Synthese (2022) 200:402
https://doi.org/10.1007/s11229-022-03870-0

ORIGINAL RESEARCH

On the role of contextual factors in cognitive neuroscience experiments: a mechanistic approach

Abel Wajnerman-Paz[1] · **Daniel Rojas-Líbano**[2,3]

Received: 15 November 2021 / Accepted: 22 August 2022 / Published online: 21 September 2022

Abstract

Experiments in cognitive neuroscience build a setup whose set of controlled stimuli and rules elicits a cognitive process in a participant. This setup requires researchers to decide the value of quite a few parameters along several dimensions. We call ‘’contextual factors” the parameters often assumed not to change the cognitive process elicited and are free to vary across the experiment’s repetitions. Against this assumption, empirical evidence shows that many of these contextual factors can significantly influence cognitive performance. Nevertheless, it is not entirely clear what it means for a cognitive phenomenon to be context-sensitive and how to identify context-sensitivity experimentally. We claim that a phenomenon can be context-sensitive either because it is only triggered within a specific context or because different contexts change its manifestation conditions. Assessing which of these forms of context-sensitivity is present in a given phenomenon requires a criterion for individuating it across contextual variations. We argue that some inter-level experiments that, within the mechanistic approach to explanation, are required to identify relations of constitutive relevance between a phenomenon and a mechanism, are also necessary for individuating the phenomenon across its contextual variations. We articulate a criterion according to which behavioral variations across contexts indicate different phenomena if and only if the mechanistic activities, components and/or organizational properties recruited in each context are different. We support this approach by showing how it is applied in paradigmatic studies addressing cognitive performance differences resulting from contextual variations of task features, such as stimulus type and response modality. Finally, we address the challenge that a form of context-sensitivity possessed by the so-called ‘multifunctional mechanisms’ is incompatible with our proposal because it entails that the same mechanism can be recruited in different contexts to produce different phenomena. We examine key cases of multifunctionality and argue that they are consistent with our proposal because a single mechanism can have different components, activities and/or organizational properties in different contexts. Thus, these modifications may

---

Extended author information available on the last page of the article

not affect the identity of a mechanism, and they could explain how it produced different phenomena in those contexts.

Keywords Context · Neuroscience experiments · Mechanism · Cognitive neuroscience

1 Introduction

The cognitive sciences have built most of their results from experimental settings where participants repeatedly execute a procedure in which a cognitive process occurs (i.e., is elicited by the procedure). Monitoring behavioral variables from the participant during the experiment is the regular assessment and confirmation that the cognitive processes under study are taking place. Thus, the phenomenon researchers seek to explain is constituted by these behavioral responses and the environmental conditions (e.g., stimuli) through which it is produced or modulated during the experiment. Concurrently with task execution, cognitive neuroscience proposes monitoring biological variables and analyzing the results to relate biological and behavioral processes. In this way, a mechanistic explanation of cognitive processes proceeds by identifying the relevant biological components, the activities they perform and their organization, from which the observed behavior results.

Experimental procedures to elicit cognitive processes had been around even before the early behavioristic approaches, were then merely refined by the cognitive sciences, and are still used in most cognitive neuroscience setups nowadays. To design and implement an experiment, researchers must decide on quite a few parameters, along several dimensions (as explained in what follows), that specify the concrete conditions wherein the procedure will take place. Any design must choose parameters such as inter-trial interval, physical properties of stimuli, feedback and response modalities, and the like. Intentionally or not, researchers also set the physical parameters of the experiment's setup (e.g., humidity, temperature, illumination). We can also conceive the social interactions between experimenter and participant as parameters to choose from: the way the instructions are delivered, the number of people present and their interactions, and similar factors. Therefore, the experiment requires designing a whole situation, from the physical surroundings to the actual task environment, including the human interactions that precede and accompany the procedure. Finally, and in addition to these aspects, participants bring to the experiment their internal context, constituted by the general state of all their physiological systems and their previously acquired behavioral capacities.

Most of these parameters can change (move in parameter space, as it were), and researchers would still conceive that the procedure assesses and elicits the same cognitive process. In fact, people have analyzed and documented that these parameters can vary widely across experiments and laboratories (Sullivan, 2008). Consider, as an example, an experiment designed to study "visual working memory". Typical designs here involve retaining, during a brief period, a transiently visualized stimulus to later report about it. The room illumination, the stimuli size, the response modality, the social environment of the laboratory, and other factors, can vary widely between or even within repetitions of the experiment. However, we would still consider that the procedure elicits the same process, thus implying that the process is to a reasonable extent independent of these factors. Therefore, we call *contextual factors* to all those elements of the experimental procedure that researchers must choose to build an experiment but are often assumed not to change the cognitive process being elicited and assessed and are therefore supposedly free to vary across the experiment's repetitions.

However, it has been shown that many of these contextual factors do influence the cognitive processes under study, as expressed in across-contexts changes in behavioral performance and its underlying neurobiological processes (see Sect. 2). For instance, the so-called "facial feedback" effect in the study of emotions seems to depend on the participants being aware they are observed (Noah et al., 2018). Likewise, reward expectations shape the neural activity subserving a cognitive task (Correa et al., 2018). Nevertheless, it is not entirely clear what it means for behavior to be context-sensitive as described by these examples.

The context-sensitivity of cognitive capacities has been addressed in the philosophical literature. For example, in recent discussions about brain structure-function mapping, it has been suggested that mechanisms can implement different cognitive functions or produce different cognitive phenomena in different (either environmental or neural) contexts. For instance, there is evidence suggesting that diverse functions recruit overlapping cortical areas (e.g., Anderson 2010), and that brain regions perform different functions depending on the networks in which they participate, i.e., their "neural context" (Klein, 2012). This is a potential problem for mechanistic functional analysis (see Sect. 4), which aims to explain any given behavioral phenomenon by identifying the functions of its underlying neurobiological components. If components' functions vary from context to context, they may not be useful for this kind of analysis[1].

Here, we approach context-sensitivity differently. Our focus is not structure-function mapping and mechanistic decomposition but rather the related issue of the individuation of cognitive phenomena. We would like to understand *when contextual variations of behavior entail a difference in the kind of phenomenon being elicited and when they do not*. We need to distinguish between the situations in which contextual variations merely modulate how the same phenomenon is manifested from situations in which contextual variations produce entirely different phenomena. We endorse a mechanistic approach to this problem. Specifically, we affirm that monitoring changes in underlying biological variables is critical to determine whether the phenomenon is preserved through contextual behavioral changes. We argue that key experiments addressing phenomena individuation entail that different phenomena are produced in different contexts if and only context-to-context variations in the mechanism(s) underlying those phenomena are detected.

However, determining *which* mechanistic changes entail a difference in the elicited phenomenon can be challenging. A first possibility is that if and only if *different kinds of mechanisms* are recruited in different contexts, then different phenomena are

---

[1] see Burnston (2020) for an alternative approach showing that context-sensitivity is consistent with decomposition.

elicited in those contexts. That is, the individuation of cognitive phenomena would depend on the individuation of their underlying mechanisms. This strong interpretation of our proposal is problematic. Crucially, if the same mechanism can produce different phenomena in different contexts (as the existence of the aforementioned multifunctional mechanisms suggests), then the context-to-context variation of the phenomenon can occur without any change in the kind of mechanism being recruited.

Here, we argue that the individuation of phenomena must be only tied to the individuation of the *components, activities or/and organizational properties* (CAOs) of the underlying mechanism. Different functions are performed in different contexts if and only if different CAOs are recruited in those contexts. Whether *the mechanism itself* (and not only the CAOs) is different in the contexts in which different functions are performed *is a different individuation problem* which we do not need to address. Thus, even in the case of multifunctional mechanisms, the detection of neural level variations is relevant to assessing whether a mechanism is implementing different functions. We contrast context-sensitive multi-functionality with domain-general context-sensitive mechanisms, in which the exact same solution or strategy (i.e., the same CAOs) works in different contexts (and therefore the same kind of phenomenon is produced), but the context modulates how the phenomenon is manifested.

We have organized the article as follows. We start with a brief description of the structure of the experimental procedures in cognitive neuroscience that presents the nature of the problem, then propose a characterization of contextual factors and present some key findings showing that these factors can impact cognitive phenomena (Sect. 2). We then articulate a mechanistic approach to phenomena individuation (Sect. 3.1) and argue that experimenters address context-sensitivity by following this criterion (Sect. 3.2). We end with a discussion of the challenge posed by multifunctional mechanisms and (Sect. 4) and a brief reflection on the individuation of CAOs (Sect. 5).

