frontiers in PSYCHOLOGY

IMPACT 2.8 FACTOR

Consciousness Research

A complete and scientifically-reliable framework for theory-discrimination in consciousness research requires specific, experimentally-testable, dynamical couplings between natural phenomena.

Nicholas Rosseinsky

Journal Name:	Frontiers in Psychology
ISSN:	1664-1078
Article type:	Hypothesis & Theory Article
First received on:	30 Aug 2014
Revised on:	25 Dec 2014
Frontiers website link:	www.frontiersin.org

A complete and scientifically-reliable framework for theory-discrimination in consciousness research requires specific, experimentally-testable, dynamical couplings between natural phenomena.

Nicholas M. Rosseinsky[1,*]

[1]Department of Neuroscience, Center for Dialog in Science, London, U.K.

*** Correspondence:** Nicholas M. Rosseinsky, Department of Neuroscience, Center for Dialog in Science, 71-75 Shelton Street, London, WC2H 9JQ, U.K.
rosseinsky.nicholas.m@cfdis.org

Keywords: measurement, consciousness, scientific method, causal closure, dynamical closure, interactionism, phenomenology.

12,000 Words; 1 Table; 13 Figures

Abstract

Although the study of consciousness via neuroimaging has been scientifically respectable since at least the 1990s, this emerging modality currently lacks an explicit foundation *e.g.* akin to the measurement theory that underpins physical and biophysical sciences. In the absence of theoretical justifications for methods, radically-skeptical criticisms have been advanced, doubting the methodological validity of *any* empirical appeals in consciousness research. Here, a comprehensive analysis is given of the combined capacity of currently-envisioned experimental and theoretical methods to provide scientifically-reliable discriminations between competing hypotheses in consciousness research, leading to two major results. First, it is shown that scientifically-reliable measurements of consciousness sufficient for complete resolution of theoretical uncertainty can in principle be made, but only if certain kinds of dynamical-coupling relationships govern natural phenomena. (Conversely, the absence of these relationships definitively excludes a complete and reliable framework, thus supporting critical views, but only in a contingent manner.) Second, experimental paradigms are described that can test whether couplings required for completeness-and-reliability pertain, potentially offering decisive empirical arbitration on the conditional clause of the first result. Novel symbolism introduced to label the phenomenal degrees-of-freedom in conscious experience facilitates analysis of methodological issues in a philosophy-neutral way, thus freeing results from dependence on any particular metaphysical stance. Consequently, central results depend only on physical logic and a small number of explicitly-stated assumptions. Philosophy-neutrality also emphasizes that completeness-and-reliability does not require dualism; instead, equations currently taken to fully account for brain-dynamical evolution must be (non-trivially) extended, as is viable under almost every metaphysical view.

1. Introduction

All scientific endeavors require a clear and consistent methodological framework, to ensure that claims for arbitration between competing theoretical alternatives are credible. The present objective is to identify conditions under which a *complete* and *scientifically-reliable* framework for consciousness research is achievable:

> **D1**. A *complete framework for theory-discrimination in consciousness research* is a collection of theoretical and experimental methods sufficient in principle to conclusively arbitrate between all competing hypotheses that are basically-coherent.

(*Basic-coherence* in D1 and elsewhere emphasizes that hypotheses included [*e.g.* as arbitration-candidates, in D1] are stated with sufficient precision, internally-consistent, and consistent with any overarching contextual presumptions.)

> **D2**. A framework is said to be *scientifically-reliable* if it does not employ any method conventionally excluded on methodological grounds from analogous application in physical or biological sciences.

D2 is framed in negative terms because of the arguable absence (Popper, 2002; Lakatos, 1970; Kuhn, 1996) of a unanimously-agreed, everywhere-applicable, positive, methodological framework for physical and biological sciences (whose methodological *integrity* will be taken as the minimal benchmark for consciousness research). Despite this absence, physical and biological sciences achieve in-practice progress by (in principle, consensually-amendable) agreement concerning modes-of-reasoning that are *not unreliable*; D2 adopts this "reliable-provided-not-demonstrably-unreliable" approach to methodological integrity.

Supplementary Image 1 and the accompanying *Supplementary Table 1* provide schematic guides to the present overall approach to completeness-and-reliability. Schematics are best reviewed alongside the next Section, which frames central concerns as specific questions and establishes certain basic definitions that serve to limit and then divide (for expository purposes) the space of foundational metaphysical possibilities relevant to enquiries. As will become clear, a vital process-goal will be to develop demonstrations that are as free as possible from metaphysical assumptions, including the assumption that no decisive future metaphysical advances will be made.

2. Methodological controversies and metaphysical contexts

2.1. Comprehensive accuracy of first-person-report as consciousness-meter

In relation to D1-completeness and D2-reliability, the particular methodological controversy of central concern here (Fig. 1) originates from the construal of first-person report as fulfilling the role of contents-of-conscious-experience-*meter* (Chalmers, 1998). (Equivalently, concerns here relate to the view that first-person report *stands in* for the function that a meter would typically perform in physical or biological science.) Well-known problems (Wilkes, 1988; Lamme, 2006; Cohen and Dennett, 2011) arising from this commonly-accepted construal typically lead to the entanglement of

research methodology with metaphysical view, because of claims that methodological problems can be eliminated under certain metaphysical stances [notably identity-theory (Kim, 2000) and certain functionalist stances (Chalmers, 1996)].

More specifically, the acute focus here (at least initially) will be on problems related to the degree of *comprehensive accuracy* attributable to the first-person "meter". [Equivalently, acute concerns can be viewed as originating in the possibility that certain aspects-of-conscious-experience might be inaccessible (Block, 2007). Defining two kinds of consciousness (Block, 1995) can only avoid inaccessibility problems if access consciousness is characterized purely in functional terms (and the existence of phenomenal consciousness is denied). However, use of the term consciousness while denying validity of the term conscious-experience (Dennett, 1991) is arguably meaningless: "consciousness"-functions presumably derive their description from association of function *with* experience; otherwise they are simply functions *e.g.* as of a conventionally-assumed-non-conscious computer, or any machine.]

2.2. The naturalistic scientific context

To begin central investigations requires an approach that can navigate potential problems associated *e.g.* with solipsism and inverted qualia (even though these can be viewed as also originating in the field's apparent dependence on first-person report). Accordingly, we first define (and then generally assume) a *naturalistic scientific context*:

> **D3**. A *naturalistic scientific context* for consciousness research is an otherwise philosophy-neutral setting that excludes psychophysical parallelism by assumption, and adopts a minimal *e.g.* abduction-based avoidance of problems associated with solipsism and inverted qualia.

(D3 defines a perhaps broader-than-usual conception of naturalism. Notably, D3 does not explicitly exclude dualist approaches, an exclusion sometimes taken as necessary or even definitive for naturalism. The present approach prefers maximal philosophy-neutrality, and thus avoids exclusions not absolutely required: exclusion of parallelism *is* conventionally taken to be essential, to evoke and frame focused consideration of neurobiologically-grounded theories-of-consciousness.)

D3 allows restatement of the present primary objective in terms of finding answers to three questions:

> **Q1**. After establishment of D3-naturalism but without making any further metaphysical assumptions, is it possible to establish a complete-and-reliable framework?

> **Q2**. If Q1-completeness-and-reliability *is* possible, what are the *minimal* D3-additional conditions necessary?

One *a priori* possibility is that additional conditions are essentially "nothing" (*i.e.* D3 is already complete-and-reliable), although controversies indicate otherwise, as will be definitively confirmed by later developments here. Note that the "no further metaphysical assumptions" constraint in Q1 means that (non-trivial) "conditions" in Q2 should be framed as aspects of natural order and therefore

be experimentally-testable, at least in principle, leading to:

> **Q3**. Assuming Q2-conditions are non-null, what experiments can test whether they actually pertain in our Universe?

2.3. Varieties of philosophy-neutrality

Q1's "no further metaphysical assumptions" constraint requires further definition. For example, a researcher who finds arguments for functionalism (say) completely convincing might not consider the presumption of a functionalist setting to be a metaphysical *assumption*, but rather a metaphysical *truth*. Progressing Q1 while treating such possibilities evokes definitions D4-D6 concerning *philosophy-neutrality*:

> **D4**. A *philosophy-neutral setting* in consciousness research will denote the view that no *current* reasoning establishes the dominance of a particular metaphysical stance over other (basically-coherent) stances; consequently, the adoption of a specific metaphysical stance is *currently* a matter of subjective preference.

[In present usage, a *metaphysical stance* is any of a wide variety of explanations for "*why*" certain spatiotemporal brain-dynamical structures (Gamez, 2014) are associated with (Chalmers, 1998) conscious experience, for example, by virtue of identity, function, information-integration, and so forth. In contrast, a *theory-of-consciousness* is a detailed account of "*which*" brain-dynamical structures are associated with consciousness in general and with certain conscious experiences in particular. Although not entirely separable – for example, certain kinds of functionalist metaphysics constrain then-valid theories-of-consciousness – the distinction is roughly analogous to delineation between neural-correlates empiricism and philosophy-of-mind, and is only used for expository succinctness, not for result-establishment.]

> **D5**. The *maximally-inferred philosophy-neutral setting* extends the D4-setting by additionally asserting that no *future* reasoning will be able to establish the dominance of a particular metaphysical stance; *i.e.* the maximal set of metaphysical inferences from present *e.g.* empirical information has already been made, and any future advances in *information* will inform only theory-of-consciousness arbitration, rather than radically reframing *e.g.* the basic character of consciousness research. Accordingly, consciousness research must always be independent of any particular metaphysical appeals (beyond any necessary for D3-naturalism).

The contrast between the philosophy-neutral setting and the maximally-inferred variant will be of the following expository use. The basic point is that D5-setting proofs are easier to establish (because any appeal to metaphysically-based cases are definitively excluded on grounds of subjectivity), but general-applicability sought for results will require going beyond a D5-limitation. In a little more detail, temporarily adopting the maximally-inferred D5-setting permits complete neglect of the possibility that functionalism, say, *might* be provably correct, thus enabling focus on the capacity of field-conventional empirical methods (including first-person report) *alone* for reliable-arbitration between various theories-of-consciousness. Subsequently, after the development of transferable

insights, the maximally-inferred setting will be dropped, to move on to a (slightly-restricted) D4-context.

2.4. Target context-of-validity for present theorems: "philosophy-neutral proofs"

An explanation of *slight-restriction* in the final setting for proofs requires a fuller account of the views that D4-settings are meant to accommodate. Specifically, despite apparently definitive limitations, the D4-setting is meant to accommodate every (basically-coherent) currently-held view in the field, as follows. First, consider the stance that identity-theory (say) is true, has been proven to be exclusively true, and all that remains is for the field-as-a-whole to appreciate the truth of already-existent proofs. Truth-of-identity-theory might seem to be precluded under D4 by its definitional component "no dominant single metaphysical stance". Explicitly and emphatically, we clarify here that the "no *currently-proven* dominant …" D4-qualification is precisely meant to accommodate *all* single-stance-views of the proven-but-not-yet-appreciated kind. (Thus, the only single-stance view not accommodated under D4 is that everyone agrees to the dominance of some particular metaphysics. Given ongoing debates, the present approach judges any such claim not basically-coherent.) Next, D4 includes D5 when the latter is characterized as a true-but-not-yet-agreed single-stance. Finally, D4 allows for the future possibility of new-proofs-for-existing-stances, existing-proofs-applied-to-new-stances (*i.e.* stance-modification in existing proofs), and new-proofs-for-new-stances.

Although intentionally accommodating a maximal multiplicity of views, D4 is not completely view-free: it references not-yet-defined *consciousness research*, and in this respect might allow for one of a number of views, or the simultaneous accommodation of many not-yet-arbitrated-between views. Relatedly, one might distinguish in kind between philosophical advances that radically alter some (not-yet-here-articulated) definition of consciousness research, and those that take place within, and relative to, some fixed definition. For definiteness, it will be useful to term the former kind of advances *zeroth-order innovations*, and various forms of the latter kind as first-order, second-order, and so forth (generically, *non-zeroth-order*).

The target outcome of the overall approach can now be clarified as follows. First, concerning the definition of consciousness research, we initially adopt a rough *de facto* view that the field comprises arbitration between various theories-of-consciousness framed as association-relationships between brain-dynamical phenomena and contents-of-conscious-experience phenomena. (Other field-definitional views will be discussed [10.1], notably the definition of consciousness as *function* without reference to *experience*.) In this broad initial characterization, arbitration may in general draw on empirical data (including first-person report), and both metaphysical and physics-theoretical advances. Second, a sub-goal of developments will be to further articulate this rough initial definition into a phenomenon-symbol characterization that will be claimed maximally concept-free *in the absence of zeroth-order innovations*, relative to any sufficiently field-delineating view of consciousness research. Third, the target character for proofs of central results will be that they are immune to (*i.e.* hold under) non-zeroth-order innovations, but not necessarily to zeroth-order advances. This finally establishes the slightly-restricted-relative-to-D4 character of the ultimate setting (*i.e.* after relinquishing initially-adopted D5-limitations): clarifying now that D4 is meant to *accommodate* the possibility of future zeroth-order innovation relative to any currently-tentative

definition of consciousness research (thus establishing D4-settings as maximally-inclusive), proofs of central results will be claimed to hold in the D4-subset that *excludes* (only) zeroth-order advances.

The target standard for proofs, notably reflecting D4-subset limitations, is now encapsulated in:

> **D6**. A *philosophy-neutral proof* in consciousness research is a proof that holds under a presently maximally-concept-free definition of consciousness research, and then under all metaphysical stances that are either known now, or derive from future non-zeroth-order innovation.

2.5. Operational settings for the philosophy-neutral advancement of consciousness research

For expository clarity, it will be helpful to elaborate D5-neutrality into an explicit statement of its operational consequences for consciousness research (Fig. 1):

> **D7**. The *standard scientific-empiricism stance* labels the view that the current situation in consciousness research is the maximally-inferred philosophy-neutral setting, and that theory-arbitration in consciousness research therefore is fully equivalent to an orthodox dependent-variable/independent-variable empirical problem directly analogous to those in conventional physics and biology (specifically, with contents-of-conscious-experience as the dependent variable, and detailed-brain-dynamics as the independent variable).

As mentioned in the preceding subsection, D5-neutrality and its D7-expression will ultimately be relinquished to construct D6-proofs, which can be operationally characterized via:

> **D8**. The *standard-empiricism-plus-insight stance* labels the view that standard scientific-empiricism in consciousness can, at least in principle, be non-trivially extended via metaphysical or theoretical insights. Notably, standard-empiricism-plus-insight allows for the possibility that theory-arbitration in consciousness research need not be precisely equivalent to orthodox dependent-variable/independent-variable *empirical* problems, because at least some dependent/independent-variable relationships (*i.e.* theories-of-consciousness) might be excluded on *metaphysical or physics-theoretical* grounds.

Note that D8 is inclusive of (rather than complementary to) D7, in the sense that D8 might be reduced to D7 if future D8-feasible advances show that the maximally-inferred philosophy-neutral D7-setting is the *only* consistent approach. Further, D8 does not limit *future* inferential settings to philosophy-neutrality of any kind (as D7 does), because advances might just as well establish the dominance of a single metaphysical stance (possibly even a presently-undiscovered view).

2.6. Approach

As preceding comments suggest, theoretical completeness-and-reliability considerations are

structured into two separate phases, respectively beginning from D7 and D8 assumptions.

2.6.1. D7-setting approach

D7-developments will concern completeness-required arbitration between two specific hypotheses "theory-I" and "theory-II" (introduced in the next Section and then later defined in detail). The capacity of any D7-framework to *reliably* perform this arbitration will be framed in terms of meter-reliability requirements in physical science, conventionally expressed as the necessary mathematical properties of a term H_{int} appearing in the Hamiltonian description of meter-system interaction. In order to analyze arbitration-reliability using physics-conventional mathematical expressions, it will be necessary to develop notation that links (*e.g.* visual) stimuli to encoding-dynamics, experience, and report, so that a Hamiltonian description of first-person-report-as-meter in consciousness research can be established. Critically (for both D7- and D8-settings), this notation will also both confer fundamental philosophy-neutrality to central reasoning and provide a vehicle for succinct general encapsulation of arbitrary theories-of-consciousness.

2.6.2. D8-setting approach

Whereas D7-developments address the capacity of any given experiment to perform specific theory-arbitrations in a scientifically-reliable way, D8-developments will concern the capacity of the field to define consciousness in a scientifically-reliable way. Further, whereas the D7-setting is definitively incapable of D6-philosophically-neutral-proofs (because D7-definition precludes consideration of possible metaphysical advances), results establishing conditions for the reliable definition of consciousness might inherently support certain philosophically-neutral proofs, because satisfaction of conditions for *un*reliability-of-consciousness-definition would preclude even the statement of various metaphysical stances whose possible advance-to-proven-status neutrality-of-proof must accommodate. Despite the consequently radically-different foundational status of D7- and D8-*results*, D8-*proofs* will draw on notation and concepts (notably H_{int}-defined meter-reliability) from the D7-approach. Methods will be transferred from the D7-setting to the D8-setting *without* dependence on any D7-assumptions.

3. Theory-I/II arbitration in the D7-setting and orthodox dynamical-closure

This Section introduces theory-I and theory-II and defines the centrally-important concept of orthodox dynamical-closure, relevant at this point for an orienting discussion of theory-I/II arbitration problems in the D7-setting.

3.1. Brain-dynamics and e-causality, conscious-experience and association

Transmission of physical information will frequently be referred to using *e-causal* terminology

defined in (Gamez, 2014), to which readers are referred for more details. The *e-causal* terminology precisely reflects the central conception here of physical law as encapsulation of inter-phenomenon dynamical regularities, and provides a carefully-established complementary means for expressing these regularities in causal rather than dynamical language.

Orderly relationships between brain-dynamical states and conscious experience will be described using *associated-with* terminology (Chalmers, 1998): the association of some aspect of conscious experience with a certain brain-dynamical state encapsulates order in a theory-of-consciousness without specification of metaphysical basis.

3.2. Theories I and II: granularity of external-environment encoding

Theories I and II are two alternative hypotheses concerning detail-level in conscious experience, and, correspondingly, concerning spatiotemporal brain-structures associated with conscious experience. To establish a concrete context, we will generally assume an imaginary future in which perfect and complete neuroanatomical and neurophysiological data is available (including the ability to image human brain activity *in vivo e.g.* to the molecular level, as well as tentatively-complete understanding of all the various levels of biophysics underlying such brain activity.) Let S0, S1, and S2 label a hierarchical sequence of neural structures involved in visual information-processing. For example, S0 might correspond to the retina, S1 to V1, and S2 to an area such as V2 or V3. Let U2 label some other (not-V1-identical) area or system hypothetically thought to be associated with conscious experience, for example the precuneus, or some other part of parietal or frontal cortex. Theory-I (Fig. 2A) is then the hypothesis that conscious experience is associated with S1 (*i.e.* V1), and that conscious experience reflects fine-grained spatial detail associated with relatively narrow S1/V1 receptive-fields. In contrast, theory-II (Fig. 2B) proposes that conscious experience is associated with U2, that U2 receptive-fields (or information-representations more generally) are coarsely-grained relative to those of S1/V1, and that conscious experience itself is therefore relatively coarse-grained, compared to that hypothesized under theory-I. (The simplifying attribution of a clear division in receptive-field scale between S1 and other cortical areas is purely expository; central theorems do not depend on this discursive simplification, as long as there is some clear division in some aspect of representation for some degree-of-freedom. This must hold unless every brain location encodes every degree-of-freedom at the same detail-level, which would itself preclude *any* theory-arbitration.) Note that the identification of S1 with V1 is purely to increase expository definiteness; all that matters for central arguments is that theory-I and theory-II propose distinct encodings-and-associated-experience.

3.3. Orthodox dynamical-closure

Under both D7- and D8-assumptions, in-principle possibilities for complete-and-reliable methodologies will depend directly on whether *orthodox dynamical-closure* is the correct description of natural order:

D9. *Orthodox dynamical-closure* is the stance that the dynamical evolution of all physical

observables relevant to macroscopic brain-function is fully determined by currently-known microphysical laws.

Relatedly, D9-dynamical-closure will help precisely track theory-I/II mechanisms for information transmission from S1 or U2 to report-governing areas (generically labelled W), discussed next.

3.4. U2-constrained and non-U2-constrained information-transmission contexts

Initially, D7-setting completeness-and-reliability will be investigated assuming that unarguable inferences from perfect future information conclusively limit neural information-transmission as follows:

> **D10.** The *U2-constrained information-transmission context* for theory-I/theory-II discrimination contains empirical (*e.g.* neurophysiological and neuroanatomical) observations that definitively exclude orthodox-e-causal transmission of information (Shannon and Weaver, 1964) at the detail-level of S1 to report-governing brain-areas W; in contrast, orthodox-e-causal transmission of all U2-details to W-areas is definitively supported.

[Note that D10 is a field-typical assumption, in that hierarchical information-processing with inter-level-bottlenecks is a prevalent model *e.g.* Fig. 4 in (Gamez, 2014)]. Although reasoning will be developed further below, the significance of D10 is perhaps self-evident: if report is limited to U2-level information, no report can be made of S1-level conscious-experiential details, leading to inaccessibility-based problems (Figs. 3 and 4) for theory-I/theory-II discrimination.

Note that D10 by itself *only* limits report to U2-level information *if D9-orthodox-closure is true* (because D10 refers to exclusion of *orthodox* information-transmission). Later developments will show that, even relaxing D10 to allow non-U2-constrained information-transmission, D9-orthodox-closure still leads to theory-I/II arbitration problems (Fig. 5). Detailed examination of these observations will be the core of D7-setting analyses.

