Chapter 8

Free Will, Language, and the Causal Exclusion Problem

Bernard Feltz and Olivier Sartenaer

1 Introduction

The starting point of the present paper is the rather commonsensical idea that mental causation *qua* mental – that is, at the very least, some minimal degree of irreducible mental causation – is a necessary condition, though perhaps not a sufficient one, for free will (see e.g. List & Menzies 2017). In other words, it is the thought that one can only feel entitled to consider human actions as being freely willed insofar as, among other things, their ultimate origin – what ultimately caused them – is not wholly microphysical in nature (on the model of, let’s say, a bunch of interrelated neurons firing). Much has been said in recent literature about how exactly to make sense of such a thought, essentially by addressing the question of how mentality could be said to be causally potent in spite of its undeniable dependence on a neural basis. Here we venture into a different kind of exploration, more specifically related to the question of what irreducible mental causation can be or, more particularly, where it could come from. In this perspective, the particular question we would like to address is the following: could it be the case that what makes mental causation apparently so special or unique is that it is deeply rooted in complex forms – perhaps only to be found in human communication – of language? Put differently, could some form of irreducible linguistic causation be at the basis of the kind of mental causation that would be appropriate, among other things, for having free will?

Here is how we plan to deal with such a question. First, we briefly introduce the causal exclusion argument, essentially as an excuse to allow for mapping the space of the possible ways in which linguistic causation could in principle

---

1 A recurrent strategy in this respect is to embrace some form of causal pluralism by considering that mental causation is not of the same kind as physical causation. See e.g., mental causation as “fact causation” (Lowe 2008) or “criterial causation” (Tse 2013).
be related to physical causation (Section 2). We then identify one (family) of these ways that would be appropriate to ground the kind of mental causation necessary for free will (Section 3). On the basis of this purely formal exercise, we then turn to the more ambitious objective of trying to provide plausible empirical support for the targeted (family of) schema(s) of linguistic-to-physical causation, building on recent neuroscientific work on neural plasticity (Section 4).

2 Linguistic-to-Physical Causation: Mapping the Conceptual Landscape

As of today, the causal exclusion argument certainly is one of the main challenges that any proponent of irreducible mental causation has to face. The moral of the argument is indeed essentially the following: between irreducible mental causation and a minimally monistic stance any neuroscientist should be attached to, a choice has to be made. This notoriously leads to a rather dramatic dilemma. Either one has to give up on genuine mental causation and, with it, free will, or one has to embrace some variation of Cartesian-style dualism, which prima facie doesn’t fit that nicely into the current scientific picture of how minds are usually supposed to work in our world.

Besides vindicating this important moral, the causal exclusion argument has a useful pedagogical advantage. Insofar as it is structured around premises that usually act as demarcation lines between competing views on the mind-body relationship, it allows for giving an interesting overview of the conceptual landscape of the mind-body problem. Even more than that, for the argument can easily be generalized to any type of “higher-level” or “non-physical” causation, it can help to appreciate the diversity of available positions on any mode of articulation between any causal realms. Obviously, the ones that will be of interest here are the physical and the linguistic, from which the notations employed below naturally follow.

One possible way of framing the argument is to point out that the following set of four statements is inconsistent, \( L_i \) and \( P_i \) being different events understood as property instances occurring at times \( t_i \):\(^2\)

\[
[\text{Dependence}]. \text{Every } L_i \text{ synchronically depends on an underlying } P_i. 
\]

\(^2\) There currently exist several formulations of the argument. The present one is close in spirit to Kim’s (2005, 39–45), where the dependence premise is cashed out in terms of nomological supervenience.
[Distinctness]. Every \( L_i \) is distinct from any underlying \( P_i \), in the sense that \( L_i \) has some causal potency of its own, “over and above” any \( P_i \)’s causal potency. More precisely, \( L_i \) is a cause of some \( P_j \) (with \( t_j > t_i \)).

[Closure]. Every \( P_j \) that has a sufficient cause has a sufficient cause \( P_i \) that is part of the same “P-level”.

[Exclusion]. No single event can have more than one sufficient cause at any time.