2 Experiments in cognitive neuroscience

Since the beginnings of experimental psychology, research programs such as those of Cornelius Donders in the nineteenth century devised ways to produce stereotyped behaviors in a consistent manner (Donders, 1969). A key ingredient in these protocols was the repetition of a procedure, which allowed comparisons within and between participants. The development continued in the twentieth century, as in the case of Edward Thorndike, involving the construction of highly specific “Puzzle Boxes”, which were wooden boxes with complicated mechanisms constituted by strings, levers, and doors, that animals learned to operate (Burnham, 1972; Chance, 1999). The repeated exposure of the animal to the box allowed reliable observations of the behavioral changes associated with the learning process. A relevant and related theme is that this methodological stance assumes that the procedure elicits a distinguishable capacity or function (i.e., we can design different puzzle boxes targeted to study other capabilities). Thus, the box, a physical device, was simultaneously an opportunity or scenario for expressing a general ability (e.g., learning) and a constraint for that capacity, attached to a specific set of contingencies. These experimental devices were further developed and refined in the Skinnerian tradition, where the addition and manipulation of rewards ("reinforcements") allowed increased repetitions of the studied behavior (Ferster, 1953; Heron & Skinner, 1939; Skinner, 1930).

Cognitive science researchers kept the same basic methodological scheme of repeated "trials" during which human participants produced stereotyped behaviors contingent on specific stimuli (Garner, 1953; Pollack, 1952; Sternberg, 1966). In addition, devices increasingly included electronic components, which allowed newer and more precise methods to deliver stimuli and register responses. However, the basic scheme was the same as Thorndike's boxes: a specifically designed mini-environment that promoted the unfolding of a behavioral/cognitive process. Finally, the advent of cognitive neuroscience simply added on top to these tasks the simultaneous recording of neural activity or monitoring of physiological processes at different levels, ranging from single cells to circuits, nuclei, and entire neural systems (Evarts, 1966; Jasper, H., Ricci, G. F., Doane, B., 1958; Petersen et al., 1988; Posner et al., 1988). In the words of Vernon Mountcastle, the methods of the then-emerging cognitive neuroscience were such that:

"…one trains an animal to emit repetitively at signal an item of behaviour, and records the electrical signs of neural activity thought relevant thereto, seeking causal relations between simultaneous variables on other grounds as well, […]. This method […] opens the way to study of the neural mechanisms of complex behaviour.“ (Mountcastle, 1978).

This "puzzle-box" approach, which we continue to use today to study cognitive processes, is a methodological cornerstone of the mind sciences. It is a specific physical scenario designed to allow the repeated and reliable production of the cognitive process one intends to study, which is monitored through the subject's behavior and physiological activity. That the procedure does elicit a consistent and well-delimited cognitive process across laboratories is, however, far from clear, as some surveys of the experimental field have shown (Sullivan, 2010).

From a methodological point of view, the stimuli's physical features, timings, and procedure instructions constitute the experiment's independent variables, i.e., those under the experimenter's control. The dependent variables, i.e., what we control through the manipulation of independent variables, are of two types. One type is constituted by the *behavioral* variables, which correspond to the participant's responses to the different stimuli and trials throughout the experiment. Typical examples of these variables are the accuracy and the reaction times. A second type is the set of *biological* variables, which are any measurement of biological processes during the trials. These variables either constitute or allow to indirectly identify the organized components and activities that define a mechanism explaining a phenomenon. Again, examples vary widely, including the BOLD signal as an indirect measure of brain activity, brain waves directly picked up by EEG electrodes, single-unit and multielectrode recordings, pupil size and eye movements, and peripheral physiological measures such as electrocardiography, electrogastrography, respiration, and measures of autonomic nervous activity, among others.

To a reasonable extent, experiments have been a mere means to obtain the behavioral responses that allow us to examine the cognitive process (e.g., whether the participant perceived the stimulus or not, classified it appropriately or not, and the like).

The experimental procedure is a means to get Mountcastle's "item of behaviour" that provides the time frame within which to examine the biological variables. Once they obtain that temporal frame, researchers often forget about behavior and focus on analyzing the biological data within the frame, assuming it contains the cognitive process of interest, defined in terms of the diversity of tasks used. However, from the very beginnings of cognitive research, it was clear that even in basic perceptual tasks, several extra-sensory factors play a role in the behavioral readout, and that they constitute additional independent variables whose manipulation can significantly alter the process. For example, researchers that initially developed and applied signal detection theory to psychophysical procedures acknowledged some of these as "bias factors" that required specific methods to take them into account (Swets, 1973).

More generally, there are reasons to believe that minor, usually unnoticed, or unacknowledged contextual variations that belong to the puzzle box features affect the phenomenon under study. Complex causal chains interweave agents with aspects of their environments, which can include more than the set of independent variables the experimenter normally focuses on. Mind phenomena display a great deal of context-sensitivity, meaning that responses or acts contingently fit the task at hand and constantly change, allowing agents to adapt to an enormous variety of situations and challenges. Most probably, all adaptive human behavior would involve some degree of context-sensitivity, expressed in capacities such as cognitive flexibility, concept formation, goal pursuing, cognitive control, and the like. They all involve some degree of adaptation to the task (or context) at hand. Therefore, our cognitive processes most likely display a complexity that would not be easy to grasp by single experiments, however well designed and executed.

An example may be of help in describing these issues. Fritz Strack, Leonard Martin, and Sabine Stepper reported evidence for a phenomenon called "facial feedback", in which subjects artificially producing smile gestures tended to rate cartoons as funnier than their control counterparts (Strack et al., 1988). The interpretation was that manipulations that induce or inhibit the facial movements associated with smiling, impact positive or negative emotions and affect. This is, therefore, a potentially important finding for theories of emotions and their relation to body processes. A registered replication report of the same experimental protocol from 17 different laboratories failed to find the effect (Fritz Strack 2016; Wagenmakers et al., 2016), leaving the field concluding that the original findings were probably spurious. However, that was not the end of the story. Tom Noah and collaborators set up a new replication attempt. This time they manipulated a variable that was not consistent in previous reports: the participants' awareness that they were being monitored or observed (Noah et al., 2018). In an experiment designed to compare these conditions, the results depended on them: when participants were observed, the effect vanished, prevailing only in the absence of external monitoring. Later, independent replication attempts have provided evidence for the original feedback hypothesis (Marsh et al., 2019).

In addition to parameter manipulation in experimental tasks, other dimensions have also proved relevant in this regard. Physical components (i.e., the environmental physics of the experimental setting) also affect behavioral results (Ashworth et al., 2021; Freiberger et al., 2016; Tian et al., 2021). Social interactions surrounding the experiment's execution also impact the processes studied in the experiment (Müller-Pinzler et al., 2015; Steinborn & Huestegge, 2020; Wahn et al., 2020). Also, the participants' internal context, given by the state of all physiological systems, may potentially bias and influence the brain systems subserving cognition during experimental tasks (Al et al., 2020; Azzalini et al., 2019; Park et al., 2014; Perl et al., 2019). Additionally, developmental (Khalidi, 2017a, 2017b), aging (De Brigard, 2017), and training/learning history (Viola, 2021) contexts have been discussed.

Our present question regarding context-sensitivity is: how do the particular contingencies of a given "puzzle box" affect the very process we study through it? More specifically, how does it affect the individuation of cognitive phenomena? In other words: how do we know whether *contextual variation changes the phenomenon elicited* in different contexts or whether it *only modulates how the same phenomenon is manifested*?

In the next sections, we propose a "mechanistic" approach to this question and argue that the top-down exploration of the mechanism underlying a given phenomenon can be used to determine how contextual variations affect its individuation.

3 Mechanistic context

3.1 On context's methodological role. A mechanistic approach

In this section, we will try to glean insights from current ideas in the philosophy of neuroscience about the role and contribution of context to explanation. To this end, we will apply the key elements of a dominant approach to neuroscientific explanation, the so-called 'new mechanism' or simply 'mechanism'. Specifically, we will focus on contextual elements that define task conditions, such as stimulus and response modality.

Most of the discussion on neuroscientific explanation during the last two decades has been focused on *constitutive* explanation (Bechtel & Abrahamsen, 2005; Craver, 2014; Machamer et al., 2000; Piccinini & Craver, 2011). In contrast with etiological or causal explanation, which explains a phenomenon by describing its antecedent causes, constitutive or componential explanation explains the behavior of a mechanism as a whole by describing the organized activities of its individual components. Distinguishing between causal and constitutive relations is important because they enable different kinds of experimental interventions that are relevant for explanation (Craver, 2007).

An important distinction for our purposes is that the CAOs of a mechanism are at a *lower* level than the phenomenon they produce. In a cognitive neuroscience experiment, this means that the neural processes investigated are at a lower level than the participant's behavior. Within the mechanistic viewpoint we are adopting, levels are conceived as connected by a part-whole relation: Cortical areas and their capacities are at a higher level than the columns or nuclei that compose them, which are in turn at a higher level than the neuronal populations composing them, and so on. Therefore, the levels as we understand them here vary case by case, depending on the mechanism being analyzed, and are not fixed entities within an unchanging natural hierarchy. Importantly, the experimental manipulation of the relata of mechanistic explanations (i.e., CAOs on the one side and the phenomenon they constitute on the other) depends on inter-level interventions.

