**A Mechanistic Guide to Reductive Physicalism**

Tudor M. Baetu

*(forthcoming in The European Journal for Philosophy of Science, DOI: 10.1007/s13194-022-00489-3)*

# Abstract

Causal mediation mechanisms are well supported by available experimental evidence and provide a practicable way to reductive physicalism. According to the causal mediation account of mechanistic explanation, descriptions as diverse as ‘black-box’ phenomena, mechanistic sketches and schemas mixing physically interpreted and operationalized biological, psychological and social variables, and detailed descriptions of mechanisms refer to the same causal structure circumscribed within the spatiotemporal boundaries of a replicable experimental setup. The coreference of coarser- and finer-grained descriptions of causal structures opens new possibilities for testing the reductive physicalism conjecture. I discuss experimental designs supporting the causal mediation account and show how recent studies demonstrating the biological mediation of mind-mind causal processes can provide evidence for reductive physicalism.

# 1. Introduction

For the most part, explanations in the life sciences, from molecular biology to neuroscience, reveal how the organization and interaction of the parts of a physical system are responsible for biological and psychological phenomena (Bechtel 2006; 2008; Craver 2007; Craver and Darden 2013; Darden 2006). Such systems of parts changing over time are commonly known as ‘mechanisms,’ while the more or less detailed descriptions of these systems and the series of changes they undergo are referred to as ‘mechanistic explanations’ (Bechtel and Abrahamsen 2005; Glennan 1996; 2002; 2017; Illari and Williamson 2012; Machamer et al. 2000).

In practice, scientists and philosophers often focus on the explanatory and epistemic aspects of mechanistic research, without worrying too much about metaphysical implications. From this perspective, one can talk about psychological mechanisms, composite mechanisms mixing neural circuits with box-and-arrow descriptions or even biopsychosocial mechanisms in the same way biologists talk about molecular mechanisms. Without questioning the epistemic and pragmatic value of multifactorial and multidisciplinary explanations, it is nevertheless reasonable to ask, “What is the physical or metaphysical interpretation of psychological variables appearing in these explanations? What is being measured when pain is reported on a numerical scale and what is being intervened upon when subjects are exposed to hypnotic suggestion?” Some researchers answer, “Psychological reality, naturally. Pain is measured by verbal reports in the same way the relative length of polynucleotides is measured by electrophoresis in agarose gels. Hypnosis manipulates a psychological thing called ‘expectation’ in the same way surgery manipulates the corpus callosum in epileptic individuals.” This is one possible answer. But there is another. It is possible that, among the variables measured and manipulated in an experiment and the experimentally demonstrated causal dependencies between these variables, some ultimately refer to the same or overlapping realities, being duplicated solely as a matter of differences in the techniques of measurement and intervention.

In this paper, I explore this second option, namely the extent to which currently available mechanistic explanations support reductive physicalism, broadly construed to encompass individual-specific (token-like) or taxon-generalizable (type-like) ontological mappings of psychological causation onto biological causation and of uninterpreted psychological variables onto physically interpreted biological variables.[[1]](#footnote-1) Two assumptions underlie the discussion conducted in this paper: the first is that the characterization of phenomena and the elucidation of mechanisms are empirical questions addressed by experimental science (Baetu 2019a; Bechtel and Richardson 2010; Craver 2007; Darden 2006); the second, that, to our best knowledge, mechanistic explanations describe reality (Craver and Darden 2013; Illari and Williamson 2011). If these two assumptions are granted, then mechanistic explanations can inform us whether certain ontological reductions are justified or not. For instance, if experimental evidence demonstrates that a neural mechanism causes a psychological phenomenon, where causality is construed as a relationship between ontologically distinct ‘cause’ and ‘effect’ items, then there are no obvious ways in which variables appearing in the description of the phenomenon could be ontologically mapped onto variables describing the mechanism, its parts, interactions, or organizational features. In contrast, if the relationship between mechanism and phenomenon takes the form of identity or some type of constitution, then a reduction of the psychological phenomenon to its underlying mechanism is warranted or at least possible.

As it turns out, the vast majority of scientific studies involved in the elucidation of mechanisms rely on correlation and causal relevance tests, which are not designed to and cannot generate conclusive evidence for identity, constitution, or any other type of non-causal relationships. Prima facie, this seems to indicate a strong bias of experimental science towards an antireductionist ontology. The view defended in this paper is that this conclusion is premature. An important subset of studies doesn’t aim to generate evidence for causal relevance simpliciter, but rather for causal mediation. While these studies don’t yield evidence for identity or constitution, they don’t test strictly etiological relationships either, favouring instead a ‘mediating mechanism’ type of explanation detailing intermediary causal steps via which an outcome occurs given an initial state. I argue that despite being framed in causal terms, mediation explanations support reductive physicalism. Since causal mediation experiments don’t test for identity or constitution relationships, ontological reduction cannot be driven by a direct mapping of variables onto other variables. Nevertheless, reductionism can proceed indirectly, via the discovery of biophysical mediators of psychosocial causal processes. Using an example from pain research, I show how evidence for mediation can justify collapsing some instances of psychological causation onto biophysical causation and how, under the assumption of parsimony, certain psychological variables can be ontologically mapped onto biophysical variables.

The paper is organized as follows: In Section 2, I discuss etiological explanations and their implications for physicalism. Section 3 covers the mechanistic constitution account, the mutual manipulability criterion, and the experimental and methodological challenges facing the latter. In section 4, I discuss scientific evidence in favour of a causal mediation account and the implications of this account. In Section 5, I explore the possibility of implementing reductive physicalism based on evidence for biological mediation of psychosocial causation. Finally, Section 6 summarizes the main findings of the paper.

# 2. The etiological account

Etiological explanations take mechanisms to be responsible for phenomena in the sense that the former cause the latter. An example cited in the philosophical literature is the etiological mechanism of Huntington’s disease (Craver 2007, 107-08). The phenomenon is construed as an outcome, in this case, the presence of a set of behavioural and psychological symptoms (chorea accompanied by cognitive impairments, depression and compulsive behaviour). The mechanism is viewed as the causal pathway initiated by an excess number of trinucleotide CAG repeats in the HTT gene and terminating with the manifestation of symptoms. In a similar vein, Melzack (2001, 1378) attempted to synthesize available knowledge about the many causal factors impacting on pain by conjecturing that pain “is produced by the output of a widely distributed neural network in the brain,” the ‘neuromatrix,’ which “is the primary mechanism that generates the neural pattern that produces pain. Its output pattern is determined by multiple influences, of which the somatic sensory input is only a part, that converge on the neuromatrix.” Prima facie, this model, too, proposes an etiological explanation.

The use of causal language reflects the nature of the evidence on which such explanations are based. As Bickle (2006, 425) points out in his discussion of studies on long-term potentiation, “in cellular and molecular cognition, the approach […] is to ‘intervene cellularly/molecularly and track behaviorally,’ i.e., intervene *causally* at the level of cellular activity or molecular pathways within specific neurons (e.g., via genetically engineered mutant animals […]); then track the effects of these interventions under controlled experimental conditions using behavioral protocols well accepted within experimental psychology.” Bickle (2006, 426) equates evidence for the causal relevance of biological factors and mechanisms to psychological task performance with successful explanation and reduction: “One only claims a successful *explanation*, a successful *search for a cellular or molecular mechanism*, or a successful *reduction*, of a psychological kind when one successfully intervenes at the lower level and then measures a statistically significant behavioral difference.”