Although these four premises are mutually inconsistent, holding only three of them allows us to define a coherent picture of the linguistic/physical articulation. Four families of positions are thereby available:³

- Varieties of dualism deny [dependence], making higher-level causal events essentially independent from any lower-level basis. Such a view can come in an interactionist version, where [distinctness] involves, as mentioned above, some form of downward causation that crosses the inter-level ontological gap. Some versions are non-interactionist or parallelist, as they eschew any commitment to the clause of downward causation present in [distinctness].
- Varieties of reductionism deny [distinctness], making higher-level events either non-existent (eliminativist reductionism), causally redundant (retentive reductionism) or causally impotent (epiphenomenalism).⁴
- Strong emergentism denies [closure], to the effect that the recognition of irreducibly causal higher-level events renders (most versions of) physicalism false. In this kind of scenario, linguistic events, though dependent on the physical world, are ontologically different from – and are able to causally influence the course of – physical events.
- Weak emergentism denies [exclusion], construing \( L_i \)’s causal action on \( P_j \) as essentially redundant with \( P_i \)’s causal action on \( P_j \). One way of cashing out this idea is through non-reductive physicalism built under the perspective of the “subset account” of powers, according to which \( L \) and \( P \) are different properties insofar as \( L \)'s powers form a subset of \( P \)'s powers.⁵

³ See Wilson (2015) for similar analysis, though more precise and complete.
⁴ Of course, that epiphenomenalism is here classified as a variety of reductionism is indicative of a causal realist stance that turns out to be a widespread background assumption in this kind of debate. We are well aware that epiphenomenalism can be considered as a form of antireductionism in other contexts (see e.g. Baysan, forthcoming).
⁵ See again Wilson (2015). Here we also adopt her terminology, keeping in mind that both the weak and strong varieties of emergence are to be considered in an ontological way. For a taxonomy that takes this into account, see Guay & Sartenaer (2016).
3 Linguistic-to-Physical Causation: The Strongly Emergentist Option(s)

Though rather coarse-grained, this brief overview of the ways in which linguistic and physical events can come to be causally related is enough for our present purpose of identifying a schema that could be conductive to free will. Although the importance and the very formulation of each of the four premises of the causal exclusion argument can be discussed – they are actually vividly debated in the literature – we suppose here that [dependence] and [distinctness] are non-negotiable ingredients of any reasonable account of free will. As a result, and insofar as the package [dependence] + [distinctness] is coextensive with emergence (see e.g. Sartenaer 2016), we take it that some form of emergentism is a necessary condition for free will.

Of course, emergentism can come in many varieties, each one with its own commitments, stakes and problems. So a next step is to further identify the kind of emergentism that would be appropriate for free will. We already showed in the previous section that emergentism can come in (at least) two varieties – weak and strong – depending on whether it is consistent or not with [closure] or [exclusion]. Such a way of distinguishing between both families of emergentism hides what actually is their real demarcation criterion, which can be formulated through the following question: on the occasion of \( L_i \) emerging from \( P_i \), are the causal powers of \( L_i \) also (possibly) exercised by \( P_i \)? Put differently, are the emerging higher-level powers also (possibly) present at the basal level?

If one takes the answer to these questions to be affirmative, then higher-level events, though existing and causally potent, don’t add anything causally new to the world’s ontology. They simply happen to do something that is in any case already done, so to speak. This first case, corresponding to weak emergentism, is consistent with Kim’s “causal inheritance principle”, according to which instances of realized properties have the same powers as the instances of their realizers (see e.g. Kim 1993, 326). Accordingly, weak emergentism can generally be seen as encapsulating the “preformationist” claim that powers of emergent entities are always already preformed in their bases (Sartenaer 2018). Because of this, this first option has to be discarded, as one cannot imagine making sense of the authorship condition of free will on the basis of the

---

6 In a nutshell, our motivation for this claim is rather commonsensical: an ideal account of free will should encapsulate (some degree of) irreducible non-physical causation – hence [distinctness] – in a way that is consistent with the widespread recognition that higher-level events are always somehow grounded in physical events – hence [dependence].
assumption that, whatever an agent freely does, his neurons (let's say) have already done it. In that kind of case, there would indeed be no principled reason for considering the agent as being free while denying that her neurons also be. If one is free, so must be the others, but since the others clearly aren't, then the first cannot be either. It thus seems, albeit somewhat unsurprisingly, that free will doesn't get along with causal redundancy.

