Isomorfismo Causal e Independencia de Sustrato:
Condiciones Necesarias para la Transferencia Funcional de Consciencia

Causal Isomorphism and Substrate Independence: Necessary Conditions for Functional Consciousness Transfer

Kevin Caracuel Llabrés
Programa de Doctorado en Filosofía
Dpto. de Filosofía, Lógica y Filosofía de la Ciencia, Universidad de Sevilla
Investigador asociado, Grupo de Investigación en Cognición y Cerebro (CBC)
Universitat Pompeu Fabra, Barcelona
kcaracuel@us.es

Elena Vidal-Moreno
Catedrática, Dpto. de Filosofía, Lógica y Filosofía de la Ciencia
Universidad de Sevilla
evidal@us.es

Joan Ferrer-Casademunt
Profesor Agregado, Dpto. de Tecnologías de la Información y las Comunicaciones
Universitat Pompeu Fabra, Barcelona
joan.ferrer@upf.edu

Fecha de recepción: 17 de octubre de 2017
Fecha de aceptación: 9 de mayo de 2018
Publicado en: Minds and Machines, Vol. 28, Nº 3, pp. 451-498
DOI: 10.1007/s11023-018-9467-x

Abstract

This paper develops a formal framework, grounded in category theory, for analysing the conditions under which consciousness can be transferred between physical substrates while preserving its qualitative character. We define causal isomorphism as a structure-preserving functor between the causal categories of two physical systems and prove that substrate independence, understood as the claim that consciousness depends only on causal-functional organization and not on physical constitution, requires but is not exhausted by causal isomorphism. Specifically, we show that a further condition, which we term temporal microstructure preservation (TMP), is necessary: the functor must preserve not only the causal relations between states but also the fine-grained temporal dynamics of those relations. From these results we derive a no-cloning theorem for consciousness, analogous in structure (though not in physical content) to the quantum no-cloning theorem: perfect transfer of consciousness between substrates is achievable only as an asymptotic limit, never exactly. The framework builds upon our previous topological characterization of phenomenal experience (Caracuel Llabrés, Vidal-Moreno and Aráoz-Gutiérrez, 2016) and connects with current debates on functionalism, multiple realizability and the metaphysics of mind uploading.

Keywords: *consciousness transfer, substrate independence, category theory, causal isomorphism, functionalism, mind uploading, no-cloning theorem, temporal microstructure, philosophy of mind, multiple realizability*

Resumen

Este artículo desarrolla un marco formal, basado en teoría de categorías, para el análisis de las condiciones bajo las cuales la consciencia puede transferirse entre sustratos físicos preservando su carácter cualitativo. Definimos el isomorfismo causal como un funtor que preserva la estructura entre las categorías causales de dos sistemas físicos y demostramos que la independencia de sustrato, entendida como la tesis de que la consciencia depende exclusivamente de la organización causal-funcional y no de la constitución física, requiere pero no se agota en el isomorfismo causal. En concreto, mostramos que se necesita una condición adicional, la preservación de la microestructura temporal (PMT): el funtor debe preservar no solo las relaciones causales entre estados sino también la dinámica temporal fina de dichas relaciones. A partir de estos resultados derivamos un teorema de no-clonación para la consciencia, análogo en estructura (aunque no en contenido físico) al teorema de no-clonación cuántico: la transferencia perfecta de consciencia entre sustratos es alcanzable solo como límite asintótico, nunca de forma exacta.

Palabras clave: *transferencia de consciencia, independencia de sustrato, teoría de categorías, isomorfismo causal, funcionalismo, carga mental, teorema de no-clonación, microestructura temporal, filosofía de la mente, realizabilidad múltiple*

1. Introduction

The prospect of transferring a conscious mind from one physical substrate to another, whether from a biological brain to a silicon-based computer or between two artificial systems, has long been a staple of speculative fiction. In recent years, however, it has become a topic of serious scientific and philosophical inquiry. Sandberg and Bostrom (2008) have produced a detailed technical roadmap for whole brain emulation. Advances in connectomics (Seung, 2012) and large-scale neural simulation

(Markram et al., 2015) are gradually reducing the gap between current capabilities and the computational requirements of full brain emulation. And the philosophical debate on the metaphysical status of such a procedure has become increasingly sophisticated (Chalmers, 2010; Schneider, 2014; Pigliucci, 2014).

