frontiers in
COMPUTATIONAL NEUROSCIENCE

Multi-subject settings restrict and define physically-orthodox neural codes for consciousness

Nicholas Rosseinsky

Journal Name:	Frontiers in Computational Neuroscience
ISSN:	1662-5188
Article type:	Hypothesis & Theory Article
First received on:	10 Mar 2014
Frontiers website link:	www.frontiersin.org

Frontiers in Human Neuroscience Hypothesis & Theory
23 February 2014

Multi-subject settings restrict and define physically-orthodox neural codes for consciousness

Nicholas M. Rosseinsky[1,*]

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

*** Correspondence:** Nicholas M. Rosseinsky, Department of Neuroscience, Center for Dialog in Science, 14 Rosebarn Avenue, Exeter, EX4 6DY, U.K.
rosseinsky.nicholas.m@gmail.com

Keywords: neural codes, biophysics, consciousness, spatially homogeneous theories, causality, other minds, unity of consciousness, binding problem.

Word count: 11996

Abstract

The "problem of many minds" and the "unity of consciousness" are usually considered to be philosophical topics, without direct or concrete implications for neuroscience. Notably, mainstream explanations of consciousness as more or less synonymous with certain brain dynamics implicitly assume that the distinct conscious experiences of a multiplicity of human subjects are simply properties of the corresponding multiplicity of different physical objects (*i.e.* brains), just as different physical objects can have unique and distinct temperatures. Here, in contrast, it is shown that the multi-subject setting has direct consequences for neural codes employable as the informational basis for consciousness, under the assumption that consciousness is a physically-orthodox phenomenon. Multi-subject consequences are of two kinds. First, if conscious-experiential space is a *sub-domain of orthodox physical space*, then certain encoding schemes viable in the single-subject setting can be excluded, thus reducing the *number* of viable sub-domain approaches. Second, if conscious-experiential space is a *property of brain dynamics*, then codes must vary by subject, thus directly altering the *nature* of viable property-based codes. These results extend previous single-subject analyses, thus constructing a complete and detailed theoretical catalog of viable consciousness-encodings under physical orthodoxy. This theoretical catalog in turn constitutes a maximal basis for empirical testing of the "physically-orthodox consciousness" hypothesis: if the encoding basis of consciousness is not of one of the relatively small number of kinds identified here, consciousness cannot be a physically orthodox phenomenon. Conversely, because viable physically-orthodox codes for consciousness differ distinctively from codes known to support behavioral computations, empirical discovery of any catalog member would be strong evidence towards the precise physical basis of consciousness in an orthodox setting.

1. Introduction

This is the final paper in a series (Rosseinsky, 2014a, 2014b) that aims to establish objective tests within conventional neuroscience for physically-orthodox theories-of-consciousness. Previous papers showed that spatially-distributed and temporally-extended codes cannot be the informational basis for consciousness under physical orthodoxy (Rosseinsky, 2014a), and that topographic codes for consciousness (again under physical orthodoxy) must differ from those for behavior (Rosseinsky, 2014b). These conclusions result respectively from locality and spatial homogeneity of causation, both of which are necessary features of a physically orthodox setting; both conclusions lead to empirical tests for physically-orthodox theories-of-consciousness. Whereas previous papers restricted themselves to single-subject (*i.e.* single-brain) settings, the present paper addresses further restrictions on physically-orthodox neural codes for consciousness that arise in multi-subject settings, and empirical tests derived from these further restrictions.

To illustrate the specific issues that arise for physically-orthodox consciousness-encodings in the multi-subject setting, consider the following (extremely simplified) example (**Figure 1**). Purely for simplicity of illustration, assume that conscious visual experience is comprised only of black or white pixels. Let B denote a measure of a brain-dynamical state that causally encodes components-of-consciousness, with B = 0 encoding a black pixel and B = 1 encoding a white pixel. Further, let a finite set of various locations within conscious experience be labelled by various real numbers in the range zero to one (non-inclusive), and extend the zero/one B-code so that the decimal portion labels location (**Figure 1A**). For example, B = 0.1 now encodes black at the location labelled by 0.1, whereas B = 1.9 encodes white at the location labelled by 0.9. Now consider a two-subject setting (**Figure 1B**), and label the B-values that occur in the two brains by B_1 and B_2. Thus, B_1 = 0.1 means a B-value of 0.1 occurring in the brain of the first subject, whereas B_2 = 0.1 means a B-value of 0.1 occurring in the brain of the second subject. If B_1 = 0.1, then there must be a black pixel in the first subject's experience at the location labelled 0.1. Similarly, if B_2 = 0.1, then there must be a black pixel in the second subject's experience at the location labelled 0.1. But these two outcomes – "… in the *first* subject's experience …" and "… in the *second* subject's experience …" – are different, and arise from the same physical state B = 1. Spatial homogeneity of causality means that B = 1 must have the same effect wherever it occurs: two different effects from a single cause contradicts spatial homogeneity. Therefore, simple codes of this kind are not viable informational bases for consciousness under physical orthodoxy.

Another way of illustrating this same issue is to explicitly recognise that B_1 = 0.1 and B_2 = 0.1 (say) are the *same* cause and must therefore have the *same* effect, under spatial homogeneity of causation. This can be achieved by dropping the condition that B_n = … creates a contribution-to-consciousness "in the n-th subject's experience", which is the origin of the multiplicity of *different* effects. Then B_1 = 0.1 generates a black pixel at the location labelled 0.1, and B_2 = 1.2 generates a white pixel at the location labelled 0.2. But without the requirements that B_1-states contribute to the first subject's experience and B_2-states contribute to the second subject's experience, the black pixel at 0.1 and the white pixel at 0.2 both occur within a *single* experience (*i.e.* within a single conscious-experiential space). More generally under this approach, *all* brains generate contributions to a single experience, which is empirically counterfactual (**Figure 1C**).

Problems of spatial homogeneity of causation and empirical accuracy can be resolved by emphasizing that locations in the experiences of the two subjects must be given different coordinates (**Figure 1D**). To illustrate this, let values in the range zero to one-half (exclusive) label locations in the first subject's experience, and values in the range one-half to one (exclusive) label locations in the second subject's experience. Furthermore, let corresponding locations in the two experiences be labelled by numbers differing by one-half, so that, for example, the location labelled 0.1 in the first subject's experience corresponds to the location labelled 0.6 in the second subject's experience. Then conscious experience is encoded for the first subject by B-values in the ranges (0,0.5) (for black pixels at various locations) and (1,1.5) (for white pixels at various locations). Similarly, the second subject's experience is encoded by B-values in the ranges (0.5,1) and (1.5,2). Now, for example, B_1 = 0.1 and B_2 = 0.6 generate black pixels at corresponding locations in the two subjects' experiences.

This approach recognizes that generation of a given contribution-to-consciousness in two different brains constitutes two different effects, and attributes two explicitly different causes to these effects, so that spatial homogeneity of causality is preserved. Notably, in this approach, the encoding of consciousness varies by subject (because numerically different *B*-conditions apply within different brains).

The example illustrates why encoding states for consciousness must vary by subject under physical orthodoxy, if locations within conscious experience are *encoded by B*-dynamics [the "property" approach to conscious-experiential space (Rosseinsky, 2014a)]. In an alternative approach, locations within conscious experience can instead *correspond to locations of B*-dynamics within orthodox physical space [the "sub-domain" approach (Rosseinsky, 2014a)]. This alternative approach can avoid the variation of encoding by subject, because each subject's experience occurs within their own brain (**Figure 2**), thus creating the required segregation of location. However, the multi-subject setting has distinct implications for sub-domain encodings. Previous papers discussed two different variants of sub-domain approaches [an "uber-map that recapitulates subjective geometry" (Rosseinsky, 2014b), on the one hand, and "dual-metric" conceptions (Rosseinsky, 2014a), on the other], and noted that the inclusion of dual-metric conceptions within the class of physically orthodox theories is of debatable validity. In the multi-subject setting, it will be shown that dual-metric approaches cannot be physically orthodox, because they require either an effect without a cause, or a non-orthodox property of physical brains. Thus, the multi-subject setting will restrict sub-domain approaches to the uber-map-recapitulating-subjective-geometry kind.

These illustrative discussions indicate that analysis of the multi-subject setting depends critically on earlier results concerning viable sub-approaches. Thus, it will be necessary as a preliminary to recapitulate critical aspects of earlier work, which is accomplished in Sections 2 and 3. Section 2 uses a central numerical example, complemented by more general formal statements, to review the spectrum of codes theoretically available for the informational basis of consciousness. [This pedagogical approach differs from earlier papers (Rosseinsky, 2014a, 2014b), that stressed details of formalism. Formal symbols originally defined in (Rosseinsky, 2014a) are summarised in Table 1.] Section 3 rehearses how previously derived limits on this spectrum result from the assumption of physical orthodoxy in the single-subject setting. Section 4 then provides a complete and rigorous demonstration of the multi-subject results informally discussed above, with reference both to classifications developed in the central numerical example and to complementary formalism.

As pointed out both by Barlow [in the general context of establishing a basic understanding of the brain's activity (Barlow, 1961)] and by Tononi and Koch [in the more specific context of understanding the brain's role in consciousness (Tononi and Koch, 2008)], theoretical principles can be invaluable in analysing what could otherwise become a mass of impenetrable data. Restrictions on consciousness-encoding established in the present series are precisely of the invaluable-principle kind: as discussed in Section 5, they lead to narrow and distinctive empirical signatures for the physically-orthodox class of implementations of a specific brain function (the generation of consciousness), a class whose existence might otherwise be invisible.

2. Varieties of encoding schemes

This Section describes a variety of encoding schemes that might in principal be employed as the informational basis for consciousness in a single-subject setting. These schemes are labelled "Type I"

[2.1] to "Type VI" [2.6], and range from single-value encodings of the entire conscious scene [2.1] to multiple-topography, vector-encoding approaches [2.6].

For each scheme, the encoding approach is first illustrated with reference to a simplified numerical example (Table 2), and then defined using formal symbols. Symbols used are fully defined in (Rosseinsky, 2014a). In summary: $s_{a,b}$ labels the b-th instance of the a-th feature type in the external environment, and $<s_{a,b}>$ labels the corresponding component of conscious experience; B is a measure of brain-dynamical activity whose values determine contributions-to-consciousness from a given brain-dynamical state; coordinates $\mathbf{r}$ label locations in orthodox physical space (including brain locations), whereas coordinates $\boldsymbol{\rho}_j$ label a finite set of neighborhoods within conscious experience at which exteroceptive experience is generated. (Table 1 summarizes these and other symbols used in the present series. **Figure 3** provides a schematic overview of the associated formal equations that can describe any theory-of-consciousness.)

2.1. Type I codes: dedicated values for each global scene

2.1.1. Illustrative example

Consider an illustrative example in which global conscious experience consists of colored, oriented, edges at just two locations (Table 2). At each location, edges can be either green or red, and (independently) either vertical or horizontal. For simplicity (*e.g.* excluding the possibility of "empty space" at either location), the total conscious experience is then one of sixteen combinatorial possibilities (Table 2). In a "Type I" code, there is a dedicated value for each global scene, *i.e.* for each combinatorial possibility. Thus, sixteen different values are used. This code is (maximally) *non-topographic*, because it employs only a single value. [Topographic maps employ a multiplicity of dynamically-relevant encoding locations, with each brain-location encoding a particular neighborhood of the external environment (Kaas, 1997).]

2.1.2. Formal symbolism for the general case

Formally, let $T(m)$ be the m-th member in an arbitrarily-ordered list of all combinatorially possible sets of (a,b,j) values [where a indexes feature-type, b indexes feature-instance, and j indexes location, *i.e.* an (a,b,j) triple defines a component $<s_{a,b}>(\boldsymbol{\rho}_j)$ of conscious experience]. For each m, define a distinct B-value $B_{T(m)}$ such that $B_{T(m)} \neq B_{T(n)}$ unless $m = n$, and let the occurrence of the physical state $B = B_{T(m)}$ (at any location $\mathbf{r}$) create the conscious experience corresponding to $T(m)$, *i.e.*

$$B(\mathbf{r}) = B_{T(m)} \Leftrightarrow \{<s_{a,b}>(\boldsymbol{\rho}_j), (a,b,j) \in T(m)\} \quad (1)$$

Note that in this paper, as in related works (Rosseinsky, 2014a, 2014b), no explicit theory-of-consciousness is given: B-contingencies such as Eq. 1 simply state details of the final brain-dynamical encoding of information employed in the generation of consciousness.

More generally, each $\{<s_{a,b}>(\boldsymbol{\rho}_j), (a,b,j) \in T(m)\}$ can be encoded by a multi-member *set* of B-values $\beta_{T(m)}$, so that the appropriate conscious experience is generated when any B-value in $\beta_{T(m)}$

occurs, *i.e.*

$$B(\mathbf{r}) \in \beta_{T(m)} \Leftrightarrow \{<s_{a,b}>(\rho_j), (a,b,j) \in T(m)\} \quad (2)$$

Here $\beta_{T(m)}$ and $\beta_{T(n)}$ must be non-overlapping sets for $m \neq n$, i.e. there is no B-value in both $\beta_{T(m)}$ and $\beta_{T(n)}$ unless $m = n$ (in which case the two sets are identical).

2.2. Type II codes: dedicated values for each feature combination at each location

2.2.1. Illustrative example

Continuing the example of [2.1.1], in a “Type II” code there are dedicated values for each *feature combination* at each location. A “feature combination” is *e.g.* “red, vertical” or “green, horizontal”. Because there are four feature combinations in the example, only four values are needed to encode a feature combination. However, if values encode both feature combination *and* location, eight values are needed (for the two locations in the example). The sixteen possible conscious scenes then derive (Table 2) from each of four values for feature-combination at the first location with each of four values (distinct from the previous four) for feature-combination at the second location. Thus, a type II code uses *two* numbers to encode each scene, where a type I code employed just one. This code is *non-topographic*, because locations at which values occur do not alter the features that are encoded. [For example, both “1,8” and “8,1” (Table 2, row 6) are unambiguous encodings of “Red-Vertical; Green-Horizontal”.]

2.2.2. Formal symbolism for the general case

Formally, let $S(n)$ be the n-th member in an arbitrarily-ordered list of all combinatorially possible sets of (a,b) values [where a indexes feature-type and b indexes feature-instance *i.e.* an (a,b) pair defines a component $<s_{a,b}>$ of conscious experience). For each n and each j, define a distinct B-value $B_{S(n),j}$ such that $B_{S(n),j} \neq B_{S(m),k}$ unless $n = m$ and $j = k$. Then let the occurrence of the physical state $B = B_{S(n),j}$ (at any location $\mathbf{r}$) generate the conscious experience corresponding to $S(n)$ at ρ_j, *i.e.*

$$B(\mathbf{r}) = B_{S(n),j} \Leftrightarrow \{<s_{a,b}>(\rho_j), (a,b) \in S(n)\} \quad (3)$$

More generally, each $\{<s_{a,b}>(\rho_j), (a,b) \in S(n)\}$ can be encoded by a multi-member *set* of B-values $\beta_{S(n),j}$, so that the appropriate conscious experience is generated when any B-value in $\beta_{S(n),j}$ occurs, *i.e.*

$$B(\mathbf{r}) \in \beta_{S(n),j} \Leftrightarrow \{<s_{a,b}>(\rho_j), (a,b) \in S(n)\} \quad (4)$$

Here $\beta_{S(n),j}$ and $\beta_{S(m),k}$ must be non-overlapping sets if $n \neq m$ or $j \neq k$, *i.e.* there is no B-value in both $\beta_{S(n),j}$ and $\beta_{S(m),k}$ unless $n = m$ and $j = k$ (in which case the two sets are identical).

2.3. Type III codes: dedicated values for each feature instance at each location

2.3.1. Illustrative example

In a "Type III" code there are dedicated values for each *feature instance* at each location, and values also encode location. A "feature instance" is one of "Red", "Green", "Vertical" or "Horizontal". Because there are four feature instances in the example, only four values are needed to encode a feature instance. However, if values encode both feature instance and location, eight values are needed (for the two locations in the example). Each of the sixteen possible conscious scenes is then encoded (Table 2) by *four* numbers, two for color and orientation at the first location, and two for color and orientation at the second location. This code is *non-topographic*, because locations at which values occur do not alter the features that are encoded.

2.3.2. Formal symbolism for the general case

Formally, let distinct values B_{abj} encode each $<s_{a,b}>$ at each ($\boldsymbol{\rho}_j$), so that the occurrence of the unique physical state $B = B_{abj}$ (at any location $\mathbf{r}$) creates the conscious experience of the feature $<s_{a,b}>$ at $\boldsymbol{\rho}_j$, *i.e.*

$$B(\mathbf{r}) = B_{abj} \Leftrightarrow <s_{a,b}>(\boldsymbol{\rho}_j) \quad (5)$$

More generally, each $<s_{a,b}>(\boldsymbol{\rho}_j)$ can be encoded by a multi-member *set* of B-values β_{abj}, so that $<s_{a,b}>(\boldsymbol{\rho}_j)$ is generated when any B-value in β_{abj} occurs, *i.e.*

$$B(\mathbf{r}) \in \beta_{abj} \Leftrightarrow <s_{a,b}>(\boldsymbol{\rho}_j) \quad (6)$$

Here β_{abj} and β_{cdk} must be non-overlapping sets if $a \neq c$ or $b \neq d$ or $j \neq k$, i.e. there is no B-value in both β_{abj} and β_{cdk} unless $a = c$ and $b = d$ and $j = k$ (in which case the two sets are identical).

2.4. Type IV codes: topography with dedicated values for feature instances

2.4.1. Illustrative example

In a "Type IV" code there are dedicated values for each feature instance (as in type III codes) but locations of feature instances are encoded *topographically* (instead of by value, as in type IV codes). Thus, only four different values are needed (for the four feature instances), but their order has encoding-meaning.