The explanatory (and pragmatic) relevance of causal manipulability is certainly easy to grasp, but ontological reduction is by no means obvious. How can a task performance effect ontologically reduce to its biological causes? Following Velmans (2009, 45-49), Price and Barrell, two well-known pain researchers, point out that research in neuroscience aims to demonstrate that “neural activity causes phenomenal experience (e.g., pain).” But “(1) Causes and correlations do not establish ontological identity, and (2) causes and correlations between two phenomena occur when they are not ontologically identical in every other instance in nature that we know about” (2012, 22). (1) is taken to rule out the possibility of inferring identity from evidence for correlation or causation, thus emphasizing a lack of evidence for reductive physicalism. (2) further highlights the fact that the relata of a causal relationship are ontologically distinct items, emphasizing the fact that evidence for causation is also evidence for ontological antireductionism.

If mechanistic explanation amount to a form of causal explanation, then, from a metaphysical standpoint, it seems reasonable to conclude that a mechanism must be ontologically distinct from the outcome-phenomenon it causes. In the case of Huntington’s disease, the implication is that a biological cause has ontologically distinct psychological effects. But if this is the case, then the etiological explanation doesn’t tell anything about the biological substrates of these psychological effects. Likewise, pain is not identical with the neuromatrix mechanism that produces it. Under an etiological interpretation, the recently discovered pattern of fMRI activity predicting the reported painfulness of a heat stimulus (Wager et al. 2013) refers to the mechanism causing pain, not the pain outcome. It specifies which structures in the brain should be monitored to measure pain or targeted by interventions to alter pain experience, yet it reveals nothing about the nature of pain itself and its presumed identity, constitution or, more generally, non-causal supervenience on a biophysical state.[[2]](#footnote-2)

The above considerations have direct implications for physicalism. In a somewhat hasty move, Velmans, Price and Barrell take absence of evidence for psychoneural identities as evidence for their absence, concluding that psychophysicalism, the antireductionist view that reality is irreducibly physical and psychological in nature, is most likely true. Strictly speaking, nothing here conclusively rules out the possibility that psychological phenomena are not identical with or cannot be reduced to a physical reality other than that of their causes. Nevertheless, it is still the case that etiological mechanisms, even when firmly grounded in biology, don’t tell us anything about the physical substrates of psychological outcomes. It follows from here that while studies aiming to demonstrate causation neither refute nor support physicalism, evidence for causation, in conjunction with the metaphysical assumption that effects cannot be ontologically reduced to their causes, unambiguously supports psychophysicalism. This evidential asymmetry favours psychophysicalism over physicalism as the most promising working hypothesis.

# 3. The constitutive account

*3.1 Higher-level phenomena consist of lower-level mechanisms*

As it turns out, not all phenomena are characterized as outcomes and not all mechanistic explanations amount to etiology. Most phenomena are characterized as patterns or behaviours associated with biological systems (Bechtel and Richardson 2010; Craver 2007; Machamer et al. 2000), or as causal or correlational regularities, such as statistically significant patterns amidst epistemically uninteresting data variation (Baetu 2019a; Bogen and Woodward 1988; Hacking 1983; McAllister 1997). For instance, in phototransduction, exposure of retinal cells to light triggers changes in the electrical conductance of cell membranes. Since this stimulus-response sequence can be systematically reproduced when experiments with cells extracted from animals of the same and different species are replicated, it cannot be an accidental association, but a phenomenon demanding an explanation.

Explanations of stimulus-response and, more generally, of association phenomena are often described as an elucidation of ‘lower-level’ mechanisms underlying ‘higher-level’ behaviours or regularities (Craver 2007; Craver and Bechtel 2007; Craver and Darden 2013). Behaviour-like explananda are sometimes further associated with ‘macro-level’ sciences, such as psychology, or descriptive disciplines, such as cytology and embryology. Mechanistic explanantia, in contrast, are associated with ‘micro-level’ explanatory sciences, such as neuroscience and molecular biology. Talk of ‘macro-’ and ‘micro-levels’ suggests that the ‘underlying’ relationship linking explanans and explanandum could be one of constitution, understood as a form of physical composition. In line with this suggestion, Craver and Bechtel propose that the phenomenon of phototransduction is mechanistically explained as follows:

“At the highest level, the eye transduces light into a pattern of neural activities […]. This process can be decomposed into lower-level components and their activities. The light enters the eye, it is inverted and focused by a lens, and it is projected onto the retina […]. The conversion of light into patterns of neural activity by the retina can itself be decomposed into diﬀerent components: in particular, the rods and cones that change their electrical state depending on speciﬁc features of the light stimulus […]. Another level down, rod cell activation is also sustained by a mechanism. Light is absorbed by and activates rhodopsin, which then stimulates G-proteins. These G-proteins activate cyclic GMP phosphodiesterase, which catalyzes the conversion of cyclic GMP [cGMP] to 5’-GMP. Lowering the concentration of cyclic GMP causes sodium channels to close, reducing the inward sodium current and thereby hyperpolarizing the cell […].” (2007, 549)

Based on this reconstruction, Craver and Bechtel (2007, 555) conclude that the phenomenon of phototransduction consists of, as opposed to being caused by, a hierarchical structure of mechanisms, much in the same way the temperature of a gas is said to consist of, and not to be caused by, the mean kinetic energy of gas molecules.

*3.2 The part-whole mutual manipulability criterion*

Unlike etiology, constitution allows for the reduction of behaviours of systems (phenomena) to organized activities of the parts of that system (operating mechanisms). However, the notion of mechanistic constitution is of no use to scientists unless a methodology capable of generating evidence for constitution is specified. Price and Barrell endorse a strictly causal model of explanation precisely because they are acutely aware of the fact that the experimental methodology deployed in the life sciences is designed to generate evidence for correlation and causation. There are no clear methodological guidelines, let alone a gold standard, when it comes to non-causal relationships such as identity or constitution.

Craver proposes to address this caveat by arguing that constitution is experimentally demonstrated when the requirements of part-whole relationship and part-whole mutual manipulability are satisfied–that is, when “(i) X is part of S; (ii) in the conditions relevant to the request for explanation there is some change to X’s φ-ing that changes S’s ψ-ing; and (iii) in the conditions relevant to the request for explanation there is some change to S’s ψ-ing that changes X’s φ-ing” (2007, 153).[[3]](#footnote-3) In the case of phototransduction, the whole *S* could be a rod cell; S’s behaviour *ψ*, phototransduction, which amounts to cell membrane hyperpolarization in response to light exposure; part *X*, cGMP; and X’s behaviour *φ*, cGMP hydrolysis, which is measured by changes in intracellular cGMP levels. The fact that interventions manipulating the levels of cGMP result in changes in the electrical conductance of the cell membrane (hyperpolarization) is construed as a ‘bottom-up’ intervention on a part affecting the whole, while light exposure is interpreted as a ‘top-down’ intervention on the cell-whole affecting its cGMP-parts.