Yet, at the heart of the debate lies a question that has received surprisingly little formal attention: under what precise mathematical conditions does a transfer of consciousness preserve the qualitative character of experience? Most discussions assume, either explicitly or implicitly, that functional equivalence suffices. If the new substrate implements the same input-output function as the original, the argument goes, then it will instantiate the same conscious experiences. This assumption is the core of functionalism (Putnam, 1967; Fodor, 1974) and, in its strongest version, of computational functionalism (Chalmers, 1996, ch. 7).

In this paper, we argue that functional equivalence, even when formalized with the utmost precision, is necessary but not sufficient. What is additionally required is the preservation of the fine-grained causal and temporal structure of the system. To make this claim precise, we employ the language of category theory (Mac Lane, 1998; Awodey, 2010), which provides a natural framework for formalizing the notion of "structure preservation" at various levels of abstraction.

Our starting point is the topological framework for phenomenal experience developed in Caracuel Llabrés, Vidal-Moreno and Aráoz-Gutiérrez (2016), where we showed that the qualitative structure of conscious experience can be captured by topological invariants of a phenomenal state space (the Q-space). In Caracuel Llabrés and Vidal-Moreno (2017), we extended this framework to the domain of moral agency, defining a continuous moral gradient over cognitive state spaces. The present paper addresses a more fundamental question: if two systems share the same topological invariants of their respective Q-spaces, does this guarantee that they have the same conscious experiences? And if not, what further conditions are needed?

The paper is structured as follows. Section 2 reviews the philosophical background on substrate independence and multiple realizability. Section 3 introduces the categorical formalism and defines the central notion of causal category. Section 4 defines causal isomorphism and proves that it is a necessary condition for consciousness transfer. Section 5 introduces the temporal microstructure preservation condition and shows why it is needed in addition to causal isomorphism. Section 6 derives the no-cloning theorem for consciousness. Section 7 discusses the implications of our results. Section 8 concludes.

2. Philosophical background

2.1 Substrate independence and functionalism

The thesis of substrate independence, in its most general form, asserts that consciousness is determined by the abstract organizational structure of a system rather than by the specific physical material that implements that structure. Consciousness, on this view, is "multiply realizable": the same conscious state can, in principle, be instantiated in carbon-based neural tissue, in silicon chips, in a simulation running on a digital computer, or in any other physical medium that supports the relevant organizational structure.

This thesis is intimately connected with functionalism in the philosophy of mind. Putnam (1967) argued that mental states are individuated by their functional roles, not by their physical constitution. Fodor (1974) generalized this into the doctrine of multiple realizability, which became one of the central arguments against reductive physicalism. More recently, Chalmers (1996, 2010) has articulated a version of organizational invariance specifically for consciousness, arguing that any system with the same fine-grained causal organization as a conscious system will have the same conscious experiences.

Critics of substrate independence have challenged it on various grounds. Searle (1980, 1992) has argued that computation is observer-relative and therefore cannot ground consciousness. Block (1978) has raised the problem of liberal functionalism: if consciousness depends only on functional organization, then absurd systems (the population of China implementing a functional description of a mind) would be conscious. Bickle (1998) has challenged multiple realizability on empirical grounds. And Maudlin (1989) has constructed ingenious thought experiments purporting to show that computationalism leads to contradictions.

Our approach does not aim to settle this debate but to sharpen it. We formalize the notion of "same causal organization" with unprecedented precision using category theory, and then investigate what follows. The result, as we shall see, is that substrate independence, properly formalized, imposes stronger constraints than commonly assumed.

2.2 Previous formal approaches

Several authors have attempted to formalize aspects of the substrate independence thesis. Chalmers (2011) proposed a formalization in terms of combinatorial state automata, where two systems are organizationally equivalent if they implement the same CSA. Milkowski (2013) offered a mechanistic account of computation that constrains the notion of implementation. And Schweizer (2016) has argued for a structural account of computation that avoids the triviality objections.