This leaves us with the second option, available as soon as one is keen to answer the aforementioned demarcation question negatively. In that perspective, higher-level events have causal powers that are neither “inherited from” nor “preformed in” their physical dependence bases, so that they really bring causal novelty to the world's ontology – of course at the expense, as we've seen, of the causal closure of the physical world. This position has received many labels over the years, like, e.g., “property dualism” or “interactionist monism”. We settle here for “strong emergentism”, capturing the claim that higher-level events strongly emerge from lower-level physical events as soon as the first ones exert causal powers that the second ones cannot possibly exert [“strong” distinctness], while the very existence of the firsts depends on the existence of the seconds [dependence]. It is such a strong variety of emergentism that we take to constitute a necessary, though not sufficient, condition for free will, as it certainly allows for capturing the authorship condition in a robust way.\footnote{While focusing directly on mental causation rather than linguistic causation, other philosophers have already defended the idea that strong emergence is necessary for free will. See for instance O'Connor (2000) and Lowe (2008) in a libertarian, agent-causalist setting. For a compatibilist take on the issue, see Kistler (2010), where the emergence involved is somehow weaker (without being weak in the sense described here), insofar as it is associated with downward causation construed as the action of a constraint.}

This being said, two refinements can be made. First, it has now become rather standard to make use of another conceptual distinction in order to further compartmentalize the conceptual landscape of emergentism. As such, some variants of strong emergence would be classified as “synchronic”, for the emergent higher-level property is supposed to be instantiated at the same moment as his lower-level, basal property. This is the canonical way of looking at emergence, which is implicitly at stake in the usual formulation of the causal exclusion argument as discussed in the previous section. Another variant of emergentism, recently come to the fore especially in philosophy of science, construes the emergence relation as essentially diachronic, the emergent being instantiated later than its basis. More than a mere detail, such distinction is of the utmost importance when it comes to metaphysical discussions, as the dependence relations at play in both cases are very different. Typically, while synchronic versions of emergence construe [dependence] in terms of...
constitution, composition, realization or supervenience, diachronic varieties interpret it as being some sort of causation. Since we will later take as paradigmatic examples of linguistic downward causation cases where linguistic occurrences have been *produced* or *generated* by human beings, it is the diachronic version of (strong) emergence that will be of interest here, along the lines of, for example, O’Connor & Wong’s (2005) dynamical theory of emergence (see Figure 8.1).

As a side remark, one can point out that recent variants of diachronic strong emergence eschew any commitment to the idea that the emergent should be of a higher-level than its basis, making emergence a concept that is not necessarily holistic in character (Humphreys 2016; Sartenaer 2018; Guay & Sartenaer 2018). Though it could be interesting to probe this line of thought further, it is not the place to do it here.

A second refinement is based on the plausible hypothesis that, in a putative case of linguistic, downwardly oriented causal influence, the emergent linguistic event is not causally sufficient for bringing about the corresponding physical effect, for it is reasonable to expect that physical determinants are also at play. Accordingly, although strong emergentism tolerates that some physical effects are *wholly* determined by higher-level emergent causes, such a radical claim is not mandatory. As it appears in figure 8.1, it can also be the case that it is the conjunction of $P_i$ and $L_j$ that causes $P_k$, provided that — and this is the non-trivial matter on which strong emergentism essentially rests — the powers of $L_i$ are not the same as the ones of $P_i$, to the extent that [closure] cannot possibly hold. Put differently, $L_j$ is at least a necessary cause of $P_k$.

Before closing the present section, two remarks are in order, the second being incidentally the occasion of summing up our previous discussion. First, up to now, linguistic causation and higher-level linguistic events have been only conceived of as “linguistic” by way of stipulation. At this stage, we are totally noncommittal about any possible specificity of language that would make it fare in an idiosyncratic way in the debates about non-physical causation. As it has been clear, we simply exploited the high degree of generality of these debates, where the very nature of the higher-level realm under consideration is usually left unspecified.

Second and more importantly, our discussion has been purposively confined so far to the domain of “armchair” metaphysics, to the effect that the schema of linguistic-to-physical causation that we identified only describes a possible and coherent mode of articulation between both these causal levels.

---

8 The various levels of “linguistic event” will be detailed in section 4.2. At this stage, it can be considered that the concept of “utterance” is relevant.
As such, our endeavour has therefore been mainly descriptive and speculative. We only contended that the diachronic, strong emergence of higher-level linguistic events out of underlying physical events, together with the recognition that the associated, irreducible linguistic causation is only responsible for the advent of subsequent physical events to the extent that it acts together with underlying physical causes, is an adequate schema for making sense of the commonsensical authorship condition of free will. It is another question whether such a schema refers to any real causal dynamics in our world. This is up to science to tell us, and this is the issue to which we turn now.