The proposal faces various conceptual problems (Baumgartner and Gebharter 2016; Harinen 2014; Leuridan 2012; Romero 2015), which I will not detail here, and remains to be subjected to rigorous scrutiny in the methodological literature. The methodological value of the proposal resides in the fact that mutual manipulability–i.e., the satisfaction of conditions (ii) and (iii)–is an indicator of a possible non-causal relationship between measured variables. A simple scenario illustrating this possibility is when a variable *A* is measured and operationally defined by a set of experimental procedures *α* typically employed in psychology, while a variable *B* is measured and operationally defined by another set of procedures, *β*, typically employed in neuroscience, yet, unbeknown to researchers, *α* and *β* measure the same factor *X*. In this case, an intervention on *X* setting the value of *A* to *a*, as measured by *α*, causes measurement reading *b* when *B* is measured by technique *β*, and vice versa. In this scenario, the causal relationship between the variables *A* and *B* reflects solely the causal nature of the intervention and measurement techniques employed, not the relationship between the factors to which these variables refer, which is one of identity (both *A* and *B* refer to *X*). In a similar way, if measured variables refer to parts and wholes, then part-whole constitution may manifest itself as mutual manipulability relationships arising because of the causal nature of the interventions and measurements involved in experiments.[[4]](#footnote-4)

On the downside, the mutual manipulability test lacks specificity, and therefore cannot conclusively demonstrate the presence of non-causal dependencies. It could be that two mutually manipulable variables stand in a relationship of identity or composition, but it could also be that they are embedded in a circular causal structure, such as a feedback loop. Craver and Bechtel argue that by combining the mutual manipulability desiderata (ii) and (iii) with the composition requirement (i), causal interpretations are ruled out. Unfortunately, this is not the case. (i) concerns part *X* and whole *S*, while the variables manipulated in (ii) and (iii) are the activities *φ* and *ψ*. Violations of the parthood relationship between X and S don’t prohibit causal relationships between *φ* and *ψ*. For instance, secreted interferon molecules are not parts of a lymphocyte, yet the interferon’s binding of cell-surface receptors causally contributes to the lymphocyte’s autoactivation. The same would have been the case if, as conjectured initially (Hagins and Yoshikami 1974), the transport of Ca2+ ions across the cell membrane would have turned out to be the messenger involved in phototransduction.

# 4. The causal mediation account

*4.1 There is only evidence for causal relevance and causal mediation*

A more immediate concern is that even if conditions (i)-(iii) would amount to a reasonably accurate mechanistic constitution test, no such tests were conducted in the studies discussed in the mechanistic literature. This entails that the proposal violates the principle of descriptive adequacy, namely the commitment of the new mechanistic philosophy to an accurate description of scientific practice (Kaiser and Krickel 2017). In the case of phototransduction, studies have shown that: (a) light stimulation of rod cells decreases the concentration of cGMP and triggers a hyperpolarization response, while these responses do not occur in the absence of light stimulation (Miller et al. 1971); and (b) the depletion of intracellular cGMP triggers a hyperpolarization response simulating the effect of light exposure, as compared to a non-depleted control (Miller and Nicol 1979). Craver and Bechtel construe experiment (a) as a top-down intervention in which a change to *S’s ψ-ing* (light-induced cell hyperpolarization) changes *X’s φ-ing* (cGMP hydrolysis, as measured by intracellular cGMP/GMP concentrations), and (b) as a bottom-up intervention in which a change to *X’s φ-ing* (cGMP depletion resulting in less GMP being produced by hydrolysis) changes *S’s ψ-ing* (an alteration of the light-induced hyperpolarization behaviour, which now occurs in the absence of light stimulation). Experiments (a) and (b), in conjunction with the fact that cGMP is a physical part of rod cells, are taken to demonstrate the mechanistic relevance of cGMP to the phenomenon of phototransduction.

Yet this interpretation of experimental results is nowhere to be found in the scientific literature. The reason is that the interpretation is spurious. Since the independent variable targeted in (b) is *intracellular cGMP concentration* and the dependent variable measured is *cell membrane hyperpolarization*, (b) may be construed as demonstrating the causal relevance of *X’s φ-ing* to *S’s ψ-ing*. However, the experiment merely identifies cGMP as a cause of hyperpolarisation. Nothing here demonstrates the relevance of cGMP to the phenomenon of phototransduction (Figure 1). This experiment cannot differentiate between a scenario in which cGMP hydrolysis causally mediates, and thus contributes to light-induced hyperpolarization (i.e., the phenomenon of phototransduction), and one in which cGMP hydrolysis triggers or mediates an alternate, light-independent hyperpolarization response, which is obviously not phototransduction, but a different phenomenon. As for experiment (a), the independent variable targeted by the intervention is *light exposure*, not *S’s ψ-ing* (induced hyperpolarization response). The experiment therefore only demonstrates the causal relevance of *light exposure* to *S’s ψ-ing* (hyperpolarization) and to changes in *X’s φ-ing* (cGMP hydrolysis resulting cGMP depletion). The experiment provides no grounds for concluding that *S’s ψ-ing* (induced hyperpolarization response) cause *X’s φ-ing* (cGMP hydrolysis), but only that both correlate with the intervention condition. It may be that *ψ* and *φ* are divergent effects of a common cause (*light exposure*), that *ψ* causes *φ*, or that *φ* causes *ψ* (Aldrich 1995). Moreover, the claim that induced hyperpolarization response manipulates cGMP hydrolysis amounts to backward causation; it is like claiming that a person’s dying manipulates a tumour’s growth.

Not only there is no evidence for mechanistic constitution, but mechanistic explanations documented in the scientific literature, as well as the experiments leading to their elucidation are systematically interpreted in strictly causal terms. The very textbook Craver and Bechtel quote as a source for their case study tells us that phototransduction “occurs in three stages: (1) Light activates visual pigments; (2) these activated molecules stimulate cGMP phosphodiesterase […]; and (3) the reduction in cGMP concentration closes the cGMP-gated channels […]” (Kandel et al. 2000, 511). This is not mechanistic constitution, but a series of intermediary stages causally mediating exposure to light (a stimulus applied to a biological system) and the generation of an action potential (the response of the system). Under the constitutive account, condition (i) of the mutual manipulability criterion dictates that cGMP must be part of cells and cells must be parts of retinas. If this requirement is not satisfied, Craver and Bechtel cannot claim that the phototransducing retina consists of hyperpolarizing rod cells, which in turn consist of cGMP hydrolyzing signalling cascades. Causal mediation doesn’t make this stipulation. It is not required that cGMP is part of rod cells and that these cells are parts of retinas for cGMP to be a causal mediator in phototransduction. This allows for the possibility that the messenger involved in phototransduction may have turned out to function like a secreted extracellular messenger triggering auto-hyperpolarization or the hyperpolarization of neighbouring retinal cells, as initially postulated by the Ca2+ hypothesis.

The strictly causal language in which the explanation is framed reflects both the nature of the phenomenon explained and that of the experimental results based on which the mechanistic explanation was developed. To elucidate a mechanism, a phenomenon must first be consistently reproduced in the context of one or more replicable experimental setups (Baetu 2019a; Bechtel and Richardson 2010; Hacking 1983). Most of the seminal experiments conducted in the 1970s and 1980s investigated ‘dark’ and ‘light currents’–that is, the electrical activity of retinal cells in dark (control) and light (test) conditions, as measured in retinal preparations extracted from freshly sacrificed animals using microelectrodes inserted at various depths in the retina. These studies are the same as (a) discussed above, minus the subsequent identification of cGMP as a correlate of light-induced hyperpolarization. Hagins et al. (1970, 380) found that “in darkness a steady current flows inward through the plasma membrane of the rod outer segments,” while “[f]lashes of light produce a photocurrent which transiently reduces the dark current.” The latter corresponds to a previously identified hyperpolarizing voltage response (Tomita 1970), which was identified as “the primary sensory consequence of light absorption by rhodopsin.”[[5]](#footnote-5) Thus, the phenomenon investigated by Hagins’ laboratory and many other research groups was the systematic reproduction of a “light-sensitive electrical conductance,” as demonstrated in a replicable in vitro experimental setup. This phenomenon was eventually reproduced in rod and cone cells from a variety of vertebrates and the elucidation of various aspects of its mechanism led to a recharacterization of phototransduction as the “absorption of light by visual pigments in the retinal rod and cone photoreceptors and its conversion into an electrical signal” (Luo et al. 2008, 9855).