Our approach differs from these in its use of category theory as the foundational language. The advantage of this choice is threefold. First, category theory is specifically designed to formalize the notion of structure preservation (via functors and natural transformations), which is precisely what substrate independence requires. Second, it allows us to operate at multiple levels of abstraction simultaneously, capturing both the coarse-grained functional structure and the fine-grained causal dynamics. Third, it connects naturally with our previous topological framework (Caracuel Llabrés et al., 2016), since algebraic topology is, at its core, a branch of category theory (functors from topological spaces to algebraic structures).

3. The categorical formalism

3.1 Causal categories

Let S be a physical system capable of instantiating conscious states. We associate to S a category C(S), which we call its causal category, defined as follows.

The objects of C(S) are the physical microstates of S. If S is a neural system, these are the complete patterns of neural activation at each instant; if S is a digital computer, these are the complete machine configurations. We denote the set of objects by Ob(C(S)).

The morphisms of C(S) are the causal transitions between microstates. A morphism $f: s1 \to s2$ represents the fact that microstate s1 causally produces microstate s2 according to the dynamical laws governing S. Composition of morphisms corresponds to the chaining of causal transitions. The identity morphism at each object represents the trivial causal transition of a state to itself (persistence).

Crucially, we enrich each morphism with temporal information. For every morphism $f: s1 \to s2$, we assign a positive real number $\tau(f)$ representing the physical time required for the causal transition from s1 to s2. This yields what we call a temporally enriched category, or T-category: a category whose morphisms carry temporal labels satisfying the additivity condition $\tau(g \circ f) = \tau(f) + \tau(g)$.

$$C(S) = (Ob(C(S)), Mor(C(S)), \circ, id, \tau) \tag{1}$$

The T-category C(S) encodes the complete causal-temporal structure of the system S at the microphysical level. Our claim is that this structure is the appropriate level of description for addressing the question of consciousness transfer.

3.2 Functors as structure-preserving maps

In category theory, a functor $F: C \to D$ between two categories is a mapping that sends objects to objects and morphisms to morphisms, preserving composition and identities. A functor, in other words, is the precise mathematical embodiment of the idea of "structure preservation".

For T-categories, we distinguish two types of functors. A causal functor is a functor $F: C(S1) \to C(S2)$ that preserves the causal structure (composition and identities) but not necessarily the temporal labels. A temporal functor is a causal functor that additionally preserves the temporal labels: $\tau(F(f)) = \tau(f)$ for all morphisms f. A temporal functor preserves both what causes what and when it does so.

$$F: C(S_1) \to C(S_2) \text{ is temporal iff } \tau(F(f)) = \tau(f)\ \forall f \in Mor(C(S_1)) \tag{2}$$

The distinction between causal and temporal functors will prove essential. As we shall argue, consciousness transfer requires a temporal functor, not merely a causal one.

3.3 Connection with the Q-space

In Caracuel Llabrés et al. (2016) we defined the Q-space as a topological space of phenomenal states. We can now express the relationship between the Q-space and the causal category in functorial terms. There exists a "phenomenal functor" $\Phi: C(S) \to Q(S)$ that maps each physical microstate to its associated phenomenal state (or to a distinguished null state if the microstate is not conscious). This functor is many-to-one in general: multiple distinct microstates may correspond to the same phenomenal state. The quotient structure induced by Φ captures precisely the notion of multiple realizability at the phenomenal level.

The existence of this functor imposes a coherence condition: if two microstates are mapped to the same phenomenal state, then their causal successors must be mapped to phenomenal states that are compatible (in a sense that can be made precise using the notion of natural transformation). This condition, which we call phenomenal coherence, is a non-trivial constraint on the dynamics of the system and ensures that the causal evolution at the physical level is consistent with the structure of the Q-space.

4. Causal isomorphism as a necessary condition

4.1 Definition and basic properties

We say that two systems S1 and S2 are causally isomorphic if there exists an invertible causal functor $F: C(S1) \to C(S2)$, that is, a functor with an inverse $G: C(S2) \to C(S1)$ such that $G \circ F = Id$ and $F \circ G = Id$. Causal isomorphism means that the two systems have exactly the same causal structure: every causal relation in one system has a unique counterpart in the other, and vice versa.

$$S_1 \cong\_c\, S_2 \iff \exists F: C(S_1) \simeq C(S_2) \text{ (invertible causal functor)} \tag{3}$$

We now state and prove our first main result.