4 Does Neuroscience Support the Identified Schema?

We have just seen that the schema of linguistic-to-physical causation could be a theoretical case of downward causation in a diachronic strong emergence perspective. We would like now to show how such a schema could be empirically relevant in relation with a neuroscientific point of view concerning language learning processes. Such a question has a double dimension. First of all, we would like to analyse learning processes in general, and secondly language learning processes specifically. Afterwards, in reference to philosophy of language, we would like to show how language-use in such a relation with the brain, could be the activity which permits the emergence of free will.
4.1  **Neuroscience and the Mechanisms of Learning**

Learning is a process strongly linked to memory. In all learning processes, memory has an important part. Moreover, in our study of the relation between free will and causality, the relation with neuroscience is of great importance. This is why, before introducing the processes specific to language learning and downward causation, we would like first to look at the cellular mechanisms of memory.

The work of Nobel Prize winner Eric Kandel (2006) on this topic is classic. Without entering into details of the experiments, we can explain that it was in work on the mollusk Aplysia that Kandel’s team first of all found evidence of two types of learning in behaviour. Certain behavioural reactions are learned for the short term – of the order of a day – while others are learned and retained for several weeks.

The researchers then focused on the mechanisms of memory at the cellular level. By studying the synaptic connections between sensory neurons and motor neurons, they found that short-term memory was linked to an increase in the release of glutamate in the synaptic cleft; this is a functional modification of the synapse. The synapse becomes more sensitive, leading to a faster stimulation of the motor neuron.

Long-term memory, on the other hand, involves an anatomical modification of the synapse, as the sensory neuron develops extensions towards the motor neuron. This process involves synthesis of proteins which requires the activity of the cell nucleus. As for short-term memory, this generation of synaptic boutons is accompanied by an increased release of glutamate, which corresponds to an increased sensitivity for the organism. Since this sensitivity takes the form of a new anatomical structure, it takes place over a much longer time-scale.

These fundamental mechanisms discovered in the context of Aplysia will be observed throughout the animal kingdom, in particular in mammals and human beings. The processes of learning vary in time. And, as in humans, short term memory and long term memory are accompanied by functional modifications and anatomical modifications, respectively.

As an example of this, Eric Kandel cites Thomas Ebert’s work on musicians. With his colleagues at the University of Konstanz, Ebert’s team compared images of the brains of violinists and cellists with those of non-musicians. This study revealed that the cortex area reserved for the right-hand fingers did not differ between musicians and non-musicians, whereas cortex area representations of the left-hand fingers were greatly extended in a proportion ranging from one to five in the brains of musicians. The size of a body part representation in the cortex thus depends on the intensity and complexity of its use.
Kandel draws interesting general consequences on the anthropological level: the architecture of each person’s brain is unique. As soon as each behaviour has such an impact on the brain’s architecture, each brain integrates the history of each individual. In consequence, even identical twins with identical genes have different brains because of their different life experiences.

Moreover, as we will see in the next section, such a mechanism can also be interpreted as a mechanism of downward causation. That concerns immediately the problematic of emergence and free will. Before this work of interpretation, we would like to refer to other mechanisms involved in learning; these mechanisms are linked to selectionist explanations.

In Neuronal Group Selection Theory (ngst), Gerald Edelman (1990, 1992, 2006), another Nobel Prize laureate, proposes a learning mechanism belonging to the logic of selectionist explanations. The genetic program induces a process of redundant connections linking sensory, motor and emotional centers within the brain. This is what Edelman calls the global cartography. This connectivity allows a multiplicity of various behaviours. An important mechanism here is selective stabilization: that implies that a nerve circuit used in the brain will be stabilized.

In this context, learning is linked to a trial and error-based strategy. Redundant structure allows for a multiplicity of behaviours. Selective stabilization mechanism in global cartography, which links sensory, motor and emotional centers, leads to the reinforcement of circuits which permit adapted behaviours. The circuits most used are reinforced to the detriment of other circuits, which remain very unstable, or even degenerate.

These second types of learning mechanism lead to the same conclusions as Kandel’s mechanisms: the structure of the brain is deeply marked by personal history. The brain’s fine structure is the result of the personal history of each individual.

