mechanisms should be understood as dynamical systems—as defined for example by sets of differential equations—rather than computational processes. However, dynamical systems can also be analyzed with the tools of information theory and described from a