Mechanistic models explain by describing the mechanism underlying a phenomenon. A mechanism produces or sustains a phenomenon through three key elements, namely its CAOs (Bechtel & Abrahamsen, 2005). The phenomenon is a crucial aspect of mechanistic explanation because, as we will see below, a CAO is part of a mechanism only if it has an appropriate difference making relation with a phenomenon. A phenomenon is multifaceted, involving many different kinds of conditions. On one side, we have the precipitating conditions (the input and background conditions) that initiate a given behavior, and the inhibitory, modulatory and nonstandard conditions, which describe the factors that can inhibit or alter behavior. On the other hand, the phenomenon is constituted by its manifestation conditions, which characterize behavior itself or its development (e.g., its different components and their temporal characteristics).

Given its multifaceted nature, the individuation of phenomena can be challenging. In addition to providing an incomplete description that fails to capture some of the mentioned conditions, there are two mistakes one can make. One can commit a lumping error, assuming that several distinct phenomena are actually one, or a splitting error, which involves incorrectly assuming that one phenomenon is many (Craver, 2007, Chap. 4). These possibilities are crucial for characterizing context-sensitivity because, as mentioned in Sect. 1, we are concerned with how context affects the individuation of cognitive phenomena. That is, we would like to assess when contextual variations force us to split a given phenomenon and when they do not.

As we saw, mechanisms are multi-level entities, and each level of a mechanism explains the capacities of the immediately higher level components (Boone & Piccinini, 2016). Thus, we have neurocognitive phenomena at any level of a mechanistic hierarchy. However, the kind of cognitive neuroscience experiments we consider in this paper mainly focus on explaining the participant's behavior. They aim at explaining how the organized activities of a set of relevant neurocognitive components produce behavior within an experimental task. In what follows, we will use 'phenomenon' to refer exclusively to the behavioral manifestation of a neurocognitive capacity plus its dependence on specific independent variables defining its precipitation, inhibitory, modulatory, and non-standard conditions.

The idea that characterizing the phenomenon is a central part of setting up an experiment and providing an explanation is key to our proposal. The contextual elements we will focus on in the following section (i.e., stimulus and response modality), which define aspects of the cognitive/behavioral task, are part of the phenomenon. Our goal is to show that the mechanistic framework should be extended to explain how we can experimentally determine when contextual variation affects a given phenomenon and in which sense. Specifically, we argue that inter-level experiments that involve a phenomenon and its underlying mechanism are not only necessary to determine the boundaries of a mechanism (which CAOs belong to the mechanism that explains a phenomenon and which do not) but also *to determine the contextual boundaries of the phenomenon* (which contextual variations change the phenomenon we want to study and which do not). We claim that experiments in cognitive neuroscience that aim to individuate a phenomenon across different contexts critically depend on the individuation of CAOs of its underlying mechanism(s). We can explicitly state in the following way the "mechanistic phenomenon" (MP) criterion we are proposing:

> MP: When turning a context C_1 into a context C_2 modulates the manifestation conditions of a putative phenomenon, then different kinds of phenomena P_1 and P_2 are produced in those contexts if and only if the set of components, activities and/or organizational properties, CAOs_1 $\{c_1,c_2\ ,...c_n;\ a_1,\ a_2\ ,...a_n;\ o_1,o_2\ ,..o_n\}$, recruited in C_1 is different from the set CAOs_2 recruited in C_2.

Therefore, we commit a lumping error when we affirm that phenomena elicited in different contexts are manifestations of the same capacity, despite the fact that the underlying mechanism(s) producing them recruits different CAOs in each context. Conversely, we commit a splitting error when we judge that across-contexts manifestations of the same phenomenon are different phenomena, despite the fact that the underlying mechanism(s) recruits the same CAOs for producing them in each context.

Of course, not every change to a CAO constitutes a change in the phenomenon. There are two situations in which this does not happen. Firstly, as we will argue in Sect. 5, the *type* to which a CAO belongs is often very flexible, i.e., the CAO can undergo many identity-preserving changes. These significant changes in a CAO do not change the phenomenon's identity.

More importantly, there are situations where a modification in the *kind* of CAO does not change the phenomenon either. The kinds of mechanistic changes we are focused on in this paper are those that are correlated with context-dependent changes at the behavior-level phenomenon. Our proposal is a mechanistic approach for disambiguating between alternative interpretations of phenomena exhibiting context-sensitivity. Thus, we claim that mechanistic changes can split a phenomenon *only if the phenomenon exhibits this context-sensitivity*. A CAO-type change entails a phenomenon-type change only if the first produces a significant modulation in a putative phenomenon. Otherwise (i.e., if the phenomenon does not exhibit any context-dependent modulation), the different CAOs can be taken to produce the same kind of phenomenon (i.e., the phenomenon is multiply realized).

We can now characterize the role that inter-level experiments have within the mechanistic approach and then explain how this role should be broadened to account for the mechanistic individuation of phenomena. Within this view, it is often assumed that after the phenomenon is correctly identified, we must characterize which CAOs constitute its underlying mechanism. This is accomplished through three main kinds of *interlevel interventions*, that is, interventions on the mechanism that affect the phenomenon and interventions on the phenomenon that affect the mechanism.

Firstly, interference experiments are bottom-up inhibitory experiments in which one intervenes to diminish, disable, or destroy some putative component in a lower-level mechanism and then detects the results of this intervention for the explanandum phenomenon. Lesion experiments are a paradigmatic example of these interventions. Secondly, stimulation experiments are bottom-up, excitatory experiments in which one intervenes to excite or intensify some component in a mechanism and then detects the effects of that intervention on the explanandum phenomenon. A classic example is determining the role in the motor cortex by observing how its stimulation affects motor behavior, as in the classic experiments of Eduard Hitzig and Gustav Fritsch (Hagner, 2012). Thirdly, activation experiments are excitatory, top-down experiments in which one intervenes to activate, trigger, or augment the explanandum phenomenon and then detects the properties or activities of one or more putative components of its mechanism. Typical examples are PET, fMRI, single- and multi-unit recording experiments, in which one engages the experimental subject in some task while monitoring associated neural activity patterns. Finally, deprivation experiments are top-down inhibitory experiments in which one prevents or suppresses the occurrence of the explanandum phenomenon and detects changes in the activities of the underlying mechanism.

Each of these four kinds of interventions is insufficient, by itself, for determining the constitutive relevance of a component, activity, or organizational property regarding a given phenomenon. For instance, top-down experiments could modulate mere correlates of the actual underlying mechanism. Similarly, bottom-up interventions could indirectly affect the phenomenon by influencing a neural structure that in turn affects some component or activity of the actual underlying mechanism. Given difficulties of this kind, Craver (2007) argues that determining constitutive relevance requires the mutual manipulability between the phenomenon and the mechanism through the combination of top-down and bottom-up interventions. It is worth mentioning that Craver neglects deprivation experiments because they are very rare in neuroscience. In Craver's paradigmatic presentation of the criterion (Craver, 2007 p. 146) they are not even listed as part of those that constitute mutual manipulability. We will not discuss here whether deprivation experiments are relevant to mutual manipulability. However, we do claim that these inter-level experiments do play a crucial role in the individuation of the phenomenon.

As we mentioned above, our proposal is that some of these interlevel interventions are not only used for determining constitutive relevance but are also required for phenomena individuation. Specifically, as we will see in the next section, some deprivation experiments are required to test the boundaries of the phenomenon, that is, to determine which variations of the task context preserve the phenomenon and which do not. These are deprivation interventions in which the phenomenon is prevented from happening by changing contextual features such as stimulus type or response modality. These interventions count as genuine interlevel experiments because, following MP, experimenters consider that a different phenomenon was produced by contextual variation (and that the original phenomenon was prevented from happening) if and only if the intervention causes a variation in the neural-level CAOs.

Of course, before exploring its underlying mechanism, individuating a phenomenon through its contextual variations requires identifying these variations. In the following section, we show that specific *intra-level behavioral interventions* (i.e., monitoring and manipulating some of the behavior-level dependent variables through interventions on behavior-level independent variables, such as precipitation conditions) also play a central role in mechanistic explanations. These are required for identifying potential contextual factors related to a phenomenon.

However, we claim that the contextual variations detected through intra-level behavioral experiments need to be complemented with the inter-level experiments described above. Specifically, we need top-down activation and deprivation experiments to determine how the variation of contextual factors affects the underlying mechanism. This is because evidence of contextual variation is consistent with two different ways in which a phenomenon or function could be context-sensitive. This combination of inter and intra level interventions will constitute our proposal for experimentally assessing the context sensitivity of a neurocognitive function.