4.2 Learning Language: Philosophy of Language and Neuroscience

Having introduced the neuroscientific mechanisms of learning, we would like to analyse the particular dimensions of language learning from the point of view of philosophy of language. This is in order to show how the proposed neurological mechanisms can meet these specific constraints. We will next propose an analysis which links these results to the topic of emergence and downward causation. Finally, we will demonstrate what the link between downward causation and language can contribute to our understanding of free will.

Habermas (2005) stresses at the outset that human cognition involves two dimensions: linguistic and social. Learning a language implies social
interactions. It’s in this context that Habermas distinguishes “subjective mind” from “objective mind”. Objective mind is “a collective knowledge preserved in symbolic form”. It includes grammar, logic, semantics but also systems of meanings culturally shared. Habermas speaks about a “space of symbolically structured reasons”. “The rational motivation of convictions and actions takes place in this dimension and follows logical, linguistic and pragmatic rules that are not reducible to natural laws”. (Habermas, 2005, 173) (91) In this context, “subjective mind” refers to individual activity: each individual who participates in “objective mind” is on the one hand capable of understanding and using a common language, can participate in a shared conversation, can share proposed meanings and values; and on the other hand, each individual becomes able to nurture this conversation through innovation and the creation of new concepts. The learning of language is precisely the intersubjective process by which the individual appropriates “objective mind” and becomes able to implement his “subjective mind”, becomes about to keep his own word and give himself his own system of meaning.

Habermas advances the hypothesis that “objective mind”, the rules that structure language and the fundamental concepts of a culture, could have a structuring effect on the brain itself. In some ways, we could talk about an implementation of “objective mind” in the brain. “The meaning’s systems can, in turn, influence the brains of participants through the grammatically regulated use of symbols. [...] In the course of ontogeny, the individual brain apparently acquires the dispositions required to “access” the programs of society and culture”. (Habermas 2005, 175)

It’s precisely this hypothesis that we would like to put forward with reference to the learning mechanisms described (among others) by Kandel and Edelman. Indeed, whether it is the functional and structural modifications or whether it is the processes of selective stabilization, learning leads to modifications of the cortex that we can interpret in Habermasian terms as an implementation of “objective mind”.

This implementation allows access to “subjective mind”, the individual’s capability to participate in conversation which characterizes his culture. The active entry of the individual into language over the course of ontogenesis requires the learning of rules of grammar, logical rules, and the meaning of concepts. Language learning gives access to shared meanings within objective mind and makes it possible to progressively take part in the process of elaboration of meanings which characterizes all culture. In this sense, the neuroscientific mechanisms of learning constitute the empirical basis for understanding the implementation of language in an individual.

A reference to more specific linguistic studies will permit an easier relation to causation and the cellular mechanisms of learning. In a recent introduction
to *The Handbook of Language Emergence* (Wiley, Blackwell, 2015), Brian MacWhinney distinguishes “six major, partially independent hierarchies: auditory phonology, articulatory phonology, lexicon, syntax, embodied roles, and communicative structure” (MacWhinney, 2015, 4). Each of these levels concerns partially distinct neuronal areas and all these levels are interconnected. Language learning is particularly complex because learning is always linked with action. Auditory phonology is the most passive activity: a baby registers the phonemes present in his environment. That is important while it prepares the utilization of such phonemes. Articulatory phonology is already more active because it proceeds not only to repetitions of phonemes, but to articulations of phonemes conducting to the production of words. With words, we participate to the lexicon level. The integration of words in sentences implies the integration of grammar constraints, thus directly concerns syntax. A more elaborated relation to language implies an embodied role and social relations. “At the most elementary level, communicative structures involve speech acts that can then be grouped into adjacency pairs from which higher-level structures such as topic chains and narrative structure can emerge. Each of these hierarchies is tightly linked to others” (MacWhinney 2015, 4)

Habermassian implementation of “objective mind” concerns all these levels. By integrating progressively all these levels in interaction with his social environment, a child becomes able to participate to social conversations, first in a passive manner, later in an active manner. Subjective mind refers to the acquired progressive ability to have its own discourse and to participate to the social production of innovative discourses. Moreover, different meaning’s systems are present in “objective mind”. Each individual has to choose between a plurality of meaning’s systems to organize his personal existence. These precisions in more specific linguistic terms will facilitate the articulation to the molecular mechanisms of language learning.