4. Symbol-expressed notation for conscious experience and related brain states

This Section initiates symbol-notation for describing theories I and II that will ultimately enable central H_{int}-analyses. (Table 1 summarizes symbolism developed in this and later Sections.)

4.1. Notations s_{ab} for stimuli and $\text{x}\mathbf{r}_j$

Modern neuroscience has established various relationships between external stimuli and brain-dynamics (Churchland and Sejnowski, 1992; Rieke *et al*., 1997; Quiroga and Panzeri, 2013). To encapsulate these, the notation s_{ab} is introduced to label the *b*-th instance of the *a*-th stimulus class. For example, in visual neuroscience (Werner and Chalupa, 2004; Marr, 2010), stimulus *classes* include color, oriented edges, and motion. For the class of colors, a color *instance* is a particular

combination of wavelengths that produce a stereotypical neural response. For s_{ab} symbols, let the value a = 1, say, index the class of color stimuli; then various symbols s_{1b} label various neurally-distinguishable colors.

Stimuli occurring at different locations in the external environment give rise to brain dynamics at characteristic brain locations [*e.g.* within topographic maps (Kaas, 1997)]. Thus, describing the relation between brain dynamics and external stimuli requires symbolism for stimulus *location*. For present developments, it suffices to assume that sensory organs sample stimuli at a finite set of locations labeled ${}_{\mathrm{X}}\mathbf{r}_i$ (where i is an integer index). Thus, the entire neurally-relevant external environment can be described by defining a set $\{s_{ij}({}_{\mathrm{X}}\mathbf{r}_k)\}$ of stimuli instances at specific locations (Fig. 6).

4.2. $\langle\ldots\rangle(\boldsymbol{\rho}_j)$-notation for conscious experience

We assume that presentation of a particular stimulus s_{ab} at a particular external location ${}_{\mathrm{X}}\mathbf{r}_k$ e-causes certain stereotypical brain-dynamics that are in turn associated with a particular component-of-conscious-experience $\langle s_{ab}\rangle$ at a specific location (Fingelkurts *et al.*, 2010; Wagner, 2006) within conscious experience that will be labeled $\boldsymbol{\rho}_k$. Thus, the symbols $\{\langle s_{ij}\rangle(\boldsymbol{\rho}_k)\}$ comprise a detailed and exhaustive labeling of *e.g.* the (realized) degrees-of-freedom in conscious visual experience (Fig. 6).

The expository significance of Figs. 6 and 7 can hardly be over-emphasized, because they establish philosophy-neutral properties of $\langle s_{ij}\rangle(\boldsymbol{\rho}_k)$-symbols. First, solid-blue double-headed arrows in Fig. 6 *defining* $\langle s_{ij}\rangle(\boldsymbol{\rho}_k)$-symbols as the notational correspondents of degrees-of-freedom in conscious-experience are structurally analogous to similar arrows defining the symbol A that is part of conventional physics. This points to an *ontology-independence* shared by $\langle s_{ij}\rangle(\boldsymbol{\rho}_k)$-symbols and conventional physical-theory notation. Second, dashed-black double-headed arrows *associating* $\langle s_{ij}\rangle(\boldsymbol{\rho}_k)$-symbols with certain brain-dynamical structures in Fig. 7 can be given an unlimited number of interpretations according to metaphysical stance. This points to a *philosophy-neutrality* inherent in definition and inherited in usage of $\langle s_{ij}\rangle(\boldsymbol{\rho}_k)$-symbols.

4.3. Degrees-of-freedom and the phenomenon/symbol approach

One conception of $\langle s\rangle$-symbols is that they label *degrees-of-freedom in conscious experience*:

> **D11**. The *degrees-of-freedom in conscious experience* are components of (phenomenal) subjective experience that vary systematically with presentations of distinct objective stimuli in the external environment; for example, various color experiences at a particular conscious-experiential location are a "degree-of-freedom in conscious experience".

Degrees-of-freedom terminology is simply an extension of the concept of phenomenal qualities inherent in conscious experience, that points specifically to various orderly varieties (color, edges, motion, and so forth). Defining symbols to correspond to degrees-of-freedom in natural phenomena (and then stating and testing hypotheses for inter-symbol relationships as abstractions of inter-

phenomenon natural law) is methodologically-foundational in physics (Fig. 8); it will be referred to as the *phenomenon-symbol* approach.

The generic ⟨*s*⟩-symbol can be construed as a label for the *whole class* of degrees-of-freedom (in contrast to specific elaborations such as ⟨s_1⟩ and ⟨s_2⟩ that label certain *specific realization*, say conscious-experience-of-red and conscious-experience-of-green). This construal of ⟨*s*⟩-symbolism allows succinct differentiation between the characterization of spatiotemporal structures in terms of conventional physical properties (spatial configuration, energy *etc.*) and the ⟨*s*⟩-labeled component-of-conscious-experience *associated with* structures. Emphatically, characterizing a conscious-experience–associated structure in terms of both orthodox physical symbols and ⟨*s*⟩-labeling *neither* requires any metaphysical separation between physical and conscious-experiential properties, *nor* asserts that conscious experience cannot be explained within orthodox physics. As a simple and direct example, suppose that some neuron whose firing-rate is above a certain level is identical-with (Smart, 1959; Place, 2012) conscious-experience-of-red. When firing-rate is sufficiently high, the neuron is described both by orthodox-physical (*e.g.* electrochemical) measures related to firing-rate, and by an ⟨s_1⟩-label capturing conscious-experiential association, under two metaphysical assumptions: first, neuron-firing-above-rate "is" *both* orthodox-physical-description *and* ⟨s_1⟩-labeled conscious-experiential-component; and second, 'thing'-labeled-by-orthodox-physical-symbols "is" 'thing'-labeled-by-⟨s_1⟩-symbol.

Nor does ⟨*s*⟩-labeling limit us to identity- or even physicalist-theories. For example, consider instead the substance-dualist view that the same neuron, when firing vigorously enough, "produces" a component-of-consciousness made out of "mind stuff". Then ⟨s_1⟩ labels the conscious-experiential association of the mind-stuff-object so produced (whose description within *physical* theory might appropriately involve a distinct, new, symbol). And so on. Thus, critically, ⟨*s*⟩-symbols (and their various elaborations) are consistent with *philosophy-neutral proofs* (D6), because they accommodate every metaphysical stance. (This claim will be further supported by ongoing developments: of course, the use of ⟨*s*⟩-symbols does not *by itself* suffice for philosophy-neutral-proofs.)

4.4. Notation *A* for brain-dynamical basis of encoding and conscious experience

Let *A* denote a measure of brain-dynamical state (*e.g.* electric-field potential) measured at various encoding-relevant brain locations ${}_{\mathrm{A}}\mathbf{r}_j$ (Fig. 6). The orderly e-causal relationship between stimuli and brain dynamics can be encapsulated in a zero-one–valued function C_{abi} on brain dynamics, with

$$P\,[s_{ab}({}_{\mathrm{X}}\mathbf{r}_i)] \rightarrow_{\mathrm{e}} C_{abi}\,[\{A({}_{\mathrm{A}}\mathbf{r}_j)\}] = 1 \qquad (1)$$

and

$$N_P\,[s_{ab}({}_{\mathrm{X}}\mathbf{r}_i)] \rightarrow_{\mathrm{e}} C_{abi}\,[\{A({}_{\mathrm{A}}\mathbf{r}_j)\}] = 0 \qquad (2)$$

Here, $\rightarrow_{\mathrm{e}}$ denotes e-causality in a typical context, P and N_P respectively denote the presence or non-presence of the particular stimulus $s_{ab}({}_{\mathrm{X}}\mathbf{r}_i)$ in the external environment, and C_{abi} is a function that encapsulates *brain-encoding of* the presence of s_{ab} at ${}_{\mathrm{X}}\mathbf{r}_i$. The statement "$C_{abi} = 1$" means that "brain

state contains the stereotypical dynamics encoding $s_{ab}({}_{X}\mathbf{r}_i)$" (The mathematical form of C_{abi} delineates those $\{A({}_{A}\mathbf{r}_j)\}$-combinations that constitute these stereotypical dynamics.) Using "$\rightarrow_a$" to denote D3-naturalism-evoked lawful-association between components-of-experience and certain encoding-dynamics, simple theories-of-consciousness then take the general form

$$C_{abi}\,[\{A({}_{A}\mathbf{r}_j)\}] = 1 \rightarrow_a S\,[\langle s_{ab}\rangle(\boldsymbol{\rho}_i)] \tag{3}$$

and

$$C_{abi}\,[\{A({}_{A}\mathbf{r}_j)\}] = 0 \rightarrow_a N_S\,[\langle s_{ab}\rangle(\boldsymbol{\rho}_i)] \tag{4}$$

[although later elaborations will account *e.g.* for dependency of conscious-experiential association on aspects of brain state other than just *C*-classified *encoding*-dynamics; *S* and N_S denote the *subjective* presence or non-presence respectively of $\langle s_{ab}\rangle(\boldsymbol{\rho}_i)$ as a component-of-conscious-experience]. Notation thus far simply symbol-expresses detailed relationships that are typically assumed in naturalistic consciousness research. [For example, Eq. 1 states that a particular stimulus at a particular location puts some subset of brain activity into a stereotypical state, and Eq. 3 says that when the brain is in this stereotypical state, that activity-subset is associated with a particular component-of-consciousness at a particular conscious-experiential location (Fig. 7).] To reiterate, using symbol-expressed rather than verbal analyses has certain critical benefits here: it enables later H_{int}-analyses, and provides both philosophical (ontological and consciousness-metaphysical) neutrality, and theory-of-consciousness generality. (Further, Eqs. 1-2 and their generalizations abstractly represent the entire neuroscientific encoding-literature, thus linking first steps firmly to empirical accounts.)

5. Symbol-expressed notation for hierarchy of environmental spatial scales

This Section extends s-$C[A]$-$\langle s\rangle$ symbolism to facilitate labeling of different granularities of resolution (spatial scales) associated with theories I and II, a vital step towards H_{int}-analysis.

5.1. Hierarchical description of the external environment

We now extend s_{ab} notation by defining ${}_{l}s_{ab}$ symbols that describe the external environment at the *l*-th level of granularity (Fig. 9). Here, the finest level of description will be denoted by ${}_{0}s$, and the coarsest (the entire exteroceptive environment) by ${}_{4}s$. The appropriate level of description for ${}_{0}s_{ab}$ symbols can be set *e.g.* by neurophysiology of sensory organs. Successively higher-level ${}_{l}s$ symbols ($l > 0$) are then successively coarser-grained descriptions corresponding to expansions of receptive field in successively higher cortical areas. For definiteness, the ${}_{0}s$ level will be taken to correspond to retinal-cell spatial-resolutions, the ${}_{1}s$ level to correspond to S1/V1, and the ${}_{2}s$ level to U2.

A basic assumption (Fig. 9F) here is that the external environment can be described in mutually consistent ways by sets of symbols $\{{}_{0}s_{ab}({}_{X}\mathbf{r}_{0,j})\}$, $\{{}_{1}s_{cd}({}_{X}\mathbf{r}_{1,k})\}$, and so on (symbols ${}_{X}\mathbf{r}_{i,j}$ label the set of point-locations relevant to the ${}_{i}s$-level). (Although stimulus-environments can be constructed in which potential conflicts occur between different level-descriptions, consideration of settings without

such conflicts suffices for present purposes.)

5.2. Symbol-expressed statements of theories I and II

Introduction of $\{{}_{l}s_{ab}({}_{X}\mathbf{r}_{l,\ j})\}$ symbols evokes notation $\boldsymbol{\rho}_{l,\ j}$ for conscious-experiential locations corresponding to the *l*-th environmental-description-level. Theory-I can then be stated as

$$C_{1,abc}[\{A({}_{A}\mathbf{r}_{1,\ j})\}] = 1 \rightarrow_{a} S\ [\langle {}_{1}s_{ab}\rangle(\boldsymbol{\rho}_{1,c})] \tag{5}$$

and theory-II as

$$C_{2,def}[\{A({}_{A}\mathbf{r}_{2,\ j})\}] = 1 \rightarrow_{a} S\ [\langle {}_{2}s_{de}\rangle(\boldsymbol{\rho}_{2,\ f})] \tag{6}$$

Eq. 5 states that conscious experience contains ${}_{1}s$-level detail, and this conscious-experiential detail is associated with certain brain-dynamical activity in S1 at locations $\{{}_{A}\mathbf{r}_{1,\ j}\}$. Eq. 6 states instead that conscious experience occurs at the ${}_{2}s$-granularity-level, associated with certain brain-dynamical activity in U2 at locations $\{{}_{A}\mathbf{r}_{2,\ j}\}$. (Note that theory-II allows an e-causal role for S1-dynamics in generating U2-dynamics, but limits the conscious-experiential association to U2.)

6. The virtue-of-⟨*s*⟩-coupling alternative to orthodox dynamical-closure

Earlier discussions (“*U2-constrained and non-U2-constrained information-transmission contexts”*) provided preliminary orientation to completeness-and-reliability problems under orthodox dynamical-closure. This Section defines *virtue-of-⟨s⟩-coupling*, which will be the D9-orthodox-closure alternative critical to completeness-and-reliability.

6.1. Violation of orthodox dynamical-closure

D9-alternative accounts will be described as *violations* of orthodox dynamical-closure, although the general tenor of D9-alternatives here is not to deny microphysical law, but rather to hypothesize hitherto-unexamined extensions, leading *e.g.* to the addition of terms to orthodox theories. In this sense, one might speak instead of *variations to* orthodox-closure conceptions.

6.1.1. Plausibility of violation

Questions of “plausibility of” or “justifications for” D9-violations will not be given detailed attention, because the stance here is that experimental tests supersede subjective preferences in a scientific approach, and experimental tests for alternatives to orthodox closure will be identified. The justification for *experimentally-testing* D9-closure will become self-evident: after central theorems,

subjective assertion (arguably unscientific, given the existence of experiments) that orthodox dynamical-closure "must" be true would simply establish the impossibility of creating a complete-and-reliable framework for consciousness research. Accordingly, serious consideration of D9-violations is evoked by any serious desire for completeness-and-reliability.

6.1.2. Violation does *not* require non-physical entities or influences

Conventional interpretation of "causal closure of the physical" means that its violation implies the existence of "non-physical" entities, putatively acting as causal antecedents for closure-violating observations. Note that *violation of orthodox dynamical-closure does not require "non-physical" causes*. Specifically, microphysical laws can be extended or amended so that certain spatiotemporal-structures acquire novel dynamical-couplings not present in current statements, while physical law still refers only to a pre-fixed set of labels for physical entities (*e.g.* Standard Model particles, and their aggregations *e.g.* into atoms, molecules, cells *etc.* that constitute spatiotemporal brain structures of present focus).

6.1.3. Why violation of orthodox dynamical-closure differs from violation of causal closure

Directly, orthodox dynamical-closure violation simply does not require causal-closure violation: dynamical evolution can be more differentiated than the account in orthodox laws, without new non-physical entities as causes for differences. In more detail, causal closure: presumes *both* a known-to-be-fixed set of entities termed "the physical" *and* already-complete knowledge of laws governing dynamical evolution; *conflates* these presumptions by incorrectly inferring that observations violating the latter (dynamical laws) necessitates variation in the former (entities); and, *imposes* the constraint that any new entity must be called "non-physical". Notably, *none* of these features of causal closure is shared by serious modern approaches to physical theory, whose basic character is the target of the present approach: D9 *identifies* (or labels) a particular set of laws governing dynamical evolution, rather than presuming their fixity, and the present approach suggests that whether D9-orthodoxy or some other basically-coherent view pertains is appropriately an *experimental* question.

6.2. "Virtue-of-$\langle s \rangle$" violation of orthodox dynamical-closure

Given notational developments thus far, one can speak of $\langle s \rangle$-labeled structures (those associated at a particular time with components-of-conscious-experience) as opposed to non-$\langle s \rangle$-labeled structures. (A particular brain area or structure will generally be $\langle s \rangle$-labeled in different ways at different times. For example, no structure is $\langle s \rangle$-labeled in deep sleep. The firing-rate-dependent association of the neuron discussed above means that an $\langle s_1 \rangle$-label pertains only at sufficiently high firing rates. And so forth.) A theory-of-consciousness is then a detailed proposal concerning the orthodox-physics-expressible conditions under which specific $\langle s_j \rangle$-labels pertain to particular brain-dynamical structures. (Somewhat analogously, a theory-of-temperature-T can be construed as a detailed proposal concerning the orthodox-physics-expressible conditions under which specific T-labels

pertain to particular physical locales.) With these understandings established, we now state:

> **D12.** *Virtue-of-⟨s⟩-label coupling* extends currently-accepted microphysical laws of dynamical evolution by bestowing non-orthodox and non-trivial e-causal influence on (and only on) ⟨*s*⟩-labeled spatiotemporal structures that are associated with realized conscious-experiential degrees-of-freedom.

6.3. Three illustrative examples of virtue-of-⟨*s*⟩-coupling realization

Three examples now illustrate more concretely how virtue-of-⟨*s*⟩-coupling can occur: first, substance-dual conscious-experience-associated entities might couple with orthodox-materialist brain-dynamics; second, "downward causation" (Chalmers, 2008) can accommodate *e.g.* a conscious-experience-associated sensory-encoding brain-*area* coupling with a report-involved-*neuron*; third, unconventionally-monist interpretations of (Eccles, 1990) can extend orthodox physical theory to admit interactions between conscious-experience-association–dedicated-particles and Standard-Model-orthodox-particles. Emphatically, these three *orienting examples* do not exhaust realization-possibilities for virtue-of-⟨*s*⟩-coupling (nor does their choice signify "leading candidate" status; rather, examples are selected to indicate the wide physical/metaphysical range of possibilities).

6.4. Completeness-and-reliability significance for virtue-of-⟨*s*⟩-label coupling

D12 defines a particular violation of D9, directly linked to the conscious-experiential associations of physical structures. Central results will show that the possibility of complete-and-reliable frameworks depends on whether orthodox-dynamical-closure or virtue-of-⟨*s*⟩-coupling pertains in our Universe. As the next step towards these results, the following Section constructs two different symbol-expressions for first-person report, respectively consistent with D9-orthodox-dynamical-closure or D12-virtue-of-⟨*s*⟩-coupling.

7. Symbol-expressed notation for different types of report

7.1. Decoding of brain dynamics and three distinct behavioral effects of brain dynamics

The most experimentally-fundamental way to understand the connection between environmental stimuli and brain-dynamics is to construct an explicit account of the function $C_{l,abj}[\{A({}_{\mathrm{A}}\mathbf{r}_{l,k})\}]$ (Adrian, 1928; Haynes and Rees, 2006). Given such accounts together with detailed direct measurements of activity $\{A({}_{\mathrm{A}}\mathbf{r}_{l,k})\}$, and writing "$\rightarrow_{\mathrm{e,I}}$" to denote third-party inferential-*decoding* of brain-dynamics, these analyses can be expressed as

$$C_{l,abj}[\{A({}_{\mathrm{A}}\mathbf{r}_{l,k})\}] = 1 \rightarrow_{\mathrm{e,I}} P[\,{}_{l}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{l,j})] \qquad (7).$$

("$\rightarrow_{\mathrm{e,I}}$" notation derives its subscripts from implicit characterization of third-party decoding as an e-

caused activity together with an explicit *I*-acknowledgement that *P*-notation on the RHS is a typical-context-based *inference*; *e.g.* artificial stimulation might satisfy *C*-conditions without *P*-presence of stimuli).

At the other end of the experimental spectrum, a system (*e.g.* a human subject) may make a statement such as "I consciously experience $\langle {}_{l}s_{ab}\rangle$ at $\boldsymbol{\rho}_{l,\,j}$". Report of this kind will be termed *first-person report* (*i.e.* report-data relating explicitly to conscious-experiential contents). One way to explain first-person report is that it originates in brain-dynamical activity encoding the *stimulus* ${}_{l}s_{ab}$ at ${}_{\mathrm{X}}\mathbf{r}_{l,\,j}$. If R_{ω} labels report, this explanation can be written as

$$\{A\}\text{: } (C_{l,abj}[\{A({}_{\mathrm{A}}\mathbf{r}_{l,k})\}] = 1) \;\rightarrow_{\mathrm{e}}\; R_{\omega}\,[\text{“}\langle {}_{l}s_{ab}(\boldsymbol{\rho}_{l,\,j})\rangle\text{”}] \qquad (8)$$

(The ω-subscript on R denotes a report that is made by an orthodox material system; the ω-symbol is a generic placeholder for ω_j symbols labeling various physically-fundamental entities *e.g.* in the Standard Model of particle-physics. Phenomenon-symbol descriptions of certain metaphysical accounts [*e.g.* (Chalmers, 1996)] require beyond-Standard-Model ω'-labeled entities; note both that ω'-entities need not be "non-physical", and that *any* brain-dynamical-coupling with ω'-entities is *excluded* by D9-orthodoxy.)