The first thing we may note about the phenomenon of light-sensitive conductance (or hyperpolarisation) is that it is causal in nature, describing how rod cells respond (the observed outcome, or dependent/measured variable) in the presence and absence of light stimulation (the experimental intervention, or independent/manipulated variable) when all other aspects of the experimental setup are known to be comparable (controlled). Current methodological standards (Leighton 2010; Shadish et al. 2002), theoretical modelling (Pearl et al. 2016; Spirtes et al. 1993) and philosophical analyses (Baetu 2020; Craver 2007; Woodward 2003) dictate that such an experiment demonstrates the causal relevance of the stimulus (light) to the response (electrical conductance). Causal characterizations of phenomena invite further experimental probing in the hope that specific changes in the experimental setup may disturb the stimulus-response sequence describing the phenomenon (Baetu 2019a; Bechtel and Richardson 2010; Craver 2007; Craver and Darden 2013). One such experiment showed that (c) “papaverine, a phosphodiesterase inhibitor, increases both cyclic GMP levels and the dark permeability of the plasma membrane; and *β*,*γ*-methylene ATP increases the effectiveness of light in suppressing both permeability and cyclic GMP levels” (Woodruff et al. 1977, 667).[[6]](#footnote-6) Once again, a strictly causal interpretation was offered. The experiment doesn’t show and is not designed to show that cGMP activity at a molecular level is constitutive of the phototransduction activity at the cellular level. Rather, the experiment simply shows that intracellular cGMPacts somewhere along the path linking stimulus and response. The overall logic of the study is that of a pair of knockout/overexpression experiments (Baetu 2012; 2019a), whereby a causal intermediary along the pathway linking stimulus and response is disturbed in the test arm of a controlled experiment (bathing retinas in papaverine/*β*,*γ*-methylene ATP solutions) and results in an enhancement/reduction of the expected response (e.g., dark current in darkness), as measured in the control arm of the experiment (untreated retina), while all other aspects of the experimental setup are kept comparable. Hence, the consensus reached at the time and still in place today is that cGMP is a “diffusible messenger that mediates excitation between the rhodopsin that absorbs a photon and the conductance channels of the outer segment plasma membrane” (Brown and Waloga 1981, 369).

*4.2 No levels, no part-whole constitution, just causal pathways*

The above discussion reveals that, in the scientific literature, the ‘underlying’ relationship linking mechanism and phenomenon is not understood as mechanistic constitution, but rather as causal mediation. A mechanistic explanation remains causal in the sense that it describes a causal structure, namely the physical process linking stimulus and response. However, the causal relationship is not understood in the etiological sense of bringing about changes in the values of a variable–that is, as an event or outcome–but rather in that of linking measured variables in relationships of statistical dependence (i.e., not allowing them to vary independently of one another), thus constraining the behaviour of the system under investigation. This is consistent with the fact that many phenomena investigated in the life sciences are not defined as outcomes simpliciter (e.g., changes in electrical conductance), but rather as outcomes in a specific experimental setup involving a particular kind of biological system subjected to specific initial conditions (e.g., changes in electrical conductance in vertebrate retinas subjected to light exposure).

Under a causal mediation account, there are no higher-level phenomena-items consisting of concerted lower-level mechanism-items, but rather coarser- and finer-grained descriptions of the same causal structure. Both the ‘higher-level’ description of the phenomenon and that of its ‘underlying’ mechanism refer to the same causal structure circumscribed within the spatiotemporal boundaries of a well-defined experimental setup. For example, experiments measuring the phenomenon of light-sensitive conductance and the causal effect of changes in cGMP levels on conductance document behaviours of the same experimental setup consisting of sliced retinas bathed in a physiological solution and impaled with electrodes. The difference between the two descriptions lies in the fact that the phenomenon is described by means of interventions and measurements probing only the input and output of this structure–that is, the darkness/light-pulse stimulus and the electrical conductance response of the rod cells–while intermediary causal stages remain unknown and unprobed. This is consistent with the fact that descriptions of phenomena are often assimilated to input-output ‘black boxes’ (Craver 2007; Craver and Darden 2013; Darden 2006; Machamer et al. 2000). A mechanistic description, on the other hand, is based on additional interventions and measurements targeting other variables describing the experimental setup. If the manipulation of these variables disturbs the stimulus-response causal relationship, these variables are taken to refer to intermediary stages of the mechanism (Baetu 2019b).[[7]](#footnote-7) In the experiment by Woodruff et al., phosphodiesterase and cyclase activity were targeted by pharmaceutical interventions resulting in an increase/decrease of cGMP levels inside rod cells and a modulation of the dark current. Changes in the stimulus-response phenomenon, as contrasted to the control arm of the experiment, showed that these interventions disrupted the mechanism responsible for light-sensitive conductance, thus demonstrating the causal relevance of cGMP as intermediary acting somewhere along the causal pathway linking stimulus and response. In this way, causal determinants initially hidden in the black box description of the phenomenon were revealed, generating knowledge about the inner workings of the mechanism of phototransduction.

While the constitutive account assumes a hierarchical ontology of system behaviours consisting of mechanistic entities, activities and organizational features, the causal mediation account is compatible with both a minimalist ontology of operationally defined variables [e.g. (Woodward 2002; 2011)], as well as richer physical interpretations, including the commonly assumed mechanistic ontology of entities, activities/interactions and organizational features [e.g. (Glennan 2017; Machamer et al. 2000)]. For example, it is not known what the measured variable ‘pain’ measured by verbal reports ultimately refers to (e.g., whether it is a physical part or process in the brain or some psychological entity). The best researchers can do for now is to rely on operationalizations of the sort, “Pain is whatever the experiencing person says it is, existing whenever he says it does” (McCaffery 1968, 95). Notwithstanding, pain, as measured by verbal reports, has been investigated in an experimental setup analogous to Hagins’ retinal preparations since the 19th century. Human subjects are exposed to noxious stimulation (e.g., placing a hand in cold water) and a verbal response is elicited (e.g., reporting perceived pain intensity on a numerical scale). Using this experimental characterization of the phenomenon of acute pain, it was possible to establish not only that the stimulation condition has a causal impact on reported pain intensity, but also that nociceptor (e.g., C-fiber) activity is a causal intermediary between noxious stimulation and pain reports (Handwerker 1996). Thus, depending on how many variables are taken into consideration and the physical interpretation of each variable, a wide variety of causal descriptions may be generated, ranging from black-box correlations and stimulus-response phenomena, to coarse-grained mechanistic sketches mixing operationalized and physically interpreted variables, to fine-grained mechanistic explanations referring to entities, activities and organizational features of the world (Adolphs and Anderson 2018; Frith 1992).

Finally, descriptions of phenomena and mechanisms under a causal mediation account are not mere epistemic constructs. Just like etiological and constitutive mechanistic explanations (Craver 2007; Illari 2013; Illari and Williamson 2011), they retain an ontic dimension in the sense that both the coarser phenomenological descriptions and the finer-grained mechanistic descriptions are measurement outcomes of physical procedures generating physical effects (data) informative of the causal structure of the world (Baetu 2019a; Trout 1998). However, these effects are not necessarily the same thing as the mechanisms they describe. Both the characterizations of phenomena and mechanistic descriptions may be incomplete, selectively reflecting only those aspects of mechanisms that can be experimentally probed or inferred from experimental results and background theoretical knowledge, as well as include experimental artefacts that have nothing to do with the measured mechanisms, but with limitations and inaccuracies of the intervention and measurement techniques. If no experimental artefacts enter the description of mechanisms, it may be concluded that an overall causal stimulus-response phenomenon consists, at least in part, of causal subprocesses corresponding to the mechanism responsible for the phenomenon (Baetu 2012; 2019a; Craver et al. 2021; Harinen 2014).