### 4.3 Learning Language and Causation

The moment has come to take up again the question of causation. We advance the hypothesis that language learning process mobilizes a type of downward causation in accordance with the metaphysical schema described in Section 3. The link we propose above between Habermas’s conceptions of language learning and the neuroscientific mechanisms of learning indeed suggests a strong diachronic emergentism. This argument involves two dimensions: one at an ontogenetic level and one at the level of the activity of the language-user herself (Feltz 2013).

At the level of ontogenesis, things are clear: language plays a role of downward causation because these are the intersubjective interactions which lead to learning, and lead to the implementation of rules in the brain. Participation
in “objective mind” leads to learning, that is to say, the implementation of rules in the brain of the subject who becomes capable of participating in the conversation, who is then provided with “subjective mind”. These processes of learning clearly reveal a dynamic which is characterized by downward causation.

More precisely, the first stages of language learning, auditory phonology, articulatory phonology, lexicon and syntax could approximately be considered as a specific form of perception. It is the case for auditory phonology. It is also partially the case for articulatory phonology and the other levels, but, in these last levels, learning is not only repeating and imitating behaviour present in environment. Learning language conducts to the ability to produce new sentences, and progressively leads an individual to be able to construct his own meaning’s system. Learning language is an active process, in interaction with cultural environment, which renders able to actively participate to social discussions. That is what Habermas means by “subjective mind”.

Regarding the activity of the language-user herself, things are more complex. As figure 8.2 shows, linguistic activity presupposes the learning of language (P₀); this, as we’ve just argued, consists in the implementation within brain connectivity of rules of grammar and meanings at time t₀. We thus presuppose that all the six levels of language learning have been realized. The brain state at time t₁ causes the sentences L₂ expressed at time t₂. Linguistic logics lead to reasoning and conceptual inventions that have an effect on the structure of the brain at P₃ so that, at time t₄, the individual expresses these new propositions L₄. In fact, we need to speak here of a continuous back and forth between language and the brain.

The image of a computer can be equally helpful here: the work of calculation takes time, the implementation of algorithms takes time, before the

![Figure 8.2](image-url)
production of a result that appears on the display. Linguistic activity cannot be performed independently of the activity of the brain, but the activity of language continually modifies the fine structure of the brain. "Subjective mind" is implemented by "objective mind", but permits novelty via the rules of language themselves. The brain state $P_3$ depends on the brain state $P_1$, but not entirely. It's necessary to integrate linguistic activity and its own logic in order to fully explain how state $P_3$ gives rise to expressions $L_4$.

Linguistic activity therefore takes the form of an interaction between language and the fine structure of the brain. The characteristics of language which involve the capacity for the production of innovation according to its own rules imply this continuous interaction. The impact of language via downward causation is therefore difficult to dispute in the processes of learning. To learn a language is to integrate "objective mind" into the connectivity of one's nervous system. Language has an impact on the structure of the nervous system and therefore downward causation is implied here. The place of downward causation in the use of language is more complex since we have to think of a constant interaction between language and the structure of the brain. The use of language involves the activity of the brain, but at the same time it also implies modifications in the connectivity of the brain.

Language thus has an impact on the connectivity of the brain. Language has its own logic of functioning. In a culture, language belongs to what Habermas calls "objective mind" which includes lexicon, syntax, but also "communicative structure" which includes meaning's systems. By "subjective mind", individual can participate in the evolution of "objective mind". In this sense, language can be considered as a mean to escape to a strict biological determinism. That doesn't automatically imply free will. Language could be another deterministic constraint. That is the position of many philosophers (e.g. Lacan, Levi-Strauss, Althusser or Atlan). By contrast, Habermas defends a philosophy of language where language is open to diversity in the "objective mind". Each individual has thus to make a choice between various meaning's systems present in a culture. That is "subjective mind". If language has its own logic, it can lead to escape to strict biological determinism. It can conduct to production of novelty. It can conduct each individual to adopt a specific meaning's system.9

---

9 This position could be compared with Dennett's view, which articulates free will with the activity of language and the property of linguistic reprogramming. On this point, we agree with Dennett's position. We are a little more distant concerning determinism. Dennett is compatibilist, while our approach to downward causation in relation with the neurological mechanism of learning opens the way to an incompletely deterministic world. Addressing these issues properly would require an extensive treatment that would go far beyond the scope of this chapter.
5 Conclusion