For example, in type IV code shown in Table 2, (1,3,1,4) in row 2 encodes Red-Vertical at $\boldsymbol{\rho}_1$ and Red-Horizontal at $\boldsymbol{\rho}_2$ whereas (1,4,1,3) in row 3 encodes Red-Horizontal at $\boldsymbol{\rho}_1$ and Red-Vertical at $\boldsymbol{\rho}_2$, so that order (or "location") is relevant to decoding. [The respective type III encodings for these two scenes are (1,3,5,8) and (1,4,5,7). Any permutations of these encodings *e.g.* (1,8,5,3) and (1,7,5,4) are still unambiguous for the type III non-topographic code, but shuffling typically changes meaning for type IV topographic encodings.]

2.4.2. Formal symbolism for the general case

Formally, let distinct values B_{ab} encode each $\langle s_{a,b} \rangle$, and let the *j*-index of $\boldsymbol{\rho}_j$ be determined by the *j*-index of the location ${}_{\mathrm{B}}\mathbf{r}_j$ at which a *B*-value occurs. Then the occurrence of the physical state $B = B_{ab}$ at ${}_{\mathrm{B}}\mathbf{r}_j$ creates the conscious experience of the feature $\langle s_{a,b} \rangle$ at $\boldsymbol{\rho}_j$, *i.e.*

$$B({}_{\mathrm{B}}\mathbf{r}_j) = B_{ab} \Leftrightarrow \langle s_{a,b} \rangle(\boldsymbol{\rho}_j) \qquad (7)$$

Note that Eq. 7 contains for the first time a *j*-subscript on **r**, which is the formal manifestation of the topographic nature of the type V code.

More generally, each $\langle s_{a,b} \rangle$ can be encoded by a multi-member set of *B*-values β_{ab}, so that $\langle s_{a,b} \rangle(\boldsymbol{\rho}_j)$ is generated when any *B*-value in β_{ab} occurs at ${}_{\mathrm{B}}\mathbf{r}_j$, *i.e.*

$$B({}_{\mathrm{B}}\mathbf{r}_j) \in \beta_{ab} \Leftrightarrow \langle s_{a,b} \rangle(\boldsymbol{\rho}_j) \qquad (8)$$

Here (exactly as for β_{abj} and β_{cdk} following Eq. 6) β_{ab} and β_{cd} must be non-overlapping sets if $a \neq c$ or $b \neq d$.

2.5. Type V codes: topography with dedicated values for feature combinations

2.5.1. Illustrative example

In a "Type V" code there are dedicated values for each feature *combination* (like a type II code), but these values do not change by location for a given feature combination (unlike a type II code). Instead, the type V code uses brain-location-of-value to encode environmental-location-of-feature. Hence, the type V code is a *topographic* variant of the type II code, because locations at which values occur are relevant to meaning.

For example, in the type V code shown in Table 2, (1,2) encodes Red-Vertical at $\boldsymbol{\rho}_1$ and Red-Horizontal at $\boldsymbol{\rho}_2$ whereas (2,1) encodes Red-Horizontal at $\boldsymbol{\rho}_1$ and Red-Vertical at $\boldsymbol{\rho}_{2.}$ [The respective type II encodings for these two scenes are (1,6) and (2,5). The shuffled values (6,1) and (5,2) are still unambiguous for the type II non-topographic code, but shuffling changes meaning for the type V topographic code.]

2.5.2. Formal symbolism for the general case

Formally, the type V code is a version of the type II code (Eqs. 3-4) in which the *j*-subscript of $\boldsymbol{\rho}_j$ derives from the particular *location* ${}_{\mathrm{B}}\mathbf{r}_j$ at which a *B*-condition is satisfied rather than from the *value* that *B* takes:

$$B({}_{\mathrm{B}}\mathbf{r}_j) = B_{S(n)} \Leftrightarrow \{\langle s_{a,b} \rangle(\boldsymbol{\rho}_j), (a,b) \in S(n)\} \qquad (9).$$

Eq. 9 amends Eq. 3 by dropping the *j*-subscript from $B_{S(n)}$ and adding it instead to **r**, thus making the topographic nature of the code formally explicit.

Similarly, the type V version of Eq. 4 is

$$B({}_{\mathrm{B}}\mathbf{r}_j) \in \beta_{S(n)} \Leftrightarrow \{<s_{a,b}>(\boldsymbol{\rho}_j),\ (a,b) \in S(n)\} \qquad (10)$$

(where, as in Eq. 4, $\beta_{S(n)}$ and $\beta_{S(m)}$ must be non-overlapping sets if $n \neq m$).

2.6. Type VI codes: maximally topographic dedicated-values-for-feature-instances

2.6.1. Illustrative example

In a "Type VI" code there are dedicated values for each feature instance (as in type III and type IV codes) but these values do not index environmental-location-of-feature (as they do in a type III code), nor are they distinct across feature type (as they are in a type IV code). Each of the sixteen possible conscious scenes is then encoded (Table 2) by *four* numbers, as in type III code and type IV codes. However, in the type VI code all four numbers are drawn from the same set (in Table 2, the set {1,2}), whereas type III codes use non-overlapping sets of values for all four entries (in Table 2, the set {1, …, 8}) and type IV codes uses distinct values for different feature types (in Table 2, the set {1, 2} is used for edge-orientations, and the set {3, 4} for colors). The type VI code is the *maximally topographic* variant of the type III code, because locations at which values occur are relevant to meaning, and the encoding significance of location is used to reduce the number of different encoding values to the smallest possible (binary) set.

For example, in type VI code shown in Table 2, (1,2,2,1) encodes Red-Horizontal at $\boldsymbol{\rho}_1$ and Green-Vertical at $\boldsymbol{\rho}_2$ whereas (2,1,1,2) encodes Green-Vertical at $\boldsymbol{\rho}_1$ and Red-Horizontal at $\boldsymbol{\rho}_2$. [The respective type III encodings for these two scenes are (1,4,6,7) and (2,3,5,8). Any permutations of these encodings *e.g.* (4,1,7,6) and (3,2,5,8) are still unambiguous for the type III non-topographic code, but shuffling changes meaning for the type VI topographic code.]

2.6.2. Formal symbolism for the general case

Formally, let distinct values B_b encode each b-value for $<s_{a,b}>$, and let the a-index of $<s_{a,b}>$ and the j-index of $\boldsymbol{\rho}_j$ be determined by the corresponding index of the location ${}_{\mathrm{B}}\mathbf{r}_{a,j}$ at which a B-value occurs. Then the occurrence of the physical state $B = B_b$ at ${}_{\mathrm{B}}\mathbf{r}_{a,j}$ creates the conscious experience of the feature $<s_{a,b}>$ at $\boldsymbol{\rho}_j$, *i.e.*

$$B({}_{\mathrm{B}}\mathbf{r}_{a,j}) = B_b \ \Leftrightarrow <s_{a,b}>(\boldsymbol{\rho}_j) \qquad (11)$$

More generally, each b-value can be encoded by a multi-member set of B-values β_b, so that $<s_{a,b}>(\boldsymbol{\rho}_j)$ is generated when any B-value in β_b occurs at ${}_{\mathrm{B}}\mathbf{r}_{aj}$, *i.e.*

$$B({}_{\mathrm{B}}\mathbf{r}_{a,j}) \in \beta_b \ \Leftrightarrow <s_{a,b}>(\boldsymbol{\rho}_j) \qquad (12)$$

(where, as in Eqs. 6 and 8, β_b and β_d must be non-overlapping sets if $b \neq d$).

3. Physical orthodoxy restricts viable consciousness-encodings in a single-subject setting

Section 2 illustrated a variety of potential (single-subject) encodings of consciousness. In this Section, the condition that resultant theories-of-consciousness be physically orthodox is imposed, leading to consequent encoding restrictions [derived previously in detail in (Rosseinsky, 2014a, 2014b)].

The two primary aspects of physical orthodoxy that restrict encoding of consciousness are locality (Rosseinsky, 2014a) and spatial homogeneity (Rosseinsky, 2014b) of causality. The two aspects impose restrictions on encodings of consciousness [3.2-3.7] that do not apply to "behavioral encoding" (*i.e.* the computational neuroscience of behavior: [3.1]), because physically-orthodox theories-of-consciousness cannot hypothesize intermediary physical structure between final encoding of consciousness and components-of-experience, whereas the dependence of behavior on sensory encoding is mediated by a massive amount of physical structure (*i.e.* the "non-sensory brain"). Physical structure supporting behavioral computations ensures that these processes inherently obey requirements of physical orthodoxy, whereas the absence of such structure in physically-orthodox theories-of-consciousness (Rosseinsky, 2014a) means that explicit restrictions must be applied to ensure that physical orthodoxy is always obeyed (Rosseinsky, 2014a, 2014b).

3.1. Wide variety of viable codes for behavioral computations

As will be discussed in [3.1.1-3.1.3], behavioral computations (as opposed to encodings-of-consciousness) can employ any of the stereotypical codes described in Section 2 *without* restriction, as well as other types of code *not* discussed in Section 2 (see [3.1.4]).

3.1.1. "Grandmother cell" and single-scalar schemes: type I codes

Although topographic codes are prevalent in early human sensory processing (Kaas, 1997), it is in principle possible that a *single* later cell might integrate information from multiple earlier topographies and thus encode the *entirety* of the external environment in its dynamical behavior. For example, the type I column in Table 2 can be interpreted as sixteen different firing-rates of a single cell. This interpretation of the type I code will be termed a "single-scalar" code, because it uses various values of a single-location physical measure, as opposed to a vector of many locations.

Alternatively, the sixteen type I values can be interpreted as different locations of cells, each dedicated to encoding a particular external environment. A code in which individual cells are dedicated to encoding highly-specific complex combinations of features is termed a "grandmother cell" approach (Gross, 2002). Debates in the literature concerning the computational merits and biological plausibility of "grandmother cell" approaches (Bowers, 2009) *vs*. parallel and distributed alternatives (Plaut and McClelland, 2010) are neglected here, because the present objective is to describe the full set of possibilities for encoding-of-consciousness, rather than to perform a comparative assessment of behavioral codes. The relevant distinction for encoding-of-consciousness is between a single-scalar approach that uses many different values at a single location, and a "grandmother cell" approach that uses a single value at many different locations: these two sub-types of type I codes have different viabilities under physical orthodoxy [3.5.3].

3.1.2. Uber-map topographic schemes: type II and type V codes

Two qualitatively different forms of topographic code discussed in previous papers (Rosseinsky, 2014a, 2014b) are relevant to present considerations. The first is the "uber-map" approach, in which a single topographic map encodes all features of the environment. If there are N_E distinct locations in the external environment that are sampled for the purposes of sensory encoding, then an uber-map contains N_E dynamically-relevant locations (neglecting the possibility of distributed encoding). Thus, by inspection of Table 2, uber-maps can only employ type II or type V codes, because these are the only codes with two encoding locations (corresponding to the two environmental locations, *i.e.* N_E = 2, in the illustrative example). Moreover, type II and type V codes can both be interpreted as topographies, because in each case (as written in Table 2), the first number encodes features at ρ_1 and the second number encodes features at $\rho_{2,}$ and the two numbers are independent. (Type II codes were described as *non*-topographic in [2.2], which is strictly correct, because these codes are insensitive to permutation. However, they can be employed in a topographic map in a redundant manner, wherein both *B*-value and brain-location encode conscious-experiential location.)

3.1.3. Many-map topographic schemes: type III, type IV and type VI codes

The second type of approach to topographic encoding is a many map scheme, in which there is a topographic map dedicated to each feature type (Rosseinsky, 2014a). If there are N_E distinct locations in the external environment and N_F feature types, then a many-map scheme contains $N_E \times N_F$ encoding locations. Examination of Table 2 shows that type III, type IV, and type VI codes can all be interpreted as many-map topographic encodings (although type III codes contain redundant information, if interpreted topographically).

3.1.4. More complex codes available for behavior

Whereas the codes illustrated in Section 2 are of a very simple kind, behavioral codes can in principle be arbitrarily complex (Churchland and Sejnowski, 1992; Tsodyks *et al*., 1996; Rieke *et al*., 1997; Reinagel and Reid, 2000; Haynes and Rees, 2006; Pillow *et al*., 2008; Quiroga and Panzeri, 2009; Panzeri *et al*., 2010), provided that codes are amenable to decoding by biophysically feasible means. For example, behavioral codes can in principle use a variety of different physical measures simultaneously [such as power at various frequencies and a wide variety of phase relationships (Makeig *et al*., 2004)], as well as employing complex spatiotemporal patterns. Thus, feasible behavioral codes range far beyond the set illustrated in Section 2.

However, more complex codes of these kinds are *not* viable as final codes for consciousness *under physical orthodoxy*, because they must be integrated or assimilated to create a direct, physically local, spatiotemporally point-like code (Rosseinsky, 2014a), and no machinery can exist to perform this integration in a physically-orthodox theory-of-consciousness. (In a physically-orthodox theory, the brain is the only apparatus available to perform integration.) Thus, feasible codes for consciousness are limited in the first instance to the set illustrated in Section 2 (or close variants). Moreover, even this set is subject to further restrictions under physical orthodoxy, as discussed in [3.4-3.7].

3.2. Physical orthodoxy excludes distributed codes for consciousness

A distributed code uses activity at many locations to encode a single fundamental piece of information (Rosseinsky, 2014a). Distributed codes are ubiquitous in early sensory processing, but cannot be used in physically-orthodox theories-of-consciousness because they violate locality of physical action (Rosseinsky, 2014a). Thus, distributed codes can be thought of as another form of "complex code" available to behavior but not consciousness [3.1.4].

3.3. Physically orthodox theories use sub-domain or property approaches

Any complete theory-of-consciousness must propose a definitive relationship between orthodox physical space and conscious-experiential space (Rosseinsky, 2014a). Two approaches to this relationship are available for physically-orthodox theories. First, in "sub-domain" schemes, conscious-experiential space is a sub-domain of orthodox physical space. Second, in "property" schemes, conscious-experiential space is a property of brain dynamics, and is both ontologically subordinate to, and topologically disconnected from, orthodox physical space. [A variety of other, non-orthodox, possibilities are discussed in (Rosseinsky, 2014b).] Viable coding approaches for consciousness depend on whether a sub-domain or property approach is employed, as discussed respectively in [3.4] and [3.5].

3.4. Sub-domain approaches recapitulate subjective geometry or require dual metrics

Sub-domain approaches can be of two kinds. If an uber-map (a single topographic map for all features simultaneously [3.1.2]) recapitulates subjective geometry (Rosseinsky, 2014a), then orthodox physical space can directly contain conscious experience, without further elaboration. Alternatively, if features are encoded either in a many-map scheme or in an uber-map that does not recapitulate subjective geometry, a second metric must exist on orthodox space that ensures the arrangement of features-in-experience corresponds to subjective geometry. [As noted in an earlier discussion (Rosseinsky, 2014b), the attribution of physically-orthodox status to dual metric approaches is at best debatable.]

3.4.1. Sub-domain uber-maps: type II and type V codes viable

For behavioral codes, both type II and type V codes are viable approaches in an uber-map setting [3.1.2]. For consciousness-encoding in a sub-domain approach employing an uber-map, this continues to be true (but does not continue to hold for uber-map consciousness-encoding in a *property* approach [3.5.1]). However, of later significance [4.2] for the multi-subject setting will be the fact that sub-domain uber-maps not directly recapitulating subjective geometry must employ a dual metric.

3.4.2. Sub-domain many-maps: type III and type IV but not type VI codes viable

For behavioral codes, type III, type IV, and type VI codes are all viable approaches in a many-map scheme [3.1.3]. For consciousness-encoding in a sub-domain approach employing many-maps, type III and type IV but *not* type VI codes are viable, under physical orthodoxy. The exclusion of type VI codes for consciousness under physical orthodoxy follows from spatial homogeneity of causality (Rosseinsky, 2014b). Consider, for example, the entry "1,1,1,1" in the first row of the "Type VI" column in Table 2. The first "1" in this code has the effect "conscious-experience of *Red* at ρ_1", while the second "1" has the effect "conscious-experience of *Vertical-edge* at ρ_1": two identical physical conditions ("B = 1") have two different effects, depending on the location at which they occur, which is not consistent with spatial homogeneity of causality (Rosseinsky, 2014b).

Of later significance [4.2] for the multi-subject setting will be the fact that *every* sub-domain *many*-map code must employ a dual metric. (For sub-domain uber-maps [3.4.1], only non-recapitulating geometries are subject to this constraint – a many-map approach cannot directly recapitulate subjective geometry, because feature types are inherently spatially segregated).

3.5. Property approaches require dynamical encoding of ρ coordinates

If conscious-experiential space is a property of brain dynamics in a physically-orthodox theory-of-consciousness, then the ρ-location of a brain-encoded feature must also be encoded by brain *dynamics*, rather than being entailed by brain-*location* of encoding (Rosseinsky, 2014b). This follows from spatial homogeneity of causality: otherwise, identical local B-states at two different locations would have different consequences (*i.e.* different ρ-coordinates).

3.5.1. Property uber-maps: type II but not type V codes viable

For behavioral encoding, both type II and type V codes are viable approaches in an uber-map setting [3.1.2]. However, for consciousness-encoding in a property approach employing an uber-map, only type II codes are viable, because type V codes do not value-encode the ρ-coordinate of feature-combinations. Consider, for example, the entry "1,1" in the first row of the "Type V" column in Table 2. The first "1" has the effect "conscious-experience of Red-Vertical *at* ρ_1", and the second "1" has the effect "conscious-experience of Red-Vertical *at* ρ_2", thus violating spatial homogeneity of causality.