Of central importance will be another first-person-report explanation available under virtue-of-$\langle s\rangle$ coupling. Eq. 8 is to be read as attributing the e-causal origin of R_{ω} to the physically-orthodox *A*-state that encodes ${}_{l}s_{ab}(\boldsymbol{\rho}_{l,\,j})$. ("$\{A\}$: …" notation in this reading denotes "those *A* such that …".) Alternatively, the e-causal origin of R_{ω} ["$\langle {}_{l}s_{ab}(\boldsymbol{\rho}_{l,\,j})\rangle$"] can explained via

$$\langle s\rangle\text{: } (C_{l,abj}[\{A({}_{\mathrm{A}}\mathbf{r}_{l,k})\}] = 1 \rightarrow_{\mathrm{a}} S\,[\langle s\rangle]\,) \;\rightarrow_{\mathrm{e},\langle s\rangle}\; R_{\omega}\,[\text{“}\langle {}_{l}s_{ab}(\boldsymbol{\rho}_{l,\,j})\rangle\text{”}] \qquad (9),$$

emphasizing a (hypothetical) e-causal pre-requisite for R_{ω} as the $\langle s\rangle$-labeled property associated with *A*-states satisfying *C*-condition. Critically, "$\rightarrow_{\mathrm{e},\langle s\rangle}$" notation declares a virtue-of-$\langle s\rangle$ *novelty* in dynamical coupling, so that Eq. 9 is not simply a restatement of Eq. 8. Identity-theorists – and others – might protest it is not meaningful to separate *C*-satisfying *A*-states from $\langle s\rangle$-labels. But a basic feature of the phenomenon-symbol approach established in Figs. 6 and 7 is that no such separation is implied. To see this, restate Eq. 9 as

$$\{A\}\text{: } (C_{l,abj}[\{A({}_{\mathrm{A}}\mathbf{r}_{l,k})\}] = 1) \;\rightarrow_{\mathrm{e},\,\langle s\rangle}\; R_{\omega}\,[\text{“}\langle {}_{l}s_{ab}(\boldsymbol{\rho}_{l,\,j})\rangle\text{”}] \qquad (10)$$

leaving the "$C = 1 \rightarrow_{\mathrm{a}} S\,[\langle s\rangle]$" component of Eq. 9 to follow from Eq. 3.

Report of stimulus but not of conscious experience will be written as R_{ω} ["${}_{l}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{l,\,j})$"], here termed *stimulus report*. The conceptual distinction between R_{ω} ["${}_{l}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{l,\,j})$"] and R_{ω} ["$\langle {}_{l}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{l,\,j})\rangle$"] is clear. Consider, for example, a camera/computer system capable of distinguishing between various stimuli. This system's reports are of an R_{ω} ["${}_{l}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{l,\,j})$"]-kind (under the conventional assumption of no conscious-experiential association). The explanation of R_{ω} ["${}_{l}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{l,\,j})$"] stimulus-report in terms of brain-dynamical information-encoding can be written as

$$\{A\}: C_{l,abj}[\{A({}_{\mathrm{A}}\mathbf{r}_{l,k})\}] = 1 \rightarrow_{\mathrm{e}} R_{\omega}\,[\text{“}{}_{l}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{l,j})\text{”}] \tag{11}$$

Note that the LHS of Eqs. 8 and 11 are *identical*, whereas the LHS of Eq. 9 and its "→" differ in kind; this points to problems in disentangling first-person and stimulus-reports under orthodox dynamical-closure (Eq. 8) that might be avoidable under virtue-of-⟨*s*⟩-coupling (Eq. 9).

Finally, non-verbal behavior (*e.g.* pupillary dilation and skin conductance) can be influenced by a stimulus, without even an in-principle ability to make a verbal report of either stimulus or conscious experience. These kinds of observable-data will be termed *non-verbalizable report*, and denoted by $R_{\omega}\,[{}_{l}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{l,j})]$. (Note that both first-person and stimulus report can in principle be made via non-verbal means, such as a button push. The key defining factor of non-verbalizable report here is that it *cannot* be made by verbal systems.) Explanation of non-verbalizable report in terms of brain dynamics can be written

$$C_{l,abj}[\{A({}_{\mathrm{A}}\mathbf{r}_{l,k})\}] = 1 \rightarrow_{\mathrm{e}} R_{\omega}\,[{}_{l}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{l,j})]. \tag{12}$$

[Eqs. 7 and 12 are given for completeness, as contextual clarification-by-contrast of centrally-relevant first-person (Eqs. 8-10) and stimulus (Eq. 11) reports.]

7.2. Symbol-expressed statements of theories I and II including report assumptions

R_{ω}-notation enables further detail in symbol-statements of theories I and II. Under D10 (assumed until stated otherwise) both theories attribute the e-causal origin of first-person report to information encoded in $\{A({}_{\mathrm{A}}\mathbf{r}_{2,j})\}$ states (*i.e.* the states of U2). Adding this fact to Eq. 5 (reproduced as Eq. 13) gives an extended statement of theory-I as

$$C_{1,abc}[\{A({}_{\mathrm{A}}\mathbf{r}_{1,i})\}] = 1 \rightarrow_{\mathrm{a}} S\,[\langle {}_{1}s_{ab}\rangle(\boldsymbol{\rho}_{1,c})] \tag{13}$$

$$C_{1,abc}[\{A({}_{\mathrm{A}}\mathbf{r}_{2,i})\}] = 1 \rightarrow_{\mathrm{e}} C_{2,def}[\{A({}_{\mathrm{A}}\mathbf{r}_{2,j})\}] = 1, \quad {}_{1}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{1,c}) \subset {}_{2}s_{de}({}_{\mathrm{X}}\mathbf{r}_{2,f}) \tag{14}$$

and

$$\{A\}: C_{2,def}[\{A({}_{\mathrm{A}}\mathbf{r}_{2,j})\}] = 1 \rightarrow_{\mathrm{e}} R_{\omega}[\text{“}\langle {}_{2}s_{de}\rangle(\boldsymbol{\rho}_{2,f})\text{”}]. \tag{15}$$

Eq. 14 clarifies that activity in $\{A({}_{\mathrm{A}}\mathbf{r}_{1,i})\}$ encoding ${}_{1}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{1,c})$ e-causes activity in $\{A({}_{\mathrm{A}}\mathbf{r}_{2,j})\}$ encoding ${}_{2}s_{de}({}_{\mathrm{X}}\mathbf{r}_{2,f})$, where ${}_{2}s_{de}({}_{\mathrm{X}}\mathbf{r}_{2,f})$ is the ${}_{2}s$-level stimulus that includes ${}_{1}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{1,c})$ [written symbolically as ${}_{1}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{1,c}) \subset {}_{2}s_{de}({}_{\mathrm{X}}\mathbf{r}_{2,f})$;. Fig. 9] Eq. 15 clarifies that $\{A({}_{\mathrm{A}}\mathbf{r}_{2,j})\}$ e-causes report of conscious experience at the ${}_{2}s$-level, despite the fact that conscious experience contains ${}_{1}s$-level detail (Eq. 13). (For precision, we declare that Eq. 15 is a complete account of *all possible* first-person reports in theory-I.)

Similarly, Eq. 6 for theory-II (reproduced below as Eq. 17) can be complemented by two equations respectively clarifying relationships between $\{A({}_{\mathrm{A}}\mathbf{r}_{1,i})\}$ and $\{A({}_{\mathrm{A}}\mathbf{r}_{2,j})\}$ dynamics and between ${}_{l}s$-

symbols encoded in these dynamics (Eq. 16), and clarifying that first-person report employs information encoded in $\{A({}_{\mathrm{A}}\mathbf{r}_{2,j})\}$ dynamics (Eq. 18):

$$C_{1,abc}[\{A({}_{\mathrm{A}}\mathbf{r}_{1,i})\}] = 1 \rightarrow_{\mathrm{e}} C_{2,def}[\{A({}_{\mathrm{A}}\mathbf{r}_{2,j})\}] = 1, \quad {}_{1}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{2,c}) \subset {}_{2}s_{de}({}_{\mathrm{X}}\mathbf{r}_{3,f}) \tag{16}$$

$$C_{2,def}[\{A({}_{\mathrm{A}}\mathbf{r}_{2,j})\}] = 1 \rightarrow_{\mathrm{a}} S\,[\langle {}_{2}s_{de}\rangle(\boldsymbol{\rho}_{2,f})] \tag{17}$$

$$\{A\}\colon C_{2,def}[\{A({}_{\mathrm{A}}\mathbf{r}_{2,j})\}] = 1 \rightarrow_{\mathrm{e}} R_{\omega}[\text{“}\langle {}_{2}s_{de}\rangle(\boldsymbol{\rho}_{2,f})\text{”}] \tag{18}$$

(Just as for Eq. 15, Eq. 18 is to be read as a full theory of first-person report. Additionally, Eq. 17 must be read as a *full* theory-of-conscious-experience, notably excluding Eq.-13-style associations with $\langle {}_{1}s\rangle$-components.)

Note that Eqs. 15 and 18 have been written in the form of Eq. 8 rather than of (virtue-of-$\langle s\rangle$-coupling) Eqs. 9-10, because Eqs. 13 to 18 are intended as theory-I/II statements *under D9-dynamical-closure*.

8. Report, “*l*-resolution”, and the Hamiltonian formulation of meter-reliability

8.1. First-person report as a meter, and *l*-resolution requirements for D7-completeness

In notation developed thus far, the values $l = 1$ and $l = 2$ denote conscious-experiential detail-level pertaining respectively to theories I and II. We adopt the following definition:

D13. The act of determining the level l of detail actually present in conscious experience will be termed *l-resolution*.

D7-demonstrations start from an assumption that theories I and II are basically-coherent, *i.e.* specifically here, well-defined and field-relevant. [Concerning *definition*, preceding symbol-expressions and contextual discussions suffice. Concerning *field-relevance*, debates concerning *straightforward generalizations of* the theory-I/theory-II contrast are certainly ongoing.] Explicitly, we have:

A1. Theory-I and theory-II are basically-coherent in the D7-setting.

[It is not possible at this point to exclude *e.g.* theory-I on the grounds that it is not consistent with first-person report, because comprehensive-accuracy of first-person report is the point at issue here. Further, the D7-setting precludes metaphysical pre-empirical arbitrations such as functionalism-based rejection of theory-I.] D1 and A1 imply our first Lemma:

L1. A D1-complete framework in the D7-setting must be able to conclusively arbitrate between theory-I and theory-II, at least in principle.

L1 and D13 imply L2:

L2. In the D7-setting, every complete framework must be capable of *l*-resolution with respect

to theories I and II.

By definition, D7-settings conceive of *l*-resolution as an empirical problem (Fig. 1) of measuring dependent-variable states (whereas under D8 certain *l*-resolutions might be achieved solely by conceptual means). Because first-person-report is the sole D7-means for D7-dependent-variable measurement, we have:

> **L3**. In the D7-setting, theory-I/theory-II arbitration depends on first-person-report acting as an *l*-resolving meter.

In conjunction with L1, L3 establishes a necessary *l*-resolving-meter role for first-person report in a D1-complete D7-framework. Our general interest is in D1-complete-*and-D2-reliable* frameworks, so first-person report must also meet conditions minimally required in orthodox scientific fields for scientifically-reliable meter function; these conditions are conventionally expressed in Hamiltonian terms.

8.2. Scientific reliability of meters in orthodox physical-science: Hamiltonian formulation

8.2.1. Hamiltonian formulation of dynamical-coupling and dynamical-evolution

Briefly, the Hamiltonian formulation for the dynamical evolution of any physical object (particle, collection-of-particles, system, collection-of-systems) first defines object-state with respect to various *canonical coordinates*, then writes the *Hamiltonian function* for the object in terms of these coordinates, and finally computes temporal state-evolution using the Hamiltonian function within an object-independent *Hamiltonian formalism*. [Further details can be found in standard treatments *e.g.* (Goldstein *et al*., 2001)]. The only details of further present relevance are that canonical coordinates occur in pairs conventionally termed canonical locations p_i and associated canonical *conjugate-momenta* Π_i.

8.2.2. Conventional Hamiltonian formulation of scientifically-reliable meter function

The Hamiltonian description of meter function was originally developed as a basis for quantum measurement methodology (von Neumann, 1996), but trivially extends to classical measurements (of present central concern). The standard formulation writes a total Hamiltonian H_{tot} for the dynamical evolution of a combined measured-system/meter apparatus as

$$H_{\text{tot}} = H_{\text{system}} + H_{\text{meter}} + H_{\text{int}} \tag{19}$$

where H_{system} and H_{meter} describe the behavior of isolated system and meter, respectively, and H_{int} describes the system-meter interaction. Briefly, the standard approach (Bohm, 1951) says that some meter-canonical-coordinate p must interact with the to-be-measured system-property (here l), leading to a change $\delta\Pi$ in p's conjugate-momentum Π. Meter-readings are e-caused by $\delta\Pi$, and l-values are then inferred from these readings. A2(a) below reproduces the standard form of H_{int} that encapsulates basically-required properties; A2(b) adds obviously-necessary conditions on making experimental

preparations:

> **A2**. (**a**) The Hamiltonian description of an *l*-resolving meter must contain a meter/measured-system interaction term of the form
>
> $$H_{\text{int}} = -\, g(t)\, .\, f\,(p,\, l) \tag{20}$$
>
> where $g(t)$ describes time-evolution of system-meter interaction, and p is a canonical coordinate in the Hamiltonian description of the meter.
>
> **A2.** (**b**) Further, to meter-resolve two *l*-values, it must be possible to make preparations whose Hamiltonian description requires those *l*-values to appear on the RHS of Eq. 20.

Details of time-evolution $g(t)$ (and the role of conjugate-momentum Π) will not be of central concern. The key point from A2(a) is the intuitively-reasonable requirement that some meter-aspect p must interact with measured-property l in a way that leads to different meter-states. The key point more generally is that, in any D2-*scientifically-reliable* framework, every employment of first-person report must be conducted in an appropriately A2-consistent manner.

9. Central results in the D7-setting

Central results for the D7- and D8-settings will now follow rather rapidly, respectively in this Section and the next.

9.1. U2-constrained scientific-unreliability of first-person *l*-resolution under D9-closure

As noted following D10 ("*U2-constrained and non-U2-constrained information-transmission contexts*"), in a U2-constrained information-transmission context *and* assuming D9-orthodox-dynamical-closure, *only* $l = 2$ information is available to report-systems, so A2(b) cannot be satisfied. Thus:

> **L4.** In the U2-constrained information-transmission context (D10), orthodox dynamical-closure (D9) implies that first-person report is not a scientifically-reliable *l*-resolving meter.

More rigorously, L4 follows from the $l = 2$ suffix in Eqs. 15 and 18: in the present phenomenon-symbol setting (Figs. 6-8), R_ω-symbols in these equations label H_{int}-originating output of meters, thus establishing a precise link between H_{int}-analyses (A2) and symbol-based expressions of theory-I/II. Certainly, $l = 2$ restrictions formally apparent in Eqs. 15 and 18 (as the 2-subscripts in R_ω-expressions) can clearly be identified verbally. However, later D9-relaxing alternatives must show precisely how A2-requirements can be *satisfied*: symbolism will be able to establish this via Eq.-9-style amendments of currently-Eq.-8-based characterizations of report (*i.e.* via amendments of Eqs. 15 and 18).

L4 is less trivial than it might first appear, because it shows that, *e.g.* identity-theoretic, arguments

that report is e-causally determined by dynamics that "are" conscious experience do *not* lead to theory-I/theory-II-resolving reports of conscious-experiential states. L4 emphasizes that the critical factor for theory-I/theory-II arbitration is specifically *l*-resolution, *not* simply the ability to resolve and report different experiences at a fixed (and unknown) *l*-value.

9.2. Relaxation of U2-constraints on D9-scientific-reliability analyses

One limitation to arguments built on L4 is that future discoveries might falsify D10-style constraints (for example, by identifying novel thalamocortical paths enabling report-area access to S1-level detail). A demonstration that L4-scientific-unreliability pertains in non-constrained settings will both establish a more general and robust form of L4, and provide a useful further insight into the kinds of H_{int}-couplings generally required for completeness-and-reliability.

Directly, consider a new theory-III in which conscious experience is associated with *either* S1 *or* U2, according to some brain-dynamical state classified by a new function *X*, say. (Whereas *C* classifies encoding-dynamics, *X* classifies some *further* aspect of brain-state directly implicated in dynamically-switching the informational-basis and associative-relationships of conscious experience.) Specifically, we assume the following relationships:

$$C_{1,abc}[\{A({}_{\mathrm{A}}\mathbf{r}_{1,j})\}] = 1,\ X\,[\{A({}_{\mathrm{A}}\mathbf{r}_{y,z})\}] = 1 \rightarrow_{\mathrm{a}} \{A({}_{\mathrm{A}}\mathbf{r}_{1,j})\} : S\,[\langle {}_{1}s_{ab}\rangle(\boldsymbol{\rho}_{1,c})] \tag{21}$$

$$C_{2,abc}[\{A({}_{\mathrm{A}}\mathbf{r}_{2,j})\}] = 1,\ X\,[\{A({}_{\mathrm{A}}\mathbf{r}_{y,z})\}] = 2 \rightarrow_{\mathrm{a}} \{A({}_{\mathrm{A}}\mathbf{r}_{2,j})\} : S\,[\langle {}_{2}s_{ab}\rangle(\boldsymbol{\rho}_{2,c})] \tag{22}$$

$$X\,[\{A({}_{\mathrm{A}}\mathbf{r}_{y,z})\}] = 0 \qquad \rightarrow_{\mathrm{a}} N_S[\ldots] \tag{23}$$

In Eqs. 21-23: $\{{}_{\mathrm{A}}\mathbf{r}_{y,z}\}$ are a set of widely-distributed *A*-locations, for example including S1, U2, and W sites; "$\{A({}_{\mathrm{A}}\mathbf{r}_{l,j})\}:\langle {}_{l}s_{ab}\rangle(\boldsymbol{\rho}_{l,c})$" extends $\rightarrow_{\mathrm{a}}$ notation to specifically declare *which* $\{A({}_{\mathrm{A}}\mathbf{r}_{n,j})\}$ are directly associated-with conscious experience; and *X*=0-states correspond to the absence of conscious-experience-as-a-whole (written as $N_S[\ldots]$). Further, we assume that when *X*=1, *W*-report-systems utilize S1-level information, and when *X*=2, *W*-report-systems instead utilize U2-level information (leading to obvious amendments, not shown here, to R_ω-statements). For example, we might conceive of *X* as a measure of phase-coupling (synchrony or coherence): when S1 and W are phase-coupled (*X*=1), W accesses S1-information; when instead U2 and W are phase-coupled (*X*=2), W is limited to U2-information. (Other, more general, *X*-evoking conceptions exist.)

Although both $l = 1$ (*X*=1) and $l = 2$ (*X*=2) H_{int}-preparations are now possible, thus satisfying A2(b), scientifically-reliable complete arbitration still *cannot* be established, as will be shown first by a specific theory-III-related example, and then by a general consideration.

First, concerning theory-III specifically, consider a competing theory-IV (other alternatives can be constructed leading to the same conclusion), in which conscious-experience always contains *l*=1 detail and is always associated with $\{A({}_{\mathrm{A}}\mathbf{r}_{1,j})\}$,

$$X\,[\{A({}_{\mathrm{A}}\mathbf{r}_{y,z})\}] = 1 \rightarrow_{\mathrm{a}} \{A({}_{\mathrm{A}}\mathbf{r}_{1,j})\} : S\,[\langle {}_{1}s_{ab}\rangle(\boldsymbol{\rho}_{1,c})] \tag{24}$$

$$X\,[\{A({}_{\mathrm{A}}\mathbf{r}_{y,z})\}] = 2 \rightarrow_{\mathrm{a}} \{A({}_{\mathrm{A}}\mathbf{r}_{1,j})\} : S\,[\langle {}_{1}s_{ab}\rangle(\boldsymbol{\rho}_{1,c})] \qquad (25),$$

but W-*report*-system information-access depends on X precisely as in theory-III. In terms of the phase-coupling interpretation, in theory-IV phase-coupling plays only an information-transmission role and has no explicit effect on conscious-experiential associative-relationships or detail-level. Directly (Fig. 5), theory-III/IV arbitration is not possible under orthodox dynamical-closure, even though l-coupling requirements of both A2(a) and A2(b) now pertain.

To identify general conditions for resolution of problems of the theory-III/IV kind, we now further define l_s-resolution as the capacity to resolve *encoded-information* detail-level (that corresponds to *stimulus*-report: R_ω ["${}_{l}s_{ab}({}_{\mathrm{X}}\mathbf{r}_{l,\,j})$"]; Eq. 11) and $l_{\langle s\rangle}$-resolution as the capacity to resolve *conscious-experiential* detail-level (that corresponds to *first-person*-report: R_ω ["$\langle {}_{l}s_{ab}(\boldsymbol{\rho}_{l,\,j})\rangle$"]; Eqs. 8-10). Theory-III/IV problems originate in the fact that theories-of-consciousness with various $l_{\langle s\rangle}$-detail can be stated relative to (independently of) any l_s-stimulus-reporting conception. This underlying, generic, issue is D7-soluble only if $l_{\langle s\rangle}$ appears in H_{int}, and this is precisely precluded by orthodox-dynamical closure (and precisely rendered in-principle possible by virtue-of-$\langle s\rangle$-coupling).

9.3. First-person report scientifically-unreliable under orthodox dynamical-closure

Immediately-preceding discussions establish our first Theorem:

> **T1.** Under orthodox dynamical-closure, first-person report is not a scientifically-reliable $l_{\langle s\rangle}$-resolving meter.

Note that T1 is actually D7-independent (although its utility in D8-settings is compromised by the in-principle possibility of purely *conceptual* modes-of- $l_{\langle s\rangle}$-resolution that do not use first-person report).

9.4. Every complete framework scientifically-unreliable in D7-settings

In the D7-setting, L1, L3 (restated to require $l_{\langle s\rangle}$-resolution), and T1 imply:

> **T2.** In the D7-setting under orthodox dynamical-closure, every complete framework is scientifically-unreliable.

Trivial restatement of T2 gives a version directly responding to Q1 and Q2:

> **L5.** In the D7-setting under orthodox dynamical-closure, there are no complete and scientifically-reliable frameworks.

9.5. D7-extensions to orthodox dynamical-closure for completeness and reliability

Because theory-I/theory-II arbitration *must* use first-person-report-as-meter in D7-settings (L3), T1

further implies:

> **L6.** In D7-settings, if there is to be at least one complete and scientifically-reliable framework, orthodox dynamical-closure must be false.