*4.3 Note on the ‘matched interlevel experiments’ account*

Craver et al. (2021) recently proposed a revised set of criteria for mechanistic constitution, arguing that three “interlevel experiments” are sufficient to justify the conclusion that a factor plays a causal role along a mechanistic process: (1) “Bottom-up Inhibitory Experiments.Delete or inhibit a component (*X’s φ-ing*). Intervene to establish startup conditions *ψin*. Measure *ψout*. Evaluate thereby whether *X* and its *φ-ing* are necessary for *ψin* to produce *ψout*”; (2) “Bottom-up Excitatory.Intervene to stimulate *X* to *φ*. Measure *ψout*. Evaluate whether one can control *ψout* by manipulating *X* or its *φ-ing*. Evaluate thereby whether *X* and its *φ-ing* are causes of *ψout*”; and (3) “Top-down Activation. Bring about startup conditions *ψin*. Measure *X’s φ-ing*. Also measure *ψout*. Evaluate thereby whether *X* and its *φ-ing* are changed when *ψin* produces *ψout*.”

According to this new account, mechanistic constitution is nothing else but causal mediation, and the proposed experimental trio amounts to a schematic description of the studies discussed in Section 4.1: (1) describes experiment (c), (2) experiment (b), and (3) experiment (a). There are, however, two respects in which the account is misleading. First, Craver et al. claim that these three experiments are sufficient to demonstrate causal mediation. They are indeed sufficient, but not minimally sufficient. (3/a) and (2/b) are exploratory experiments that play a role in the discovery of putative causal mediators, which in turn drives the formulation of mechanistic hypotheses. However, these experiments don’t test causal mediation hypotheses. As discussed in Section 4.1 and illustrated in Figure 1, type (3/a) experiments only demonstrate that *Xφ-ing* is a correlate of the phenomenon of interest (*ψin*🡪*ψout*, *Xφ-ing*), while type (2/b) experiments only demonstrate that *Xφ-ing* is a cause of the measured outcome (*Xφ-ing*🡪*ψout*). (3/a) cannot rule out the possibility that *Xφ-ing* is a divergent effect of *ψin* (*ψin*🡪*Xφ-ing*🡪*Ωout*), while (2/b) cannot rule out the possibility that *Xφ-ing* is a convergent cause of *ψout* (*αin*🡪*Xφ-ing*🡪*ψout*). The only conclusive experiment is (1/c), which tests whether the manipulation of *Xφ-ing* (the test condition; *ψin*, *Xφ-ing*🡪*ψout*) completely or partially screens out the effect of *ψin* on *ψout* (as observed in the control condition; *ψin*🡪*ψout*, *Xφ-ing*) (Baetu 2012; Pearl et al. 2016). Since (2/b) and (3/a) are the bottom-up and top-down experiments previously assumed to demonstrate mutual manipulability, their bundling with the only experiment that really matters, namely (1/c), appears to be an attempt to smuggle back mutual manipulability under the veil of sufficiency.



Figure 1. Mediating, divergent and convergent causal pathways

The second issue is the interlevel glossing ostensibly missing from the scientific literature. None of the studies on phototransduction, nor those cited by Craver et al. assume levels at any stage of experimental design and execution or draw any conclusions about levels from experimental results. In all these experiments, the magnitude of a causal effect, direct or mediated, is estimated by the difference between the baseline outcome in the control arm of the experiment and the outcome in the test arm (Hernán and Robins 2020; Shadish et al. 2002). If there is no difference and the magnitude of the causal effect is zero, this usually means that there are no causal links between the tested variables (although there are exceptions, such as redundant pathways generated by gene duplication). If the magnitude is non-zero, then causation can be inferred (assuming an unbiased intervention targeting only the independent variable). The additional information that the variable *Xφ-ing* (e.g., cGMP, ALML neuron activity) is lower-level while the variables *ψin* (light exposure, tapping the worm’s head) and *ψout* (hyperpolarization, worm’s reversal) are higher-level is irrelevant to causal inference.

Not only the interlevel glossing lacks any methodological substance, but it is also a confused post-hoc reconstruction of experimental practice. Experimenters can plan to conduct and be aware that they are conducting interlevel experiments only if the variables manipulated and measured in these experiments are previously assigned higher and lower levels. For instance, experimenters would need to know beforehand that *ψin* is higher-level (light exposure, tapping the worm’s head) while *Xφ-ing* (cGMP, ALML neuron activity) to correctly label (3/a) as ‘top-down,’ and not ‘bottom-up’ or ‘intra-level.’ Yet not a word about how the level-status of variables is determined prior to conducting experiments is to be found in the primary scientific literature, the methodological literature, or Craver et al. In fact, Craver’s (2007) proposal that levels are defined relative to mechanisms entails that level assignments are impossible prior to the elucidation of mechanisms. If so, then it is only after scientists conducted their experiments and inferred a mediating mechanism that Craver et al. can proceed to label the intermediary stages of the mechanism as ‘lower-level’ and the mechanisms’ inputs and outputs as ‘higher-level.’ These labels are then retrospectively projected onto an experimental design that has nothing to do with levels, creating the illusion that researchers planned and conducted a carefully matched combination of top-down and bottom-up experiments with the explicit purpose of probing interlevel mechanistic relationships.

Finally, no level-labelling method is ever specified, justified, and systematically applied. For instance, type (3/a) experiments are branded ‘top-down’ based on the intuition that the manipulated variable *ψin* is ‘higher-level,’ while the measured outcome *Xφ-ing* is ‘lower-level.’ But these experiments also measure the *ψout* outcome; if they didn’t, there would be nothing tying *Xφ-ing* to the phenomenon of interest (*ψin*🡪*ψout*). But since *ψout* is part of the description of a ‘higher-level’ phenomenon, (3/a) could also be labelled ‘top-top.’ Likewise, since that the control of type (1/c) experiment is the test condition of a ‘top-down’ (3/a) experiment, (1/c) could be construed as either or both ‘bottom-up’ or/and ‘top-down,’ depending on how we choose to look at it. In the end, nothing methodologically valuable is lost when sufficiency claims and interlevel labels are removed, and nothing but confusion is gained when they are added.

# 5. Reduction under a causal meditation account

*5.1 Biological mediators of psychological causation*

As discussed in Section 2, the methodological assumption that experimental science as currently conducted can only generate knowledge about the causes of psychological phenomena, combined with the metaphysical assumption that effects must be ontologically distinct from their causes, led some researchers to conclude that evidence from experimental science is biased in favour of ontological antireductionism, or psychophysicalism (Price and Barrell 2012; Velmans 2009). Whether for this or some other reason, antireductionism is indeed the dominant trend in contemporary pain research.[[8]](#footnote-8) Evidence for causal relevance resulted in a proliferation of multifactorial causal models according to which pain experience “is determined by the interaction among biological, psychological (which include cognition, affect, behaviour), and social factors (which include the social and cultural contexts that influence a person’s perception of and response to physical signs and symptoms)” (Asmundson and Wright 2004, 42).[[9]](#footnote-9) Since biopsychosocial models are construed as putative etiological mechanisms of pain, it is difficult to see how the measured variable pain could be ontologically mapped or reduced to any of the biopsychosocial risk factors involved in its etiology. Nevertheless, as I will show in this section, even if pain is not reduced, evidence for biological mediators of psychological causes of pain allows for a collapsing of some instances of psychological causation onto biological causation and, under the assumption of parsimony, of some physically uninterpreted psychological variables onto physically interpreted biological variables. Thus, compared to the etiological account, the causal mediation account is better equipped to support physicalism.

An important source of evidence supporting a partial collapsing of psychological causation onto biological substrates comes from multivariate pattern analysis, a data analysis technique allowing researchers to correlate structural elements of stimuli, distributed patterns of neural activation evoked by these stimuli and structural elements of reported conscious experience.[[10]](#footnote-10) In a groundbreaking study, Wager et al. (2013) identified an fMRI-based biomarker, dubbed the Neurologic Pain Signature (NPS), accurately predicting acute pain experience in response to noxious thermal, mechanical and electrical stimulation. The strength of the NPS response and the reported intensity of pain were substantially reduced upon the administration of an opioid agonist analgesic, thus demonstrating the causal relevance of NPS as a mediator of noxious pain. In agreement with biopsychosocial models postulating psychosocial causes in addition to biological determinants, NPS did not correlate with the variation in pain experience when the stimulus intensity is held constant, nor could account for the pain-modulating effects of placebo treatment, cognitive self-regulation and perceived control, all of which were previously shown to be psychological determinants of pain.