3.2 Individuating phenomena through task context variations

Many aspects that constitute the experimental context include parameters that define the task itself, such as the number of trials, inter-trial intervals, feedback type, stimuli features, and response modality. We will examine context-sensitivity as is manifested through the modulation of behavior through these parameters. These elements may play different roles within a mechanistic explanation, but they clearly are part of (or are somehow related to) the mechanistic *phenomenon*. For instance, stimulus features are related to the precipitating conditions of the phenomenon that initiate or modulate the task, whereas response modality is related to its manifestation conditions (the specific way in which the task is performed). We can assess whether a mechanism is sensitive to these contextual task conditions through intra-level behavioral experiments, i.e., experiments in which both the intervened and monitored variables are at the behavioral level.

Regarding stimulus conditions, we can examine their role by determining whether changing them produces changes in behavioral response properties (that is, the manifestation conditions of the phenomenon). A paradigmatic example can be found within the debate about the so-called ‘grounded cognition’. In contrast with amodal views, according to which reasoning or thinking about a given category depends on a single modality-independent capacity, empiricist or grounded approaches to cognition are a family of theories which share the thesis that cognition depends on a variety of modality-specific capacities for perception, action, and emotion. For instance, in a task that requires thinking about dogs, grounded views would predict that the brain will recruit different modality-specific capacities such as the ability to process auditory information about dog barks, the ability to process visual information about dog shapes and/or the ability to process olfactory information about dog scent, etc. (e.g., Barsalou et al., 2003; Barsalou, 2016; Reilly et al., 2016, Binder, 2016).

Key behavioral-level evidence supporting grounded views comes from the determination of switching costs in conceptual tasks (i.e., tasks that require employing conceptual knowledge about a given category) involving different modalities, which can also be seen as evidence for the context-sensitivity of capacities involved in these tasks. For instance, Pecher, Zeelenberg, and Barsalou have shown switching effects during a verification task, in which participants are asked whether or not a particular property is true about a given category (e.g., CAT–has a head). They examined pairs of trials that were either from the same modality (LEAVES–rustling followed by BLENDER–loud) or different modalities (CRANBERRIES–tart followed by BLENDER–loud). They found longer reaction times for the second trial in a pair of different modality (switch) trials than for the second trial in a pair of the same modality (no-switch) trials (Pecher et al., 2003). Thus, changing the modality of a stimulus affects performance in this verification task.

This kind of experiment shows that a phenomenon is sensitive to the variation of specific stimulus features. That is, these features define part of its context-sensitivity. However, this kind of evidence is consistent with two different forms of context-sensitivity. One possibility is that the target phenomenon is only triggered by the stimulus conditions present in one particular context. That is, different stimulus conditions trigger *different* cognitive functions, i.e., verification of auditory properties vs. verification of gustatory properties. As we saw, this is the kind of context-sensitivity that grounded approaches entail: different capacities or cognitive functions are recruited for conceptually processing information from different modalities.

According to the MP criterion, this would happen only if these different stimulus conditions triggered different mechanistic CAOs. Experimenters indeed distinguish between cognitive phenomena by tracking these neural-level variations. For instance, the inter-level experiments that motivate the grounded view show that different modality-specific regions dedicated to vision and action are recruited during conceptual tasks that require processing information about different modalities. For instance, early brain-imaging studies showed that retrieving the name of the typical color of an object elicited activity near a region in the occipital cortex which is activated during color perception (Martin et al. 1995). Similarly, the word ‘kick’ causes the activation of the motor representation of the leg (Hauk et al. 2004) and saying ‘hammer’ to a picture of a hammer activates information about how to manipulate the object (Chao and Martin 2000). Other examples include activation of motor-processing regions by reading about motion (Deen and McCarthy 2010; Saygin et al. 2010) and activating somatosensory cortex by viewing pictures of graspable objects (Smith and Goodale 2015)[2].

The idea that this constitutes evidence for the grounded view that different phenomena are produced when cognitive tasks involve different modalities suggests that experimenters are following the reasoning articulated by MP. Crucially, being able to split phenomena in this way can help to move forward in their characterization. By distinguishing between these modality-specific capacities, we may be able to build a more accurate profile for each of them, by discovering other relevant aspects in which they may differ in addition to their neural basis, such as differences related to reliability, degradation, habituation, learning, inhibition and interference conditions, etc.

Alternatively, it is also possible that the observed variations in performance do not indicate that different phenomena are produced in different contexts, but rather that the manifestation of the same phenomenon is modulated by the modification of its precipitation conditions. According to MP, this would be the case if the same set of neural CAOs is recruited in those different contexts. If this were the case, affirming that contextual variations indicate that different phenomena are produced in different contexts would constitute a splitting error. We saw that a mechanistic phenomenon is multifaceted and can include variations in its input-output shape (or pairs of

---

[2] See Wajnerman Paz (2018) for a discussion of a more fine-grained (neural-coding level) approach to the mechanistic basis of grounded cognition.

precipitating-manifestation conditions) that depend on contextual factors (e.g., how the phenomenon occurs in artificial laboratory conditions vs. more naturalistic conditions). Therefore, it is possible that contextual variations in a putative phenomenon are not due to the fact that different CAOs are operative in different contexts but rather the very same set of CAOs works differently in different contexts, modulated by contextual factors.

Within the described debate, this line of reasoning is intended to capture the way in which experimenters opposing the grounded view interpret neural-level evidence. For instance, Piazza and colleagues used fMRI to compare brain responses to a numerosity estimation task and to an exact counting task, using visual and auditory stimuli. They first segregated the functional structures involved in estimation and counting and then showed that they could perform these tasks using both visual and auditory stimuli. The same neural structure partially located in the right and left intraparietal sulcus is involved in estimating numerosity independently of the stimuli's perceptual modality (Piazza et al., 2006). In this case, even if we find that differences in stimulus modality modify the manifestation conditions in the way specified by Pecher et al., (2003) (e.g., greater response delay when we switch stimulus modality), these different pairs of precipitating-manifestation conditions may not constitute different phenomena (as grounded theorists claim) but rather contextual modulations of a single phenomenon produced by a single set of neural CAOs. The phenomenon would be context-sensitive not because it is only triggered by particular stimulus features (e.g., only responds to quantities when they are visually encoded), but rather because variations in those features could change how it is manifested.

These considerations entail that to determine which forms of context-sensitivity are taking place, we need inter-level experiments. We need top-down interventions to assess whether behavioral-level variations result from variations in the constitutively relevant neural CAOs. If different CAOs are recruited in different contexts, then the phenomena produced in each context will also differ. Alternatively, if the same CAOs are recruited in different contexts, then different manifestations of the same phenomenon would be produced in those contexts. As we mentioned in Sect. 1, this proposal could be challenged by the existence of multifunctional mechanisms. These are mechanisms that can produce different phenomena in different contexts. If there are such mechanisms, then the fact that the same mechanism is recruited in different contexts does not entail that the same phenomenon is produced in each of them. We will discuss this objection in Sect. 4.

We can apply our line of reasoning to identifying the *manifestation* conditions through the variation of the response modality by which participants deliver their responses during an experiment. In this case, we must also first identify contextual variations in the phenomenon by designing an intra-level behavioral experiment in which we change the response modality. For instance, Frederick et al. compared, among other variables, reaction time during an olfactory discrimination task performed under two different response modalities (Go/No-Go: GNG; Two-Alternative Choice: TAC) (Frederick et al., 2011). They found that the tasks differ in response times, with GNG showing shorter periods than TAC.

As in the previous case, it has been hypothesized that this difference is due to implementing different cognitive functions. Specifically, each response modality would require a different cognitive strategy for optimizing reward rate through speed-accuracy trade-off (Rinberg et al., 2006). According to our criterion, behavior-level variations indicate that different cognitive functions are recruited only if different CAOs can be found at a neural level. In this case, the phenomenon would be context-sensitive because it would only be triggered when a specific response modality is required for a given olfactory task.

However, it is also possible that contextual differences are not due to the recruitment of different mechanistic CAOs but rather the same CAOs are being modulated by different task contextual factors. As we saw, assessing which of these two possibilities is the case requires inter-level top-down experiments for monitoring the actual neural structures activated during each version of the task. If top-down experiments determine that a single set of mechanistic CAOs implements these different strategies, then MP entails that these contextual variations related to response modality will constitute different aspects of a single multifaceted phenomenon.