L6 is stated simply to frame the following observation: not *every* falsification of orthodox dynamical-closure leads to a complete and scientifically-reliable framework – of central interest (*e.g.* Q2) is *what kinds* of variation to orthodox dynamical-closure are required. Directly, virtue-of-⟨*s*⟩ coupling (D12) can restore reliability, if fundamentally-new consequent possibilities for coupling are actually present in transmission of $l_{\langle s\rangle}$-data to report-as-meter (*i.e.* allowing restatement of theory-I/II first-person-reports in terms of Eqs. 9-10):

> **T3.** In D7-settings, a complete and scientifically-reliable framework exists *if* virtue-of-⟨*s*⟩ coupling pertains, if brain report-systems actually couple with and reportably-measure all distinct ⟨*s*⟩-states, and if sufficient resolution of brain-dynamical states is available.

Unqualified T3-completeness requires demonstration that, under virtue-of-⟨*s*⟩ coupling, it is possible to arbitrate between *every* possible (basically-coherent) theory-of-consciousness contrast, not just between theories I and II, although this is rather straightforward under T3-conditions that constitute access to all arbitration-necessary data. (The final clause in T3 is redundant under the earlier assumption of perfect future brain-data; it is included to emphasize that T3-completeness is guaranteed roughly under "the available of all possibly-relevant data of all kinds".)

T3 asserts the existence of a complete and reliable framework under virtue-of-⟨*s*⟩ coupling, but might leave open the possibility of obtaining completeness-and-reliability under some other (perhaps less radical) amendment to orthodox dynamical-closure. However, because the D7-setting *requires* use of report-as-meter (L3), and because the only A2-consistent way to reliably meter-resolve *l*-states is for $l_{\langle s\rangle}$-labels to appear in H_{int}, we have:

> **T4**. In D7-settings, a complete and scientifically-reliable framework exists *only if* virtue-of-⟨*s*⟩-label coupling pertains.

10. Central results in the D8-setting

10.1. D7-setting and D8-setting strategies compared

As stated earlier *("D8-setting approach")*, the most basic change from D7- to D8-demonstrations is to switch focus from an *experiment*'s capacity to reliably perform completeness-required theory-arbitrations to the *field's* capacity to reliably define conscious experience as a basic category (Wilkes, 1988). If it is possible to show that unreliability/reliability of *consciousness-definition* pertains respectively under orthodox-dynamical-closure/virtue-of-⟨*s*⟩-coupling, then D8-extensions of D7-theorems will have been established. [Generally here, definition of consciousness is taken to minimally require definition of conscious-*experience*, although other conventions are used. Views *e.g.* that consciousness can denote function without any reference to experience (Cohen and Dennett, 2011) lead to the following refinement: D8-theorems apply to frameworks for conscious-*experiential*

research, not to consciousness-*without-speaking-of-experience* research.]

As noted earlier, D8-reasoning will use *concepts employed in* the D7-setting, notably those related to first-person-report-as-meter and H_{int}-based meter-reliability (A2), but will not appeal to any D7-premised assumptions or inferences (because otherwise D8-proofs would self-contradictorily inherit D7-definitive limitations).

D8-completeness-and-reliability results include (subsume) D7-results, in the sense that the D8-setting includes the D7-setting, and therefore any results proven starting from the D8-setting must hold in the D7-setting. Thus, D8-proofs independent of earlier D7-proofs render those D7-proofs superfluous (D7-*results* will follow from D8-results). However, even if one does not accept details of D8-reasoning (that notably challenges "common-sense" intuition in a way that D7-reasoning likely does not), D7-results still follow from D7-reasoning, which is significant because the D7-setting *effectively* corresponds to arguably the dominantly-presumed context in modern consciousness-research empiricism.

10.2. D8-setting proofs

First, we assert that the scientific investigation of any phenomenon eventually requires some orienting framing, which must ultimately be expressible in terms of a contrast between natural-conditions-of-interest as opposed to other-non-*explananda*-conditions:

> **A3**. The scientifically-reliable investigation of any phenomenon "X" requires the scientifically-reliable definition of at least one "X"/not-"X" contrast.

[Problems here do not originate in alleged impossibilities of identifying orienting-contrasts for conscious experience. For example, in the present view, an "occurring-when-awake"-"disappearing-in-deep-sleep" framing (Gamez, 2014) is orienting-sufficient. The question will be: what D2-*scientific-reliability* can be attributed to any such framing?]

Using S (…) to denote the (*e.g.* when-awake) presence of *some* $\langle s\rangle$-labeled-phenomena and N_S(…) to denote their (*e.g.* deep sleep) absence, within the present phenomenon-symbol setting it follows that:

> **A4.** A scientifically-reliable investigation of consciousness exists only if a scientifically-reliable definition of the S(…)/N_S(…) contrast exists.

In order to link the S(…)/N_S(…)-contrast with D7-analyses employing H_{int}-reliability of $l_{\langle s\rangle}$-resolution, we define:

> **D14.** $\langle$...$\rangle$-resolution (at some particular time) is the ascertainment of whether a particular $\langle s_{ab}\rangle$-labeled component-of-consciousness exists as part of a particular subject's conscious experience, *i.e.* whether it is appropriate to write $S(\langle s_{ab}\rangle)$ or $N_S(\langle s_{ab}\rangle)$.

The view that S(…)/N_S(…)-establishment necessarily depends on establishing the S-existence of at

least one $\langle s_{ab} \rangle$-phenomenon evokes:

> **A5.** A scientifically-reliable definition of the $S(\ldots)/N_S(\ldots)$ contrast exists only if there is scientifically-reliable $\langle \ldots \rangle$-resolution.

[Note that A5 pertains in the present phenomenon-symbol setting *e.g.* as encapsulated in Figs. 6-8, and in the absence of zeroth-order innovations that might give some conceptual, *i.e.* non-report-employing, traction on $S(\ldots)/N_S(\ldots)$-establishment.]

Because orthodox dynamical-closure excludes both direct consciousness-meters and use of private, first-person-report-to-self (Chalmers, 1996) to establish the existence of $\langle s \rangle$-labeled components-of-consciousness, $\langle \ldots \rangle$-resolution must be via publically-accessible (conventional) first-person-report:

> **L7.** $\langle \ldots \rangle$-resolution is necessarily implemented by first-person report under orthodox dynamical-closure.

Just as scientifically-reliable D7-l-resolution requires $l_{\langle s \rangle}$-coupling in H_{int}, so scientifically-reliable $\langle \ldots \rangle$-resolution requires $\langle \ldots \rangle$-coupling in H_{int} (Fig. 10), *i.e.* a coupling between p (Eq. 20) and not just the $l_{\langle s \rangle}$-level-of-detail but the basic-realization-of-some-conscious-experiential-degree-of-freedom. But the latter kind of coupling is excluded by orthodox dynamical-closure, just as the former is:

> **L8.** Under orthodox dynamical-closure, $\langle \ldots \rangle$-resolution is scientifically-unreliable.

A5 and L8 imply:

> **T5.** Under orthodox-dynamical closure, there is no scientifically-reliable framework for consciousness research.

A trivial restatement of T5 in the form of a direct answer to Q1 gives:

> **L9.** Under orthodox dynamical-closure, there are no complete and reliable frameworks for consciousness research.

The D8-equivalent of T3 would be:

> **T6′.** A complete and scientifically-reliable framework *exists if* virtue-of-$\langle s \rangle$ coupling pertains, if brain-report systems actually couple with and reportably-measure all distinct $\langle s \rangle$-states, and if sufficient resolution of brain-dynamical states is available.

Because no D8-claim is made that "no currently-unforeseen problems could arise" (whereas the D7-setting was implicitly taken to preclude this), we state instead:

> **T6.** A complete and scientifically-reliable framework is *not provably impossible by present methods if* virtue-of-$\langle s \rangle$ coupling pertains.

Although T6 is weaker than T6′, arguably it establishes a basic field-integrity commensurable with that of physical and biological sciences, because in these fields too there are no *guarantees* that

current-methodology-challenging discoveries will never occur.

Finally, as the D8-equivalent of T4, we have

> **T7.** In the present phenomenon-symbol characterization of consciousness research, and in the absence of zeroth-order innovations providing a non-⟨...⟩-resolving basis for the $S(\ldots)/N_S(\ldots)$-contrast, a complete and scientifically-reliable framework exists *only if* virtue-of-⟨*s*⟩-label coupling pertains.

Qualifications in T7 not present in T4 serve to limit possibilities to a *necessary* ⟨...⟩-coupling in H_{int} as the basis for $S(\ldots)/N_S(\ldots)$-establishment.

10.3. Other views concerning assumptions

Theorems here depend on assumptions, and challenges raised by T5 in particular evoke serious examination of premises. Specifically, it might conceptually be possible to define phenomena other than via contrast, thus challenging A3. Nevertheless, no concrete A3-alternative is directly evident, and those few serious treatments of D8-setting arguments as exist in the literature [notably (Chalmers, 1996)] fail to solve H_{int}-framed problems, indicating a scarcity of A3-substitutes. *At the very least*, T5-reasoning: offers a novel formulation of (here, a D9-contingent version of) neglected minority views (Wilkes, 1988; Hawking, 2000; Cohen and Dennett, 2011); challenges sufficiency of alleged solutions to conscious-experiential definition (Chalmers, 1996); and, now places the burden-of-proof on the view that scientifically-reliable definition of conscious experience "must" be possible. Notably, a methodological account of physics can be constructed free of A3-related problems, indicating that, under orthodox dynamical-closure, the attainment by consciousness research of physics-equivalent methodological integrity presents as-yet unsolved problems. One solution might be to adopt different standards for consciousness research, although the mainstream view [*e.g.* (Block *et al*., 2014)] seems to be instead that there are no serious methodological problems and that consciousness research has already acquired a fully-scientific status.

10.4. Q1 and Q2 investigations complete

T2-T7 constitute a comprehensive answer to the Q1/Q2-aspects of central enquiries. Focus now turns to the Q3-topic of empirically determining which sub-cases of T2-T7 actually apply in our Universe.

11. Minimal experimental signals from required extensions to orthodox dynamical-closure

11.1. H_{int} requirements as variations to orthodox physical equations

The preceding Sections showed that completeness-and-reliability minimally requires virtue-of-$l_{\langle s\rangle}$ (under D7) or ⟨...⟩-coupling (under D8). We now redefine virtue-of-⟨*s*⟩-coupling (D12) to precisely

encompass these (and only these) two kinds of orthodox-dynamical-closure-violating amendments of physical law. Both kinds must lead to basically-coherent appearance of A2-sufficient H_{int} terms in relevant theoretical descriptions. Employing the Hamiltonian formalism now for the expression of fundamental physical *law* (in contrast with its earlier use to describe just particular *systems* and *meters*) at any description-level (microphysical, emergent-phenomenon, and so on), completeness-and-reliability minimally require

$$H_{actual} = H_{ODC}\,(\{p_i, q_j\}) + H_{vODC}\,(\langle s\rangle, \{p_k\}) \qquad (26)$$

where: H_{actual} is the Hamiltonian encapsulation of all *actually*-occurring dynamic couplings in nature; H_{ODC} expresses the *orthodox-dynamical-closure view* of physical couplings; H_{vODC} is the *violation*-of-orthodox-dynamical-closure term underpinning virtue-of-$\langle s\rangle$ coupling; $\{p_i, q_j\}$ are the canonical coordinates for orthodox-dynamical-closure laws-of-nature; and, $\{p_k\}$ are some $\langle s\rangle$-coupling subset of the $\{p_i\}$. (The Hamiltonian formalism is essentially the most basic, precise, and field-conventional physics-theoretic expression of fundamental laws. Although relativistic technicalities mean that Lagrangian-expressed approaches have certain presentational advantages, Lagrangian and Hamiltonian expressions are equivalent for present purposes.)

Clearly, in contrast to the completeness-and-reliability–required hypothesis of Eq. 26, orthodox dynamical-closure simply hypothesizes that

$$H_{actual} = H_{ODC}\,(\{p_i, q_j\}) \qquad (27).$$

As is the basic character of the best experiments, any well-designed test of Eq. 26 *vs*. Eq. 27 must yield significant information, because results will inform either the existence of new and fundamentally-significant physical couplings (Eq. 26) or the impossibility of a complete-and-reliable science-of-consciousness (Eq. 27).

11.2. Many ways for conserved physical quantities to be conserved

A potential objection to Eq. 26 lies in the misconception that it must lead to violation *e.g.* of energy-momentum conservation. For example, if some spatiotemporal structure couples in a novel way to increase the firing rate, say, of a particular neuron (increase, compared to orthodox microphysical law), this will affect certain calculations of the neuron-associated energy. But it is self-evidently false to claim that this requires violation of the conservation of energy: increase in the neuron's energy (relative to orthodoxy) can be precisely offset by decrease in the spatiotemporal-structure's energy (relative to orthodoxy). Certainly, violation of orthodox dynamical-closure would mean that energy conservation would take an unorthodox detailed-form (Burns, 1999), but that is not violation of the conservation law. [This view differs from that of (Wilson, 1999), which neglects *e.g.* the possibility of new particle-states in energy-accounting.] In order to exclude relevant D9-violations on grounds of inconsistency with existing empirical observations (a scientifically-valid reason), we would have to possess extensive, conclusive, and detailed data supporting the conservation-of-energy *by orthodox means* in every domain and at every scale of the brain. Current computations of energy-balances in the brain (Attwell and Laughlin, 2001; Shulman *et al*., 2004), in contrast, are approximate and global

in character: relative to current data, there is ample room for non-orthodox conservation mechanisms.

11.3. The "idealized" test

An idealized test of Eq. 26 *vs*. Eq. 27 (Fig. 11A-C) compares a detailed *in silico* simulation (Markram, 2006) of a specific human subject's brain with the actual physical evolution of brain dynamics. If virtue-of-⟨*s*⟩-coupling pertains to actual brain-dynamics (Eq. 26), and the computer simulation accurately replicates the dynamically-orthodox aspect of brain-dynamical evolution (Eq. 27), differences between actual and simulated evolution must occur. Of course, not *every* deviation is evidence for Eq. 26: deviations must be logically relatable to brain-dynamics involved in report and to contents of conscious-experience, and must survive usual tests for statistical significance to rule out a simple noise origin (Figure 11D-H). "Logically relatable" means that deviations must display an order sufficient to render brain-originated reports a reliably-coupled "meter" for conscious experience (A2). Practical challenges to fulfilling these and other requirements are discussed in the following subsections.

11.4. Potential difficulties for the idealized test

11.4.1. Confounds in certain metaphysical settings

One speculative possibility in certain metaphysical settings [*e.g.* (Chalmers, 1996)] is that any sufficiently-complete computer simulation of a brain would itself be associated with conscious experience. If this simulation-associated conscious experience then virtue-of-⟨*s*⟩-coupled with the simulation, simulated dynamics might match actual brain dynamics because the simulation does *not* then replicate Eq. 27. However, this speculative confound can be easily controlled for, because each computational step in the simulation can be recorded and examined at a later date: virtue-of-⟨*s*⟩-coupling *in the simulation* would be revealed as characteristic deviations from programmed behavior.

11.4.2. Disentangling virtue-of-⟨*s*⟩-coupling effects during calibration

A more difficult technical problem concerns the tuning of parameters in the simulation model. In principle, the model should be calibrated across the full hierarchy of dynamically-relevant scales (from synapses to brain areas) against physical systems that are free from virtue-of-⟨*s*⟩-coupling. Otherwise, virtue-of-⟨*s*⟩-coupling existing in the physical system might appear indirectly in the simulation via coupling-affected parameter values. However, approaches exist to mitigate this problem, for example, requiring models to closely reproduce brain-dynamics in both deep-sleep and waking states.

11.4.3. Making sufficiently-accurate measurements of initial conditions

The idealized test requires initial conditions in the model to be close enough to actual conditions in the brain that any later deviations can be reliably attributed to violation-of-orthodox-dynamical-closure (rather than to initial-condition disparity). "Initial conditions" here comprise both relatively-static brain *structure* (*e.g.* neurons, their structures, and their physical connectivity) and relatively-fluctuating brain *state* (*e.g.* firing dynamics, vesicle states, biochemical concentrations, *etc.*). What constitutes "close enough" depends critically on mathematical properties of actual brain-dynamics: if dynamics are very sensitive to initial conditions (*e.g.* dynamical chaos), then arbitrarily-small errors in simulation-model initial-conditions might lead to arbitrarily-large discrepancies in dynamics as time increases; if dynamics are relatively insensitive to initial conditions (*e.g.* stable limit-cycles) then small differences in initial conditions can be constrained (or even dampened) in their impact. If the latter, relatively-insensitive, case pertains, then multiple trial approaches can yield a signal: although measurement errors will generate trial-by-trial simulated-*vs*.-actual discrepancies, H_{vODC} effects can be observable *e.g.* via comparison of the grand trial averages of simulated and actual dynamics.

11.4.4. Initial-condition sensitivity of brain-dynamical evolution

The previous subsection pointed out that initial-condition-alignment problems are necessarily acute only if brain-dynamical evolution is initial-condition-hypersensitive. Although full discussion is beyond present scope, a central view in computational neuroscience [*e.g.* (Skarda and Freeman, 1990; Chen and Yu, 2003; Freeman and Quiroga, 2012)] is that macroscopic brain dynamics are maintained within a particular dynamical regime that precludes initial-condition-hypersensitivity. For example, it is precisely this regime that is thought to render signals visible in conventional EEG approaches [*i.e.* evoked and event-related potentials (Luck, 2014)] employing grand-trial-averaged time courses.

11.4.5. Limitations on knowledge might preclude "good-enough" simulation

Even assuming relative initial-condition insensitivity (so that initial-condition-measurement problems are surmountable *e.g.* by grand-trial averaging; Figure 11D-H), a reliable H_{vODC} signal can be established only if simulations are "good-enough", in the sense that their average is sufficiently close to the H_{ODC} component of H_{actual} in Eq. 27, thus allowing the isolation of H_{vODC} components from measurements of actual brain-dynamics. The *existence* of a simulation model good-enough in this sense can be tested for *e.g.* by randomly assigning a large number of trials into two subsets ["split-half" methodology (Kuder and Richardson, 1937)]: H_{vODC} estimates (supporting either Eq. 26 or Eq. 27) should then be similar in each subset and in the total collection of trials. Putative repeated failures of split-half tests may lead to a future conclusion that limitations on knowledge will always preclude the *construction* of a simulation good-enough for the idealized test.

Three responses can be made to observations emphasizing potential knowledge-limitation difficulties in building good-enough idealized-test models. First, it would certainly be premature to *not* undertake investigations of the kind suggested here simply because it is *possible* they won't work. [In any case, Blue Brain (Markram, 2006) and the complementary BRAIN initiative

(Alivisatos *et al*., 2013) are in-progress projects aimed precisely at idealized-test–required knowledge.] Second, and relatedly, knowledge required to advance H_{vODC} investigations is *precisely the same as* that required to advance conventional neural-correlates methodologies beyond problems currently-appreciated for a presumptive H_{ODC} context [*e.g.* (Navajas *et al*., 2014)]. Thus, it is hardly rational to support the viability of *status quo* approaches "against" experimental suggestions here on grounds of knowledge-limitations that must be overcome according to near-*status-quo*-agendas. (In any case, as noted earlier, rejecting present experimental suggestions does not help the *status quo*, according to theorems: it leads to the conclusion that *no* complete-and-reliable framework exists, *status quo* or otherwise). Third, the idealized test is not at all the sole approach to H_{vODC} investigation.

11.5. Beyond the "idealized" test: other experimental approaches

The idealized test has been introduced both as an in-principle future possibility and to explore basic issues in H_{vODC} empiricism. With basic issues in view, further *specific* examples in the two immediately-following subsections *illustrate* that experimental approaches employed or suggested as H_{ODC}-context neural-correlates probes can *generally* be modified to act as H_{vODC} tests. Just as for the idealized test, objections that H_{vODC} modifications will "never" be practicable generally backfire, *e.g.* if their intention is to support alleged-current or even potential-future completeness-and-reliability of near-*status-quo* empiricism in an orthodox dynamical-closure setting.

11.5.1. Synthetic lesion paradigms that control for orthodox-context confounds

In *orthodox* (dynamical-closure) contexts, serious problems exist for simple synthetic-lesion approaches that disrupt brain-dynamics putatively associated with some aspect of consciousness: disruption of dynamics has downstream e-causal effects that cannot be disentangled from variation in report. (Memory-based report cannot be used to compensate for this effect, because orthodox dynamical-closure would have to be violated to gain meaningful information.) However, future "perfect" synthetic-lesion experiments might avoid this problem by effectively replacing brain-activity in a certain area by apparatus, thus holding activity elsewhere in the brain constant. But perfect versions suffer another problem: as shown in Fig. 4, they cannot generate theory-I/theory-II–arbitrating signals because, although varying by condition, report must be the same under both hypotheses. However, *a perfect synthetic-lesion experiment can produce H_{vODC} signatures*: report that disappears in a lesion condition holding constant orthodox e-causal information-transmission must have relied on an H_{vODC}-contribution from the lesion-disrupted brain-dynamics (Fig. 10).

11.5.2. Back-tracing

As another example, it has been suggested (Gamez, 2014) that the method of "back-tracing" (Krichmar *et al*., 2005) might be employed in (conventional) neural-correlates empiricism. This method appears particularly suited to H_{vODC} tests, either directly or in conjunction with *e.g.* a

synthetic-lesion paradigm. For example, back-tracing the neural origins of the putative absence of report in the lesion condition might in principle distinguish between the crucial absence of an *orthodox* e-causal influence required for report-generation (in which case no reliable inferences concerning H_{vODC} can be made), and the contrasting case of orthodox e-causal pathways unaltered under intervention (in which case absence-of-report may be attributed to absence of apparently report-critical H_{vODC} effects).