3.5.2. Property many-maps: type III but not type IV or type VI codes viable

For behavioral codes, type III, type IV, and type VI codes are all viable approaches in a many-map approach [3.1.3]. For consciousness-encoding in a sub-domain approach employing many-maps, type

III but *not* type IV or type VI codes are viable, under physical orthodoxy [3.4.2]. Type IV codes are viable under the sub-domain approach (because requisite $\boldsymbol{\rho}$-coordinates follow from $\mathbf{r}$ locations) but not under the property approach (because $\boldsymbol{\rho}$-coordinates are not directly value-encoded in the type IV code). This follows from spatial homogeneity of causality: the condition $B = 1$, for example (Table 2), has two different conscious-experiential effects, namely "Red *at* $\boldsymbol{\rho}_1$" and "Red *at* $\boldsymbol{\rho}_2$". (The exclusion of type VI codes for property schemes under physical orthodoxy also follows from spatial homogeneity, just as in [3.4.2].)

3.5.3. Single-scalar type I code viable for property but not sub-domain approaches

The single-scalar version [3.1.1] of the type I code is viable for property but not sub-domain approaches if physical orthodoxy applies. Because sub-domain approaches employ orthodox physical space as the container for conscious experience, single-scalar sub-domain approaches must be non-local (activity at a single location in orthodox space has effects at many other locations in orthodox space). In contrast, under certain interpretations of the property approach (Rosseinsky, 2014b), activity at a single location in orthodox space can have effects at many locations in conscious-experiential space without incurring non-locality [because the locality requirement from physical orthodoxy applies to $\mathbf{r}$-to-$\mathbf{r}$ displacements, not to $\mathbf{r}$-to-$\boldsymbol{\rho}$ displacements, although what can be transferred from $\mathbf{r}$-to-$\boldsymbol{\rho}$ is limited by the conception of the property approach as physically orthodox (Rosseinsky, 2014a)].

However, the "grandmother cell" version [3.1.1] of the type I code is not physically orthodox, even for property approaches, because it violates spatial homogeneity of causality (homogeneous activity in different cells leads to different conscious scenes).

3.6. Spatial homogeneity severely restricts occurrence of *B* encoding-values

Consider a specific firing pattern in a specific neuron that is interpretable as encoding the existence of a particular feature in the external environment. Trivially, the occurrence of this same pattern in a different neuron *cannot* generally be interpreted as the same encoding (except under very special circumstances, *e.g.* when the two neurons always fire in the same pattern, by virtue of mutual connectivity). Thus, dynamical states that have behavioral encoding significance at a given brain-location can occur elsewhere in the brain without generating the same encoding. This is *not* true for *B*-encodings of consciousness under physical orthodoxy.

Consider a condition, $B = 1$ say, whose satisfaction generates some component-of-consciousness $\langle s_{a,b}\rangle(\boldsymbol{\rho}_j)$. Under physical orthodoxy, if $B = 1$ *anywhere in the brain* then $\langle s_{a,b}\rangle$ must be generated. This follows from spatial homogeneity of causality, and the fact that no intermediate machinery can be invoked by a physically-orthodox theory-of-consciousness. Therefore, $B = 1$ states can only occur when $s_{a,b}$ is present in the external environment. The strength of this observation cannot be overstated: *B*-states implicated in the physically-orthodox generation of components-of-consciousness must be *precisely restricted* in a global sense by brain biophysics, and both the fact of this control and the biophysical mechanisms required to implement it should be extremely powerful indicators towards the physical basis of consciousness, if physical orthodoxy pertains (Rosseinsky, 2014b). One demonstration of the power of this result is that it excludes the electromagnetic-field ("e.m.-field") activity of the brain as the final encoding substrate for consciousness under physical orthodoxy, because e.m.-field states cannot be of the precisely-restricted kind (Rosseinsky, 2014b).

3.7. Physical orthodoxy admits limited set of encoding methods for consciousness

Table 3 summarises the situations in which the various codes introduced in Section 2 can be viable bases for consciousness under physical orthodoxy, thus encapsulating much of the discussion of [3.1-3.5]. Two features are directly apparent from Table 3. First, under physical orthodoxy, type VI codes that are valid for behavioral encoding [3.1.3] are *never* viable bases for consciousness. Second, other codes (of types I to V) are only valid under the specific accompanying circumstances defined in the Table rows. These two observations indicate some of the restrictions that physical orthodoxy imposes on consciousness-encoding.

However, the most significant restrictions imposed by physical orthodoxy are not directly present in Table 3. First, Table 3 does not show (directly) the huge variety of codes viable for behavior but not for consciousness [3.1.4; 3.2; 3.5.3]. Put differently, Table 3 is remarkable for the *very few* approaches that it includes. Second, Table 3 does not show explicitly that the occurrence of *B*-values with consciousness-encoding significance must be strictly controlled [3.6], which is possibly the most powerful single restriction on encoding-of-consciousness as compared to encoding-for-behavior.

4. Additional restrictions in a multi-subject setting

Section 3 constructed the set of all possible encoding bases for consciousness in a physically-orthodox *single*-subject setting, culminating in the rather small and restricted set of approaches displayed in Table 3 and discussed in [3.7]. This Section identifies further restrictions on consciousness-encodings that occur in the physically-orthodox *multi*-subject case.

4.1. Formalism for the multi-subject setting

In a multi-subject setting, conscious-experiential coordinate systems $\boldsymbol{\rho}$ must be constructed for each subject. Let the coordinate system for the *n*-th subject be denoted by ${}_n\boldsymbol{\rho}$, so that locations in the experience of the first subject are indexed by coordinates ${}_1\boldsymbol{\rho}_1$, ${}_1\boldsymbol{\rho}_2$, … *etc*., whereas locations in the experience of the second subject are indexed by ${}_2\boldsymbol{\rho}_1$, ${}_2\boldsymbol{\rho}_2$, …, and so forth. The generic expression for the generation of conscious experience in a *single*-subject setting is (Eqs. 2,4,6,8,10, and 12)

$$B(\mathbf{r}) \in X \Leftrightarrow \langle s_{a,b} \rangle(\boldsymbol{\rho}_j) \quad (13)$$

for some set *X*, where $\mathbf{r}$ and *X* on the left-hand side ("LHS") jointly determine the values of *a*, *b*, and *j* on the right-hand side ("RHS"), in a theory-dependent manner. For example, in Eq. 2, $X \equiv \beta_{T(m)}$ and $T(m)$ determines collection of (a,b,j) values [and (a,b,j) values are independent of $\mathbf{r}$], whereas in Eq. 12, $X \equiv \beta_b$ determines the *b*-value and $\mathbf{r} = {}_B\mathbf{r}_{aj}$ determines the *a*- and *j*-values.

The generic equation corresponding to Eq. 13 for the *multi*-subject setting is

$$B(\mathbf{r}) \in X \Leftrightarrow <s_{a,b}>({}_{n}\boldsymbol{\rho}_{j}) \qquad (14),$$

which obviously has an *extra* subscript n on the RHS whose value must be set in some way by conditions on the LHS. Under physical orthodoxy, this n-value can be set in one of two ways, depending on whether conscious-experiential space is a sub-domain of orthodox space or a property of brain dynamics [3.3]. As will be shown in [4.2], for physically-orthodox *sub-domain* approaches, the n-value can follow from an n-label attached to the **r**-coordinate on the LHS of Eq. 14. That is, instead of writing ${}_{\mathrm{B}}\mathbf{r}_j$ for B-encoding locations (as in Eqs. 8, 10, and 12), in the multi-subject setting the proper formalism is to write ${}_{\mathrm{B},n}\mathbf{r}_j$ for the j-th encoding location in the n-th subject. However, as will be shown in [4.3], for physically-orthodox *property* approaches, the n-value on the RHS of Eq. 14 cannot follow from ${}_{\mathrm{B},n}\mathbf{r}_j$ because this would violate spatial homogeneity of causality. Instead, there must be an additional n-subscript on the X-set on the LHS, which means that codes for consciousness must vary by subject.

4.2. Sub-domain approach in the multi-subject setting

4.2.1. Multi-subject n-value of ${}_{n}\rho$ follows from n-value of ${}_{n}$r in sub-domain approach

In sub-domain approaches, components of conscious experience exist in the same *space* as the brain-dynamics that encode them. Under the locality required by physical orthodoxy, components of consciousness are also generated at the same *location* as their encoding dynamics. Thus, collections of components are naturally grouped together in space, by virtue of being co-located with particular brains (**Figure 2**). Hence, a particular feature (or component-of-consciousness) belongs unambiguously to a particular collection, *i.e.* to a particular subject's experience. This (sub-domain contingent) *physical* fact makes it *formally* valid for the n-value of ${}_{n}\boldsymbol{\rho}_j$ to be set by the n-value of ${}_{\mathrm{B},n}\mathbf{r}_j$.

In formal terms, the notation $\mathbf{r}(\boldsymbol{\rho})$ was introduced in (Rosseinsky, 2014b) for sub-domain approaches, to emphasize that **ρ**-coordinates are simply additional labels for locations already labelled by **r** coordinates. [In this notation, **r**(…) is a function that can only take **ρ**-coordinates as its argument. The notation *with this specific meaning* is, of course, applicable to sub-domain but not property approaches.] In the multi-subject setting, then, it is possible to write

$${}_{n}\boldsymbol{\rho}_j = \mathbf{r}^{-1}({}_{\mathrm{B},n}\mathbf{r}_j) \qquad (15).$$

where $\mathbf{r}^{-1}(\ldots)$ is the inverse function of **r**(…), that takes (certain) **r**-coordinates as argument, and evaluates to corresponding **ρ**-labels. Grouping together of the ${}_{n}\boldsymbol{\rho}_j$ for a given subject (*i.e.* a particular n) then follows from the grouping together in orthodox physical space of the ${}_{\mathrm{B},n}\mathbf{r}_j$ (for the same n) (**Figure 2**). The ${}_{\mathrm{B},n}\mathbf{r}_j$ are grouped together in orthodox physical space because they label points within the n-th brain.

4.2.2. Dual-metric approaches not physically orthodox in the multi-subject setting

Once generated at the *same point* as their encoding dynamics (in a *local* sub-domain theory), components-of-consciousness are either already arranged in a manner that generates subjective experience, or they must be construed as existing in an alternative configuration that *does* generate subjective experience. The latter construal was formally described in (Rosseinsky, 2014a) as a "dual-metric" approach, because points in orthodox space must be conceived of being arranged relatively in two different ways (*i.e.* in both objective and subjective configurations). As noted in (Rosseinsky, 2014b), whether a dual-metric theory can be considered physically-orthodox is debatable, even in a single-subject setting, because orthodox physics recognizes only one metric. However, dual-metric approaches were described initially and have been maintained thus far because they preserve the sub-domain version of an intuition that "features-in-experience need not be arranged relatively in space as are their encoding dynamics" (so to speak). Moreover, excluding dual-metric approaches significantly reduces the set of sub-domain theories, because a dual-metric is required by any sub-domain encoding approach other than an uber-map [3.1.2] that directly recapitulates subjective geometry without re-arrangement.

In a multi-subject setting, the dual-metric conception is definitively *inconsistent* with physical orthodoxy, because it supposes an effect without a cause (or attributes non-orthodox effects to physical brains). To see this, consider an **r** coordinate system whose origin is fixed relative to the earth's surface, and two subjects (labelled n = 1 and n = 2), initially at rest with respect to the **r**-origin. Let ${}_1\mathbf{r}_0$ be the coordinates for the center of subject 1's brain, and take the origin of the ${}_1\boldsymbol{\rho}$ coordinate system to be this same point (*i.e.* the point also labelled ${}_1\mathbf{r}_0$). Similarly, let ${}_2\mathbf{r}_0$ denote the center of subject 2's brain, and let the origin of the ${}_2\boldsymbol{\rho}$ coordinate system be the point labelled ${}_2\mathbf{r}_0$. Under the dual-metric concept, there is a set of points in the neighborhood of ${}_1\mathbf{r}_0$ that have a special physical property (namely, dual arrangements), and there is another set of points in the neighborhood of ${}_2\mathbf{r}_0$ that also have this special property. Note that specific details of the special property differ at the two locations, because the rearrangement depends on the locations of brain-encoding sites relative to ${}_1\mathbf{r}_0$ and ${}_2\mathbf{r}_0$, respectively (**Figure 4**). Next, consider the *movement* of subject 2 relative to subject 1, and let the new location of the center of subject 2's brain be labelled ${}_2\mathbf{r}_0'$. Now the special property applies in the neighborhoods of ${}_1\mathbf{r}_0$ and ${}_2\mathbf{r}_0'$, *i.e.* the regions in which metric-duality applies must *co-move with subjects*. This movement of special neighborhoods is an effect without a cause, unless the dual metric is caused by some aspect of the physical brain (which is a non-orthodox property). In single-subject settings, physically-orthodox characterization of dual-metric suggestions implicitly relies on a conception of metric-duality as an inherent physical property of a *subject-centric* universe. However, the multi-subject setting reveals the untenable nature of this attempt.

4.2.3. Exclusion of dual-metric approach excludes most sub-domain codes

Because "a dual-metric is required by any sub-domain encoding approach other than an uber-map that directly recapitulates subjective geometry without re-arrangement" [4.2.2], the exclusion of dual-metric approaches from the class of physically-orthodox theories-of-consciousness [4.2.2] means that most sub-domain codes are *not* physically orthodox. Specifically, many-map approaches and uber-maps that do not inherently recapitulate subjective geometry cannot be physically orthodox, thus excluding the encoding approaches in the second and third rows of Table 3.

4.3. Property approach in the multi-subject setting

4.3.1. Multi-subject n-value of $_n\rho$ follows from n-value of $\beta_{\ldots,n}$ in property approach

In the property approach, components of conscious experience are generated in a conscious-experiential space that is distinct from orthodox physical space. Under physical orthodoxy, the coordinate locations of components-of-consciousness within this space must be specified by brain dynamics, so that, for example, the B-value encoding "Red at $\boldsymbol{\rho}_1$" must differ from the B-value encoding "Red at $\boldsymbol{\rho}_2$". This follows from the spatial homogeneity of causality, together with the absence of intermediate physical structure linking **r**- and $\boldsymbol{\rho}$-locations in a physically-orthodox theory-of-consciousness (Rosseinsky, 2014b). Similarly, in the *multi*-subject setting, the B-value encoding "Red at $_1\boldsymbol{\rho}_1$" (*i.e. subject 1's* "Red at $\boldsymbol{\rho}_1$") must differ from the B-value encoding "Red at $_2\boldsymbol{\rho}_1$" (*i.e. subject 2's* "Red at $\boldsymbol{\rho}_1$"). Otherwise, a single B-value would have two different effects, creating features at both $_1\boldsymbol{\rho}_1$ and $_2\boldsymbol{\rho}_1$. (Note that $\boldsymbol{\rho}$-coordinate systems *must* be distinguished by subject in this way, otherwise all brains generate features in a single coordinate system, *i.e.* in a single collective experience, which is counterfactual.)

Formally, the variation of B-encoding values across subject means that the β-sets in Eqs. 2, 4, and 6 acquire an extra n-subscript, becoming respectively $\beta_{T(m),n}$, $\beta_{S(m),j,n}$, and $\beta_{abj,n}$. (Only Eqs. 2, 4, and 6 are extended in this way, because physically-orthodox *property* approaches can only employ codes of types I, II, and III [3.5].)

4.3.2. Details but not number of valid property codes change in the multi-subject setting

Whereas the multi-subject setting excludes certain kinds of *sub-domain* codes that were considered valid in the single subject setting, the multi-subject setting does not change the overall *number* of valid *property* codes. However, the *details* of each kind of valid property code change in the multi-subject setting, because physically-orthodox property-based codes must be subject-dependent.

4.4. Multi-subject setting restricts and defines physically-orthodox consciousness-encodings

Restrictions on valid codes resulting from physical orthodoxy in a single-subject setting were discussed in [3.7], and displayed in Table 3. Table 4 updates Table 3 to reflect further restrictions on valid codes in the multi-subject setting. Notably, Table 4 contains about half as many valid codes as Table 3, and shows a reduction in the *number* of *sub-domain* codes but a change in the *details* of *property* codes. (Restrictions discussed in [3.7] but not explicitly shown in Table 3 still apply in the multi-subject setting, *i.e.* Table 4 must be complemented by the same additional restrictions as Table 3.)

5. Discussion

5.1. Experimental predictions

Restrictions on codes are experimental predictions, because both codes themselves and the biophysical infrastructure to generate them must be observable (Rosseinsky, 2014a, 2014b).

Although brain-dynamical observations of the human brain *in vivo* at the neuronal and molecular level are not presently feasible, two currently active modes of enquiry can verify or refute the existence of restricted codes of the kinds identified here. First, detailed anatomical and biophysical mapping of brain structure (Alivisatos *et al*., 2013) can provide insight into the classes of codes that brains can generate. Second, the output from large scale brain simulations (Markram, 2006) can be parsed for the existence of candidate *B*-measures meeting restrictions. Notably, the set of viable codes for consciousness are all dramatically simpler than presently hypothesized behavioral codes, so that their occurrence and machinery for their generation should be very visible.

Table 3 summarises structural restrictions on physically-orthodox codes resulting from prior single-subject considerations (Rosseinsky, 2014a, 2014b), and Table 4 shows how these restrictions become stronger in the multi-subject setting. A notable further condition on all physically-orthodox codes is that *B*-values encoding for components-of-consciousness can only occur in severely restricted and specific circumstances [3.6], *i.e.* if and *only if* the existence of the corresponding environmental feature is encoded in brain dynamics.