As in the previous stimulus-type case, experimenters follow this line of reasoning. For instance, Frederick et al. also explored, through top-down activation experiments, the neural basis of odor discrimination in GNG and TAC response modalities (Frederick et al., 2016). Following the first hypothesis mentioned above (i.e., the recruitment of different cognitive functions for different response modalities) it has been suggested that the cognitive strategy underlying the task under TAC modality depends on local field potential (LFP) gamma ($40 - 110$ Hz) oscillations of a local network within the olfactory bulb, whereas the GNG modality would depend more strongly on beta oscillations (15–35 Hz) in a system-wide network (Beshel et al., 2007; Martin et al., 2007). If this were the case, then the behavioral differences described above would constitute manifestation conditions of different phenomena (each condition being produced by a different kind of neural activity). However, Frederick and colleagues showed that beta and gamma LFP oscillations occur in both modalities, with gamma dominating the early odor sampling period (the first 2 to 4 inhalations) and beta dominating later. Thus, gamma followed by beta oscillations represents a sequence of neurocognitive states during odor discrimination independent of response type. Thus, following the reasoning articulated by MP, they concluded that the behavioral differences associated with different response types constitute different manifestation conditions of *the same* phenomenon produced by the same mechanistic activities.

In this section, we have articulated three related theses. Firstly, intra-level behavioral interventions are necessary for identifying potential behavior-level contextual factors that can modulate the manifestation conditions of a putative phenomenon. Secondly, these contextual variations are consistent with two different forms of context sensitivity. It could be the case that a phenomenon can only be produced under specific task conditions or, alternatively, these conditions change how the same phenomenon is manifested in different situations. Finally, in determining which of these two situations obtains, we must rely on top-down inter-level interventions, monitoring potential variations in the mechanism CAOs underlying a putative phenomenon that result from the manipulation of task-context features.

4 The multifunctional mechanism challenge

We anticipated in previous sections that multifunctional mechanisms may pose a relevant challenge to our proposal. If the same mechanism can implement different psychological functions in different contexts, then the fact that we observed no context-to-context variation in a mechanism through the top-down manipulation of task features does not entail that the function or phenomenon to be explained is the same across contexts[3].

Recent discussions on structure–function mapping and mechanistic decomposition stressed the fact that, according to current research, brain areas seem to be multifunctional. For instance, Anderson showed that cortical areas are redeployed across nine cognitive domains (vision, memory, numeric cognition, etc.) (Anderson, 2010). McCaffrey describes three different strategies that have been proposed to explain (or explain away) multifunctional mechanisms (McCaffrey, 2015). The "Subdivide and Conquer" strategy consists in showing that what seems to be a single multifunctional mechanism is constituted by different (although perhaps anatomically or functionally overlapping) mechanisms implementing different functions. According to a second "Cognitive Ontology Revision" strategy suggested (Price & Friston, 2005), while at one level of description a mechanism may seem to perform many functions, it can be seen in a more abstract level of description as performing a single function. Finally, the "Networks and Context-Sensitive Mappings" strategy proposed by Klein (2012) is that brain regions perform different functions depending on the networks in which they participate, which constitute their "neural context".

We agree with the "Context-Sensitive Mappings" and the "Subdivide and Conquer" strategies that genuinely different functions can be performed in slightly different contexts either by the same mechanism or by highly overlapping mechanisms. However, we argue that these different functions are implemented in different contexts *by different CAOs, which may or may not belong to the same mechanism*. As we saw, according to MP, it is sufficient to split a phenomenon that different CAOs are recruited in different contexts. However, it is possible that these different sets of CAOs *are part of a single multifunctional mechanism*. Thus, our criterion is consistent with multifunctionality and is neutral regarding the "Context-Sensitive Mappings" and the "Subdivide and Conquer" strategies. In other words, according to MP, the individuation of phenomena is tied to the individuation of mechanistic CAOs *but not to the individuation of mechanisms themselves*. These two can be told apart under a plausible view of mechanism individuation. We can see this by considering two main kinds of multifunctional mechanisms.

If the brain is not modular and the mechanisms supporting different capacities are not largely segregated, then there are two possible ways in which multifunctionality may appear. Following Anderson (2010), we will refer to them as "reuse" and "holism". According to the reuse hypothesis, all or most neural components at different levels of organization are used to support different tasks or implement different capacities. However, for each task or capacity, there are different cooperation patterns involving *different connections and/or different sets of components*. Thus, the reuse

---

[3] We thank an anonymous reviewer for this observation.

hypothesis has an implication that is critical to MP: *components will typically not respond only to a specific kind of task*. Therefore, if we detect through a top-down experiment that a component or set of components are activated in different task contexts, we cannot affirm on that basis alone that the same phenomenon is produced in each context. However, the reuse hypothesis also entails that the differences detected in the phenomenon will result from *slight differences in the set of components that are recruited and their specific pattern of interaction*. When these differences occur in different contexts, we would say that different phenomena were produced in those contexts.

A key observation is that these differences do not entail that *different mechanisms* have been recruited in those contexts. As Levy & Bechtel (2016) among others have suggested, mechanisms can have different CAOs in different moments. That is, the identity of a mechanism does not depend on the identity of its CAOs through contextual variations. However, these modifications could explain why the same mechanism produced different functions in those contexts.

A different form of multifunctionality could be found if the brain is more holistically organized and the very same set of components that constitute a mechanism are involved in different tasks. These cases are more problematic for our proposal because the deployment of different capacities in different contexts may not be reflected by differences in the (largely overlapping) set of components and/or their connections. If components can perform different functions in different contexts, then the same set of active and organized components may produce different phenomena. Taking an example described by McCafrrey (2015), the dorsal striatum is involved both in reward learning and voluntary movement, but it contributes to these capacities through different functions, namely, temporal difference detection in reward learning and disinhibition or gating mechanism in voluntary movement (Liljeholm & O'Doherty, 2012; Suri & Schultz, 2001). Of course, given that there is no difference without a difference-maker, behavioral-level differences can be explained through differences in the way that these components perform those different functions. We posit that these differences may be enough to split the relevant phenomena.

This can be illustrated by taking another interesting example provided by McCraffey (2015) that includes a description of the underlying mechanism(s): The exact same neural populations can implement different coding schemes for encoding different kinds of stimuli. For instance, Leutgeub et al. argue that a single population of hippocampal neurons has different coding schemes for information related to spatial navigation and episodic memory (Leutgeb et al., 2005). While what subset of the population is firing signifies the rat's spatial location, the population's rate function reflects the presence of certain environmental features regardless of location.

In this case, the same set of components (that is, the same neural structure) implements different functions. However, the difference in the phenomena produced can be explained by a modification in the *kind of activities* that are performed by those components. As in the previous case, differences in the kinds of activities recruited in different contexts may not entail that different mechanisms are being deployed. Thus, this situation is also consistent with a multifunctional mechanism. According to our criterion, the recruitment of different kinds of activities in different contexts is sufficient for splitting a phenomenon even if these activities belong to the same mechanism.

A critical issue related to how our proposal can accommodate multifunctional mechanisms is whether and how different kinds of CAOs recruited in different contexts can belong to a single mechanism. After all, CAOs plausibly play some role in a mechanism individuation[4].

We agree that preserving *some* CAOs is necessary for individuating a mechanism. This is more so for multifunctional mechanisms, given that (according to our MP criterion) the different CAOs will be associated with *different* phenomena. Therefore, phenomena could not be used for linking the different CAOs to a single mechanism.

Although we will not address this issue in this paper, we bet that context-to-context preservation of *components or structures* can plausibly be a sufficient condition for mechanism identity. For example, suppose the same components produce different phenomena by acquiring a different organization (e.g., changing their pattern of causal connections) or by performing different operations (e.g., changing the frequency of oscillations). In that case, the different organizational properties and/or operations could be seen as belonging to a single multifunctional mechanism.

We also take this to be close to a necessary condition. Of course, mechanisms can acquire and lose components. However, if different and mostly non-overlapping sets of components are systematically activated in different contexts, they will plausibly be regarded as different mechanisms.

5 Neural coding and identity preserving variations

We argued in the previous section that alleged cases of multifunctional mechanisms are not problematic for our criterion because we only need to assess whether its CAOs were modified in order to split a phenomenon, independently of whether these modifications split the mechanism as well or not. However, even if we can be neutral regarding the individuation of mechanisms, we do need to say something about the individuation of mechanistic CAOs. It is reasonable to assume that the physiological activity of neural components at different levels will almost always vary in different task contexts and even between trials (Bale & Petersen, 2009; Beshel et al., 2007; Martin et al., 2007; Murphy et al., 2014; Valero-Cuevas et al., 2009; Vogel et al., 2005), and this does not necessarily mean that the component is performing different kinds of activities or operations. A very simple example is a neuron *n* increasing its firing rate in response to a stimulus *s*: the instantiation of *n*'s increasing its firing rate is consistent with many different firing patterns that *n* can have in different trials in which *s* is presented.