11.6. What can be deduced from a positive causal-coupling signal?

Experimental signatures supporting Eq. 26 over Eq. 27 would not in fact establish conclusively that virtue-of-⟨*s*⟩-coupling exists. Rather, they would indicate a non-orthodox dynamical innovation in brain-dynamics under certain conditions associated conscious experience. Further, even if signatures were interpreted as establishing virtue-of-⟨*s*⟩-coupling, this alone would not suffice to prove the existence of a complete-and-reliable science of consciousness (T3, T6). However, the virtue-of-⟨*s*⟩-coupling interpretation would certainly refute the view that completeness-and-reliability could be *provably unattainable* (T6). (As usual, all inferences from experimental data must be made with careful reference to theoretical assumptions and sensitivity-of-measurement: for example, analyses ruling out virtue-of-⟨*s*⟩-coupling at classical scales might leave open quantum modes, but this possibility would also require a theoretical context in which quantum states can be amplified to provide orderly e-causal effects on report-governing brain-dynamics.)

12. Discussion

12.1. What degree of "every-metaphysical-stance" validity has been established?

The goal here has been a maximally assumption-free consideration of possibilities for a complete-and-reliable consciousness-research framework, leading to claims that completeness-and-reliability can *only* under exist virtue-of-⟨*s*⟩-coupling. The contextually-claimed meaning of "*only*" is that completeness-and-reliability conditions apply under every metaphysical stance and associated mode-of-reasoning, including as-yet-unarticulated metaphysical views. But has "every-metaphysical-stance" generality actually been established? To answer this question, generality claims will be assessed primarily with respect to three different kinds of possible metaphysical innovation (Fig. 12, A-C). (By construction, these three kinds will exhaust all result-relevant possibilities. Structural consideration is not given below to *theory-of-consciousness* innovations, because for metaphysical-generality purposes these can broadly be treated as the product of, rather than an input to, consciousness research, provided A1 holds.)

12.1.1. Zeroth-order innovations in consciousness research

Refining earlier orienting discussions, a *zeroth-order innovation in consciousness research* (Fig. 12A) is now defined as some new view rejecting claims that Figs. 6 and 7 are a general and

maximally-assumption free characterization of the problem of consciousness research. Because the content of Figs. 6-7 is essentially that "consciousness research is the investigation and characterization of *e.g.* associative relationships between degrees-of-freedom in brain dynamics and in conscious experience", the position here is that a zeroth-order innovation corresponds to a field-redefinition, and such redefinitions are *excluded* from what is meant here by every-metaphysical-stance applicability. Thus, it is not necessary to claim immunity from zeroth-order innovations to establish every-metaphysical-stance applicability so clarified, and no such claims are in fact made.

12.1.2. First-order innovations in consciousness research

A *first-order innovation* (Fig. 12B) is defined as a new explanatory-basis for the existence of associations between spatiotemporal-brain-structures and conscious-experience (currently-proposed reasons include "by virtue of identity between …", "by virtue of function performed by …", and so on). The claim *is* made that reasoning is immune to first-order innovations. Evidence for first-order immunity can be found in philosophy-neutrality of centrally-employed symbols ($\langle\ldots\rangle(\boldsymbol{\rho}_j)$-*notation for conscious experience*; Figs. 6-7), and in the absence of direct appeal to any specific metaphysical (first-order) stance. Consequently, and more directly, proofs hold under addition of any statement of the form "conscious-experiential associations exist because …".

12.1.3. Second-order innovations in consciousness research

A second-order innovation (Fig. 12C) is defined as a line of reasoning that limits the informational basis of conscious experience to a specified subset of brain-encoded information [for example, functionalist reasoning in (Chalmers, 1996); first-order innovations can enable second-order innovations]. Validity of second-order innovations could critically compromise the central structure of earlier *D7*-reasoning, because D1-completeness is taken to require *e.g.* theory-I/theory-II arbitration that second-order innovations might render unnecessary. Immunity against second-order innovations *is* claimed here, but the basis for this claim is more complicated than for its first-order equivalent.

12.1.3.1. D7-setting and zeroth-order innovations in physical theory

D7-reasoning is partially protected from second-order innovation by construction: standard scientific-empiricism (D7) has been *defined* via the ascription of unreliably-subjective status to the adoption of any specific first-order metaphysical stance (on which second-order innovations might be based). However, there is another kind of second-order innovation dependent not on a *metaphysical* advance, but instead on very particular kinds of advance in *physical* theory. Notably, it might be established on theoretical grounds that certain physical systems and not others are associated with conscious experience, and then further (*e.g.* based on experimental observations) that only certain spatiotemporal-brain-structures are of the consciousness-associated kind. Significantly, any advances of this kind, here termed *zeroth-order innovations in physical theory*, must propose structural re-

conceptualizations of physical theory far beyond those currently envisioned (*i.e.* beyond all current concrete visions for a so-called Theory-of-Everything). Every-stance-applicability is further explicitly refined now to *exclude* applicability under zeroth-order physical theory innovations, and D7-reasoning is *not* claimed immune from these kinds of advance.

12.1.3.2. Existence and character of zeroth-order innovations in physical theory

A complete account of context-of-validity should substantiate two claims concerning zeroth-order physical innovation, namely:

> **A6**. Zeroth-order physics-theoretic innovations might in principle perform theory-I/theory-II arbitration without *any*, *direct*, dependence on first-person data.
>
> **A7**. Zeroth-order physics-theoretic innovations necessarily constitute beyond-Theory-of-Everything developments

(A6 justifies the relevance of zeroth-order physics-theoretic innovations to consciousness research that might otherwise be doubted *e.g.* on T4-T7 grounds. A7 justifies the exclusion of these innovations from "currently-envisioned science", the implicit target-validity-context here.) Both points are addressed in the following brief sketch.

In absence of virtue-of-$\langle s\rangle$-supportive data, the scientific-reliability status of even *drawing e.g.* Fig. 6 is uncertain. Nevertheless, without ultimately depending on their reliability, Figs. 6 and 8 can be used to stimulate reasoning in Fig. 13A. Whereas (Fig. 13A) $P_j(\omega)$-symbols (labelling physically-orthodox fundamental properties) conceptually have a multiplicity-of-referents (*e.g.* according to ontological conception), $\langle s\rangle$-symbols in Fig. 6 have a single, well-defined referent (*e.g. conscious-experience-of*-red, *conscious-experience-of*-this-edge, and so forth). Of course, there are multiple *theories-of-consciousness* and multiple *metaphysical stances* that can be employed to *explain the existence* of what-is-labeled-by-$\langle s_1\rangle(\boldsymbol{\rho}_1)$ (say). But the *direct-referent*-of-symbol for $\langle s\rangle(\boldsymbol{\rho})$-symbols differs basically from the direct-referent-of-symbol for $P_j(\omega)$-symbols: because each subject has direct access to their own experience (irrespective of scientifically-reliable reporting), the direct-referent of a particular $\langle s\rangle(\boldsymbol{\rho})$-combination is *absolutely certain* (in principle); because "physical reality" is in contrast hidden behind an epistemic veil (Berkeley, 1878), the direct-referent of a particular $P_j(\omega)$-symbol is *absolutely uncertain* (but the *numerical measure* of that direct-referent is of course known: in this sense, $P_j(\omega)$-symbols have two kinds of referent). (Note that no insurmountable problems arise for physical theory from this absolute-uncertainty, provided *e.g.* physics is conceived of as the explanation of meter-readings: all conceptions-of-referents predict the same meter readings, even though how the meter itself "actually exists" is itself absolutely uncertain.)

These observations are used simply to evoke a physical-theory-setting that itself contains a symbol, η say (Fig. 13A), for the number-of-epistemologically-allowed-direct-referents associated with *other* symbols. Having evoked this step, all basic references to consciousness and $\langle s\rangle$-symbols are dropped [although a trace of the latter remains, in that there must be some symbols, α_k say, with $\eta\,(\alpha_k) = 1$; in contrast, for example, $\eta\,(P_j) = \infty$]. Fig. 13B shows how a physical-theory-setting with

an appropriately-defined η-symbol might in principle effect arbitration-without-report (thus supporting A6). The fact that such a setting must include notation capable of describing epistemological-properties of other symbols shows that these settings differ radically from current Theory-of-Everything candidates that lack any such capacity (thus supporting A7).

12.1.3.3. D8-setting

Excluding both kinds of zeroth-order possibility, D8-reasoning is immune to second-order innovations because it relates to the possibility of scientific-reliability in establishing the basic meaning of $\langle s \rangle$-symbols, a problem which is independent of (and conceptually prior to) the inferential effect of second-order advances, namely limitation of conscious-experiential-*associations* to certain spatiotemporal-structures.

12.1.4. Summary: theorems immune to first-order and second-order innovations

After preceding clarifications, results here are claimed to hold in every context that excludes zeroth-order innovations of either kind. Theorems are therefore not universal in an absolute sense, but do apply in every setting in which physical theory and consciousness research take a form sufficiently close to their broadest-possible present-day characterization.

12.2. Role of case-by-case considerations in establishing general applicability

As should be clear both from the structural design of present developments and the immediately-preceding reflections on what this design has achieved, proofs based on case-by-case consideration of the existing literature would not have established the degree of general-applicability claimed here for present theorems. Nevertheless, conclusions in the literature differing from those here might evoke a due-diligence concern, in that differences might in principle indicate deficiencies in present developments. Accordingly, a wide-ranging analysis of such differences and their analytical origins is available in *Supplementary Data Sheet 1*. Emphatically, this additional material does not form part of central proofs or of first-instance claims for validity or generality: these depend instead on the inherent integrity (or otherwise) of direct modes of proof employed above.

12.3. Implications for completeness and scientific-reliability in consciousness research

Theorems here indicate that the in-principle possibility of a complete-and-reliable framework for theory-arbitration in consciousness research stands or falls with the in-nature virtue-of-$\langle s \rangle$-violation of orthodox dynamical-closure, unless zeroth-order innovations are found. This suggests that the highest methodological priority should be given to both experimental investigation of whether D9-orthodox-dynamical-closure or the D12-virtue-of-$\langle s \rangle$-coupling alternative actually describe our Universe, and to theoretical investigation of those zeroth-order innovations that might circumvent a

necessarily-methodologically-foundational role for orthodox-dynamical-closure/virtue-of-⟨*s*⟩-coupling empiricism.

Given current absence of decisive information, three distinct methodological stances are viable: the assumption that orthodox dynamical-closure is certainly correct, the assumption that virtue-of-⟨*s*⟩-coupling dynamical-closure-violations certainly pertain, or the view that either of these two assumptions (or something else) might later turn out to be true. Central theorems allow direct inference of the methodological completeness-and-reliability consequences inherent in each of these three stances. Because methodological consequences differ radically by stance, neural-correlates empiricism should fully specify the pre-experimental bridging principles invoked (Chalmers, 1998), and then carefully account for both reliability and discriminatory-reach of subsequent theory-arbitration, particularly if analyses opine that experimental data can establish allegedly-unique theoretical conclusions.

(Block *et al.*, 2014) recently claimed that "we *will*, eventually, crack this natural phenomenon [of consciousness]" (italics added). But without radical innovations in physical or metaphysical understanding, results here indicate human assertion does not govern our in-principle ability to reduce the large set of theories-of-consciousness to a unique and parsimonious candidate *in a scientifically-reliable way*; instead, we are at the mercy of specific natural-order conditions whose fulfillment in our Universe is currently an open question, although fortunately this question is itself open to experimental investigation.

Table 1. Summary of symbol-notation introduced in the main text.

SYMBOL	DEFINITION
s_{ab}	*b*-th instance of *a*-th feature-type (stimulus) in the external environment
${}_{l}s_{ab}$	*b*-th instance of *a*-th feature-type (stimulus) in the *l*-th level hierarchical description of the external environment
${}_{X}\mathbf{r}_{j}$	*j*-th location for sampling of sensory information in the external environment
${}_{X}\mathbf{r}_{l,j}$	*j*-th location for sampling of sensory information in the *l*-th level hierarchical description of the external environment
$\langle s_{ab}\rangle$	Contribution to conscious experience generated by s_{ab}
$\langle {}_{l}s_{ab}\rangle$	Contribution to conscious experience generated by ${}_{l}s_{ab}$
$\boldsymbol{\rho}_{j}$	*j*-th location in conscious experience; information sampled at ${}_{X}\mathbf{r}_{j}$ leads to a contribution-to-consciousness at $\boldsymbol{\rho}_{j}$
$\boldsymbol{\rho}_{l,j}$	*j*-th location in conscious experience at the *l*-th hierarchical level; information sampled at ${}_{X}\mathbf{r}_{l,j}$ leads to a contribution-to-consciousness at $\boldsymbol{\rho}_{l,.j}$
A	Measure of brain activity for encoding of stimuli
${}_{A}\mathbf{r}_{i}$	*i*-th brain location relevant to *A* measurement
${}_{A}\mathbf{r}_{l,i}$	*i*-th brain location relevant to *A* measurement, in a cortical area with receptive fields at the *l*-th hierarchical spatial scale
C_{abj}	Classifier function on *A*-states for encoding of $s_{ab}({}_{X}\mathbf{r}_{j})$
$C_{l,abj}$	Classifier function on *A*-states for encoding of ${}_{l}s_{ab}({}_{X}\mathbf{r}_{l,j})$
$R_{\omega}[\text{“}\langle s_{ab}\rangle\text{”}]$	First-person report of conscious experience of s_{ab}
$R_{\omega}[\text{“}s_{ab}\text{”}]$	Explicit (categorized, cognitive) report of s_{ab} stimulus-details
$R_{\omega}[s_{ab}]$	Non-verbalizable (not-categorized, non-cognitive) effect of s_{ab}
P, N_P	Presence (P) or non-presence (N_P) of a specified stimulus in the environment
S, N_S	Presence (S) or non-presence (N_P) of a specified $\langle s\rangle$-contribution in subjective experience
$X \rightarrow_{a} S\,[\langle s\rangle]$	Physical state X is associated with $\langle s\rangle$ (a component of subjective experience)
$X \rightarrow_{e} Y$	X e-causes Y (Gamez, 2014) via couplings consistent with orthodox dynamical-closure
$X \rightarrow_{e,\langle s\rangle} Y$	X e-causes Y, via a virtue-of-$\langle s\rangle$-label coupling that violates orthodox dynamical-closure
$X \rightarrow_{e,I} Y$	Some e-causal system (*e.g.* a researcher) observes X and infers Y (via couplings consistent with orthodox dynamical-closure)
${}_{l}s_{ab}({}_{X}\mathbf{r}_{l.c}) \subset {}_{m}s_{de}({}_{X}\mathbf{r}_{m.f})$	${}_{l}s_{ab}({}_{X}\mathbf{r}_{l.c})$ is a sub-component of the hierarchically higher-level stimulus ${}_{m}s_{de}({}_{X}\mathbf{r}_{m.f})$ ($l < m$); see Figure 9F
$\{\ldots\}$	Set or collection of …

1247

13. References

Adrian, E. D. (1928). *The Basis of Sensation: The Action of the Sense Organs*. New York: W. W. Norton & Co.

Alivisatos, A. P., Chun, M., Church, G. M., Deisseroth, K., Donoghue, J. P., Greenspan, R. J., McEuen, P. L., Roukes, M., Sejnowski, T. J., and Weiss, P. S. (2013). The brain activity map. *Science* 339, 1284–1285.

Attwell, D., and Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* 21, 1133–1145.

Berkeley, G. (1878). *A treatise concerning the principles of human knowledge*. Philadelphia, PA: JB Lippincott & Company.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.* 30, 481–499.

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

Block, N., Carmel, D., Fleming, S. M., Kentridge, R. W., Koch, C., Lamme, V. A., Lau, H., and Rosenthal, D. (2014). Consciousness science: real progress and lingering misconceptions. *Trends Cogn. Sci.* 20, 1–2.

Bohm, D. (1951). *Quantum Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Burns, J. E. (1999). Volition and physical laws. *J. Conscious. Stud.* 6, 27–47.

Chalmers, D. J. (1998). On the Search for the Neural Correlate of Consciousness. *Sci. Conscious. II Second Tucson Discuss. Debates*, 219–229.

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press, USA.

Chalmers, D. J. (2008). "The varieties of emergence," in *The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion*, eds. P. Davies and P. Clayton (Oxford: Oxford University Press), 244–256.

Chen, G., and Yu, X. eds. (2003). *Chaos Control: Theory and Applications*. 2003 edition. Berlin ; New York: Springer.

Churchland, P. S., and Sejnowski, T. J. (1992). *The Computational Brain*. Cambridge, MA: MIT Press.

Cohen, M. A., and Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends Cogn. Sci.* 15, 358–364.

Dennett, D. C. (1991). *Consciousness Explained*. Boston, MA: Little Brown.

Eccles, J. (1990). A Unitary Hypothesis of Mind–Brain Interaction in the Cerebral Cortex. *Proc. R. Soc. Lond. B Biol. Sci.* 240, 433–451.

Fingelkurts, A. A., Fingelkurts, A. A., and Neves, C. F. H. (2010). Natural world physical, brain operational, and mind phenomenal space-time. *Phys. Life Rev.* 7, 195–249. doi:10.1016/j.plrev.2010.04.001.

Freeman, W., and Quiroga, R. Q. (2012). *Imaging Brain Function with EEG: Advanced Temporal and Spatial Analysis of Electroencephalographic Signals*. New York: Springer.

Gamez, D. (2014). The measurement of consciousness: a framework for the scientific study of consciousness. *Conscious. Res.* 5, 714. doi:10.3389/fpsyg.2014.00714.

Goldstein, H., Jr, C. P. P., and Safko, J. L. (2001). *Classical Mechanics*. San Francisco: Addison-Wesley.

Hawking, S. W. (2000). "The Objections of an Unashamed Reductionist," in *The Large, the Small and the Human Mind*, ed. M. Longair (Cambridge: Cambridge University Press), 169–172.

Haynes, J.-D., and Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* 7, 523–534.

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. *Brain Res. Bull.* 44, 107–112.

Kim, J. (2000). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: Bradford.

Krichmar, J. L., Nitz, D. A., Gally, J. A., and Edelman, G. M. (2005). Characterizing functional hippocampal pathways in a brain-based device as it solves a spatial memory task. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2111–2116.

Kuder, G. F., and Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika* 2, 151–160.

Kuhn, T. S. (1996). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. *Crit. Growth Knowl.* 4, 91–196.

Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: Bradford.

Markram, H. (2006). The blue brain project. *Nat. Rev. Neurosci.* 7, 153–160.

Marr, D. (2010). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, MA: MIT Press.

Navajas, J., Rey, H. G., and Quian Quiroga, R. (2014). Perceptual and contextual awareness: methodological considerations in the search for the neural correlates of consciousness. *Conscious. Res.* 5, 959. doi:10.3389/fpsyg.2014.00959.

Neumann, J. V. (1996). *Mathematical Foundations of Quantum Mechanics*. Princeton, NJ: Princeton University Press.

Place, U. T. (2012). Thirty Five Years On—Is Consciousness Still a Brain Process? *Grazer Philos. Stud.* 36, 19–31.

Popper, K. (2002). *The Logic of Scientific Discovery*. New Delhi: Routledge.

Quiroga, R. Q., and Panzeri, S. (2013). *Principles of Neural Coding*. Boca Raton, FL: CRC Press.

Rieke, F., de Ruyter van Steveninck, R., Warland, D., and Bialek, W. (1997). *Spikes: Exploring the Neural Code*. Cambridge, MA: MIT Press.

Shannon, C. E., and Weaver, W. (1964). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Shulman, R. G., Rothman, D. L., Behar, K. L., and Hyder, F. (2004). Energetic basis of brain activity: implications for neuroimaging. *Trends Neurosci.* 27, 489–495.

Skarda, C. A., and Freeman, W. J. (1990). Chaos and the new science of the brain. *Concepts Neurosci.* 1, 275–285.

Smart, J. J. C. (1959). Sensations and brain processes. *Philos. Rev.* 68, 141–156.

Wagner, M. (2006). *The Geometries of Visual Space*. New York: Psychology Press.

Werner, J. S., and Chalupa, L. M. (2004). *The Visual Neurosciences*. Cambridge, MA: MIT Press.

Wilkes, K. V. (1988). "---, yishi, , duh, um, and consciousness," in *Consciousness and Contemporary Science*, eds. A. Marcel and E. Bisiach (Oxford: Clarendon Press), 16–41.

Wilson, D. L. (1999). Mind-brain interaction and violation of physical laws. *J. Conscious. Stud.* 6, 8–9.

14. Figure legends

Figure 1. The role of meters in conventional and conscious-experiential research. Detailed definitions of terminology and $\langle s\rangle(\boldsymbol{\rho})$-symbols are given in the main-text body. **(A)** *The conventional inferential problem*. For a system S, a meter M makes a measurement Π related to a phenomenon-of-interest; simultaneous measurements by meters $\{m_i\}$ establish the values of fundamental properties P_j of fundamental system-constituents ω_k. A theory-of-Π is a function f relating $\{P_j(\omega_k)\}$-values to Π-values. **(B)** *The idealized inferential problem for conscious experience*. "System" comprises brain/associated-conscious-experience (dashed-black arrow schematically depicts *association* of the contents-of-experience with brain activity). *Ideally*, an M-meter *directly dynamically-coupled with conscious-experiential-properties* (green arrow) records $\{\langle {}_k s_{ab}\rangle(\boldsymbol{\rho}_{k,c})\}$-values that encapsulate details of experience; simultaneous brain-dynamical measurements $\{A({}_{\mathrm{A}}\mathbf{r}_j)\}$ are made. (Inside-square diagonal **does *not* imply duality**, but emphasizes *two* relevant, $\langle s\rangle$- and A-labeled, *phenomena*-to-be-related.) A theory-of-consciousness is a function F relating $\{A({}_{\mathrm{A}}\mathbf{r}_j)\}$-values to $\{\langle {}_k s_{ab}\rangle(\boldsymbol{\rho}_{k,c})\}$-values. **(C)** *First-person report under orthodox dynamical-coupling*. First-person report R_ω["$\langle s\rangle$"] replaces meter M, but originates in orthodox-A-couplings of brain-activity (green arrows). First-person report will be shown not scientifically-reliable, in this scenario. **(D)** *First-person report with virtue-of-$\langle s\rangle$ dynamical coupling*. If $\langle s\rangle$-labeling (conscious-experiential-association) results in *non-orthodox* report-coupling (*additional* green arrow originating in dashed-black *association*), R_ω["$\langle s\rangle$"] can in principle obtain M-style scientific-reliability; $\langle s\rangle$-label-coupling can be conceived of as novel dynamics attributable to A-states satisfying conditions for conscious-experiential association, so that the association-originating green arrow can be viewed as lying entirely within the physical brain.