In a follow-up study, the research group identified a second fMRI biomarker, termed the Stimulus Intensity Independent Pain Signature-1 (SIIPS1), which predicted “variation in pain above and beyond noxious stimulus intensity (for example, heat temperature) and nociceptive brain processes estimated by the NPS” (Woo et al. 2017). Since “SIIPS1 explains a substantial amount of the variation in trial-by-trial pain ratings not captured by the NPS” (Woo et al. 2017), it was taken to reflect “a measure of the conscious experience of the pain, independent of its discriminative nociceptive properties” (Adolphs and Anderson 2018, 276). In addition, Woo et al. showed that the “SIIPS1 was a significant and consistent mediator of the effects of psychological interventions, including manipulations of expectancy and perceived control, whereas the NPS was not,” concluding that “SIIPS1 is likely to be influenced by psychological, ‘top-down’ influences on pain in ways that are not well captured by the NPS.”[[11]](#footnote-11)

The two studies reveal an alignment of psychological and biophysical determinants of pain with specific, largely non-overlapping neurosignatures. Pharmaceutical interventions and statistical mediation analyses further provide evidence that these neurosignatures are not mere covariates of biological and psychological interventions on pain experience, but most likely causally mediate the effects of interventions on pain experience. This is a very significant result. Under a causal mediation account, both biopsychosocial models of pain and fMRI signatures are descriptions of the same mechanism, with the biological determination of pain mapping onto the stimulus dependent NPS signature triggered by nociception pathways, and psychosocial causation onto the stimulus independent SIIPS1 signature evoked by psychological interventions. The fact that psychosocial causation passes through the bottleneck of a biological neurosignature is a first indication that what seems to be an instance of psychological (‘mental’) causation reduces in fact to biological activity, thus refuting the anti-reductionist hypothesis of a biologically-unmediated causal link between psychological variables.

As it stands, the evidence is not airtight. Woo et al. (2017) only showed that SIIPS1 partially mediates the effects of psychological interventions on pain and failed to demonstrate that all causal contributions to pain have been accounted for. However, if future studies address these caveats, this would provide a local experimental confirmation of the principles of physical causal completeness and of causal exclusion often assumed by physicalist accounts of the mind (Kim 2005; Papineau 2002). This would constitute a direct and independent piece of experimental evidence for physicalism, distinct from general theoretical justifications, such as Helmholtz’s energy conservation rationale (Bevilacqua 1993). Advances in non-invasive intervention methods, such as transcranial magnetic stimulation (Horvath et al. 2011), may further allow for ‘knockouts’ of specific neurosignatures and rule out causal overdetermination scenarios whereby psychosocial causation acts as a failsafe pathway opened only when biological pathways are blocked.

*5.2 Variable mapping*

Under the assumption of parsimony, collapsing psychological causation onto biological processes may further justify a mapping of psychological variables onto biological ones. In one of the experiments conducted by Woo et al. (2017), conditioning subjects to associate auditory cues with high- and low-intensity noxious heat was shown to have a modulating effect on reported pain. According to a non-reductive explanation of this result, a psychological conditioning intervention is assumed to target the psychological variable *expectation*, which in turn has a causal impact on the measured psychological outcome *pain*. However, if the modulating effect of the conditioning intervention passes through the bottleneck of a biological process, SIIPS1 activity, the most parsimonious interpretation of experimental results dictates that the variable *expectation* refers to *changes in SIIPS1 activity induced by a conditioning procedure*. In turn, this removes unnecessary psychosocial causal processes in virtue of which an additional, psychological referent of *expectation* would mediate conditioning-induced changes in SIIPS1 activity.

Conversely, to empirically justify the introduction of an additional referent of the variable *expectation*, it is necessary to experimentally demonstrate that it is possible to modulate pain responses by manipulating the variable *expectation* independently of *conditioning-induced changes in SIIPS1 activity*. This may, of course, turn out to be the case, and thus provide evidence against the physicalist model. However, if, in the long run, no experimental results justify postulating an additional referent, the psychological variable *expectation* can be mapped onto the biological variable *conditioning-induced changes in SIIPS1 activity*, and eventually identified with a more specific neural mechanism involved in learning and prediction.[[12]](#footnote-12)

If this scenario pans out, it will constitute a partial vindication of reductive physicalism by showing that at least some manipulated psychosocial variables, such as *expectation* are nothing else but stimulus-induced patterns of neural activity. The question whether the dependent variable *pain* can likewise be reduced to brain activity remains open to discussion. Hopefully, since *pain* itself can be manipulated by psychological techniques (Rainville 2008; Rainville et al. 1997) and its effects on other variables measured, it too may be eventually tackled in the same way as *expectation*. Whichever way things may turn out, the above discussion would have at least succeeded in demonstrating that the pessimism expressed those who believe that the quest for the biological causes of psychological phenomena cannot possibly prove reductive physicalism right is misguided.

# 6. Conclusion

If the ontological relationship between mechanisms and phenomena is a form of non-causal determination, a reduction of phenomena to their underlying mechanisms may be justified. Unfortunately, experimental evidence for non-causal determination is either absent or inconclusive. Conversely, if this relationship is strictly etiological, then mechanistic explanations don’t allow for an ontological mapping of psychological phenomena onto the biophysical mechanisms causing them. In this paper, I argue that an alternative to these two scenarios exists, namely a causal mediation account which remains compatible with the standard causal interpretation of controlled experiments accounting for experimental results in the life sciences, while also allowing for an ontological mapping of psychological causation onto biological processes and of physically uninterpreted psychological variables onto physically interpreted biological variables.

Acknowledgments

This research was supported by SSHRC Grant # 430-2020-0654.

# References

Adolphs, R., and D. J. Anderson. 2018. *The Neuroscience of Emotion: A New Synthesis*. Princeton, NJ: Princeton University Press.

Aldrich, J. 1995. "Correlations Genuine and Spurious in Pearson and Yule." *Statistical Science* 10 (4):364-76.

Asmundson, G. J. G., and K. D. Wright. 2004. "Biopsychosocial Approaches to Pain." In *Pain: Psychological Perspectives*, ed. T. Hadjistavropoulos and K. D. Craig. Mahwah, NJ: Lawrence Erlbaum Associates.

Baetu, T. M. 2012. "Filling In the Mechanistic Details: Two-Variable Experiments as Tests for Constitutive Relevance." *European Journal for Philosophy of Science* 2 (3):337-53.

———. 2019a. *Mechanisms in Molecular Biology*. Edited by Grant Ramsey and Michael Ruse, *Elements in the Philosophy of Biology*. Cambridge: Cambridge University Press.

———. 2019b. "On Pain Experience, Interdisciplinary Integration and Levels of Description, Explanation and Reality." *Synthese* 196 (8):3231-50.

———. 2020. "Causal Inference in Biomedical Research." *Biology and Philosophy* 35:43.

Baumgartner, M., and A. Gebharter. 2016. "Constitutive Relevance, Mutual Manipulability, and Fat-Handedness." *British Journal for the Philosophy of Science* 67:731-56.

Bechtel, W. 2006. *Discovering Cell Mechanisms: The Creation of Modern Cell Biology*. Cambridge: Cambridge University Press.

———. 2008. *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. New York: Routledge.

Bechtel, W., and A. Abrahamsen. 2005. "Explanation: A Mechanist Alternative." *Studies in History and Philosophy of Biological and Biomedical Sciences* 36:421-41.