5.2. Observations excluding physical orthodoxy inform theory-of-consciousness

Empirical observations *excluding* the existence of any code meeting the restrictions of Table 4 would establish the significant result that consciousness is not a physically-orthodox phenomenon. As discussed in earlier work (Rosseinsky, 2014a, 2014b), physical orthodoxy could then be relaxed in one of two ways. In one direction, otherwise-universal principles of locality and spatial homogeneity of causality might not apply to the generation of consciousness. In another direction, locality and homogeneity can be preserved at the cost of theories-of-consciousness that utilize either extra dimensions or non-orthodox particle spectra to hypothesize intermediate physical structure.

5.3. Observations within orthodox set inform theory-of-consciousness

Empirical observations *supporting* the existence of a code meeting the restrictions of Table 4 would allow persistence of the mainstream view (Anderson, 1972; Baars, 1988; Edelman, 1989; Crick and Koch, 1990; Tononi and Edelman, 1998; Tegmark, 2000; Churchland, 2005; Baars and Edelman, 2012) that consciousness is a physically-orthodox phenomenon. Moreover, such observations would provide useful information of two kinds. First, the physical substrate of the code at the molecular or neuronal level would provide strong indications of the actual physical basis of consciousness, dramatically narrowing the scope for further experimental investigations. Second, as illustrated in Table 5, the *kind* of code observed would typically narrow the leading candidates for the explanation of consciousness within the physically-orthodox setting. For example, if a type II code with recapitulation of subjective geometry and without cross-subject variation is observed (the first row of Table 5), then the central candidate must be the sub-domain approach. The only ambiguous case is a type II code that both varies by subject and recapitulates subjective geometry (fourth row, Table 5), which supports both property and sub-domain theories.

5.4. “Strictest” physical orthodoxy excludes sub-domain approaches

The definition of physical orthodoxy adopted so far has emphasized two separate aspects. First, repeated explicit appeals have been made to locality and spatial homogeneity of causality. Second, in order to exclude the possibility of intermediate physical structure beyond the orthodox quark-electron-photon constituted brain, an orthodox three-dimensional spatial setting containing standard particle types has been emphasized. These aspects of physical orthodoxy are uncontroversial.

In order to provide a complete assessment of the relationship between physical orthodoxy and theories-of-consciousness, note that the “strictest” definition of physical orthodoxy further extends the usual conception to include *ontology-independence* of theories (Russell, 1927). That is, if an *empirically unsupported* axiomatic assertion of a conventionally realist ontology is to be avoided, science must be built on foundations that admit every conceivable ontology. Certainly, although this is not usually emphasized, all theories and observations in physical (as opposed to biological) science are *explicitly* consistent with ontological-independence, because physical science is reducible to quantitative measures made by instruments and explanation of these measures by purely mathematical theories. However, not all theories-of-consciousness discussed here are ontology-independent, because the sub-domain approach relies on the existence of space in a conventionally realist (or similar) ontology. Thus, adopting the strictest view of physical orthodoxy *excludes sub-domain theories from the physically orthodox class*.

5.5. Evolutionary and developmental problems in property approaches

As noted in a previous paper (Rosseinsky, 2014b), the exclusion of the e.m.-field as the final encoding substrate for consciousness [3.6] accentuates problems (Velmans, 2012) associated with the evolutionary explanation of physically-orthodox consciousness. Evolutionary problems originate in an assumption that consciousness is not causally efficacious (Huxley, 1893), as is necessarily assumed by a physically-orthodox theory-of-consciousness. Multi-subject results here further accentuate these problems, especially for property approaches in which the encoding of consciousness must vary by subject. (For sub-domain approaches, the multi-subject setting restricts codes to uber-maps recapitulating subjective geometry. These theories are, to say the least, peculiar from an evolutionary perspective, because there are no selective pressures to ensure the appropriate geometry of encoding dynamics, if consciousness is not causally efficacious. However, evolutionary problems are *even greater* for property approaches.)

If a property approach is to explain consciousness, there must be a large number of disjoint sets of *B*-states, each of which can fully encode complete conscious experiences (*i.e.* every feature combination at every ρ-location). Moreover, each human brain must create *B*-values in *precisely one* of these sets, and evolutionary and developmental processes must exist to ensure that this happens. (Otherwise, a space indexed by a given ${}_n\rho$ coordinate system will contain experiences from more than one brain.) Although it is possible to imagine complex genetic processes that could ensure that there is no overlap between the *B*-sets in any two brains, there are again no selective pressures to create this outcome when consciousness is not causally efficacious.

5.6. The unity of conscious experience

5.6.1. New characterizations of aspects of unity offered by present approach

The "unity of consciousness" has been discussed by various commentators (Cleeremans, 2003; Bayne, 2012), primarily from philosophical and psychological rather physical or biophysical viewpoints. The present series offers an entirely new characterization of various aspects of unity, by framing properties of conscious experience in terms of *spaces* and *coordinates*. For example, an earlier discussion (Rosseinsky, 2014b) pointed out that the single-subject binding problem in exteroceptive conscious experience (sometimes described as the problem of "representational unity" of consciousness; **Figure 5A,B**) can be framed as the question of how various $<s_{ab}>$ that relate to a single location ${}_{X}\mathbf{r}_1$ (say) in the external environment come to have the same $\boldsymbol{\rho}_1$ coordinate, rather an a variety of unequal $\boldsymbol{\rho}$-coordinates (*i.e.* locations in conscious-experiential space). Notably, present formalism emphasizes that *conscious-experiential* binding problems relate to the generation of $\{<s_{ab}>\}$ at $\boldsymbol{\rho}_j$ locations, whereas the binding problem in *behavioral computation* [or the "general coordination problem" (Feldman, 2013)] concerns the very different issue of spatial convergence in $\mathbf{r}$-space of various dynamical states that encode environmental features $\{s_{ab}\}$.

5.6.2. New characterizations are naturally physical rather than philosophical

Characterizations of unity problems offered by the present approach advance the field significantly, because problems framed in terms of spaces and coordinates are expressed in a basically *physical* language: $\mathbf{r}$-coordinates and the space that they label are amongst the most fundamental components of physical theory; direct contact with physics is immediately achieved by describing unity issues in terms of $\boldsymbol{\rho}$-coordinates, conscious-experiential spaces, and the relationship between these constructs and orthodox physical space. For example, the previous discussion (Rosseinsky, 2014b) of representational unity [5.6.1] pointed out that various sub-domain- and property-based solutions can be reframed in terms of certain (physical) generation-of-$<s_{ab}>$-*at*-$\boldsymbol{\rho}_j$ properties possessed by certain (physical) *B*-states.

5.6.3. Unity of feature-type spaces

Another (single-subject) unity-of-consciousness problem, somewhat related to representational unity [5.6.1], will be termed here the "unity of feature-type spaces". Whereas representational unity concerns coordinate-values in a single $\boldsymbol{\rho}$-space, the unity of feature-type spaces asks why a multiplicity of feature types are experienced in a *single* $\boldsymbol{\rho}$-space rather than in a *multiplicity* of $\boldsymbol{\rho}$-spaces, one for each feature type (**Figure 5C,D**). This problem is broadly analogous to what has been termed the problem of "co-consciousness" (Bayne, 2001), and the present approach offers a simple resolution in terms of an empirical observation: there is only one, topologically-connected, $\boldsymbol{\rho}$-space associated with a given brain, and all components-of-experience are generated within this space. Although "why" questions can be asked with respect to this observation, these are somewhat analogous to questions of why there is "only one $\mathbf{r}$-space" (so to speak): in both cases, topologically-disconnected multiplicities of sub-spaces are conceivable, but have no empirical support.

5.6.4. Collections of $<s_{ab}>$ generated by *n*-th and *m*-th brains are disjoint

The problem of representational unity [5.6.1] concerns the generation of identical **ρ**-coordinates for multiple feature types. The unity of feature-type spaces (or problem of co-consciousness) [5.6.3] concerns the existence of a single **ρ**-space for multiple feature types. Another unity-of-consciousness issue concerns why “I” experience components-of-consciousness generated by “my” brain but not by “your” brain, and vice versa (**Figure 5E,F**). This question has been termed the problem of “subject unity” (Bayne, 2009), although in the present approach it is more naturally characterized via the *disjoint* nature of $<s_{ab}>$-collections generated by the *n*-th and *m*-th brains, *i.e.* as a property of separation between otherwise unified collections, rather than directly as a further unity problem. Put differently, the problem of subject unity is the same as the problem of privacy of conscious experience: why is it that individual brains generate contributions-to-consciousness in private, dedicated ${}_{n}\boldsymbol{\rho}$-spaces rather than in one common **ρ**-space?

The disjoint or private nature of contributions-to-consciousness was a critical aspect of the original illustrative problem discussed in the Introduction. Developments since then have shown that there are two basic mechanisms for generating subject-disjoint as opposed to completely-aggregated experience, in a physically orthodox setting. First, in sub-domain approaches, an uber-map recapitulating subjective geometry naturally groups together components-of-consciousness generated by a single brain [4.2]. In this case, the origin of disjoint conscious-experiential *spaces* is rather clear: they arise from the disjoint interior spaces of various brains. Second, in property approaches, each brain must generate its own set of encoding *B*-values that in turn label locations in various disjoint property-spaces [4.3]. However, discussions of the property approach thus far have emphasized *coordinates* over *spaces*: prior to the generation of subject-specific coordinates, the subject-disjoint property-spaces must be generated (in a manner that creates a dedicated link between subject-specific spatial locations and subject-specific *B*-encodings). Whether this kind of generation can be termed “physically orthodox” remains at best unclear, at this point.

5.6.5. Limitations in the application of present formalism to unity problems

Because the present approach began with limitation to exteroceptive consciousness and the exclusion of other aspects of consciousness, certain other unity problems discussed in the literature that are directly related either to the “conscious self” [“unity of subjectivity” (Bayne, 2009)] or to brain mechanisms for *report* of conscious experience [“unity of access” (Bayne, 2009)] are not amenable to treatment by present formalism. Nonetheless, the multi-subject setting discussed in this paper offers an entry point to a rigorous treatment of subjectivity, in the following way. Famously, William James observed that “the passing thought … is itself the thinker” (James, 1950), thus offering one resolution to potentially dualist conceptualizations of self. Analogously, in the present formalism, it might be said that “the *n*-th subject is the dynamically-changing $<s_{ab}>$ contents of ${}_{n}\boldsymbol{\rho}$-space” (although this statement naturally requires an extension of $<s_{ab}>$ definitions to include non-exteroceptive aspects of conscious experience, such as conscious interoception and conscious cognition). Notably, this approach extends the Jamesian conception to identify the phenomenon of consciousness in the first instance with *a collection of disjoint conscious-experiential spaces and their contents* (**Figure 6**).

5.7. Implications for neurobiologically-based theories-of-consciousness

The objective of the present series is to establish objective tests in conventional neuroscience for the class of physically-orthodox theories-of-consciousness, rather than to evaluate any specific theory-of-consciousness. Nevertheless, brief discussion of a single exemplar from the literature of neurobiological theories-of-consciousness may help to explain present results and their implications. For this purpose, consider the version of the prominent "dynamic core" theory discussed in (Tononi and Edelman, 1998), which, in essence, proposes that consciousness is a property of certain dynamical states of a set of neurons distinguished from others in the brain by their common mutual information. As a first point, it is important to understand that $D_{abj}[\{B(\ldots)\}]$ notation can accommodate the dynamic core hypothesis, by writing *D* as an appropriate information-theoretic function. However, it is then directly apparent that the dynamic core theory must be *non-local* (because it requires simultaneous computation across a large set of brain locations), and so *cannot be physically orthodox*. Of course, this does *not* invalidate the dynamic core proposal as a possible explanation of consciousness, but it does (importantly) clarify its scientific status: the dynamic core theory can only be true if otherwise-universal and fundamentally-significant physical principles do not apply to consciousness. Put differently, the dynamic core proposal is a "new physics" theory-of-consciousness.

Other analyses from the present series all support this characterization. For example, consider the claim that the dynamic core explains unity and privacy by virtue of the delineation of dynamics within a single brain into core and non-core categories. The present paper establishes that unified privacy is a matter of the disjoint $\{<s_{ab}>\}$ collections generated by the *n*-th and *m*-th brains [5.6.4], and demonstrates associated requirements for either subject-varying codes (under a property approach) or an uber-map recapitulating subjective geometry (under a sub-domain approach) [4.4]. Although the dynamic core approach certainly allows for neuronal coalitions contributing to consciousness to vary by subject, it does not inherently require an orderly variation of consciousness-encoding by subject, of the kind required to generate unified privacy. Accordingly, the dynamic core theory of (Tononi and Edelman, 1998) must generate unified privacy by an uber-map recapitulating subjective geometry, which also contradicts the structure of the dynamic core proposal (unless it is supplemented by a *B*-theory of the recapitulating uber-map kind). At best, the dynamic core hypothesis as stated can only create empirically-accurate unified privacy (**Figure 1**) by a dual-metric approach, again rendering the proposal physically non-orthodox.

According to Tononi and Edelman, it is a *category* error to require that certain neurons have particular consciousness-related properties *e.g.* by virtue of anatomical location. The overall approach in the present series avoids this error by explicitly constructing *C*-classifiers on *A*-states (Rosseinsky, 2014a) to describe behavioral encoding that can certainly depend on anatomical location, and *D*-classifiers on *B*-states that attribute contribution-to-consciousness properties to physical states (not to anatomical locations). The present approach points out that, if the goal is a physically-orthodox theory, it is a *theoretical* error to attribute a component-of-consciousness to spatially-distributed or temporally-extended dynamical states. Although avoiding the category error, the dynamic core proposal makes the theoretical error [at least, if the goal is indeed physical orthodoxy (Edelman, 1989)]. Because observations made here concerning the dynamic core theory apply equally to other theories in the literature, the present approach points out new theoretical challenges that have not yet been met by *any* of the leading neurobiologically-based theories-of-consciousness (Baars, 2001; Tononi, 2004; Seth et al., 2006; Klimesch, 2013; Dehaene, 2014).

5.8. Novelty of approach

Results in the present series (Rosseinsky, 2014a, 2014b) summarized in [5.1-5.7] constitute a set of strong, specific, and novel predictions concerning the encoding basis of any physically-orthodox theory-of-consciousness. Predictions here make a contribution to the neurobiology-of-consciousness literature that is different in kind from prior work, because predictions concern low-level details of codes rather than cortical areas, brain structures, or aggregate properties of dynamical states. Good scientific practice then requires responses to two questions with respect to these new results. First, why have they not been identified previously? This leads to a discussion of novel aspects in the present approach that critically support the elaboration of results, which is the subject matter of the present subsection [5.8.1-5.8.7]. Second, what assumptions do present results depend upon, and what alternative assumptions can be made? These questions are addressed in [5.9].

5.8.1. Limitation to exteroceptive consciousness

The present series is founded on the explicit limitation to exteroceptive consciousness, thus focusing developments on a relatively well-defined and easy-to-analyse subset of phenomena, and excluding difficult and controversial issues associated with a “conscious self”.

5.8.2. Agnosticism with respect to theory-of-consciousness

The present series explicitly avoids proposing an explicit theory-of-consciousness, either physical or biological. (The only assumption made is that information encoded in brain dynamics plays an essential role in generating the contents of consciousness, thus excluding psychophysical parallelism). Avoiding a proposal for the fundamental physical basis of consciousness means that present developments are unaffected by problems concerning the absence of experimental technology for the detection of consciousness (an issue discussed further in [5.9.2]). Avoiding specific proposals for the biological basis of consciousness means that, in the first instance, the object of consideration is the class of all possible theories-of-consciousness (apart from psychophysical parallelism).

5.8.3. Division of class of all theories into physically orthodox and unorthodox

After selecting the class of all possible (non-parallel) theories-of-consciousness [5.8.2], papers in the present series make a key division into physically-orthodox and physically-unorthodox sub-classes, and then seek to identify experimental tests that can distinguish between the two sub-classes. This novel approach, first including all theories and then dividing them in a very particular way into two sub-classes for theoretical and empirical comparison, is central to results here.

5.8.4. Identification and comprehensive treatment of conscious-experiential space

Another key factor in present developments is the identification of conscious-experiential space as a critical theoretical construct. Explicit discussion of this construct is necessary in order to pursue implications from locality and spatial homogeneity of causality. However, considerable uncertainty exists concerning the basic nature of conscious-experiential space. This uncertainty is handled in the present approach by pursuing analyses with respect to *all* conceptions that are logically possible within a physically-orthodox setting. Focus on the physically-orthodox class of theories [5.8.3] is essential to render the treatment of conscious-experiential space tractable, because otherwise the discussion would be ill-posed.

5.8.5. Precise definition of physical orthodoxy

Although the mainstream approach to consciousness presumes a physically-orthodox explanation [*e.g.* (Baars, 1988; Edelman, 1989; Crick and Koch, 1990; Tononi and Edelman, 1998; Churchland, 2005; Baars and Edelman, 2012)], precise definitions of physical orthodoxy and its associated requirements have been absent from the literature, to date. The novel remedy of this absence (*e.g.* [5.4]) is a significant contributor to present advances.

5.8.6. Clear and distinct treatment of both causality and correlation

Formalism introduced in (Rosseinsky, 2014a) clearly accounts for both correlates and causes of consciousness (in *A* and *B* notation respectively). In fact, present results can in large part be attributed to the novelty of a comprehensive formal approach [5.8.7] that is capable first of making causation/correlation distinctions, and then of following them through to their logical conclusions. The absence to date of an explicit treatment of causes has obscured encoding restrictions deriving from features of physical orthodoxy that must apply to *causal* encodings but need not apply to *correlated* dynamics. Moreover, the present approach offers a new indirect method for *identifying* causally-implicated structures [5.3]: those brain structures generating codes meeting the narrow requirements of Table 4 are prime candidates for the loci of physical causes-of-consciousness.