Figure 2. Schematic illustration of two competing theories-of-conscious-experience (theories I and II). (A) Theory I: Conscious experience is associated with neural activity in S1/V1 (schematically illustrated by black oval); small-scale visual features such as the grey lines are encoded in S1/V1 and occur in visual experience. **(B)** Theory II: conscious experience is associated neural activity in a hierarchically-later U2 (black oval); small-scale visual features such as the grey lines are encoded in S1/V1 but not in U2, and consequently do not occur in visual experience. (Schematic depictions *e.g.* of S1-as-V1 are not intended to be neuroanatomically precise, in part because S1 and U2 need not be identified with any specific brain areas, systems, or structures, in central proofs; the only critical points are that theory-I and theory-II hypothesize conscious-experiential-association with two distinct brain-structures, and that these two structures encode at least one feature of the external environment in distinct ways, capable in principle of leading to differences in the contents-of-conscious-experience.)

Figure 3. Neural-correlates approaches cannot arbitrate between theories I and II under orthodox dynamical-closure and U2-information-transmission constraints. (A) *Illustration of simplified e-causal and associative relationships employed in this Figure*. Edges at four stimulus locations (*s*-labeled row of boxes) e-cause neural activity in four corresponding S1 neurons (connected to *s*-boxes by green arrows; locations of S1-neurons are purely schematic, rather than neuroanatomically precise). Firing of a S1 neuron reliably induces firing of a U2 neuron (connected to S1 neurons by black lines). The informational basis of report (*R*-labeled boxes) is limited to information encoded in U2-activity (D10). Each U2-neuron's receptive field is the sum of its S1-afferents' receptive-fields, so that report has only two independent components (two *R*-boxes connected to U2 neurons by green e-causal arrows). **(B)** *Theories-of-conscious-experience* (dashed-black arrows depict associative relations). Theory I hypothesizes that conscious experience is associated with S1-activity (four E_{I}-labeled boxes). Theory II hypothesizes that conscious experience is associated with U2-activity (two E_{II}-labeled boxes). **(C-F)** Neural activity (red circles), report *R*, and experience *E*, for different stimuli *s* under different theories-of-conscious-experience. **(C)** *Non-adjacent stimuli, theory I.* **(D)** *Non-adjacent stimuli, theory II.* In panels C and D (theories I and II), reports are identical, but conscious experiences differ. **(E)** *Adjacent stimuli, theory I.* **(F)** *Adjacent stimuli, theory II.* In panels E and F (theories I and II), reports are identical, as are conscious experiences. Stimulus-report regularities (panels C-F) show that theory-I/II arbitration is impossible in this neural-correlates paradigm.

Figure 4. Synthetic-lesion experiments cannot provide in-experiment data that publicly informs consciousness research, under orthodox dynamical-closure and U2-information-transmission constraints. Schematic depictions of stimulus *s*, report *R*, conscious experience *E*, and neural dynamics follow Figure 3 in a detailed way (circles depict neurons in specific brain areas; red fill indicates neural activity; black fill indicates disrupted activity). Figure 3A-B schematically depict e-causal and associative relationships inherent in theories-of-consciousness I and II. As in Figure 3, the U2-information-transmission constraint for report is assumed (D10). *Synthetic-lesion outcomes according to theory I* **(A-C)** *and theory II* **(D-F)**: (**A,D**) Pre-intervention, for the same stimulus *s*, theories I and II predict different experiences but the same report (repeating Figure 2C-D). (**B,E**) The first intervention focally eliminates certain early-cortical neural activity (black circles) but keeps all other brain activity unchanged. According to theory I (panel B), intervention eliminates conscious experience whereas theory II (panel E) predicts unchanged conscious experience associated with cortically-later activity. Theories I and II make *identical* report-predictions. **(C,F)** The second intervention restores early activity but focally eliminates later activity (black circles) (again, with all other brain activity unchanged). According to theory I (panel C), intervention does not affect conscious experience associated with early cortical activity, whereas theory II (panel F) predicts elimination of conscious experience. Again, theories I and II make *identical* report-predictions. Because report *R* is identical in (A) and (D), in (B) and (E), and in (C) and (F), publicly available in-experiment report does not vary by theory, rendering theory-I/II arbitration impossible in this synthetic-lesion paradigm.

Figure 5. Neural-correlates approaches cannot arbitrate between theories III and IV that relax U2-constraints on information-transmission, under orthodox dynamical-closure. (A,B) Theory-III: *X* classifies dynamically-switched coupling of communication between report-governing-areas W and S1 (*X* = 1) or U2 (*X* = 2). In theory-III: **(A)** when *X* = 1, conscious-experience is associated with S1; **(B)** when *X* = 2, conscious-experience is associated with U2. **(C,D)** Theory-IV: *X* classifies dynamically-switched coupling of communication, but conscious-experience is always associated with S1, *i.e.* contains the red square plus four grey lines both when **(C)** X = 1 and **(D)** X = 2. **(A,C)** *Neural correlates experiment when X = 1.* Report and experience are identical under theory-III and theory-IV. **(B,D)** *Neural correlates experiment when X = 2.* Report is identical under theory-III and theory-IV, but conscious experience differs, rendering theory-III/IV arbitration impossible in this neural-correlates paradigm.

Figure 6. Basic symbolism for conscious experience is physical-theory-analogous and ontologically-general. The orthodox matter-energy constitution of a tree and a brain are schematically depicted as line drawings (black spikes depict brain-dynamical activity); visual conscious experience of the tree is depicted via color images. Symbols $\{s_{ab}({}_{\mathrm{X}}\mathbf{r}_c)\}$ label perceptually-relevant components of the environment (here, edges). Symbols $\{A({}_{\mathrm{A}}\mathbf{r}_j)\}$ label brain-activity (notably including stereotypical activity-patterns encoding *e.g.* edges). Symbols $\{\langle s_{ab}\rangle(\boldsymbol{\rho}_c)\}$ label components-of-conscious-experience (here, conscious experiences of edges). Note that the horizontal dashed-line **does *not* establish any kind of duality**; instead it stands as a visual reminder for two features inherent to consciousness research. First, the dashed-line can be viewed as depicting the *epistemic barrier* behind which both external-environmental and brain-dynamical above-the-line phenomena are hidden. In contrast, below-the-line phenomenological referents of $\langle s_{ab}\rangle(\boldsymbol{\rho}_c)$-symbols are directly (introspectively) accessible. Second, the dashed-line reminds us that there are *two sets-of-phenomena* to be related, *i.e.* brain-dynamical phenomena above-the-line and conscious-experiential phenomena below-the-line. Thus, if conscious experience is to be explained fully within physical theory in terms of already-existing symbols (in a manner analogous *e.g.* to explanations-of-temperature), symbols for conscious experience (analogous *e.g.* to T for temperature) may be useful. Just as a symbol for temperature does not imply any temperature-ontology, so symbols for conscious experience are ontologically-general (*i.e.* consistent with monist, dualist, and substance-silent conceptions).

Figure 7. *Every* basically-coherent naturalistic theory-of-consciousness can be given succinct *metaphysically-neutral* statement via present symbolism. This Figure depicts a generalization of symbolism introduced in the text. (Results are unaffected by which symbolism is employed). Top box, top green e-causal arrow: the presence P of a stimulus $s_{ab}({}_{\mathrm{X}}\mathbf{r}_j)$ e-causes stereotypical brain activity classified via C_{abj} $[\{A({}_{\mathrm{A}}\mathbf{r}_i)\}] = 1$. Middle box, middle green e-causal arrow: stereotypical brain activity ($C_{abj} = 1$) e-causes stereotypical effects on brain-dynamical configuration (measured by $\{B({}_{\mathrm{B}}\mathbf{r}_i)\}$; $D_{abj} = 1$ encapsulates stereotypical effects on B-measured configuration). Stereotypical brain configuration ($D_{abj} = 1$) is lawfully associated (dashed-black arrows) with a component-of-conscious-experience $\langle s_{ab}\rangle(\boldsymbol{\rho}_j)$ that is consequently S-present in subjective experience. Notation shown here generalizes the text by introducing a B-measured feature (*e.g.* biomolecule states) distinct from but coupled to A-measured (*e.g.* electromagnetic-field) activity. Although symbol-expressions for theories-of-consciousness requires further straightforward elaboration to account for *e.g.* vigilance, attention, and other more complex dependencies of conscious-experience on brain-state (see *e.g.* Eqs. 21-23), notation is both theory-of-consciousness general (elaboration-capable of expressing any proposal for orderly *association between* brain-states and conscious-experiential states) and metaphysically-neutral (in that theory-of-consciousness symbol-expressions do not rely on any metaphysical stance, and can thus accommodate every such stance). Metaphysical stances can be construed as *explanations for existence of* associative dashed-black arrows (see Fig. 12, B-C).

Figure 8. Phenomenon-symbol approach coheres with the methodological foundation of physical theory. This Figure complements Figures 6 and 7 by emphasizing that orthodox physical theory can be encapsulated in the "phenomenon-symbol approach" (vertical solid line dividing natural world of phenomena from formal world of symbols; solid-blue lines schematically encapsulate methodologically-fundamental phenomenon-symbol correspondences). **(A)** (Orthodox-reductionist) physical theory can be expressed in terms of symbols $P_j[\omega_k]$ for fundamental properties P_j of elementary particles ω_k, and collective-system properties expressed as functions $g(P_j[\omega_k])$. Thus, orthodox-physical s_{ab} and A phenomenon-symbol correspondences in Figures 6 and 7 are specific cases of a generic approach. **(B)** This panel emphasizes *ontology-neutrality of physical theory* inherited by phenomenon-symbol description of conscious-experience, and *epistemic asymmetry* distinguishing orthodox-physical and conscious-experiential phenomenon-symbol approaches. Whereas panel A depicts a naïve-realist ontology for physical reality (elementary particles "actually exist" as point-like or near-point-like objects), this panel depicts a mathematical or informational ontology (elementary objects "exist as" numbers or computations). Symbols in physical theory can describe (correspond with: solid-blue arrows) naïve-realist, mathematical, informational, and other ontologies. Equally, conscious-experiential-symbol definitions are also consistent with various above-the-dashed-line ontological-conceptions. Epistemic asymmetry is evidenced by the fact that direct-referents (*e.g.* "greenness of green") of conscious-experiential symbols do not alter with above-the-line ontology, in contrast with direct-referents of $P_j[\omega_k]$ symbols.

Figure 9. Symbolism for hierarchy of scales in the visual environment. (A) The entirety of the visual field is labeled by the single symbol ${}_4s$. **(B)** Large objects such the squares are labeled by various symbols of the form ${}_3s_i$ ($i = 1,\ldots, N$). Black cross at ${}_X\mathbf{r}_{3,1}$ depicts a coordinate location (*e.g.* the center of the relevant receptive field) for the specific object ${}_3s_1$. **(C-E)** Successively-smaller scale components are labeled by symbols ${}_ls_i$ with associated locations ${}_X\mathbf{r}_{l,i}$; symbols ${}_0s_i$ label the smallest biologically-relevant scale. **(F)** Partial illustration of relationships between hierarchies: stimuli labeled by ${}_0s$-level symbols are typically components of stimuli labeled at the ${}_1s$-level; similarly stimuli labeled by ${}_1s$-level symbols are typically components of stimuli labeled at the ${}_2s$-level; and so on (higher level relationships not shown). Thus, *e.g.* the edge of the square labeled ${}_2s_1$ in panel C can also be described via progressively more detailed collections$\{{}_1s_b\}$ and $\{{}_0s_a\}$. The theorem-relevance of extensive *l*-notation is that a framework's capacity for scientifically-reliable theory-I/II arbitration will be expressed (via A2) in terms of an interaction Hamiltonian H_{int}'s explicit dependence on (certain further refinements of) *l*-symbols.

Figure 10. Scientifically-reliable definition of conscious experience under orthodox dynamical-closure and ⟨...⟩-coupling. (A,B) *Orthodox dynamical-coupling.* U2-activity associated with (dashed-black arrows) conscious experience (panel A) is replaced by apparatus (panel B) that connects to and orthodox-e-causally influences (green arrows) non-U2 brain-areas so that non-U2 brain-dynamics are unaltered, after apparatus introduction. First-person report used to *define* conscious experience is unreliable under orthodox dynamical-closure (unless further *e.g.* functionalist metaphysical assumptions pertain; but D7-setting results preclude the use *e.g.* of introspection to establish detailed metaphysical stances in a scientifically-reliable manner). [Panels A and B reproduce reasoning in (Cohen and Dennett, 2011).] **(C,D)** ⟨...⟩-*coupling*. If the comparison of panels A and B is repeated in a setting where ⟨...⟩-coupling pertains (indicated schematically by the green *non-orthodox* e-causal arrow originating in associative dashed-black arrow), and if information transmitted by ⟨...⟩-coupling is <u>necessary</u> for first-person report to occur, then replacement of U2-activity by apparatus will lead to a change in report (panel D *cf.* panel C). In-experiment observation of panel-C/panel-D contrast-in-report would support ⟨...⟩-coupling. [Panels C and D extend (Cohen and Dennett, 2011) by consideration of a natural-law possibility neglected there.] (Note that failure to observe the panel-C/panel-D report-contrast can also occur under *e.g.* functionalist assumptions even in the presence of ⟨...⟩-coupling. But observation of panel-C/panel-D report-contrast is an unambiguous signal for ⟨...⟩-coupling, provided that considerable technical challenges in constructing apparatus sufficiently replicating e-causal effects of U2-dynamics have been met.)

Figure 11. Idealized experiment to arbitrate between orthodox-dynamical-closure and virtue-of-⟨*s*⟩-coupling. (A-C) *Idealized experiment.* **(A)** At time t_0, actual brain dynamics $\{A(t_0)\}_{act}$ are used to set values $\{A(t_0)\}_{sim}$ in a detailed computer simulation (blue cube). **(B)** At a later time t_1 (without intervening sensory stimulation, *e.g.* in deep sleep) actual and simulated dynamics are equal, establishing basic capacity of the simulation to reproduce detailed brain activity. **(C)** Both actual and simulated brains are given sensory input. If virtue-of-⟨*s*⟩-coupling pertains, simulated and actual dynamics must diverge at some later time (after controlling for possible coupling with simulation-associated conscious experience). Alternatively, identity between simulated and actual dynamics would rule out virtue-of-⟨*s*⟩ coupling. **(D-H)** *Averaging across trials.* **(D-F)** Trial-by-trial data (first three rows) and trial-average (bottom row). **(D)** Simulation of H_{ODC}. **(E)** H_{ODC} contribution to actual brain-dynamics. Actual and simulated H_{ODC} differ here due to initial-condition measurement-error; over many trials and under certain dynamical assumptions, panel-D-average = panel-E-average. **(F)** H_{vODC} contribution to actual brain-dynamics; note that H_{vODC} dynamical range is assumed ten times smaller than measurement-error variance in H_{ODC}. **(G-H)** Even though envelope of data (panel G = E +F) varies by an order of magnitude more than the signal-range (panel H), signal is detectable as: average-of-data – average-of-simulations = panel-E-average + panel-F-average – panel-D-average = panel-F-average (because panel-D-average = panel-E-average). (In practice, panel-D and panel-E averages will not precisely cancel, leading to a mixture of H_{vODC} dynamics with simulation error. Standard statistical tools such as split-half reliability can provide H_{vODC}-signal confidence-levels.) Signal-noise analyses parallel conventional ERP trial-averaging methods in EEG data analysis.

Figure 12. Zeroth-, first-, and second-order innovations in consciousness research. (A) *Zeroth-order innovations* (in consciousness research; see also Figure 13) deny (red cross) that the phenomenon-symbol conception (contents of black square, reproducing Figure 6) is the appropriate basis for consciousness research. Figure 6 simply states that consciousness research investigates relationships between brain-dynamical states and conscious-experiential degrees-of-freedom: a zeroth-order innovation in consciousness research must propose some concrete and as-yet-unapparent alternative conception. **(B)** *First-order innovations* identify new metaphysical explanations (red rectangle, extending black-rectangle-depicted pre-existing metaphysical explanations) for the dashed-black-arrow association between certain brain-dynamical states and conscious-experiential degrees-of-freedom. (This panel provides an explicit schematic depiction of metaphysical-stance-generality inherent to dashed-black-line schematics throughout present Figures, notably theory-of-consciousness-declarative Figure 7.) **(C)** *Second-order innovations* identify new modes of reasoning that can arbitrate between (red tick, red cross) competing theories-of-consciousness (black-oval alternatives for brain-area-associations) based on a *given* set of metaphysical association-explanations (contents of red rectangle). Second-order innovations may or may not depend on first-order innovations, although theory-arbitration relevance of a first-order innovation requires a consequent second-order innovation. Novel theory-arbitration methods need not originate solely in metaphysically-related first- and second-order innovations in consciousness research: Figure 13 illustrates the potential theory-arbitration capacity of a *zeroth-order innovation in physical theory*.

Figure 13. Zeroth-order innovations in physical theory. (A) *Epistemic asymmetry as a basis for novel constructions in physical theory.* As noted in Figure 8, symbols $P_i(\omega_j)$ for fundamental properties are consistent with an unlimited number of different ("≠") ontologies (different above-the-dashed line depictions of physical reality); thus, an unlimited number of distinct solid-blue lines are attributable to $P_i(\omega_j)$ symbols. Writing β_{ij} for $P_i(\omega_j)$, and η for the number of solid-blue lines, $\eta[\beta] = \infty$. In contrast, under each different above-the-dashed-line conception, the direct referent of $\langle s_{ab}\rangle(\rho_j)$-symbols is invariant ("="), so that a single solid-blue line is attributable to $\langle s_a\rangle(\rho_b)$. Writing α_{ab} for $\langle s_a\rangle(\rho_b)$, $\eta[\alpha] = 1$. A *zeroth-order innovation in physical theory* is then a fundamental setting containing α, β, and η symbols, and some approach for unifying these symbols (*e.g.* so that α and β symbols are explained via some common underlying state Ω, say). Details of unifying settings are beyond present scope. **(B,C)** *Application of α and β symbols to theory-arbitration in consciousness research.* Assume the existence of a well-constructed $\{\alpha, \beta, \eta\}$-setting, and some (non-first-person-report) method for establishing appropriate $\{\alpha, \beta\}$-descriptions of physical systems. In panel B, α-symbols in the description of S1 but not U2 indicate that S1 and not U2 is associated with conscious experience, because α-symbols can be tentatively identified with $\langle s\rangle$-labeling (on the basis of $\eta[\alpha] = 1$). Similarly, if panel C is the correct $\{\alpha, \beta\}$-description of the brain, then U2 and not S1 can be associated with conscious experience.