Thus, not all variations in physiological activity should count as variations in mechanistic CAOs. This is crucial to our proposal because although we argued that if no variation in the mechanism is detected through a top-down intervention, we should not split the phenomenon, we don't necessarily have to split it when a varia-

---

[4] We thank an anonymous reviewer for this observation.

tion is found. Thus, a key question is what contextual variations in CAOs are consistent with their identity across different contexts.

As the firing-rate example illustrates, identity-preserving physiological variations are determined by the *kind or type* of activity, component, or organizational property that is relevant to produce the target phenomenon (e.g. spike rate vs. spike time). In some cases, the type is broad enough to allow very significant variations. In other cases, very subtle modifications of neural activity entail that a different type has been recruited by a given mechanism. Thus it is critical to understand how kinds are defined in different frameworks or models in order to avoid splitting or lumping errors regarding a given phenomenon.

In this section, we will illustrate this by focusing on a type of activity, namely neural coding, whose identity is both especially flexible regarding how it can be realized by a set of components and, at the same time, specially sensitive to subtle modifications. Both the ontological flexibility and fragility of neural codes derive from the fact that they are partly constituted by rules, understood as input-output mappings. Implementing a rule is a particularly flexible kind of operation because radically different components and activities can be part of the same input-output mapping that defines a given rule. At the same time, it can be especially fragile because slightly different input-output mappings constitute different rules. For instance, the active components of a circuit that implements an OR logic gate can be completely different in different contexts insofar as they are consistent with the rule defining this operation. In turn, an OR gate becomes a different XOR gate through a minimal change of its defining rule (i.e., if two 1 inputs produce a 0 output). This general idea applies to transition functions in dynamical models, signal processing operations in computational models, coding schemes in neural coding models, and so on.

Let’s see how this works in neural codes. A common approach to neural coding is population analysis, which aims to determine coding regimes by discovering patterns in the combined activity of different neurons. A pattern of spike trains produced by a neural population is often recorded by using electrode arrays or (more recently) fluorescence microscopy and interpreted with decoding algorithms or information theory. Once its informational properties are understood, the pattern is systematically explored to determine which features of the spike trains carry the relevant information, i.e., which code the population implements (Quian Quiroga & Panzeri, 2009; Quiroga & Panzeri, 2013).

There are two aspects of population activity that are often used to characterize neural codes. The first one is density. The density of a code is determined by the average fraction of neurons of a given population that are (more or less) simultaneously triggered to represent a given stimulus. Density can vary from close to 0 to about 1/2. When density is higher than 1/2 it can be decreased without loss of information by replacing each active neuron with an inactive one, and vice versa. Codes with lowest density are local codes, in which each condition is represented by only one active neuron. The highest density is given by holographic coding. These codes represent each stimulus by the combination of activities of ½ neurons of a population. Sparse codes are a compromise between dense and local codes. Under this regime, multiple (but few) neurons participate in the representation of each condition (Foldiak & Endres, 2008).

A second aspect is the distinction between selective and distributed codes. In distributed codes, different subsets of cells in a population encode different conditions, but, crucially, these subsets share cells. In contrast, in selective coding, a population represents different conditions by using non-overlapping subsets. This implies that single cells are interpretable only if the code is selective. Under this coding regime, each unit or neuron responds highly selectively to a given condition (e.g., specific faces, objects, or words) such that the output of that unit can be interpreted unambiguously. Units of this sort are sometimes called localist units or "grandmother cells". On the other hand, in distributed representations, given that each unit responds to a wide range of different conditions, individual units cannot be interpreted without knowing the state of other units in the network (Bowers et al., 2016; Page, 2000; Thorpe, 1995).

The dense/sparse and the distributed/selective distinctions are related. A distributed code requires some degree of density (i.e., more than one cell has to respond to each stimulus) in order to form overlapping sets of neurons. Also, populations in visual areas are known to be both sparse and selective (Lehky et al., 2005, and see Sect. 5). However, these notions are relatively independent. A given population can implement a sparse and distributed code if overlapping subsets of cells fire in response to different stimuli, and each of those subsets are constituted by very few cells. In turn, a selective and dense representation could be implemented by a population in which many redundant neurons respond selectively to the same input (Bowers et al., 2016).

If a given population implements a high-density distributed code in which many overlapping cells are active for encoding different stimuli in different trials, we may be inclined to accept that, despite the slight differences in physiological activity of the population in those trials, it is responding in the same way or performing the same operation in different trials. However, the same applies to the case of a selective code, in which differences in responses to different stimuli are starker, given that *different and non-overlapping sets of neural components will be recruited for encoding different stimuli in different trials*. The operation is the same because we cannot understand what the population is doing, i.e., which coding regime it is applying unless we understand the rule that defines the scheme, and this rule is determined by the mapping between each component of a given stimulus set (the set of stimuli to which the population responds) and its corresponding population response. Thus, we would say that the population is recruiting the same operations in different trials despite radical differences in the set of active neural components.

By contrast, given that neural codes are not only defined by a coding scheme but also by a stimulus set, when the same coding scheme is used to codify completely different kinds of stimuli, even if physiological activities are the same in different contexts, we can say that the target population or mechanism is doing something different, that a different operation is being performed and therefore a different function is implemented. For instance, a relevant model of neural coding in the prefrontal cortex affirms that prefrontal populations implement what has been called 'adaptive coding', meaning that their response properties are highly adaptable to very diverse kinds of stimuli that are relevant to different tasks (e.g., Duncan 2001; Stokes et al., 2013; Woolgar et al., 2011). In a classic study, Freedman, Riesenhuber, Poggio, and Miller (2001) trained monkeys in a cat–dog categorization task. After training, the responses of many prefrontal neurons differentiated between cats and dogs, even those close together across the category boundary. Then they trained one monkey in a new task that was based on the same stimulus set but in which the cat–dog distinction was irrelevant (the animal had to sort stimuli into three new categories). After training on this new task, cat–dog information was lost from neural responses. Instead, these respected the new category distinctions relevant to the new task. Even if the intrinsic properties of the target PFC population that define its coding scheme are the same in both contexts, the code itself is different because the population codifies different kinds of properties. In these cases, we would say that the population is performing different operations in different contexts, thus (according to MP) implementing different functions.

Crucially, a recent key case of alleged multifunctionality can be analyzed in the same way. Burnston (2020) offers an interesting set of examples of components that seem to have, in terms of McCaffrey (2015), "variable roles". These are cases in which neural activity is 'multiplexed'. Multiplexing is a coding regime implemented by the relationship between the action potentials of individual cells and the background 'local field potential' (LFP). In the example provided by Burnston, a single cell or group of cells exhibits the responses to 'light' and 'dark' stimuli, but these are only encoded through a combination of these responses with their modulation by properties of the background LFP. Following Watrous et al., (2015), Burnston describes three possible coding schemes. The rate-based signals can be modulated either by frequency, phase and/or amplitude. We can briefly describe the first two. In the first case, the target unit only responds to dark stimuli when a low-frequency LFP is present, and to light stimuli only in the presence of a higher frequency. In the phase modulation scheme, signals respond to both stimuli in both frequencies, but at distinct phases of each.

If the same circuit changes its coding scheme (say, from a phase coding to a frequency coding) then what the circuit is doing is different because changes in the causal relation between stimuli, spiking rate and LFP properties reflect a change in the rule that the circuit is employing for processing the input signal. Also, if the circuit uses the same coding scheme to encode very different kinds of information, we will also say that it performs different operations. Burnston (2020) argues that this is indeed the case because LFP modulation "changes what type of function a population or brain area is contributing to […] by changing what kind of information is being processed" (pp. 16,17). That is, when a cognitive mechanism codifies and processes different kinds of information, it will be implementing different functions despite employing the same coding scheme.

Thus, the assessment of which mechanistic variations entail a variation in the phenomenon needs to be done on a case-by-case basis, taking into account the types of activities, components, and organizational properties articulated by particular frameworks or models. As the neural coding case illustrates, when the types are highly flexible, the same phenomenon will be consistent with very different patterns of neural activity. In turn, when the types are especially fragile, the modification of the phenomenon can result from very subtle modifications of neural activity.

6 Conclusion

We have explored how to interpret and experimentally approach the contextual modulation of behavior in cognitive experiments. Among the different questions related to context-sensitivity, we focused on one key issue: The meaning of variations in behavior that result from contextual differences in task features, such as stimulus and response modality. We argued that there are at least two possible interpretations of these variations: They could indicate the presence of different phenomena, or they could signal different operation regimes for the same phenomenon. In the first case, we find a context-sensitive phenomenon in the sense that it is only triggered when specific contextual factors are present, such as a specific stimulus type. Thus, changing the context entails changing the cognitive process being studied. In the second case, we find a context-sensitive phenomenon in the sense that contextual variations change how the same phenomenon is manifested in different situations.