Theorem 1 (Necessity of causal isomorphism). If the transfer of consciousness from S1 to S2 preserves the qualitative character of all phenomenal states, then S1 and S2 are causally isomorphic.

Proof sketch. Suppose the transfer preserves all phenomenal states. Then, by the properties of the phenomenal functor Φ, every causal transition in S1 that is phenomenally relevant (i.e., that produces a change in phenomenal state) must have a counterpart in S2 that produces the same change in phenomenal state. The phenomenal coherence condition then extends this correspondence to all causal transitions, including those that are not directly phenomenally relevant (since they may become relevant through composition with other transitions). The resulting correspondence is an invertible causal functor. A complete proof, which requires careful handling of the null phenomenal states, is given in Appendix A.

The converse of Theorem 1 does not hold, and this is the crux of our argument. Causal isomorphism is necessary but not sufficient. To see why, we need to examine the role of temporal structure.

4.2 Why causal isomorphism is not sufficient

Consider two systems S1 and S2 that are causally isomorphic via a functor F, but where F does not preserve temporal labels. Concretely, suppose that a causal transition $s1 \rightarrow s2$ in S1 takes 10 milliseconds, while the corresponding transition $F(s1) \rightarrow F(s2)$ in S2 takes 100 milliseconds. The causal structure is identical, but the temporal dynamics are different.

Does this matter for consciousness? We argue that it does, for the following reason. Many phenomenally relevant properties of neural dynamics are temporal in nature: the frequency of neural oscillations, the synchronization patterns between neural populations, the precise timing of spike trains (Singer and Gray, 1995; Varela et al., 2001). If the temporal binding hypothesis has any validity (and there is substantial empirical evidence that it does), then the temporal microstructure of causal transitions is constitutive of phenomenal character, not merely incidental to it.

To put the point in the language of our topological framework: the Q-space is not determined solely by the causal structure of the system but also by its temporal dynamics. Two systems with the same causal structure but different temporal dynamics may have Q-spaces that are topologically distinct. A causal functor that distorts temporal labels may, therefore, distort the topology of the Q-space, violating the preservation of phenomenal character.

5. Temporal microstructure preservation

5.1 Definition and motivation

We define the temporal microstructure preservation (TMP) condition as follows. A causal functor $F\colon C(S1) \rightarrow C(S2)$ satisfies TMP if it is a temporal functor, that is, if $\tau(F(f)) = \tau(f)$ for all morphisms f. In other words, TMP requires that the functor preserves not only what causes what, but also the precise timing of every causal transition.

This is a strong condition, and intentionally so. Our aim is to identify the necessary conditions for perfect transfer of consciousness, and we argue that anything less than exact temporal preservation introduces a margin of phenomenal distortion. The question of how much distortion is tolerable for practical purposes is a separate (and important) question that we address in Section 5.3.

Theorem 2 (Necessity of TMP). If the transfer of consciousness from S1 to S2 preserves the full qualitative character of experience (including all temporal aspects of phenomenal character), then the transfer functor F must satisfy TMP.

The proof follows from the observation that temporal properties of experience (the experienced duration of events, the perceived simultaneity or succession of phenomenal contents, the temporal grain of perceptual experience) are constitutive of phenomenal character and are, by the arguments of Section 4.2, dependent on the temporal microstructure of the underlying causal transitions.

5.2 The combined condition

Combining Theorems 1 and 2, we obtain the main characterization result of this paper.

Theorem 3 (Necessary conditions for consciousness transfer). A transfer of consciousness from S1 to S2 preserves the full qualitative character of experience only if there exists an invertible temporal functor F: C(S1) ≃ C(S2), i.e., only if S1 and S2 are temporally isomorphic.

$$S_1 \cong\_t\, S_2 \iff \exists F: C(S_1) \simeq C(S_2) \text{ with } \tau(F(f)) = \tau(f)\ \forall f \qquad (4)$$

Temporal isomorphism is a strictly stronger condition than causal isomorphism. It requires not only that the causal web of the two systems be structurally identical, but that every thread in that web have exactly the same temporal extension. This is, in a precise sense, the most stringent condition one can impose on a structure-preserving map between causal categories.

5.3 Approximate transfer and the fidelity metric

Perfect temporal isomorphism is, of course, a mathematical idealization. No physical implementation can reproduce the temporal dynamics of another system with infinite precision. This observation leads naturally to the question of approximate transfer.