5.8.7. Role of formalism

The present series depends heavily on a new formal symbolism for theories-of-consciousness. This symbolism offers precision and universality that is absent from the usual verbal discussions. Both these attributes are vital to the successful pursuit of central arguments, which simultaneously require nuanced distinctions of complex concepts (supported by precision) together with treatment of large classes of theories (supported by universality). Papers in the present series use formalism as a central language to express observations from the diverse domains of philosophy, physics and neuroscience, thus demonstrating the promise of a common formal symbolism as a powerful communication tool for future interdisciplinary investigations.

Precision and comprehensive reach inherent in the formal approach lead to new insights and understandings that together define both tactical and strategic approaches to furthering the scientific understanding of consciousness. These tactical and strategic components are made plain in [5.10]: *tactical* contributions are relatively local inferences, such as "strict physical orthodoxy requires subject-varying codes", whereas *strategic* contributions identify the approaches that should be prioritised if certain empirical conditions are not observed. The contribution of formalism to the derivation of tactical conclusions is exactly the precision and universality just discussed. Perhaps less obvious is the strategic capacity afforded by a formalism with well-specified foundational assumptions: if conclusions (in the form of empirical predictions) are falsified, then clear formal linkages from conclusions to assumptions facilitate identification of minimal alterations in assumptions that can lead to different conclusions (notably, to conclusions that might concur with empirical observations that falsified previous conclusions).

5.9. Robustness

As noted at the beginning of [5.7], the novelty and power of present results evoke a consideration of assumptions made, and alternative assumptions that might lead to different conclusions. Discussion of these points in the present subsection is divided into relatively explicit axiomatic assumptions [5.9.1], and relatively implicit assumptions inherent in the overall methodology [5.9.2].

5.9.1. Axiomatic robustness

As in related works (Rosseinsky, 2014a, 2014b), the present paper makes few explicit assumptions whose alteration might lead to different conclusions. In addition to the assumptions in earlier papers that consciousness exists and can be defined (and that psychophysical parallelism is to be excluded), the present paper assumes that many conscious experiences (or "many minds") exist. Another way of stating this assumption is that each typical human brain produces conscious experience in broadly the same manner. Certainly, making the alternative solipsistic assumption that the only conscious experience in existence is that of the present reader would lead to different conclusions, because then only restrictions from the single-subject setting (Table 3) would apply. However, solipsism itself is not consistent with the physically-orthodox approach (in which broadly homogeneous physical brains must have broadly homogeneous properties), so the stronger restrictions of Table 4 must logically apply to physically-orthodox theories-of-consciousness.

5.9.2. Methodological robustness

Attempts to offer a scientific explanation of consciousness have been criticized on a number of methodological grounds, including: unclear definitions of the phenomena under consideration (Dennett, 1991); the possibility that phenomena are not homogeneous across subjects (Shoemaker, 1982); the absence of objective experimental tests for consciousness in non-human systems (Hawking, 2000); use of detailed subjective reports that might be unreliable (Frith *et al*., 1999) or incomplete (Lycan, 1996); a current limitation to experimental methods (Chalmers, 2000) that can only identify correlates rather than causes of consciousness; and, potential confounds in experimental approaches to establishing neural-correlates-of-consciousness (Lau, 2008; Hohwy, 2009; Aru *et al*., 2012)

Although in principle any of these criticisms might be levelled at developments in this series, two powerful and distinct kinds of rebuttal can be made. First, present developments concern empirical signatures that must be associated *with a physically-orthodox theory-of-consciousness*: the very construction of physically-orthodox theories presumes that basic issues *e.g.* concerning definitions of phenomena and the validity of scientific method can be overcome. This kind of rebuttal is similar to the argument used in [5.9.1], because it relies on properties of the theories-of-consciousness that are the present focus of study.

Second, the present methodology has been constructed in a particular way that insulates it *directly* from many of the standard criticisms, thus creating a kind of rebuttal that is *independent* of the theories under consideration. For example, experimental problems associated with subjective report (Frith *et al*., 1999) and potential confounds (Lau, 2008; Hohwy, 2009; Aru *et al*., 2012) do not apply, because empirical tests suggested here are novel, objective, and straightforward. Similarly, the fact that homogeneity of component phenomena cannot be objectively established across subjects [the "inverted qualia" problem (Shoemaker, 1982)] does not invalidate reasoning in the present series, because multi-subject analyses depend only on the subject-by-subject applicability of restrictions from physical orthodoxy, and not on any cross-subject relationships between phenomena labelled by $<s_{ab}>$ symbols. Perhaps more subtly, the absence of objective experimental tests for consciousness in non-human systems [a problem that might invalidate the scientific validity of reasoning involving the phenomenon of consciousness (Hawking, 2000)] is not a problem here, because the methodological role of present experimental proposals is to delineate between two theory groups that both hypothesize the generation of consciousness from (possibly different) physical dynamics, rather than to verify or falsify a hypothesis that certain dynamics generate consciousness and other dynamics do not.

The only standard criticism for which a direct rebuttal of the second kind does not exist concerns the related issues of the existence and definition of consciousness, to which, briefly, three kinds of response can be made. First, an explicit operational definition of phenomena under study (Rosseinsky, 2014a) might be accepted as resolving both definitional and existence-related issues. Notably, in foundationally significant single-subject settings, readers can verify *for themselves* what is meant by "consciousness". Second, criticisms that the operational definition is itself "not scientific" (by virtue of reference to subjective experience) fails to recognise that the philosophical foundations of science must themselves make such reference (to delineate the subjective "evidence of the senses", so to speak, from an allegedly extant objective reality that is the topic of science itself). Third, a rigid exclusion of consciousness from scientific study on grounds of existence or definition leads to a science definitively describing a Universe *without* consciousness, raising the question of whether scientific theories can then be complete or even accurate explanations of all natural phenomena. Although this last point does not establish the scientific validity of studying consciousness, it certainly raises the stakes in a way that challenges casual dismissals.

Because the present approach is directly insulated against almost all standard criticisms, arguably it constitutes a new and useful *methodological* innovation in the scientific study of consciousness, in addition to offering new experimental tests [5.1-5.3], theoretical insights [5.6], and theoretical challenges [5.4,5.5,5.7].

5.10. Outlook

Results here show that, if the strictest form of physical orthodoxy [5.4] applies to the generation of consciousness, then encoding-for-consciousness must vary by subject. Directly, this result follows from a limitation to property-based theories-of-consciousness (under ontological independence of theory [5.4]) and from the requirement that encoding-dynamics must vary by subject for these kinds of theories [4.3]. This result constitutes an empirical test for strictly orthodox theories-of-consciousness. (Restrictions for physically-orthodox explanations derived in earlier papers must also apply. Notably, the occurrence of any physical state generating a contribution-to-consciousness must be limited by brain biophysics to only those circumstances in which neural systems encode the corresponding environmental feature. This restriction does not apply *e.g.* to the occurrence of e.m.-field states used in neural computations that generate behavior.) If variation-by-subject is empirically excluded, some relaxation must be in made strict physical orthodoxy, in order to explain consciousness.

A defining aspect of *strict* physical orthodoxy is that theories must be ontology independent. Relaxing this aspect to allow theories that are premised on a necessarily realist ontology admits only one additional coding possibility for consciousness in the multi-subject setting, in which an uber-map (a single topographic map encoding all features simultaneously) is arranged in orthodox three-dimensional space in a manner recapitulating subjective geometry. Under this relaxation, the many-subject setting provides a strong rejection of a “physically orthodox” characterization of previously-described dual-metric approaches, which provide marginal contributions to the physically-orthodox theory set in single-subject settings. If neither subject-varying codes nor uber-maps recapitulating subjective geometry are empirically observed, yet further relaxations must be made to physical orthodoxy.

One possible direction for further relaxation is to propose that spatial homogeneity of causality does not in fact govern the generation of consciousness, thus setting consciousness apart from other natural phenomena. This path would certainly broaden the set of feasible codes, but still leaves the generation of consciousness mysterious. For example, in the absence of spatial homogeneity of causality, one can propose without direct contradiction that homogeneous dynamics in distinct brains can cause contributions-to-consciousness in distinct, unified, collections-of-experience. But there can then be no explanation in terms of physical order for the origin of both difference-in-effect (dynamics in different brains have effects in different collections) and unity-of-experience (dynamics in a single brain share the common property of contributing features to the same collection). Although it is possible that the phenomenon of consciousness does in fact depend on these kinds of mysteries, the scientific path to establishing such a view is first to exhaust all alternative explanations that propose instead a rational and orderly basis for phenomena. Thus, if neither subject-varying codes nor uber-maps recapitulating subjective geometry are empirically observed, the appropriate next step is not to relax spatial homogeneity, but to investigate instead non-orthodox theories that maintain spatial homogeneity while relaxing either spatial-dimensionality or particle-spectrum assumptions implicit in the usual physical setting. As noted in previous works, these latter kinds of theories are capable of explaining relationships between brain dynamics and conscious experience without imposing *any* restriction on brain-level encodings, while maintaining over-arching physical principles of locality and spatial homogeneity. However, the promotion of such approaches to the status of serious scientific theories would require their verification by experimental signatures (*i.e.* measurements predicted by higher-dimensional or extended-spectrum approaches but not *e.g.* by alternative theories premised on relaxation for consciousness of otherwise-universal spatial homogeneity).

In summary, following theoretical observations here and in previous works, the outlook for the neurobiological basis of consciousness is as follows. If subject-varying codes are observed (that also obey previously-established restrictions from locality and single-subject spatial homogeneity of causality), then physically-orthodox property-based theories-of-consciousness are strongly supported. If the existence of subject-varying codes is ruled out empirically, but uber-maps recapitulating subjective geometry are discovered (again, obeying previous restrictions), then primacy must be given to theories which jointly hypothesize the existence of a realist ontology and the employment of a sub-domain of (realist) orthodox space as the container for the components of conscious experience. Finally, if neither subject-varying codes nor uber-maps of the correct kind are observed, the next step in establishing a scientifically-based relationship between brain dynamics and conscious experience should be to investigate the feasibility of experimental signals from rigorously-defined higher-dimensional or extended particle-spectrum approaches, that can preserve spatial homogeneity of causality and simultaneously cohere with then-established empirical facts concerning brain-dynamical encoding-of-consciousness.

Table 1. Summary of formal symbols. This Table contains brief definitions for the symbol set introduced in (Rosseinsky, 2014a) and employed throughout the present series.

SYMBOL	DEFINITION	SECTION(1)
s_{ab}	*b*-th instance of *a*-th feature-type (stimulus) in the external environment.	3.1
${}_{X}\mathbf{r}_j$	*j*-th location in the external environment for sampling of sensory information	3.3
$<s_{ab}>$	Contribution to conscious experience generated by s_{ab}	5.3
$\boldsymbol{\rho}_j$	*j*-th location in conscious experience; information sampled at ${}_{X}\mathbf{r}_j$ leads to a contribution-to-consciousness at $\boldsymbol{\rho}_j$	5.4
A	Measure of brain activity for behavioral encoding	4.2
B	Measure of brain activity for encoding and final brain-dynamical cause of consciousness	5.5.3
${}_{A}\mathbf{r}_i$	*i*-th brain location relevant to *A* measurement	4.4
${}_{B}\mathbf{r}_i$	*i*-th brain location relevant to *B* measurement	5.5.3
C_{abj}	Classifier function on *A*-states for behavioral encoding	4.5
D_{abj}	Classifier function on *B*-states for encoding and final brain-dynamical cause of consciousness	5.6
$\exists_P$	Denotes existence of a physical object, property, or phenomenon	4.5
$\nexists_P$	Denotes absence of a physical object, property, or phenomenon	4.5
{ …}	Set or collection of …	3.2

(1) References given are to section numbers in (Rosseinsky, 2014a).

Table 2. Illustrative codes for combinations of {Red (R)**, Green** (G)**} and {Vertical** (V)**, Horizontal** (H)**} edge features at two locations {ρ_1, ρ_2}.** This Table gives numerical illustrations for each of six basic codes (types I to VI) that together span the entire spectrum of potentially viable code-structures for the encoding of consciousness. Entries in the Table body are illustrative numeric codes for feature-combinations shown in the leftmost column, where *e.g.* "R,V; G,H" (row 6) means "Red Vertical edge at ρ_1; Green Horizontal edge at ρ_2".

FEATURES	**ENCODING**					
	Non-topographic			**Topographic**		
ρ_1; ρ_2	*Type I*	*Type II*	*Type III*	*Type IV*	*Type V*	*Type VI*
R,V; R,V	1	1,5	1,3,5,7	1,3,1,3	1,1	1,1,1,1
R,V; R,H	2	1,6	1,3,5,8	1,3,1,4	1,2	1,1,1,2
R,H; R,V	3	2,5	1,4,5,7	1,4,1,3	2,1	1,2,1,1
R,H; R,H	4	2,6	1,4,5,8	1,4,1,4	2,2	1,2,1,2
R,V; G,V	5	1,7	1,3,6,7	1,3,2,3	1,3	1,1,2,1
R,V; G,H	6	1,8	1,3,6,8	1,3,2,4	1,4	1,1,2,2
R,H; G,V	7	2,7	1,4,6,7	1,4,2,3	2,3	1,2,2,1
R,H; G,H	8	2,8	1,4,6,8	1,4,2,4	2,4	1,2,2,2
G,V; R,V	9	3,5	2,3,5,7	2,3,1,3	3,1	2,1,1,1
G,V; R,H	10	3,6	2,3,5,8	2,3,1,4	3,2	2,1,1,2
G,H; R,V	11	4,5	2,4,5,7	2,4,1,3	4,1	2,2,1,1
G,H; R,H	12	4,6	2,4,5,8	2,4,1,4	4,2	2,2,1,2
G,V; G,V	13	3,7	2,3,6,7	2,3,2,3	3,3	2,1,2,1
G,V; G,H	14	3,8	2,3,6,8	2,3,2,4	3,4	2,1,2,2
G,H; G,V	15	4,7	2,4,6,7	2,4,2,3	4,3	2,2,2,1
G,H; G,H	16	4,8	2,4,6,8	2,4,2,4	4,4	2,2,2,2

Table 3. Viable codes for consciousness in a single-subject setting under physical orthodoxy. This Table summarizes *all* viable code structures for a physically-orthodox theory-of-consciousness in a single-subject setting [3.7]. Type I codes are not topographic, so cannot be valid in rows 1-5 (which identify codes for various topographically-based theories-of-consciousness). Type III, IV, and VI codes that use two numbers for each environmental location (in the numerical example of Table 2) are not consistent with uber-map schemes (rows 1, 2 and 4) that definitively use a single number per location [3.1.2]. Similarly, Type II and V codes that employ only a single number for each location are not appropriate [3.1.3] for multi-map codes (rows 3 and 5). Note that Type VI codes that are viable in principle for behavioral coding are not valid for any approach to consciousness [3.4.2].

THEORY-OF-CONSCIOUSNESS			**VALIDITY OF CODES**					
Space	**Maps**	**Sub-type**	**Type I**	**Type II**	**Type III**	**Type IV**	**Type V**	**Type VI**
Sub-domain	Uber	Recap.[1]	N/A	Yes[2]	N/A	N/A	Yes	N/A
	Uber	Non-recap.[1]	N/A	Yes[2]	N/A	N/A	Yes	N/A
	Many	Non-recap.[1]	N/A	N/A	Yes	Yes	N/A	No
Property	Uber	ρ-coding	N/A	Yes	N/A	N/A	No	N/A
	Many	ρ-coding	N/A	N/A	Yes	No	N/A	No
	Non-topo.[3]	Single-scalar[4]	Yes	N/A	N/A	N/A	N/A	N/A

(1) "Recap." and "non-recap." refer respectively to maps that do and do not recapitulate subjective geometry (Rosseinsky, 2014a).
(2) Type II codes in a sub-domain approach encode more information than is necessary, because values encode ρ-locations that are also determined by **r**-locations.
(3) "Non-topo.": abbreviation for "non-topographic".
(4) "Grandmother cell" single-neuron codes are not viable physically-orthodox encodings of consciousness [3.5.3]. More complex non-topographic codes are also excluded, in the first instance because they are non-local (Rosseinsky, 2014a).

Table 4. Viable codes for consciousness in a multi-subject setting under physical orthodoxy. Fewer codes are valid in the multi-subject setting than in the single-subject setting, as a result of the reduction in valid *sub-domain* codes: whereas Table 3 displays nine valid basic codes (*i.e.* "Yes" entries), this Table shows only five. Thus, the multi-subject setting *restricts* viable physically-orthodox neural codes for consciousness. Because codes in *property-based* theories-of-consciousness must vary by subject (as denoted by "+" nomenclature), the multi-subject setting also *defines* viable physically-orthodox neural codes for consciousness.

THEORY-OF-CONSCIOUSNESS			**VALIDITY OF CODES**					
Space	**Maps**	**Sub-type**	**Type I**	**Type II**	**Type III**	**Type IV**	**Type V**	**Type VI**
Sub-domain	Uber	Recap.[1]	N/A	Yes[2]	N/A	N/A	Yes	N/A
Property	Uber	ρ-coding	N/A	Yes+[3]	N/A	N/A	No	N/A
	Many	ρ-coding	N/A	N/A	Yes+[3]	No	N/A	No
	Non-topo.[4]	Single-scalar	Yes+[3]	N/A	N/A	N/A	N/A	N/A

(1) "Recap." means "with recapitulation of subjective geometry".
(2) Type II codes in a sub-domain approach encode more information than is necessary, because values encode ρ-locations that are also determined by **r**-locations.
(3) "+" notation denotes that code must vary by subject [4.3].
(4) "Non-topo.": abbreviation for "non-topographic".