Additionally, we argued that determining which of these two possibilities obtains requires combining intra-level and inter-level experiments. Intra-level behavioral interventions are necessary for assessing whether contextual factors affect the manifestation conditions of a phenomenon. Given that these contextual variations are consistent with the two described forms of context sensitivity, we need inter-level interventions. We claimed that if we endorse a mechanistic (in the sense of inter-level) approach to phenomena individuation, then we can assess which form of context-sensitivity a phenomenon has by examining whether the CAOs of its underlying mechanism change from context to context. If interventions on task parameters are not accompanied by significant variations in the organized activities of mechanistic components, then differences in manifestation conditions are not associated with different phenomena but rather a single phenomenon with different context-dependent operation regimes.

We then addressed the challenge that our proposal may be inconsistent with the form of context-sensitivity possessed by multifunctional mechanisms. The objection is that if the same mechanism can implement different functions in different contexts, then the context-to-context variation of the phenomenon can occur with no significant variation at a mechanistic level. We argued that our proposal is consistent with this possibility. If we follow Levy & Bechtel (2016) in the idea that the identity of a mechanism does not depend on particular CAOs. According to our criterion, in order to split a phenomenon, we only need to assess whether the kinds of CAOs that a mechanism recruits in different contexts were modified and these differences are consistent with the possibility that the mechanism is the same. Finally, we emphasized that the individuation of components, activities, and organizational properties will be tied to the kinds or types articulated in different frameworks or models and exemplified this through the case of neural coding.

We think that mechanistic approaches to context-sensitivity of the kind we have explored are also useful to tackle other relevant issues in contemporary neuroscience. These include problems such as addressing inter- and intra-subject variability (Goodhew & Edwards, 2019; Seghier & Price, 2018), characterizing the sources of replicability issues (Gelman, 2015; Van Bavel et al., 2016), addressing the challenges of the ecological validity of experimental results (Adolph, 2020; Miller et al., 2019;

Shamay-Tsoory & Mendelsohn, 2019), and, in general, understanding the situatedness of the cognitive processes under study. More work is needed to understand how our proposal, and similar discussions in the philosophy of neuroscience, can contribute to clarifying these issues and distill implications for empirical neuroscience.

Acknowledgements This work was supported by Fondo Nacional de Desarrollo Científico y Tecnológico Fondecyt, Proyecto 11190604, awarded to DRL. AWP receives funding from Fondo Nacional de Desarrollo Científico y Tecnológico Fondecyt, Proyectos 11220327 (IR: Abel Wajnerman Paz), 1210091 (IR: Juan Manuel Garrido) and 1200197 (IR: Francisco Pereira). AWP thanks members of the Fondecyt project 1210091(specially Juan Manuel Garrido, José Tomás Alvarado, Jorge Fuentes Muñoz) for their generous and ongoing feedback on different versions of this manuscript. Also, a very special thanks to the reviewers from Synthese for a really insightful and constructive discussion that was critical for arriving at the final version of the manuscript. Finally, he would like to thank DRL for a very rewarding and enjoyable co-authoring process. DRL would like to thank Francisco Parada, Alejandra Rossi, and Christian Salas for enriching and insightful discussions and for making the CENHN an inspiring place to do neuroscience. He also warmly thanks to AWP for his enthusiasm and dedication to a delightful co-thinking and co-writing process. Finally, DRL also thanks the anonymous reviewers for their suggestions and comments.

Author contributions AWP and DRL contributed to the study's conception and development. AWP and DRL wrote the first draft of the manuscript and revised all subsequent versions. AWP and DRL read and approved the final manuscript.

Declarations

Statements and declarations The authors have no competing interests to declare relevant to this article's content.

References

Adolph, K. E. (2020). Ecological validity: Mistaking the lab for real life. *My Biggest Research Mistake: Adventures and Misadventures in Psychological Research* (pp. 187–190). SAGE Publications, Inc. 2455 Teller Road, Thousand Oaks California 91320

Al, E., Iliopoulos, F., Forschack, N., Nierhaus, T., Grund, M., Motyka, P., et al. (2020). Heart-brain interactions shape somatosensory perception and evoked potentials. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(19), 10575–10584

Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *The Behavioral and brain sciences*, *33*(4), 245–66; discussion 266–313

Ashworth, E. T., Cotter, J. D., & Kilding, A. E. (2021). Impact of elevated core temperature on cognition in hot environments within a military context. *European journal of applied physiology*, *121*(4), 1061–1071

Azzalini, D., Rebollo, I., & Tallon-Baudry, C. (2019). Visceral Signals Shape Brain Dynamics and Cognition. *Trends in cognitive sciences*, *23*(6), 488–509

Bale, M. R., & Petersen, R. S. (2009). Transformation in the neural code for whisker deflection direction along the lemniscal pathway. *Journal of neurophysiology*, *102*(5), 2771–2780

Barsalou, L. W. (2016). Situated conceptualization: Theory and applications. In Y. Coello (Ed.), *Foundations of embodied cognition: Perceptual and emotional embodiment, (pp* (Vol. 290, pp. 11–37). New York, NY, US: Routledge/Taylor & Francis Group, viii

Barsalou, L. W., Simmons, K., Barbey, W., A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in cognitive sciences*, *7*(2), 84–91

Bechtel, W., & Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in history and philosophy of biological and biomedical sciences*, *36*(2), 421–441

Beshel, J., Kopell, N., & Kay, L. M. (2007). Olfactory bulb gamma oscillations are enhanced with task demands. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *27*(31), 8358–8365

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychonomic bulletin & review*, *23*(4), 1096–1108

Boone, W., & Piccinini, G. (2016). The cognitive neuroscience revolution. *Synthese*, *193*(5), 1509–1534

Bowers, J. S., Vankov, I. I., Damian, M. F., & Davis, C. J. (2016). Why do some neurons in cortex respond to information in a selective manner? Insights from artificial neural networks. *Cognition*, *148*, 47–63

Burnham, J. C. (1972). Thorndike's puzzle boxes. *Journal of the history of the behavioral sciences*, *8*(2), 159–167

Burnston, D. C. (2020). Getting over atomism: Functional decomposition in complex neural systems.The British journal for the philosophy of science,000–000

Chance, P. (1999). Thorndike's Puzzle Boxes And The Origins Of The Experimental Analysis Of Behavior. *Journal of the experimental analysis of behavior*, *72*(3), 433–440

Correa, C. M. C., Noorman, S., Jiang, J., Palminteri, S., Cohen, M. X., Lebreton, M., & van Gaal, S. (2018). How the Level of Reward Awareness Changes the Computational and Electrophysiological Signatures of Reinforcement Learning. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *38*(48), 10338–10348

Craver, C. F. (2007). *Explaining the Brain*. Oxford University Press. Accessed 2 October 2021

Craver, C. F. (2014). The Ontic Account of Scientific Explanation. In M. I. Kaiser, O. R. Scholz, D. Plenge, & A. Hüttemann (Eds.), *Explanation in the Special Sciences: The Case of Biology and History* (pp. 27–52). Dordrecht: Springer Netherlands

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*. https://doi.org/10.1016/0001-6918(69)90065-1

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature reviews Neuroscience*, *2*(11), 820–829

Evarts, E. V. (1966). Methods for recording activity of individual neurons in moving animals. *Methods in medical research* (pp. 241–250). Chicago, IL: Year Book Medical Publishers

Ferster, C. B. (1953). The use of the free operant in the analysis of behavior. *Psychological bulletin*, *50*(4), 263

Foldiak, P., & Endres, D. (2008). Sparse coding. *Scholarpedia journal*, *3*(1), 2984

Frederick, D. E., Brown, A., Brim, E., Mehta, N., Vujovic, M., & Kay, L. M. (2016). Gamma and Beta Oscillations Define a Sequence of Neurocognitive Modes Present in Odor Processing. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *36*(29), 7750–7767

Frederick, D. E., Rojas-Líbano, D., Scott, M., & Kay, L. M. (2011). Rat behavior in go/no-go and two-alternative choice odor discrimination: differences and similarities. *Behavioral neuroscience*, *125*(4), 588–603

Freiberger, J. J., Derrick, B. J., Natoli, M. J., Akushevich, I., Schinazi, E. A., Parker, C., et al. (2016). Assessment of the interaction of hyperbaric N2, CO2, and O2 on psychomotor performance in divers. *Journal of applied physiology*, *121*(4), 953–964

Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. *Journal of experimental psychology*, *46*(5), 373–380

Gelman, A. (2015). The Connection Between Varying Treatment Effects and the Crisis of Unreplicable Research: A Bayesian Perspective. *Journal of management*, *41*(2), 632–643

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and cognition*, *69*, 14–25

Hagner, M. (2012). The electrical excitability of the brain: toward the emergence of an experiment. *Journal of the history of the neurosciences*, *21*(3), 237–249