Bechtel, W., and R. Richardson. 2010. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Cambridge, MA: MIT Press.

Bevilacqua, F. 1993. "Helmholtz’s Ueber die Erhaltung der Kraft: The Emergence of a Theoretical Physicist." In *Hermann von Helmholtz and the Foundations of Nineteenth-Century Science*, ed. D. Cahan. Berkeley: University of California Press.

Bickle, J. 2006. "Reducing Mind to Molecular Pathways: Explicating the Reductionism Implicit in Current Cellular and Molecular Neuroscience." *Synthese* 151:411-34.

Bogen, J., and J. Woodward. 1988. "Saving the Phenomena." *The Philosophical Review* 97 (3):303-52.

Brown, J. E., and G. Waloga. 1981. "Effects of Cyclic Nucleotides and Calcium Ions on Bufo Rods." In *Molecular Mechanisms of Photoreceptor Transduction*, ed. W. H. Miller, 369-80. New York, NY: Academic.

Campbell, J. 2008. "Causation in Psychiatry", in K. Kendler and J. Parnas (eds.), *Philosophical Issues in Psychiatry*, Baltimore: Johns Hopkins University Press, 196-216.

Corns, J. 2012. *Pain is Not a Natural Kind*. ProQuest Dissertations Publishing: City University of New York.

Craver, C. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Clarendon Press.

Craver, C., and W. Bechtel. 2007. "Top-Down Causation without Top-Down Causes." *Biology and Philosophy* 22:547-63.

Craver, C., and L. Darden. 2013. *In Search of Biological Mechanisms: Discoveries across the Life Sciences*. Chicago, IL: University of Chicago Press.

Craver, C. F., S. Glennan, et al. 2021. "Constitutive Relevance & Mutual Manipulability Revisited." *Synthese* 199:8807-28.

Darden, L. 2006. *Reasoning in Biological Discoveries: Essays on Mechanisms, Interfield Relations, and Anomaly Resolution*. Cambridge: Cambridge University Press.

Frith, C. D. 1992. *The Cognitive Neuropsychology of Schizophrenia*. Hove: Lawrence Erlbaum Associates.

Fu, Y. 2020. "Phototransduction in Rods and Cones." In *Webvision: The Organization of the Retina and Visual System*, ed. H. Kolb, E. Fernandez and R. Nelson, 631-68. Salt Lake City, UT: University of Utah Health Sciences Center.

Glennan, S. 1996. "Mechanisms and the Nature of Causation." *Erkenntnis* 44:49-71.

———. 2002. "Rethinking Mechanistic Explanation." *Philosophy of Science* 69:S342-S53.

———. 2017. *The New Mechanical Philosophy*. New York: Oxford University Press.

Hacking, I. 1983. *Representing and Intervening* Cambridge: Cambridge University Press.

Hagins, W. A., R. D. Penn, et al. 1970. "Dark Current and Photocurrent in Retinal Rods." *Biophysical Journal* 10 (5):380-412.

Hagins, W. A., and S. Yoshikami. 1974. "A Role for Ca2+ in Excitation of Retinal Rods and Cones." *Experimental Eye Research* 18 (3):299-309.

Handwerker, H. O. 1996. "Sixty Years of C-Fiber Recordings from Animal and Human Skin Nerves: Historical Notes." In *The Polymodal Pathological Pain Receptor: A Gateway to Pathological Pain*, ed. T. Kumazawa, L. Kruger and K. Mfzumura, 39-51.

Hardcastle, V. 1999. *The Myth of Pain*. Cambridge, MA: MIT Press.

Harinen, T. 2014. "Mutual Manipulability and Causal Inbetweenness." *Synthese* 195 (1):34-54.

Haxby, J. V. 2012. "Multivariate Pattern Analysis of fMRI: The Early Beginnings." *NeuroImage* 62:852-55.

Hernán, M. A., and J. M. Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

Horvath, J. C., J. M. Perez, et al. 2011. "Transcranial Magnetic Stimulation: A Historical Evaluation and Future Prognosis of Therapeutically Relevant Ethical Concerns." *Journal of Medical Ethics* 37 (2):137-43.

IASP Task Force on Taxonomy *Pain Terms and Definitions* 2020 [cited. Available from <https://www.iasp-pain.org/resources/terminology>.

Illari, P. 2013. "Mechanistic Explanation: Integrating the Ontic and Epistemic." *Erkenntnis* 78 (2):237-55.

Illari, P., and J. Williamson. 2011. "Mechanisms Are Real and Local." In *Causality in the Sciences*, ed. P. McKay Illari, F. Russo and J. Williamson, 818-44. Oxford: Oxford University Press.

———. 2012. "What is a Mechanism? Thinking about Mechanisms *across* the Sciences." *European Journal for Philosophy of Science* 2 (1):119-35.

Kaiser, M. I., and B. Krickel. 2017. "The Metaphysics of Constitutive Mechanistic Phenomena." *British Journal for the Philosophy of Science* 68 (3):745-79.

Kandel, E. R., J. H. Schwartz, et al. 2000. *Principles of neuroscience*. 4th ed. New York: McGraw Hill.

Kendler, K. S., and J. Campbell. 2009. "Interventionist Causal Models in Psychiatry: Repositioning the Mind-Body Problem." *Psychological Medicine* 39:881-87.

Kim, J. 2005. *Physicalism or Something Near Enough* Princeton, NJ: Princeton University Press.

Krickel, B. 2018. *The Mechanical World: The Metaphysical Commitments of the New Mechanistic Approach*. Cham: Springer.

Leighton, J. P. 2010. "Internal Validity." In *Encyclopedia of Research Design*, ed. N. J. Salkind, 619-22. Thousand Oaks, CA: SAGE.

Leuridan, B. 2012. "Three Problems for the Mutual Manipulability Account of Constitutive Relevance in Mechanisms." *British Journal for the Philosophy of Science* 63:399-427.

Luo, D.-G., T. Xue, et al. 2008. "How Vision Begins: An Odyssey." *Proceedings to the National Academy of Sciences* 105 (29):9855-62.

Machamer, P., L. Darden, et al. 2000. "Thinking About Mechanisms." *Philosophy of Science* 67:1-25.

McAllister, J. W. 1997. "Phenomena and Patterns in Data Sets." *Erkenntnis* 47:217-28.

McCaffery, M. 1968. *Nursing Practice Theories Related to Cognition, Bodily Pain, and Man-Environment Interactions*. Los Angeles: UCLA Students Store.

Melzack, R. 2001. "Pain and the Neuromatrix in the Brain." *Journal of Dental Education* 65 (12):1378-82.

Miller, W. H., R. E. Gorman, et al. 1971. "Cyclic Adenosine Monophosphate: Function in Photoreceptors." *Science* 174 (4006):295-97.

Miller, W. H., and G. D. Nicol. 1979. "Evidence that Cyclic GMP Regulates Membrane Potential in Rod Photoreceptors." *Nature* 280:64-66.

Papineau, D. 2002. *Thinking about Consciousness*. Oxford: Clarendon Press.

Pearl, J., M. Glymour, et al. 2016. *Causal Inference in Statistics: A Primer*. Chichester: Wiley & Sons.

Price, D. D., and J. J. Barrell. 2012. *Inner Experience and Neuroscience: Merging Both Perspectives*. Cambridge, MA: MIT Press.

Rainsford, K. 2011. "Fifty Years since the Discovery of Ibuprofen." *Inflammopharmacology* 19 (6):293-97.

Rainville, P. 2008. "Hypnosis and the Analgesic Effect of Suggestions." *Pain* 134 (1):1-2.