Table 5. Possible empirical observations for physically-orthodox codes in a multi-subject setting, and implications for the correct theory-of-consciousness. The empirically observed code (first column) provides information about the correct theory-of-consciousness (second and third columns) [5.3]. Note that codes must either recapitulate subjective geometry, or vary by subject (as indicated by the "+" nomenclature). These very strong limitations are required if consciousness is a physically-orthodox phenomenon.

Observed code	Central inference	Secondary possibility
Type II, with recap.[(1)]	Sub-domain	
Type V, with recap.[(1)]	Sub-domain	
Type I+[(2)]	Property	
Type II+[(2)]	Property	Sub-domain (if recap.[(1)(3)])
Type III+[(2)]	Property	

(1) "Recap." means "with recapitulation of subjective geometry".
(2) "+" notation denotes a code that varies by subject.
(3) A subject-varying code ("+") that also employs a sub-domain uber-map with recapitulation is redundant in the sense that subject-variation is only required for property-based codes (because multi-subject physical orthodoxy is already satisfied by a sub-domain uber-map with recapitulation: Table 4, first row).

6. References

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.

Anderson, P. W. (1972). More is different. *Science* 177, 393–396.

Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 737–746.

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Baars, B. J. (2001). *In the Theater of Consciousness: The Workspace of the Mind*. New York: Oxford University Press.

Baars, B. J., and Edelman, D. B. (2012). Consciousness, biology and quantum hypotheses. *Phys. Life Rev.* 9, 285–294.

Barlow, H. B. (1961). "Possible principles underlying the transformation of sensory messages," in *Sensory communication*, 217–234.

Bayne, T. (2001). Co-consciousness. *J. Conscious. Stud.* 8, 79–92.

Bayne, T. (2012). *The Unity of Consciousness*. New York: Oxford University Press.

Bayne, T. (2009). Unity of consciousness. *Scholarpedia* 4. doi:10.4249/scholarpedia.7414.

Bowers, J. S. (2009). On the biological plausibility of grandmother cells: implications for neural network theories in psychology and neuroscience. *Psychol. Rev.* 116, 220.

Chalmers, D. J. (2000). "What is a Neural Correlate of Consciousness?," in *Neural correlates of consciousness: Empirical and conceptual questions*, ed. T. Metzinger (Cambridge, MA: MIT press), 17–39.

Churchland, P. S. (2005). A neurophilosophical slant on consciousness research. *Prog. Brain Res.* 149, 285.

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

Cleeremans, A. (2003). *The Unity of consciousness: binding, integration, and dissociation*. New York: Oxford University Press.

Crick, F., and Koch, C. (1990). Towards a neurobiological theory of consciousness. *Semin. Neurosci.* 2, 263–275.

Dehaene, S. (2014). *Consciousness and the brain: deciphering how the brain codes our thoughts*. New York: Viking.

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

Edelman, G. M. (1989). *The remembered present: A biological theory of consciousness*. New York: Basic Books.

Feldman, J. (2013). The neural binding problem(s). *Cogn. Neurodyn.* 7, 1–11.

Frith, C., Perry, R., and Lumer, E. (1999). The neural correlates of conscious experience: an experimental framework. *Trends Cogn. Sci.* 3, 105–114.

Gross, C. G. (2002). Genealogy of the "grandmother cell." *The Neuroscientist* 8, 512–518.

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.

Hohwy, J. (2009). The neural correlates of consciousness: new experimental approaches needed? *Conscious. Cogn.* 18, 428–438.

Huxley, T. H. (1893). "On the Hypothesis that Animals are Automata, and its History," in *Collected Essays: Volume I, Method and Results* (London: Macmillan), 195–250.

James, W. (1950). *The principles of psychology*. Vol. 1. New York: Dover.

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

Klimesch, W. (2013). An algorithm for the EEG frequency architecture of consciousness and brain body coupling. *Front. Hum. Neurosci.* 7, 766. doi:10.3389/fnhum.2013.00766.

Lau, H. (2008). "Are we studying consciousness yet?," in *Frontiers of consciousness: Chichele lectures*, 245.

Lycan, W. G. (1996). *Consciousness and experience*. Cambridge, MA: MIT press.

Makeig, S., Debener, S., Onton, J., and Delorme, A. (2004). Mining event-related brain dynamics. *Trends Cogn. Sci.* 8, 204–210.

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

Panzeri, S., Brunel, N., Logothetis, N. K., and Kayser, C. (2010). Sensory neural codes using multiplexed temporal scales. *Trends Neurosci.* 33, 111.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454, 995–999.

Plaut, D. C., and McClelland, J. L. (2010). Locating object knowledge in the brain: Comment on

Bowers's (2009) attempt to revive the grandmother cell hypothesis. *Psychol. Rev.* 117, 284–290.

Quiroga, R. Q., and Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nat. Rev. Neurosci.* 10, 173–185.

Reinagel, P., and Reid, R. C. (2000). Temporal Coding of Visual Information in the Thalamus. *J. Neurosci.* 20, 5392–5400.

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

Rosseinsky, N. M. (2014a). Physically orthodox theories-of-consciousness must predict biophysical mechanisms for the integration of spatially-distributed and temporally-extended neural codes. Manuscript in submission.

Rosseinsky, N. M. (2014b). Topographic codes for consciousness are constrained to a small and distinctive set by a physically-orthodox setting. Manuscript in Submission.

Russell, B. (1927). *The Analysis of Matter*. London: Kegan Paul.

Seth, A. K., Izhikevich, E., Reeke, G. N., and Edelman, G. M. (2006). Theories and measures of consciousness: an extended framework. *Proc. Natl. Acad. Sci.* 103, 10799–10804.

Shoemaker, S. (1982). The inverted spectrum. *J. Philos.* 79, 357–381.

Tegmark, M. (2000). Importance of quantum decoherence in brain processes. *Phys. Rev. E* 61, 4194.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci.* 5, 42.

Tononi, G., and Edelman, G. M. (1998). Consciousness and complexity. *Science* 282, 1846–1851.

Tononi, G., and Koch, C. (2008). The neural correlates of consciousness. *Ann. N. Y. Acad. Sci.* 1124, 239–261.

Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J., and McNaughton, B. L. (1996). Population dynamics and theta rhythm phase precession of hippocampal place cell firing: a spiking neuron model. *Hippocampus* 6, 271–280.

Velmans, M. (2012). The evolution of consciousness. *Contemp. Soc. Sci.* 7, 117–138.

7. Figures

Figure 1. Schematic illustration of spatial-homogeneity-of-causality issues in a multi-subject setting under the property approach. (A) Simple encoding scheme used as basis of illustrations. Locations in conscious-experiential space (below the dashed line) are marked by red crosses and labelled by real numbers 0.1, 0.2, 0.3, and so forth. *B*-states at various brain locations (marked by black crosses) generate white or black pixels (if the integer part of *B* is respectively 1 or 0) at conscious-experiential locations corresponding to the decimal part of *B*-values. **(B)** Spatial homogeneity of causality is not respected if *B*-codes are identical across brains, and conscious-experiential space is a property of brain dynamics. (Sub-division of conscious-experiential space by the solid line schematically depicts two different collections of components-of-consciousness). B_1 = 0.1 must cause a black pixel at the point labelled 0.1 in the *left*-hand sub-division of conscious experiential space, whereas B_2 = 0.1 must cause a black pixel at the point labelled 0.1 in the *right*-hand sub-division. But under spatial homogeneity the effects of *B* = 0.1 cannot depend on location. **(C)** Empirical conscious experience is not reproduced if spatial homogeneity of causality is imposed on a *B*-code identical across brains. The state *B* = 0.1 in the two brains creates a single component-of-consciousness, while the states *B* = 1.2 and *B* = 0.2 attempt to create conflicting components (*i.e.* both black and white pixels; conflict depicted by grey pixel). **(D)** Spatial homogeneity of causality and empirical accuracy is achieved when locations in distinct conscious experiences are given distinct coordinate labels. Encoding-of-consciousness must also differ across brains.

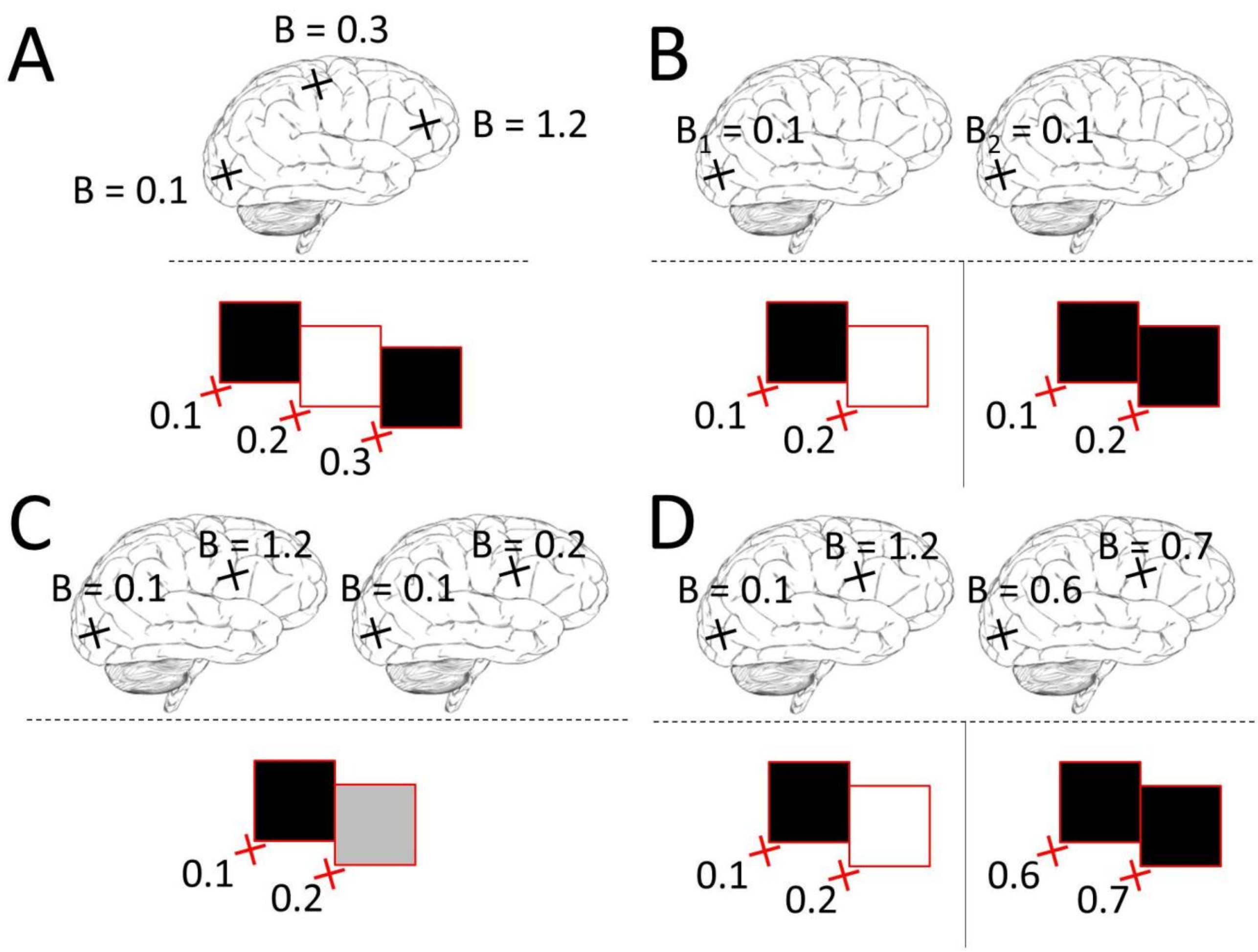

Figure 2. Schematic illustration of natural spatial-homogeneity-of-causality in a multi-subject setting under the sub-domain approach. (A) Orthodox view of two brains containing various *B*-states (occurring at locations marked by black crosses). *B* = 0 encodes a black pixel and *B* = 1 encodes a white pixel. **(B)** View showing use of sub-domains of orthodox physical space as conscious-experiential spaces [the “sub-domain” approach to conscious-experiential space (Rosseinsky, 2014a)]. Components-of-consciousness co-locate with *B*-states and empirically accurate groupings of components are generated, with a *B*-code that is identical across brains.

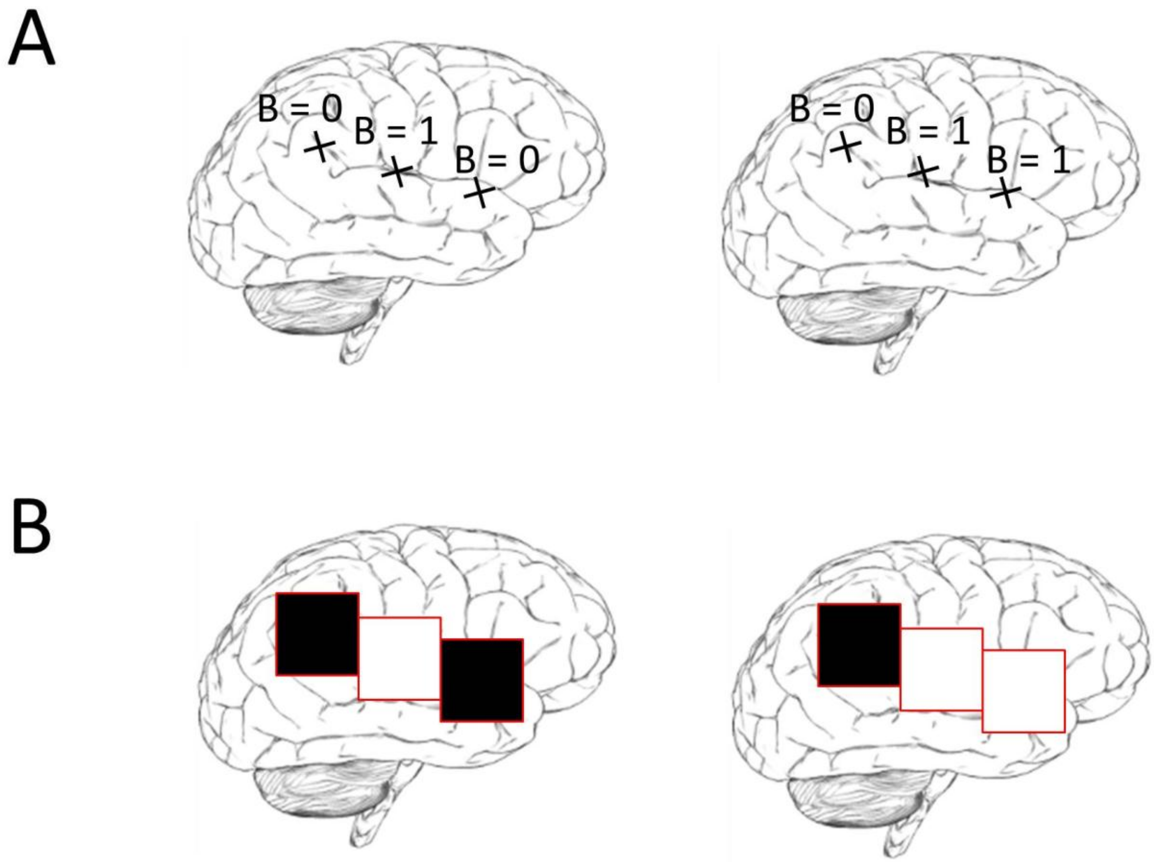

Figure 3. Schematic illustration of the formal statement of a complete theory of exteroceptive consciousness. Formal symbols introduced in (Rosseinsky, 2014a) can completely state an arbitrary theory-of-consciousness in just three basic equations. [Sensory and behavioral encoding-dynamics are schematically represented by triangular spikes; the symbol *A* (Rosseinsky, 2014a) is used to denote a physical measure of these dynamical states. Consciousness-encoding *B*-dynamics are schematically represented by black crosses. Arrows schematically depict causal relationships.] Red arrow, red equation: the existence of a stimulus instance s_{ab} at an external location ${}_{X}\mathbf{r}_j$ means that *A*-dynamics at brain locations $\{{}_{A}\mathbf{r}_i\}$ must satisfy $C_{abj} = 1$. Green arrow, green equation: physical coupling of *A*- and *B*-dynamics means that $C_{abj} = 1$ satisfaction generates *B*-dynamics that satisfy $D_{abj} = 1$. Black arrow, black equation: by definition of the causal properties of *B*-states, $D_{abj} = 1$ satisfaction causes the generation of the component-of-consciousness $<s_{ab}>$ at the conscious-experiential location $\boldsymbol{\rho}_j$.