In this paper, after having provided a general, conceptual map of the possible ways in which linguistic causation could be related to physical causation (Section 2), we proposed a metaphysically coherent picture of linguistic downward causation that would be satisfying for securing the authorship condition of free will. In particular, it was the main purpose of Section 3 to emphasize that this could be achieved under the perspective of a diachronic and strong version of emergentism, with respect to which emergent linguistic events causally depend on underlying physical events and, at the same time and in spite of such dependence, linguistic events could be the place of irreducible causal powers. We then turn in Section 4 to some recent neuroscientific works in order to assess the empirical plausibility of such a schema, supplementing the discussion with the philosophical insights of Habermas’ distinction between “objective mind” and “subjective mind”. In particular, it has been contended that Kandel and Edelman’s works on neural plasticity offer an empirical basis for the implementation of “objective mind” and the activity of “subjective mind” under an emergentist perspective adequate for free will.10

References

Baysan, U. (forthcoming). Causal Emergence and Epiphenomenal Emergence. Erkenntnis.

Edelman, G. (1990). The Remembered Present: A Biological Theory of Consciousness. New York: Basic Books.

Edelman, G. (1992). Bright Air, Brilliant Fire: On the Matter of the Mind. New York: Basic Books.

Edelman, G. (2006). Second Nature: Brain Science and Human Knowledge. New Haven: Yale University Press.

Feltz, B., (2013). Plasticité neuronale et libre arbitre. Revue Philosophique de Louvain, 110, 27–52. Reedited in Archives de philosophie du droit, 55, Paris, 145–168.

Guay, A., & Sartenaer, O. (2016). A New Look at Emergence. Or When After is Different. European Journal for Philosophy of Science, 6, 297–322.

10 We would like to thank the audience of the ARC Colloquium: Free Will, Language, and Neuroscience held in Louvain-la-Neuve in August 2017 for helpful comments on an earlier version of this paper. Olivier Sartenaer also gratefully acknowledges the financial support of the Alexander von Humboldt Foundation.
Guay, A., & Sartenaer, O. (2018). Emergent Quasiparticles. Or How to Get a Rich Physics from a Sober Metaphysics. In O. Bueno, M. Fagan, & R.-L. Chen (Eds.), Individuation, Process and Scientific Practices (pp. 214–234). New York: Oxford University Press.

Habermas, J. (2005). Freedom and Determinism, in Between Naturalism and Religion: Philosophical Essays, 151–180. Malden: Polity Press.

Humphreys, P.W. (2016). Emergence: A Philosophical Account. New York: Oxford University Press.

Kandel, E. (2006). In Search of Memory: The Emergence of a New Science of Mind. New York: W.W. Norton & Company.

Kim, J. (1993). Supervenience and Mind. Cambridge: Cambridge University Press.

Kim, J. (2005). Physicalism, or Something Near Enough. Princeton: Princeton University Press.

Kistler, M. (2010). Strong Emergence and Freedom: Comments on Stephan. In C. Macdonald & G. Macdonald (eds.), Emergence in Mind, 240–251. New York: Oxford University Press.

List, C., & Menzies, P. (2017). My Brain Made Me Do It: The Exclusion Argument Against Free Will, and What’s Wrong with It. In H. Beebee, C. Hitchcock, & H. Price (eds.), Making a Difference, 269–285. Oxford: Oxford University Press.

Lowe, E.J. (2008). Personal Agency: The Metaphysics of Mind and Action. Oxford: Oxford University Press.

MacWhinney, B. (2015). Handbook of Language Emergence, Wiley, Blackwell.

O’Connor, T. (2000). Persons and Causes: The Metaphysics of Free Will. New York: Oxford University Press.

O’Connor, T., & Wong, H.Y. (2005). The Metaphysics of Emergence. Noûs, 39, 658–678.

Sartenaer, O. (2016). Sixteen Years Later. Making Sense of Emergence (Again). Journal for General Philosophy of Science, 47, 79–103.

Sartenaer, O. (2018). Flat Emergence. Pacific Philosophical Quarterly, 99, 225–250.

Tse, P.U. (2013). The Neural Basis of Free Will. Cambridge: MIT Press.

Wegner, D. (2002). The Illusion of Conscious Will. Cambridge: MIT Press.

Wilson, J. (2015). Metaphysical Emergence: Weak and Strong. In T. Bigaj & C. Wüthrich (eds.), Metaphysics in Contemporary Physics, 251–306. Leiden: Brill.