Heron, W. T., & Skinner, B. F. (1939). An apparatus for the study of animal behavior. *The Psychological record*, *3*, 166–176

Jasper, H., Ricci, G. F., & Doane, B. (1958). Patterns of cortical neuronal discharge during conditioned responses in monkeys. In G. E. W. Wolstenholme, & C. O'Connor (Eds.), *Neurological basis of behaviour* (pp. 277–294). Boston, MA: Little, Brown and Company

Klein, C. (2012). Cognitive Ontology and Region- versus Network-Oriented Analyses. *Philosophy of science*, *79*(5), 952–960

Lehky, S. R., Sejnowski, T. J., & Desimone, R. (2005). Selectivity and sparseness in the responses of striate complex cells. *Vision research*, *45*(1), 57–73

Leutgeb, S., Leutgeb, J. K., Barnes, C. A., Moser, E. I., McNaughton, B. L., & Moser, M. B. (2005). Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science*, *309*(5734), 619–623

Liljeholm, M., & O'Doherty, J. P. (2012). Contributions of the striatum to learning, motivation, and performance: an associative account. *Trends in cognitive sciences*, *16*(9), 467–475

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of science*, *67*(1), 1–25

Marsh, A. A., Rhoads, S. A., & Ryan, R. M. (2019). A multi-semester classroom demonstration yields evidence in support of the facial feedback effect. *Emotion*, *19*(8), 1500–1504

Martin, C., Beshel, J., & Kay, L. M. (2007). An olfacto-hippocampal network is dynamically involved in odor-discrimination learning. *Journal of neurophysiology*, *98*(4), 2196–2205

McCaffrey, J. B. (2015). The Brain's Heterogeneous Functional Landscape. *Philosophy of science*, *82*(5), 1010–1022

Miller, L. C., Shaikh, S. J., Jeong, D. C., Wang, L., Gillig, T. K., Godoy, C. G., et al. (2019). Causal Inference in Generalizable Environments: Systematic Representative Design. *Psychological inquiry*, *30*(4), 173–202

Mountcastle, V. B. (1978). Brain mechanisms for directed attention. *Journal of the Royal Society of Medicine*, *71*(1), 14–28

Müller-Pinzler, L., Gazzola, V., Keysers, C., Sommer, J., Jansen, A., Frässle, S., et al. (2015). Neural pathways of embarrassment and their modulation by social anxiety. *Neuroimage*, *119*, 252–261

Murphy, P. R., Vandekerckhove, J., & Nieuwenhuis, S. (2014). Pupil-linked arousal determines variability in perceptual decision making.PLoS computational biology, 10(9), e1003854

Noah, T., Schul, Y., & Mayo, R. (2018). When both the original study and its failed replication are correct: Feeling observed eliminates the facial-feedback effect. *Journal of personality and social psychology*, *114*(5), 657–664

Page, M. (2000). Connectionist modelling in psychology: a localist manifesto. *The Behavioral and brain sciences*, *23*(4), 443–67; discussion 467–512

Park, H. D., Correia, S., Ducorps, A., & Tallon-Baudry, C. (2014). Spontaneous fluctuations in neural responses to heartbeats predict visual detection. *Nature neuroscience*, *17*(4), 612–618

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological science*, *14*(2), 119–124

Perl, O., Ravia, A., Rubinson, M., Eisen, A., Soroka, T., Mor, N., et al. (2019). Human non-olfactory cognition phase-locked with inhalation. *Nature human behaviour*, *3*(5), 501–512

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*(6157), 585–589

Piazza, M., Mechelli, A., Price, C. J., & Butterworth, B. (2006). Exact and approximate judgements of visual and auditory numerosity: an fMRI study. *Brain research*, *1106*(1), 177–188

Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese*, *183*(3), 283–311

Pollack, I. (1952). The Information of Elementary Auditory Displays. *The Journal of the Acoustical Society of America*, *24*(6), 745–749

Posner, M. I., Petersen, S. E., Fox, P. T., & Raichle, M. E. (1988). Localization of cognitive operations in the human brain. *Science*, *240*(4859), 1627–1631. Accessed 21 July 2021

Price, C. J., & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive neuropsychology*, *22*(3), 262–275

Quian Quiroga, R., & Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nature reviews Neuroscience*, *10*(3), 173–185

Quiroga, R. Q., & Panzeri, S. (2013). *Principles of Neural Coding*. CRC Press

Reilly, J., Peelle, J. E., Garcia, A., & Crutch, S. J. (2016). Linking somatic and symbolic representation in semantic memory: the dynamic multilevel reactivation framework. *Psychonomic bulletin & review*, *23*(4), 1002–1014

Rinberg, D., Koulakov, A., & Gelperin, A. (2006). Speed-accuracy tradeoff in olfaction. *Neuron*, *51*(3), 351–358

Seghier, M. L., & Price, C. J. (2018). Interpreting and Utilising Intersubject Variability in Brain Function. *Trends in cognitive sciences*, *22*(6), 517–530

Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-Life Neuroscience: An Ecological Approach to Brain and Behavior Research. *Perspectives on psychological science: a journal of the Association for Psychological Science*, *14*(5), 841–859

Skinner, B. F. (1930). On the conditions of elicitation of certain eating reflexes. *Proceedings of the National Academy of Sciences of the United States of America*, *16*(6), 433–438

Steinborn, M. B., & Huestegge, L. (2020). Socially alerted cognition evoked by a confederate's mere presence: analysis of reaction-time distributions and delta plots. *Psychological research*, *84*(5), 1424–1439

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*(3736), 652–654

Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78*(2), 364–375

Strack, F. (2016, November). Reflection on the Smiling Registered Replication Report. *Perspectives on psychological science: a journal of the Association for Psychological Science*

Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: a nonobtrusive test of the facial feedback hypothesis. *Journal of personality and social psychology*, *54*(5), 768–777

Sullivan, J. A. (2008). The multiplicity of experimental protocols: a challenge to reductionist and non-reductionist models of the unity of neuroscience. *Synthese*, *167*(3), 511

Sullivan, J. A. (2010). Reconsidering "spatial memory" and the Morris water maze. *Synthese*, *177*(2), 261–283

Suri, R. E., & Schultz, W. (2001). Temporal difference model reproduces anticipatory neural activity. *Neural computation*, *13*(4), 841–862

Swets, J. A. (1973). The Relative Operating Characteristic in Psychology: A technique for isolating effects of response bias finds wide use in the study of perception and cognition. *Science*, *182*(4116), 990–1000

Thorpe, S. (1995). Visual recognition: From retina to inferotemporal cortex. *Revue de Neuropsychologie*, *5*(4), 389–410

Tian, X., Fang, Z., & Liu, W. (2021). Decreased humidity improves cognitive performance at extreme high indoor temperature. *Indoor air*, *31*(3), 608–627

Valero-Cuevas, F. J., Venkadesan, M., & Todorov, E. (2009). Structured variability of muscle activations supports the minimal intervention principle of motor control. *Journal of neurophysiology*, *102*(1), 59–68

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(23), 6454–6459

Vogel, A., Hennig, R. M., & Ronacher, B. (2005). Increase of neuronal response variability at higher processing levels as revealed by simultaneous recordings. *Journal of neurophysiology*, *93*(6), 3548–3559

Wagenmakers, E. J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams, R. B., et al. (2016). Registered Replication Report: Strack, Martin, & Stepper (1988). *Perspectives on psychological science: a journal of the Association for Psychological Science*, *11*(6), 917–928

Wahn, B., Rohe, T., Gearhart, A., Kingstone, A., & Sinnett, S. (2020). Performing a task jointly enhances the sound-induced flash illusion. *Quarterly journal of experimental psychology*, *73*(12), 2260–2271

Wajnerman Paz, A. (2018). An efficient coding approach to the debate on grounded cognition. *Synthese*, *195*(12), 5245–5269

Watrous, A. J., Deuker, L., Fell, J., & Axmacher, N. (2015). Phase-amplitude coupling supports phase coding in human ECoG. *eLife*, *4*. https://doi.org/10.7554/eLife.07886

Woolgar, A., Hampshire, A., Thompson, R., & Duncan, J. (2011). Adaptive coding of task-relevant information in human frontoparietal cortex. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *31*(41), 14592–14599. Accessed 29 March 2022

Authors and Affiliations

Abel Wajnerman-Paz[1] · **Daniel Rojas-Líbano**[2,3]

✉ Daniel Rojas-Líbano
daniel.rojasli@mail.udp.cl

[1] Departamento de Filosofía, Universidad Alberto Hurtado, Santiago, Chile

[2] Centro de Neurociencia Humana y Neuropsicología, Facultad de Psicología, Universidad Diego Portales, Santiago, Chile

[3] Facultad de Psicología, Universidad Diego Portales, Vergara 275, Santiago, Chile