Figure 1.JPEG

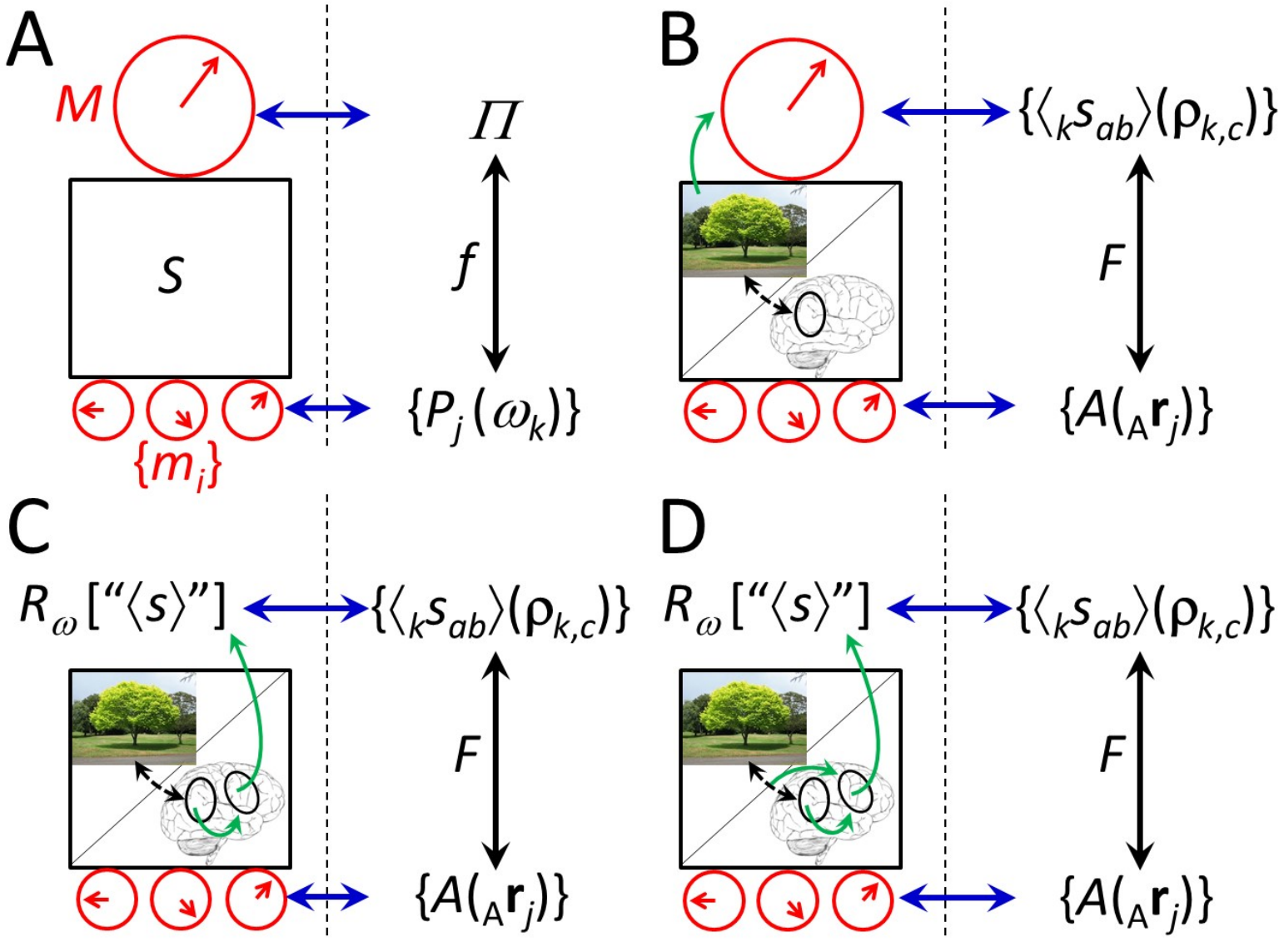

Figure 2.JPEG

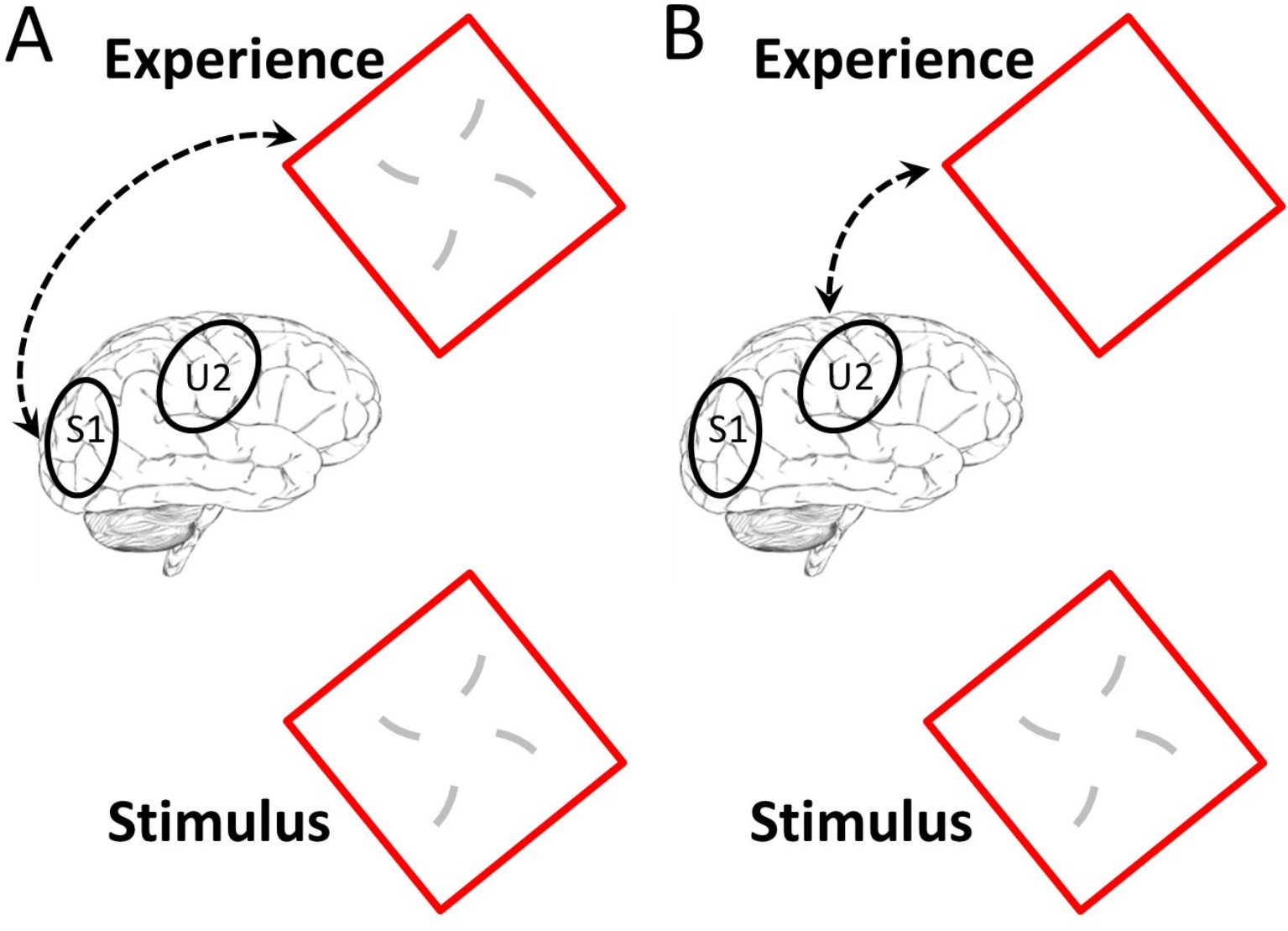

Figure 3.JPEG

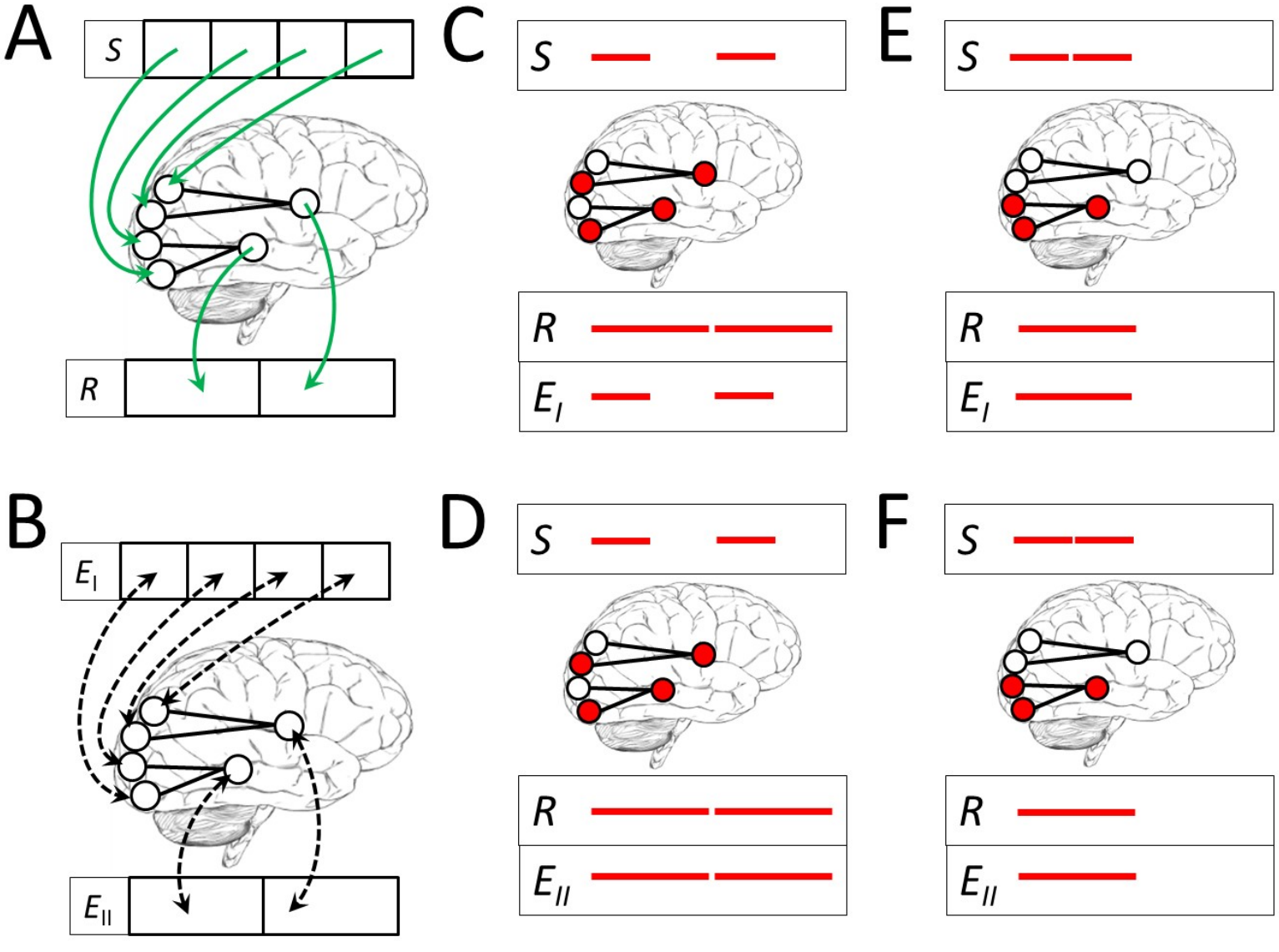

Figure 4.JPEG

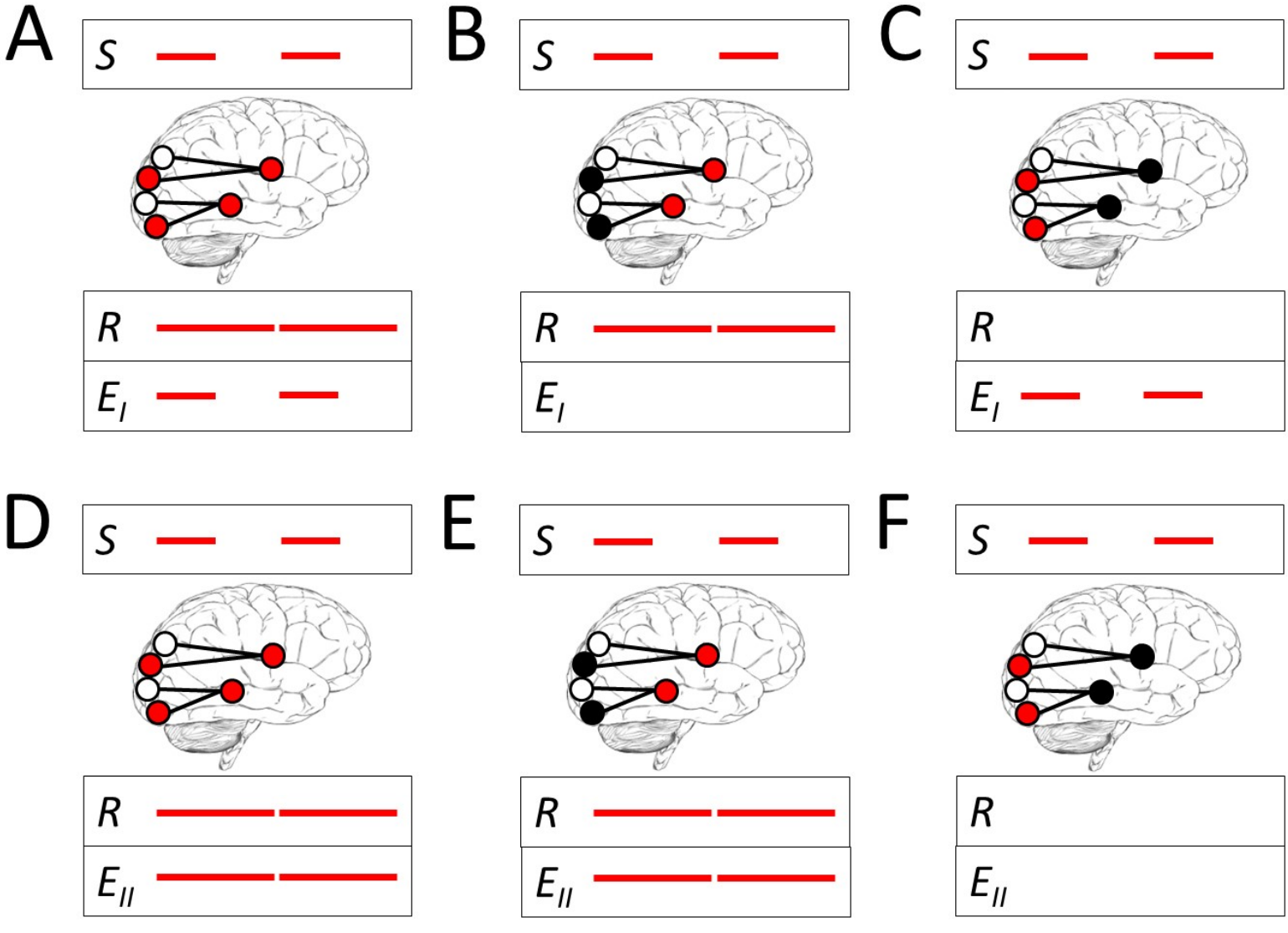

Figure 5.JPEG

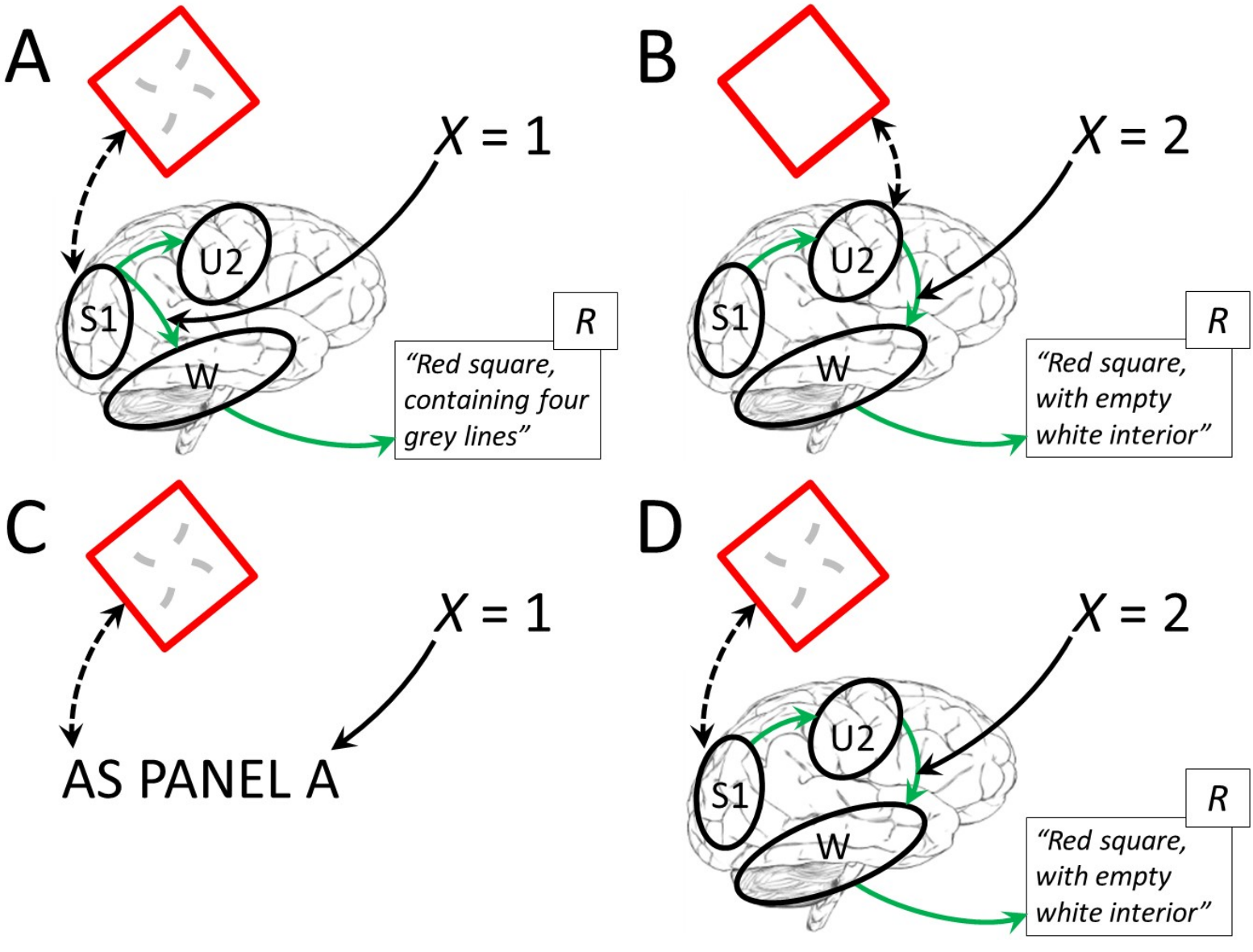

Figure 6.JPEG

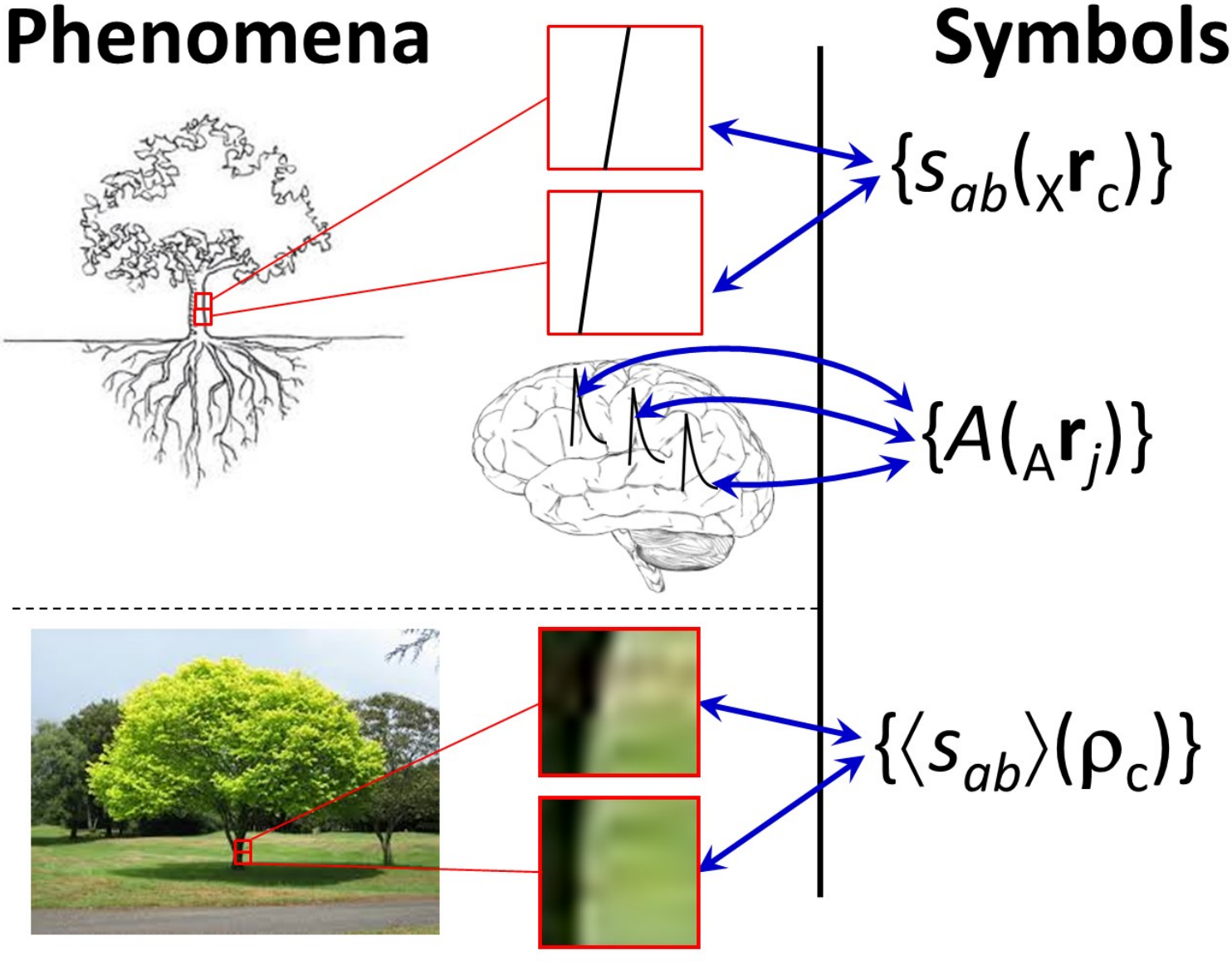

Figure 7.JPEG