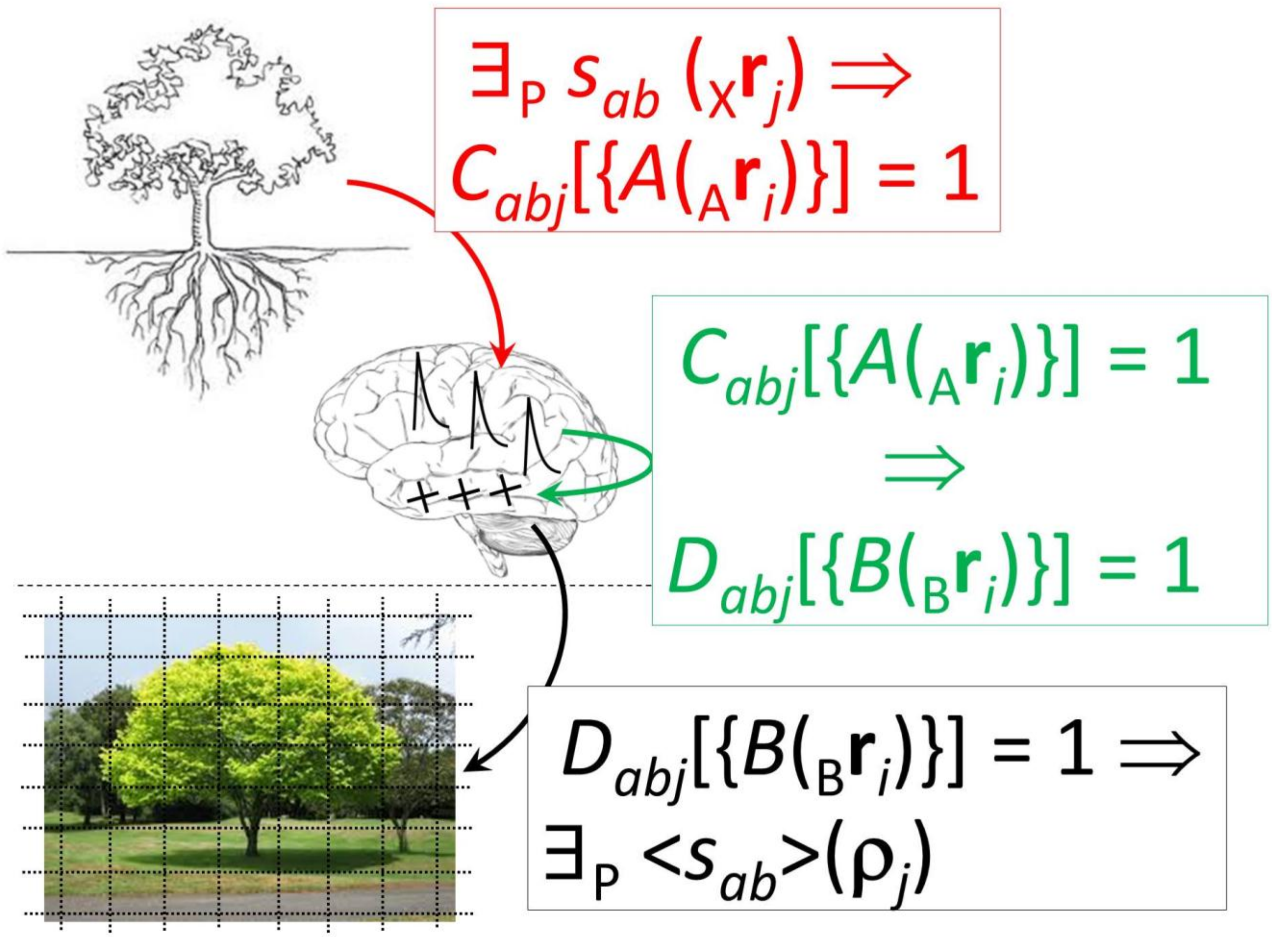

Figure 4. Schematic illustration of why dual-metric approaches are manifestly not physically orthodox in a multi-subject setting. (A) Each brain has a physically-special region (the interior of the blue ellipse) in which dual-metric relationships apply. Initial locations of brain centers are labeled ${}_1\mathbf{r}_0$ and ${}_2\mathbf{r}_0$. **(B)** When one brain moves relative to the other (so that its center is now at ${}_2\mathbf{r}_0'$), the special region must co-move with that brain. Either this movement is uncaused (violating physical orthodoxy), or some physically-unorthodox property of the physical brain is responsible for generating the dual-metric relationship, thus ensuring co-movement.

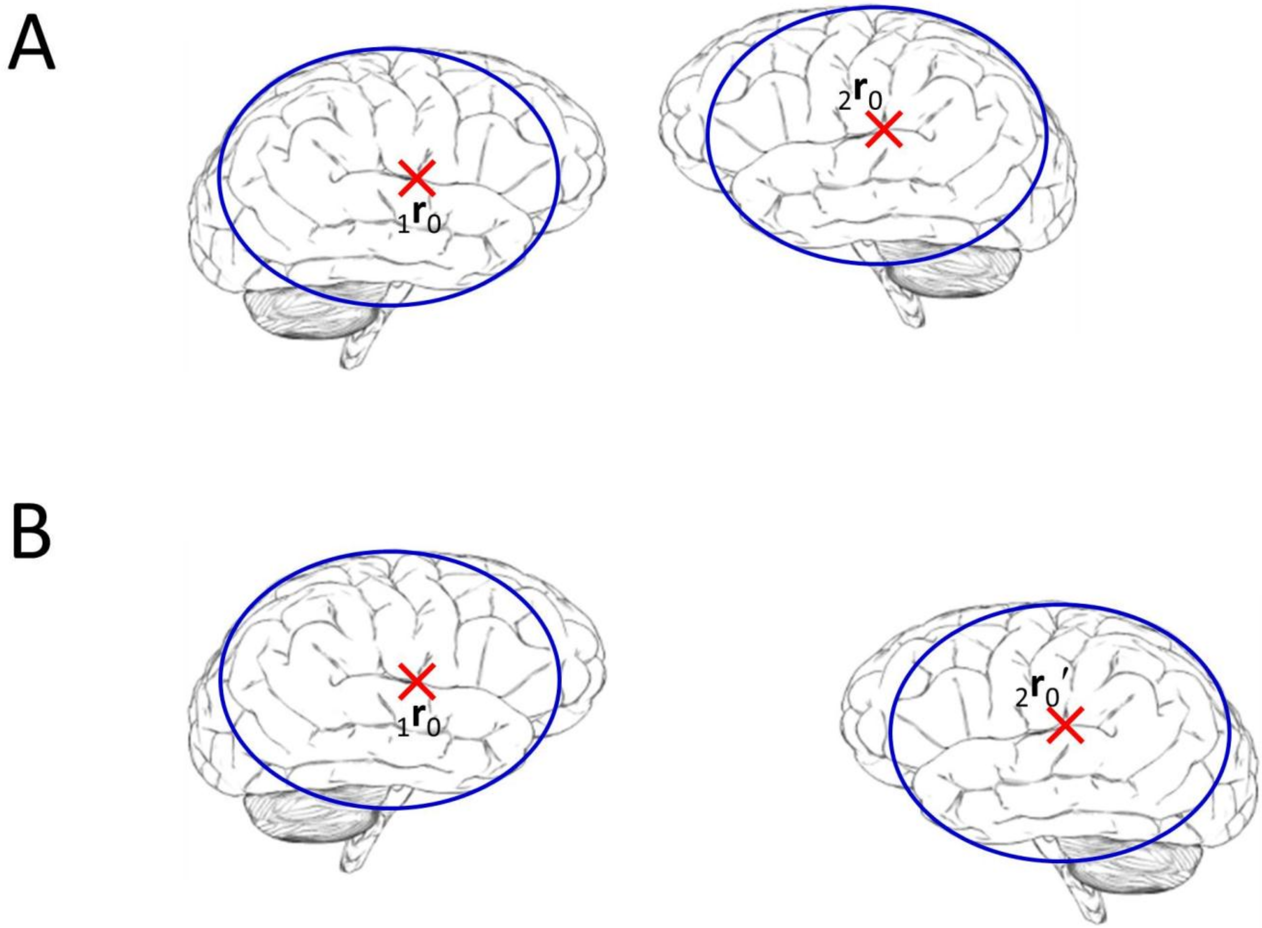

Figure 5. Schematic illustrations of various aspects of "the unity of consciousness". Black arrows schematically represent causal links between brain states and components-of-consciousness. **(A)**,**(B)** *The conscious-experiential binding problem ("representational unity")*. Motion and color at a single environmental location are encoded by *B*-dynamics at two brain locations (shown by green and purple crosses, respectively). Without specific constraints, *B*-dynamics might be expected to generate experience-of-color and experience-of-motion at two different conscious-experiential locations (panel A). In actuality (panel B), experience-of-color and experience-of-motion are "bound together" at a single location. **(C)**,**(D)** *The unity of feature-type spaces ("co-consciousness").* In principle, color experiences could occur in one conscious-experiential sub-space (left sub-division of conscious-experiential space, panel C) and motion experiences could occur in another (right sub-division, panel C; locations in second sub-space distinguished by blue crosses). In actuality (panel D), there is only one conscious-experiential space in which all feature-types co-occur. **(E)**,**(F)** *Collections of* $<s_{ab}>$ *generated by n-th and m-th brains are disjoint*. In principle, various collections of components-of-consciousness might contain contributions from more than one brain. For example, in panel E, conscious experience is divided into two collections (schematically separated by the solid vertical line), each containing contributions from both brains. In actuality, the behavior shown in panel F occurs, in which each collection contains contributions from only one brain.

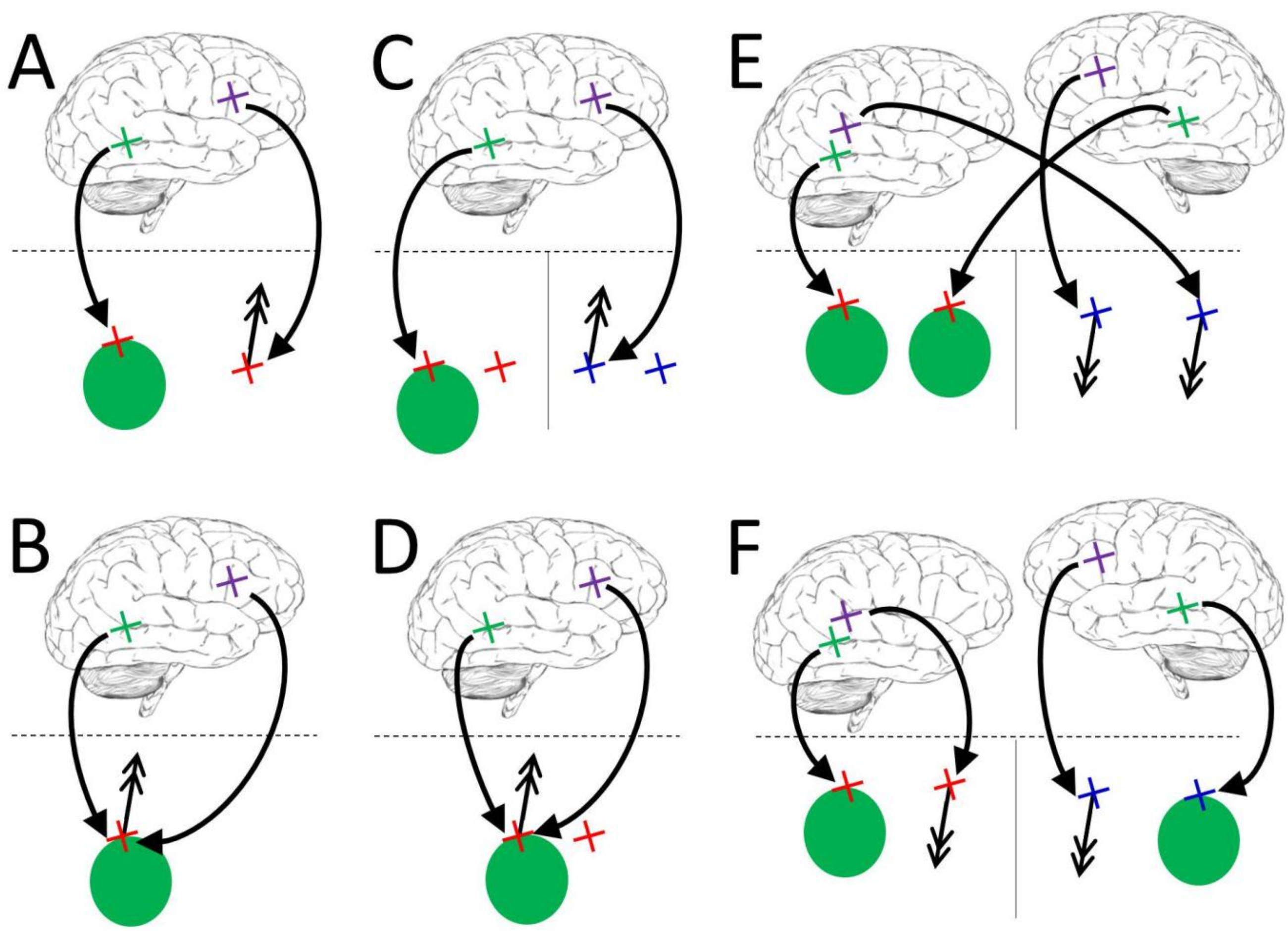

Figure 6. The phenomenon of consciousness in the multi-subject setting is a collection of disjoint conscious-experiential spaces and their contents: a schematic illustration. Three human subjects schematically depicted as geometrically-shaped bodies containing brains (upper half, above horizontal dashed line) have three separate conscious experiences of the physical environment (lower half, below horizontal dashed line). The subject on the left sees the inside of a house. The subject on the right sees a tree. The subject in the center sees the same tree, partially occluded by the brown hair and red shirt of the subject on the right. Black arrows indicate causal relationships between brains and conscious-experiential spaces and contents. Dashed black lines framing schematic depictions of conscious visual experiences (lower half, below dashed line) reflect the disjoint nature of conscious-experiential spaces. For a complete treatment, content of spaces must be extended to include non-exteroceptive components-of-consciousness.