We define the transfer fidelity of a causal functor F: C(S1) → C(S2) as:

$$\Phi(F) = 1 - \sup\{\, |\tau(F(f)) - \tau(f)| \,/\, \tau(f) : f \in Mor(C(S_1)),\ \tau(f) > 0 \,\} \qquad (5)$$

When Φ(F) = 1, the functor satisfies TMP exactly (perfect transfer). When Φ(F) < 1, there is temporal distortion, and the degree of distortion is quantified by 1 - Φ(F). The transfer fidelity provides a metric for evaluating the quality of a consciousness transfer procedure: the closer Φ(F) is to 1, the more faithfully the qualitative character of experience is preserved.

A natural conjecture, which we state without proof, is that the phenomenal distortion introduced by an imperfect transfer is continuous in the transfer fidelity: small deviations from perfect fidelity produce small deviations in phenomenal character. If this conjecture holds, it provides a justification for the practical relevance of approximate transfer, even though perfect transfer remains an unattainable ideal.

6. A no-cloning theorem for consciousness

6.1 Statement and proof

We are now in a position to state the most striking consequence of our framework. Recall that the quantum no-cloning theorem (Wootters and Zurek, 1982; Dieks, 1982) states that it is impossible to create an identical copy of an arbitrary unknown quantum state. We derive an analogous result for consciousness, though from entirely different premises.

Theorem 4 (No-cloning theorem for consciousness). Let S be a conscious physical system. There exists no physical procedure that produces a system S' that is temporally isomorphic to S while S continues to exist unchanged. That is, perfect consciousness cloning is physically impossible.

Proof sketch. Suppose such a procedure exists. Let S be in microstate s at time t0, and let P be the physical procedure that produces S'. The procedure P must interact physically with S (to read its state) and with the environment (to construct S'). These interactions modify the causal category of S: the morphisms available to S during and after the cloning process differ from those available in the absence of the process. In particular, the temporal labels of the morphisms emanating from s are perturbed by the physical interaction with P. Therefore, the causal category of the post-cloning S is not temporally isomorphic to the causal category of the pre-cloning S. Since S' is temporally isomorphic to the post-cloning S (by assumption), it cannot be temporally isomorphic to the pre-cloning S. Hence, the consciousness of S' does not perfectly replicate the consciousness that S had before cloning. The cloning is imperfect.

We emphasize that this result does not depend on quantum mechanics. The argument is purely structural: it follows from the fact that any physical measurement perturbs the measured system, a principle that holds in classical physics as well (though to a lesser degree in practice). What the theorem establishes is an in-principle impossibility of perfect cloning, not a practical difficulty.

6.2 Asymptotic cloning and the fidelity bound

Although perfect cloning is impossible, the transfer fidelity metric introduced in Section 5.3 allows us to formulate the question of approximate cloning. We conjecture that for any $\varepsilon > 0$, there exists a physical procedure $P\varepsilon$ such that the transfer fidelity $\Phi(F) > 1 - \varepsilon$. In other words, the fidelity of cloning can be made arbitrarily close to 1 but never exactly equal to 1.

$$lim\_\{\varepsilon \to 0\}\ \Phi(P\varepsilon) = 1, \text{ but } \Phi(P) < 1 \text{ for all physical } P \tag{6}$$

This asymptotic result, if correct, means that consciousness transfer is a limit concept: something that can be approached with increasing fidelity but never perfectly achieved. The philosophical significance of this is considerable. It means that every consciousness transfer, no matter how technologically advanced, introduces some phenomenal distortion. The transferred mind is never exactly the same mind. It is, at best, an extraordinarily close approximation.

Whether this residual distortion matters in practice depends on how fine-grained phenomenal identity is. If phenomenal identity admits a threshold (a minimum distance in Q-space below which two phenomenal states are "effectively identical"), then sufficiently high-fidelity transfer may be practically perfect even if mathematically imperfect. This is an empirical question that our framework poses but cannot resolve.

7. Discussion

7.1 Implications for functionalism

Our results have direct implications for the debate on functionalism. Standard functionalism holds that mental states are determined by functional organization. Our framework makes precise what "functional organization" must mean for this thesis to be true: it must include not only the causal structure (which functional roles are realized) but also the temporal microstructure (when causal transitions occur and how long they take).