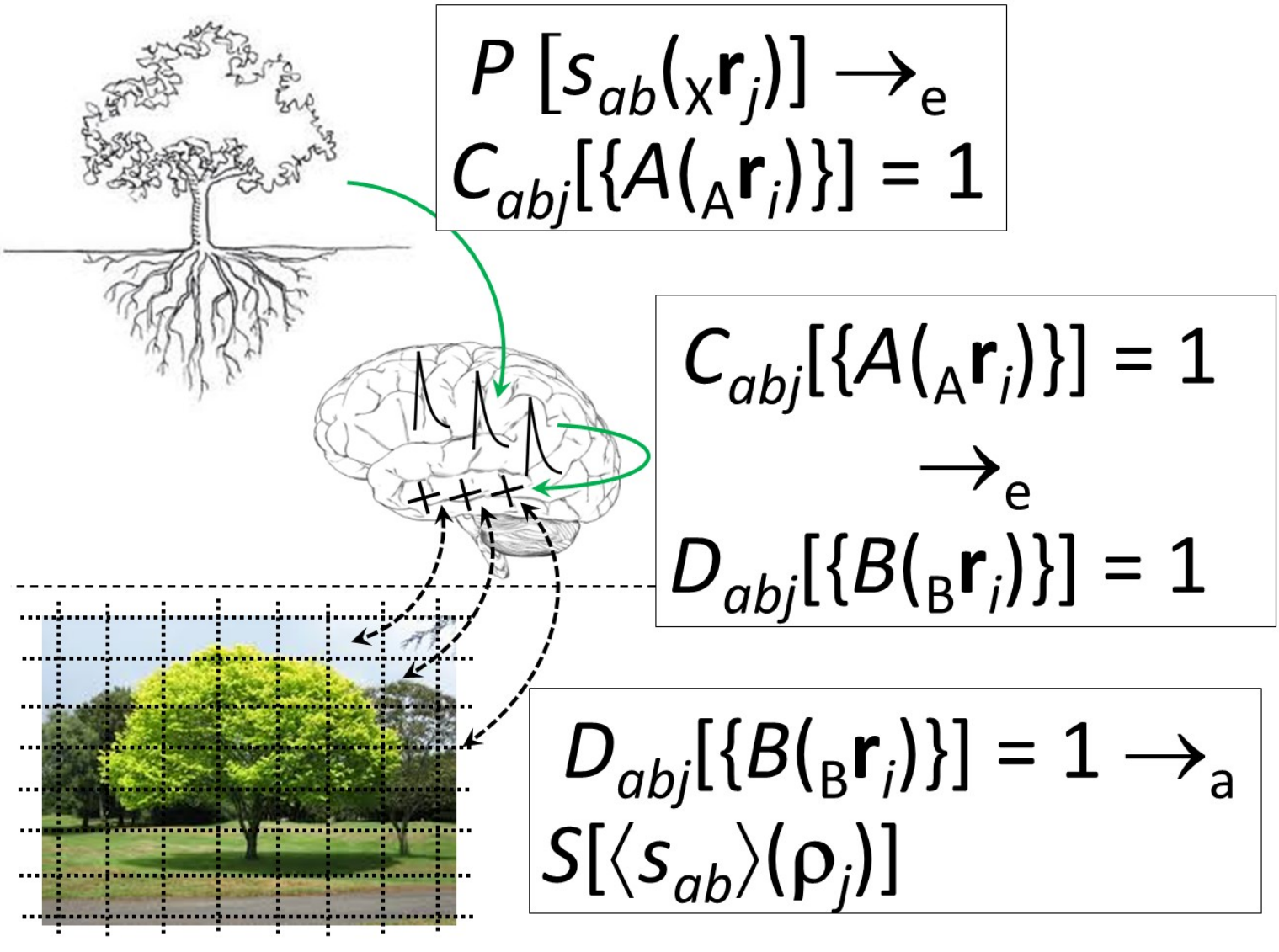

Figure 8.JPEG

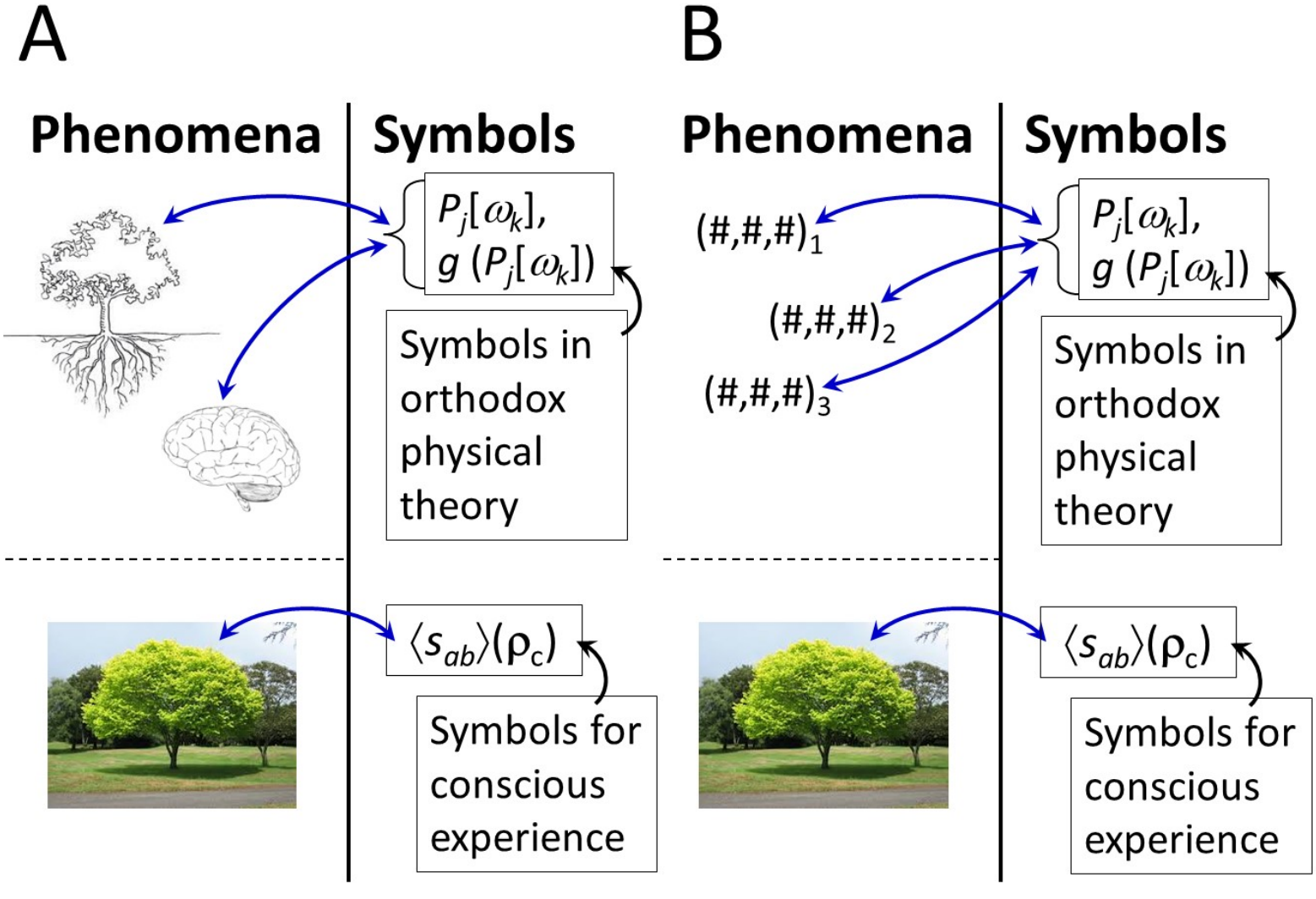

Figure 9.JPEG

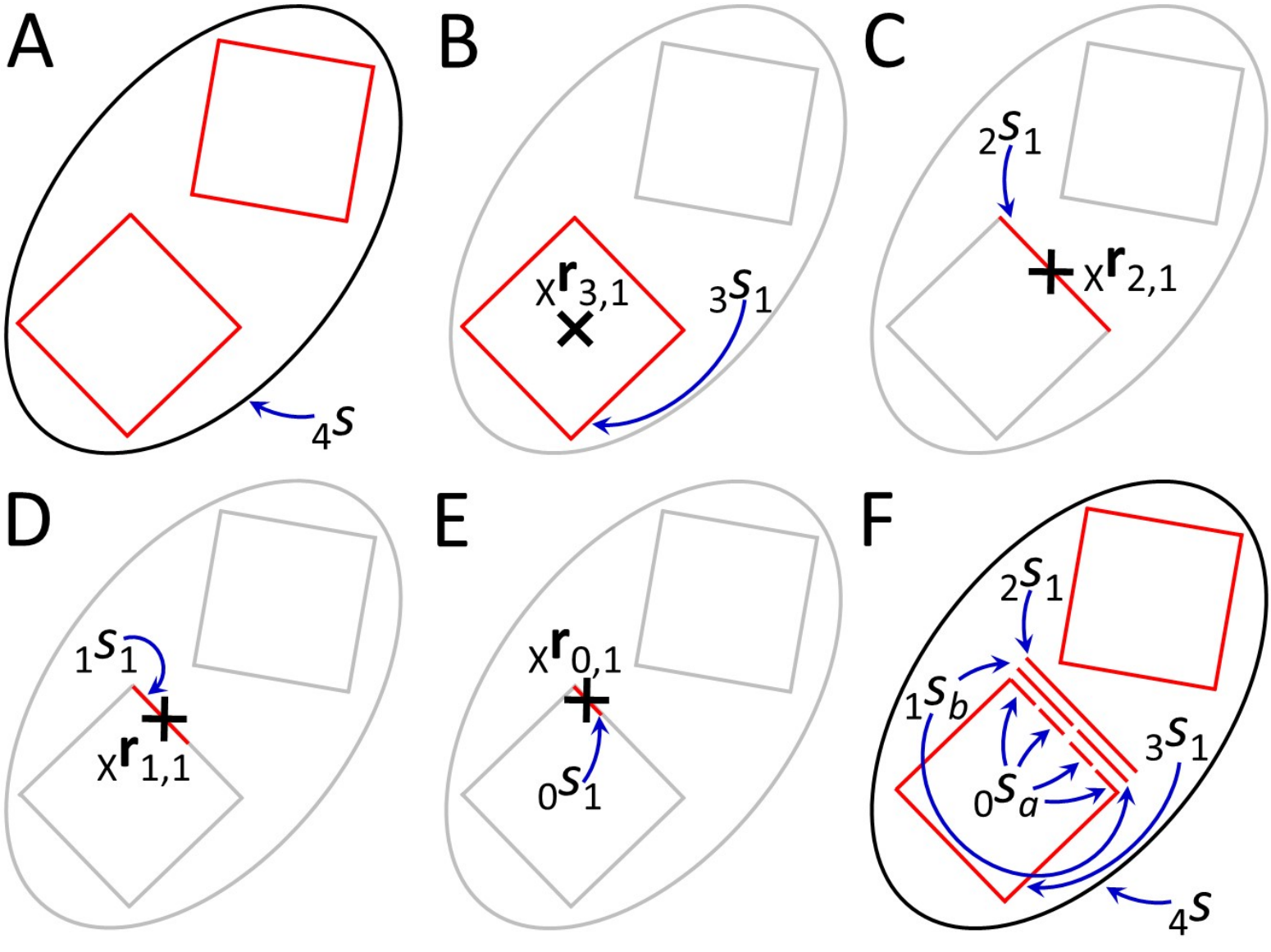

Figure 10.JPEG

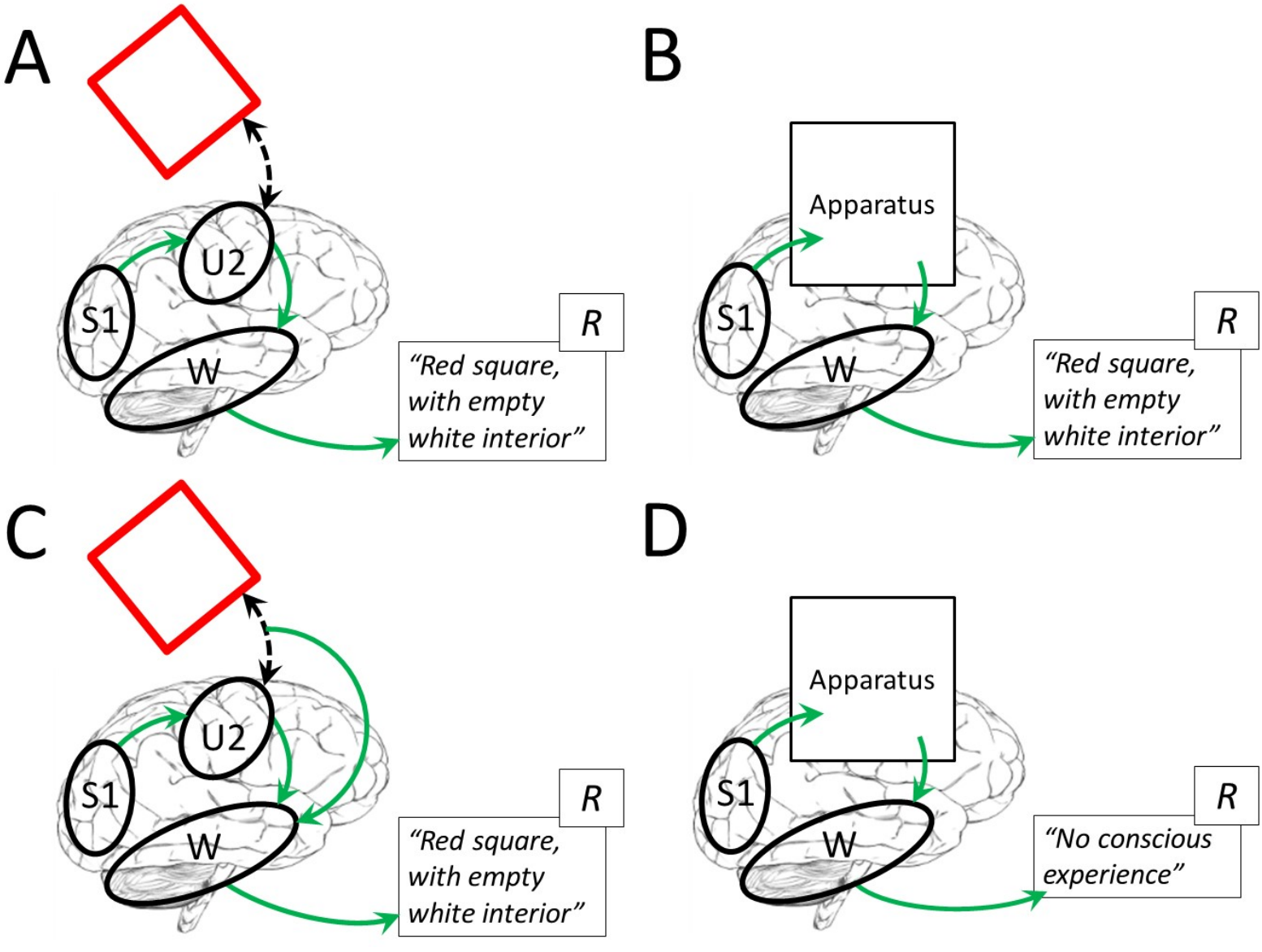

Figure 11.JPEG

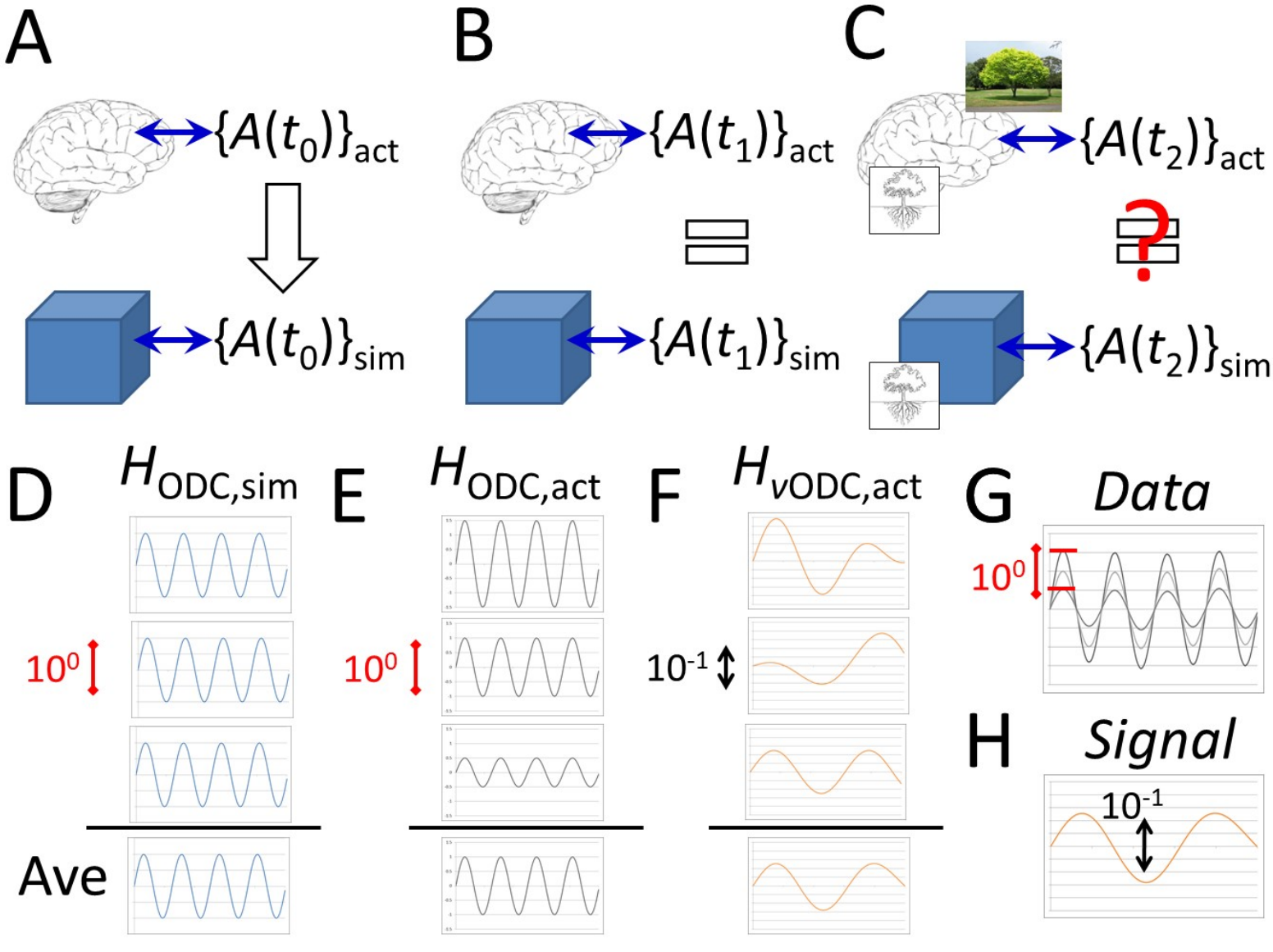

Figure 12.JPEG

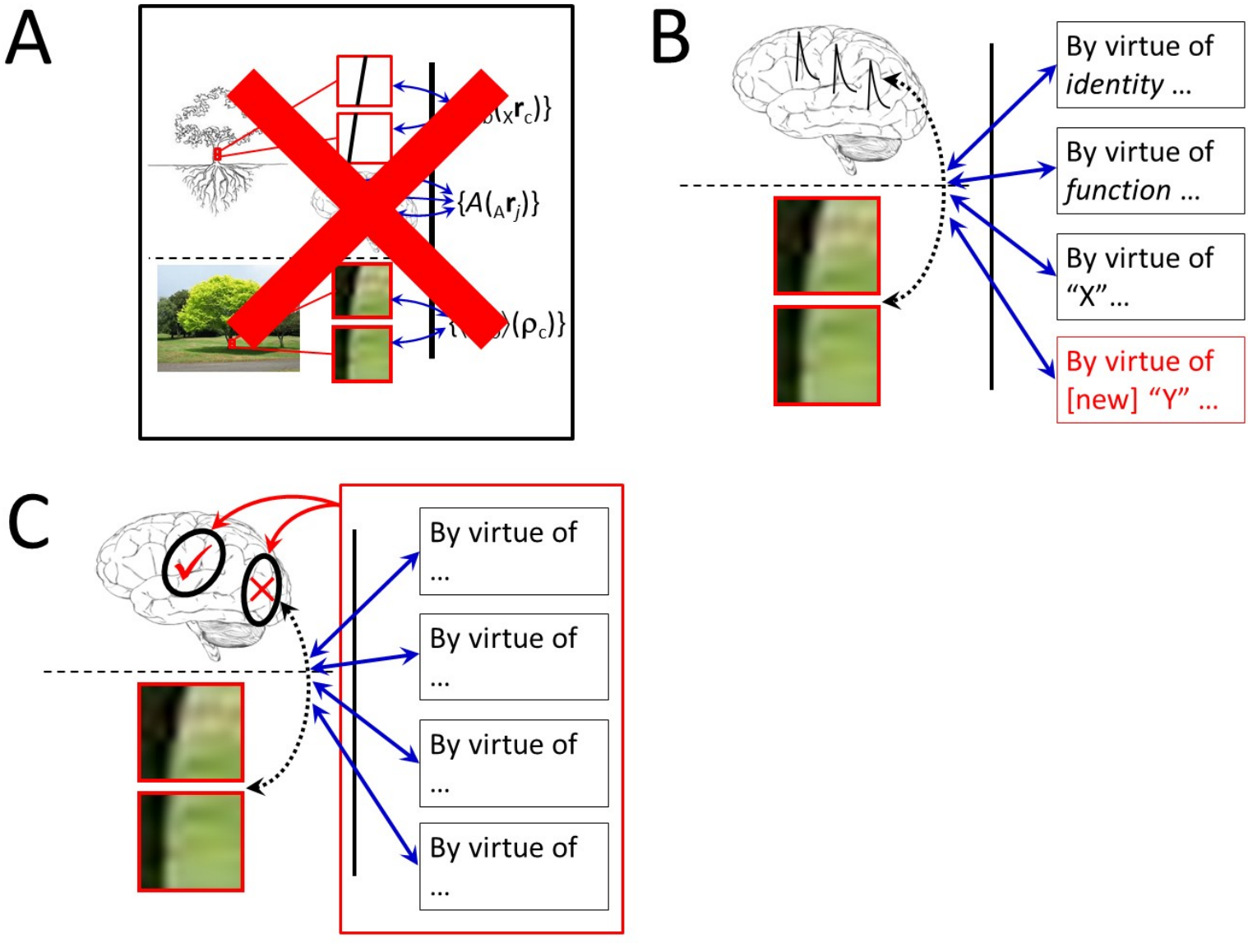

Figure 13.JPEG