Rainville, P., G. H. Duncan, et al. 1997. "Pain Affect Encoded in Human Anterior Cingulate But Not Somatosensory Cortex." *Science* 277 (5328):968-71.

Romero, F. 2015. "Why There Isn’t Inter-Level Causation in Mechanisms." *Synthese* 192:3731-55.

Seth, A. 2021. *Being You: A New Science of Consciousness*. London: Faber.

Shadish, W. R., T. D. Cook, et al. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.

Spirtes, P., C. Glymour, et al. 1993. *Causation, Prediction and Search*. New York: Springer-Verlag.

Tomita, T. 1970. "Electrical Activity of Vertebrate Photoreceptors." *Quarterly Reviews of Biophysics* 3 (2):179-222.

Trout, J. D. 1998. *Measuring the Intentional World: Realism, Naturalism, and Quantitative Methods in the Behavioral Sciences*. Oxford: Oxford University Press.

Velmans , M. 2009. *Understanding Consciousness*. 2nd ed. London: Routledge.

Wager, T. D., L. Y. Atlas, et al. 2013. "An fMRI-Based Neurologic Signature of Physical Pain." *New England Journal of Medicine* 368 (15):1388-97.

Woo, C. W., L. Schmidt, et al. 2017. "Quantifying Cerebral Contributions to Pain beyond Nociception." *Nature Communications* doi: 10.1038/ncomms14211.

Woodruff, M. L., D. Bownds, et al. 1977. "Guanosine 3',5'-Cyclic Monophosphate and the in vitro Physiology of Frog Photoreceptor Membranes." *Journal of General Physiology* 69 (5):667-79.

Woodward, J. 2002. "What is a Mechanism? A Counterfactual Account." *Philosophy of Science* 69:S366-S77.

———. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

———. 2008. "Cause and Explanation in Psychiatry: An Interventionist Perspective." In *Philosophical Issues in Psychiatry: Explanation, Phenomenology and Nosology*, ed. K. Kendler and J. Parnas. Baltimore: Johns Hopkins University Press.

———. 2011. "Mechanisms Revisited." *Synthese* 183 (3):409-27.

9931 words

1. For a nonreductive proposal, see (Krickel 2018). [↑](#footnote-ref-1)
2. This explanatory standard corresponds to what Seth (2021, Ch. 1) calls the ‘real problem of consciousness,’ the goal of which is “to explain, predict, and control the phenomenological properties of conscious experience.” [↑](#footnote-ref-2)
3. If (i)-(iii) are construed as necessary and sufficient conditions for mechanistic constitution, the trio defines the set of relationships in the world that simultaneously behave like material composition and reciprocal manipulability, or causation as understood by interventionist accounts. Since the properties of composition and causation are mutually exclusive (e.g., one is synchronic, the other diachronic), their simultaneous conjunction stipulates a counterintuitive mode of existence analogous to the wave-particle duality. Following Craver and Bechtel (2007), the simplified interpretation adopted in this paper is that mechanistic constitution is a type of material composition, but not all material composition is mechanistic composition. [↑](#footnote-ref-3)
4. In the above scenarios, there is no mechanistic constitution relationship behaving both like identity/composition and causality (manipulability). Instead, identity and composition relationships merely appear to be causal due to experimental artefacts. [↑](#footnote-ref-4)
5. Rod (and cone) cells are ‘on’ in the dark (i.e., depolarized at -40 mV, corresponding to the ‘dark current’ discovered Hagins and coworkers) and ‘off’ when stimulated by light (hyperpolarized to a potential of -60mV, as indicated by Tomita). It was subsequently discovered that cGMP levels are high in the dark and keep cGMP-gated sodium channels open, hence the inward dark current resulting in the inhibitory neurotransmitter glutamate being released in the synaptic cleft. Light results in the activation of a phosphodiesterase, which hydrolyzes cGMP, reducing its the intracellular concentration. cGMP-gated sodium channels close, causing a hyperpolarization of the cell due to a continuous efflux of potassium ions. Hyperpolarization causes voltage-gated calcium channels to close and therefore of calcium levels in the cell to drop. This reduces the amount of inhibitory glutamate released in the synaptic cleft, causing the depolarization (excitation) of bipolar neurons (Fu 2020). Note also that this explanation too is strictly about causal mediation, not mechanistic constitution. [↑](#footnote-ref-5)
6. ‘Dark permeability’ refers to the inward Na+ dark current explained in note 5. Methylene ATP analogs are resistant to hydrolytic attack by nucleotide phosphohydrolases and are thought to act as competitive inhibitors of nucleotide cyclases. [↑](#footnote-ref-6)
7. Mechanisms are described as causal pathways doesn’t entail that they are linear. For example, in the absence of measurements of intermediary stages, a cyclical metabolic pathway, such as Krebs’ cycle, appears as a linear input-output phenomenon (2 acetyl-CoA, 6 NAD+, 2 FAD, 2 ADP+Pi 🡪 4 CO2, 6 NADH, 6 H+, 2 FADH2, 2 ATP, 2 CoA). [↑](#footnote-ref-7)
8. Some authors argue that the empirical justification of antireductionism rests primarily on the notion that causal inference from controlled experiments is indifferent to the nature of the variables manipulated in the experiment (Baetu 2019b; Campbell 2008; Kendler and Campbell 2009; Woodward 2008). For instance, if we agree that ibuprofen, a drug known to target the molecular mechanisms of inflammation, has a measurable analgesic effect on pain (as measured by verbal reports) because randomized controlled trials demonstrate a statistically significant difference in pain ratings between test patients exposed to ibuprofen and comparable control patients (Rainsford 2011), then we must concede that hypnotic suggestion targeting expectations must also have a real analgesic effect, since methodologically equivalent randomized trials likewise demonstrate a statistically significant difference in reported pain between test patients exposed to hypnotic suggestion and comparable control patients (Rainville 2008). [↑](#footnote-ref-8)
9. The biopsychosocial consensus is also reflected in the IASP (2020) definition of pain. [↑](#footnote-ref-9)
10. Haxby (2012, 853) explains the technique as follows: “The idea was straightforward and based on a concept from conventional statistics, namely split-sample cross-validation. If a given stimulus category evoked a distinct pattern of activity, then independent observations of the response to that category should be more similar to each other than to responses to different categories. Correlation of patterns was the chosen measure of similarity, and I made independent observations by dividing the data for each subject into two halves–even-numbered and odd-numbered runs. Thus, I predicted that within-category correlations would be higher than between-category correlations.” [↑](#footnote-ref-10)
11. In the absence of experimental techniques selectively interfering with the activity of specific brain areas in humans, causal mediation was assessed by statistical methods. Psychological manipulations (conditioning individuals to associate auditory cues or options denoted by abstract symbols with high- and low-intensity noxious heat) were the independent variable, trial-by-trial pain ratings the dependent variable, and trial-by-trial SIIPS1 and NPS responses during pain were the mediating variables transmitting the influence of the independent variable to the dependent variable. [↑](#footnote-ref-11)
12. Reference may hold true only locally–that is, only as far as pain modulating effects are concerned, without being generalizable to other experimental setups or phenomena involving measurements or manipulations of the variable *expectation*. If so, this would suggest that psychological function is underpinned by a variety of related or unrelated biological mechanisms [e.g. (Corns 2012; Hardcastle 1999)]. This is certainly true of many biological functions (e.g., gene expression regulation, immune responses, metabolic pathways) and dysfunctions (e.g., types of diabetes, genetic disorders), a situation often reflecting an evolutionary process of duplication and coopting of mechanistic components. It is also worth emphasizing that the fact that an operationally defined measured/manipulated variable *expectation* is shown to have a biological referent does not automatically entail the elimination of a theoretically defined construct *expectation* associated with this variable. The theoretical construct may survive, although it is now associated with a biological activity. [↑](#footnote-ref-12)