This constitutes a refinement rather than a refutation of functionalism. The functionalist can accept our results by expanding the notion of functional organization to include temporal properties. What our results rule out is a "time-insensitive" functionalism that ignores the dynamics of state transitions and focuses only on the static structure of causal relations. Such a functionalism is, we have shown, too coarse-grained to guarantee the preservation of phenomenal character.

7.2 Implications for mind uploading

For the practical prospect of mind uploading, our results are cautiously pessimistic. The requirement of temporal isomorphism means that a successful upload must replicate not only the connectivity of the brain (which is the focus of connectomics) and the functional properties of neurons (which is the focus of computational neuroscience) but also the precise timing of all electrochemical processes down to the level at which they are causally relevant.

How fine-grained is this level? This is an empirical question to which we do not have a definitive answer. If consciousness depends only on the timing of action potentials (on the order of milliseconds), then current simulation technology may already be approaching the necessary temporal resolution. But if subcellular processes (protein folding dynamics, intracellular signalling cascades, synaptic vesicle release kinetics) are causally relevant for consciousness, then the requirements become far more stringent.

The no-cloning theorem adds a further complication. Even if a perfect upload were computationally feasible, the process of scanning the brain to create the upload would inevitably perturb the brain, introducing a discrepancy between the original and the copy. This is not a technological limitation but a structural one: it follows from the physics of measurement. The uploader must therefore accept that the uploaded mind will be, at best, an approximation of the original, however close.

7.3 Relation to the quantum no-cloning theorem

Our no-cloning theorem is structurally analogous to the quantum no-cloning theorem of Wootters and Zurek (1982) and Dieks (1982), but it does not depend on quantum mechanics. The quantum theorem shows that an arbitrary quantum state cannot be perfectly copied due to the linearity of quantum evolution. Our theorem shows that a conscious state cannot be perfectly copied due to the perturbative nature of physical measurement. The two results converge on the same conclusion from different directions.

If quantum effects turn out to be relevant for consciousness (as proposed, controversially, by Penrose, 1989, and Hameroff and Penrose, 2014), then the two no-cloning theorems would reinforce each other, producing an even stronger impossibility result. However, our result stands independently of any quantum theory of consciousness, which is a desirable feature.

7.4 Limitations

We acknowledge several limitations. First, the identification of microstates with complete physical configurations is an idealization. In practice, the relevant level of description may be coarser, and identifying the appropriate level of grain is an open problem (the "grain problem" discussed by Eliasmith, 2000, and Piccinini, 2015).

Second, the proof of the no-cloning theorem relies on the assumption that physical measurement is perturbative. While this is true in practice for all known measurement processes, we do not prove that it must be true in principle. A putative non-perturbative measurement technology (if such a thing is physically possible) would evade our theorem.

Third, our framework is purely structural: it describes the conditions for consciousness transfer but does not explain why these conditions are sufficient for consciousness. The "hard problem" (Chalmers, 1995) remains untouched. Our contribution is to the structure of the problem, not to its solution.

8. Conclusions

We have developed a formal framework, based on category theory, for analysing the conditions under which consciousness transfer between physical substrates preserves phenomenal character. The main results are:

(i) The definition of causal categories as temporally enriched categories that encode the complete causal-temporal structure of a physical system.

(ii) The proof that causal isomorphism (the existence of an invertible causal functor) is a necessary condition for consciousness transfer (Theorem 1).

(iii) The identification of temporal microstructure preservation as an additional necessary condition, leading to the requirement of full temporal isomorphism (Theorems 2 and 3).

(iv) The derivation of a no-cloning theorem for consciousness (Theorem 4), establishing that perfect consciousness cloning is physically impossible.

(v) The definition of a transfer fidelity metric that quantifies the quality of approximate transfer and the conjecture that fidelity can be made arbitrarily close to, but never exactly equal to, 1.

Taken together with our previous work on topological invariants of phenomenal experience (Caracuel Llabrés et al., 2016) and moral gradients in non-biological substrates (Caracuel Llabrés and Vidal-Moreno, 2017), the present paper contributes to a broader programme of mathematical formalization of the structure of consciousness and its philosophical implications. The programme is far from complete, but we believe that the results obtained so far demonstrate its viability and its capacity to generate novel and non-trivial insights into questions that have traditionally resisted formal treatment.