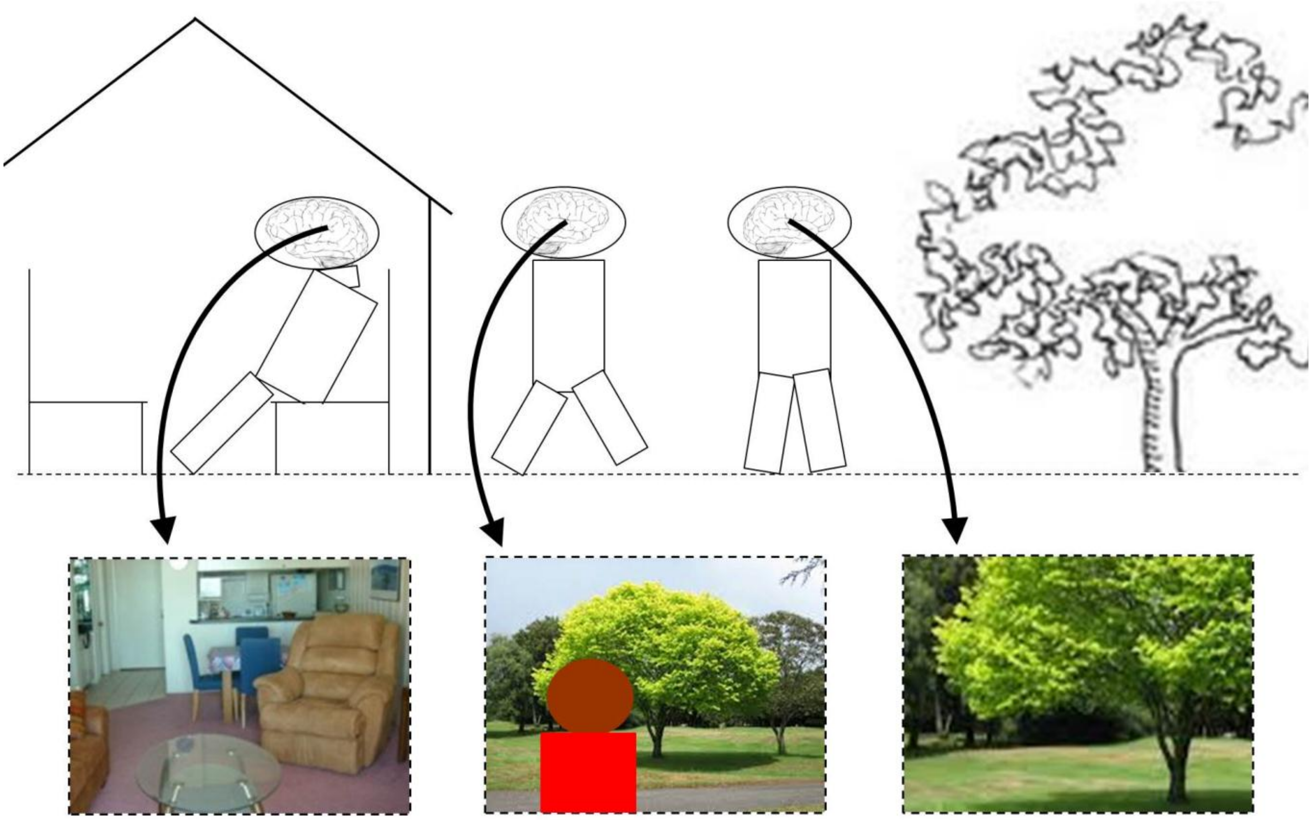

Figure 1.JPEG

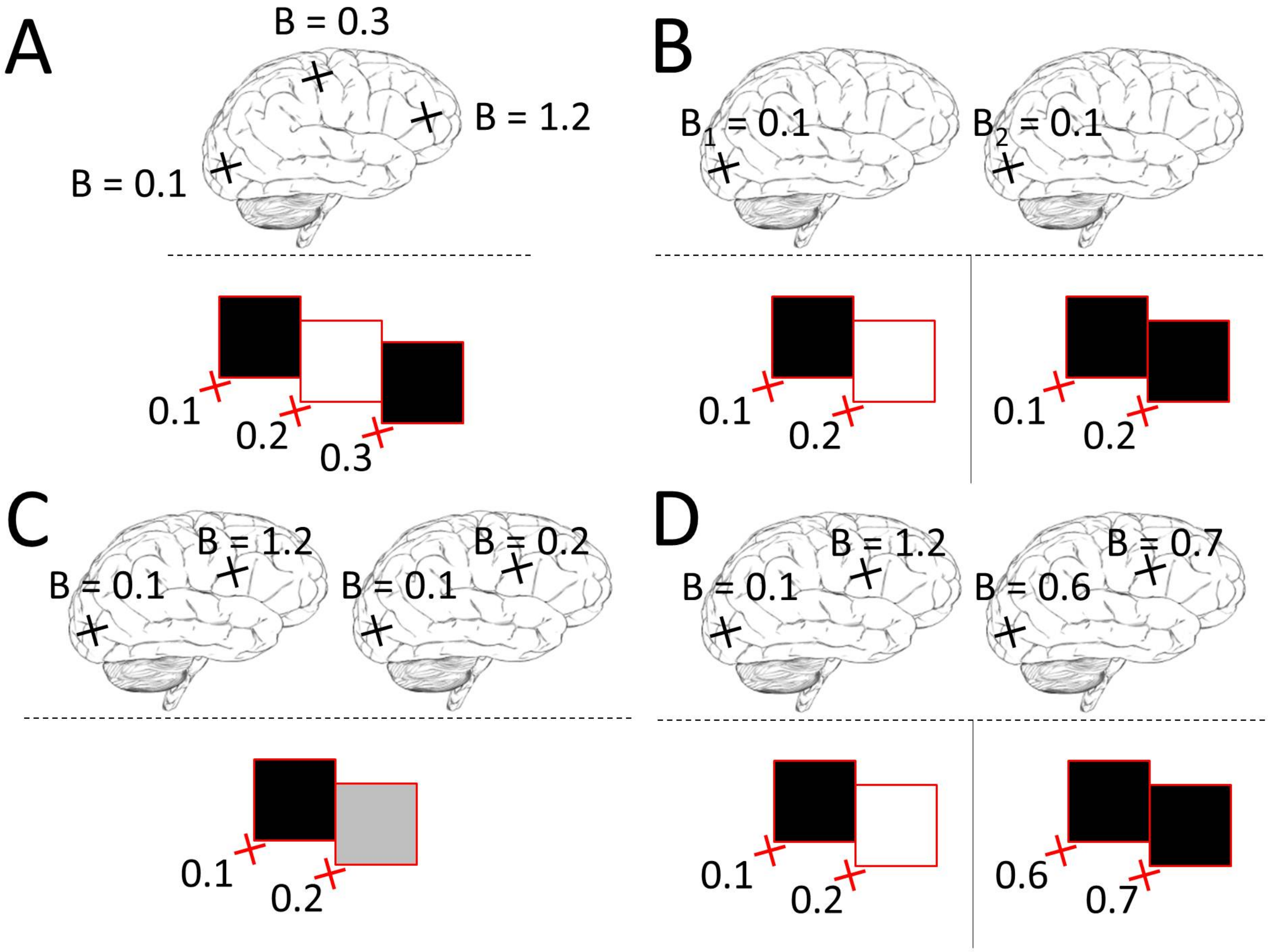

Figure 2.JPEG

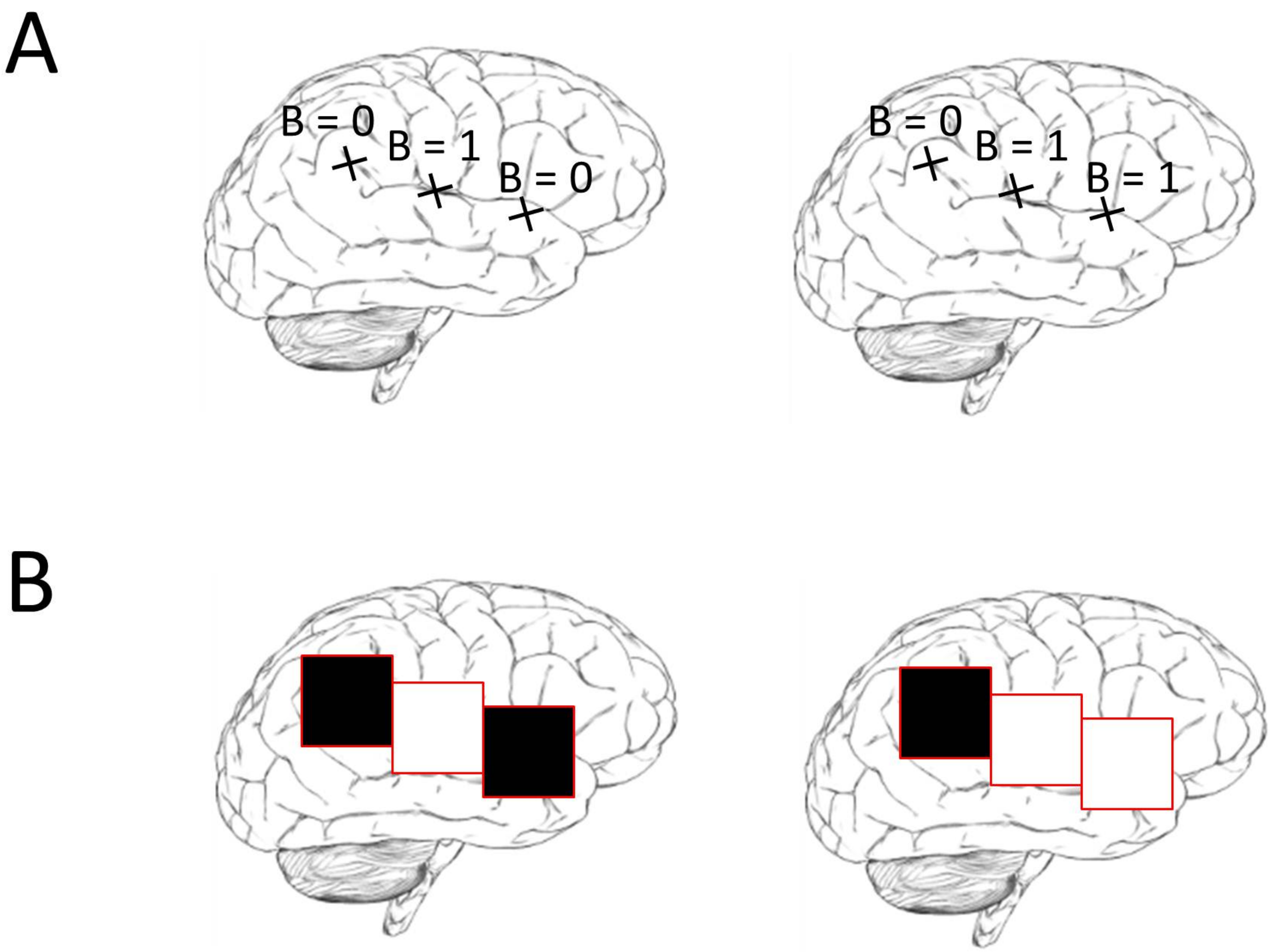

Figure 3.JPEG

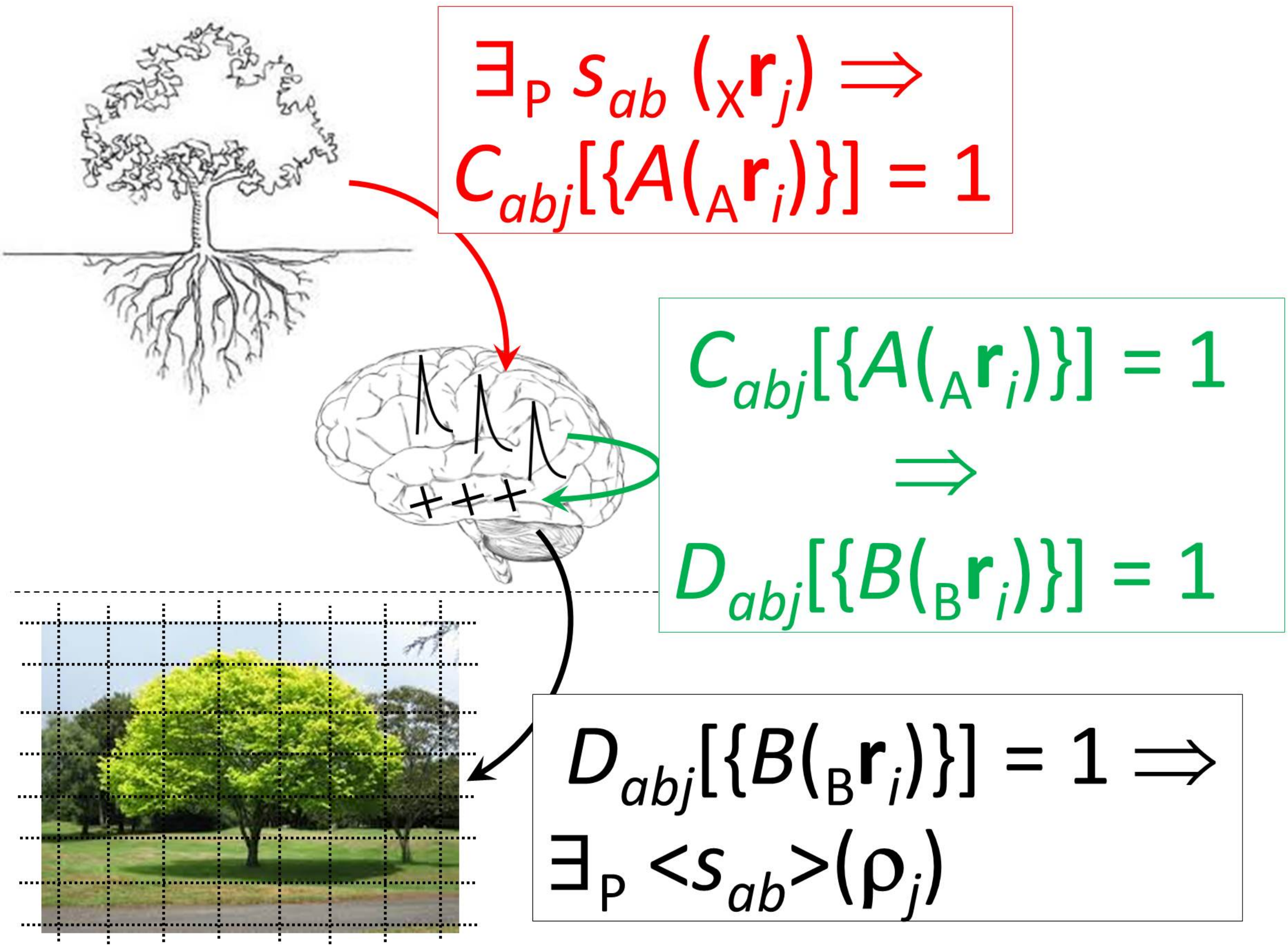

Figure 4.JPEG

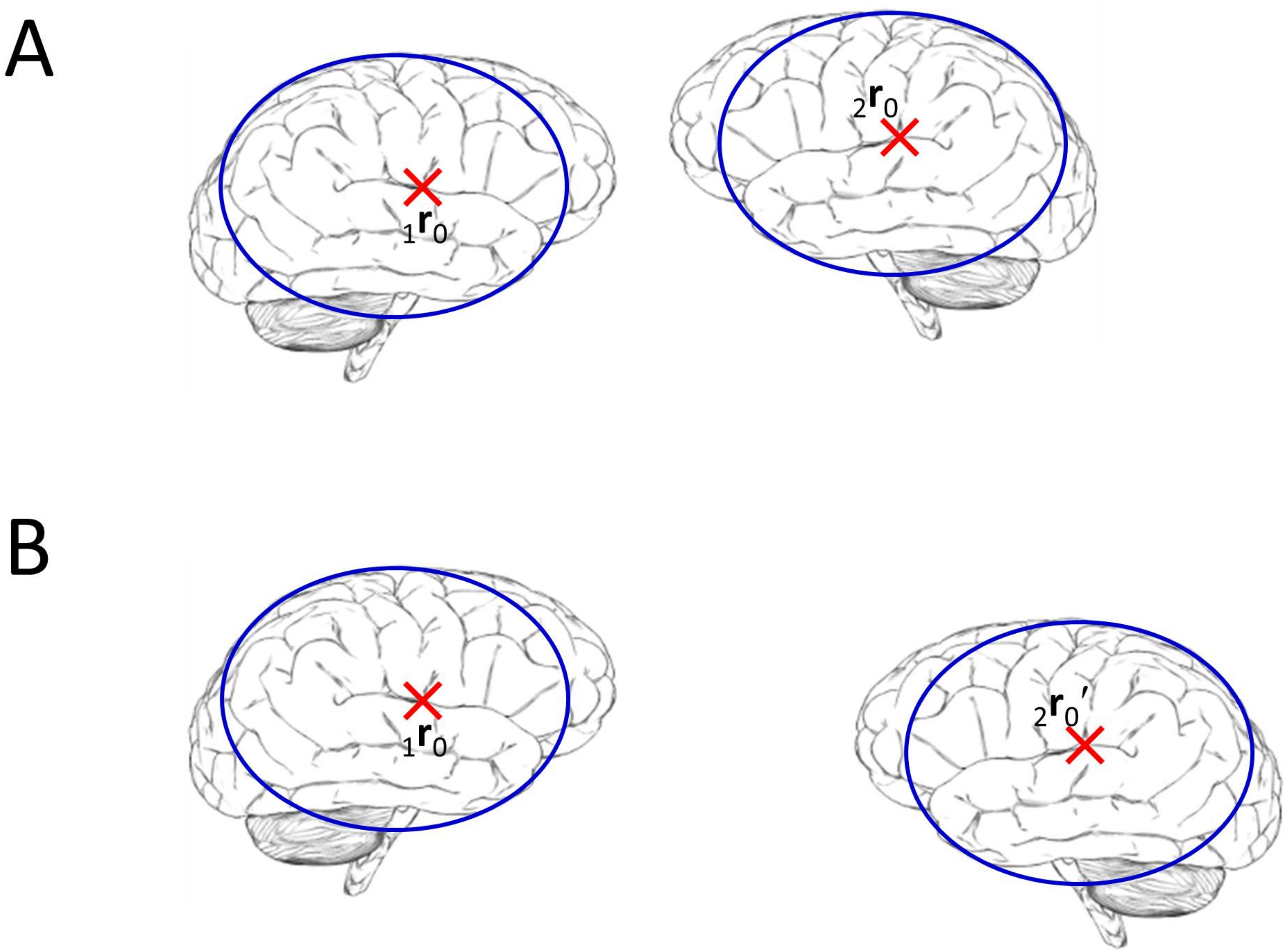

Figure 5.JPEG

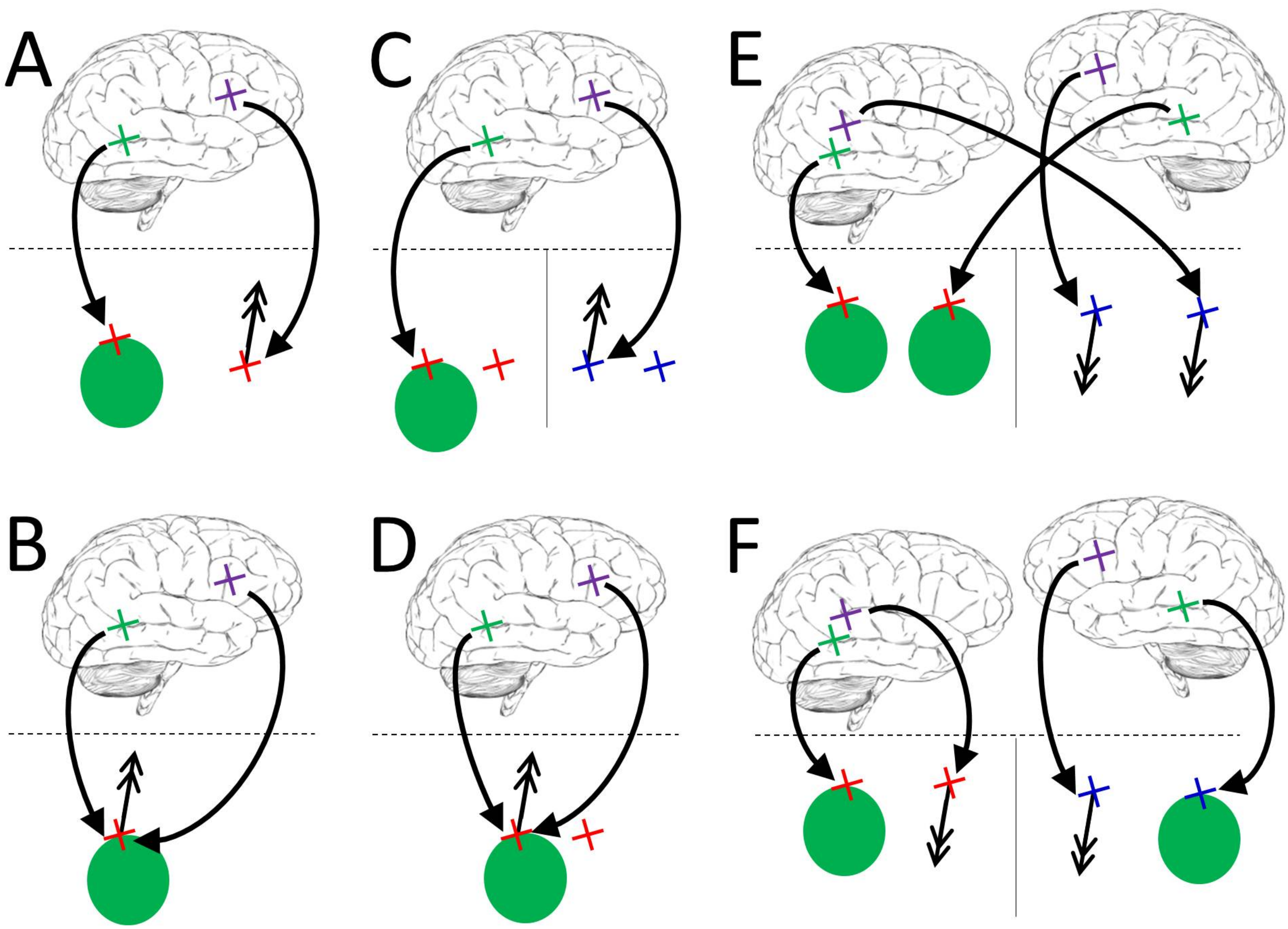

Figure 6.JPEG