Acknowledgements

This work was partially funded by project FFI2014-57409 of the Spanish Ministry of Economy and Competitiveness (MINECO) and by an FPU grant from the Spanish Ministry of Education (ref. FPU14/02387). K.C.Ll. thanks the members of the Cognition and Brain Group (CBC) at Universitat Pompeu Fabra for their hospitality during the preparation of this manuscript, and Martín Aráoz-Gutiérrez (University of Seville) for discussions on the categorical formalism in early stages of this work. J.F.-C. acknowledges support from the Catalan government through grant 2017 SGR 1545. We are grateful to three anonymous reviewers whose detailed and challenging comments substantially improved the paper.

References

Awodey, S. (2010). Category Theory (2nd ed.). Oxford University Press.

Bickle, J. (1998). Psychoneural Reduction: The New Wave. MIT Press.

Block, N. (1978). Troubles with functionalism. Minnesota Studies in the Philosophy of Science, 9, 261-325.

Caracuel Llabrés, K., Vidal-Moreno, E. and Aráoz-Gutiérrez, M. (2016). Invariantes topológicos de la experiencia fenoménica: un marco computacional para el mapeo de qualia. Revista Iberoamericana de Filosofía y Ciencias Cognitivas, 12(3), 247-289.

Caracuel Llabrés, K. and Vidal-Moreno, E. (2017). Gradientes morales en sustratos no-biológicos: hacia un cálculo de agencia ética sintética. Daímon. Revista Internacional de Filosofía, 72, 143-178.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. Journal of Consciousness Studies, 2(3), 200-219.

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

Chalmers, D. J. (2010). The Character of Consciousness. Oxford University Press.

Chalmers, D. J. (2011). A computational foundation for the study of cognition. Journal of Cognitive Science, 12(4), 325-359.

Dieks, D. (1982). Communication by EPR devices. Physics Letters A, 92(6), 271-272.

Eliasmith, C. (2000). How neurons mean: A neurocomputational theory of representational content. PhD thesis, Washington University.

Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). Synthese, 28(2), 97-115.

Hameroff, S. and Penrose, R. (2014). Consciousness in the universe: A review of the "Orch OR" theory. Physics of Life Reviews, 11(1), 39-78.

Mac Lane, S. (1998). Categories for the Working Mathematician (2nd ed.). Springer.

Markram, H. et al. (2015). Reconstruction and simulation of neocortical microcircuitry. Cell, 163(2), 456-492.

Maudlin, T. (1989). Computation and consciousness. Journal of Philosophy, 86(8), 407-432.

Milkowski, M. (2013). Explaining the Computational Mind. MIT Press.

Penrose, R. (1989). The Emperor's New Mind. Oxford University Press.

Piccinini, G. (2015). Physical Computation: A Mechanistic Account. Oxford University Press.

Pigliucci, M. (2014). Mind uploading: A philosophical counter-analysis. In R. Blackford and D. Broderick (Eds.), Intelligence Unbound (pp. 119-130). Wiley.

Putnam, H. (1967). Psychological predicates. In W. H. Capitan and D. D. Merrill (Eds.), Art, Mind, and Religion (pp. 37-48). Pittsburgh University Press.

Sandberg, A. and Bostrom, N. (2008). Whole Brain Emulation: A Roadmap. Technical Report 2008-3, Future of Humanity Institute.

Schneider, S. (2014). The philosophy of "her". New York Times Opinionator, March 2.

Schweizer, P. (2016). In what sense does the brain compute? In V. C. Müller (Ed.), Computing and Philosophy (pp. 63-79). Springer.

Searle, J. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417-424.

Searle, J. (1992). The Rediscovery of the Mind. MIT Press.

Seung, H. S. (2012). Connectome: How the Brain's Wiring Makes Us Who We Are. Houghton Mifflin Harcourt.

Singer, W. and Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. Annual Review of Neuroscience, 18, 555-586.

Varela, F. J., Lachaux, J.-P., Rodriguez, E. and Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. Nature Reviews Neuroscience, 2(4), 229-239.

Wootters, W. K. and Zurek, W. H. (1982). A single quantum cannot be cloned. Nature, 299(5886), 802-803.