A Non-Reductionist Solution
to the Problem of Social Causation

Rodrigo A. dos S. Gouvea

University of Leipzig
Abstract

The thesis of the causal closure of the physical world renders mental and social causation philosophically problematic. In *The Construction of Social Reality*, John Searle offers a partial solution to the problem of the causal efficacy of social and institutional facts by an appeal to the notion of the Background, or, as I will argue, by an appeal to its physical components. Since Searle's solution refers to physical facts in order to explain social causation, it does not seem to differ from the solution offered by reductive physicalists to the problem of mental causation. In this paper, I will discuss both responses to these two problems of higher-order causation. As a result of this investigation, the paper offers an account of how and to what extent does Searle's solution solve the problem of the causal efficacy of social facts, without implying their reducibility to physical facts.

*Keywords*: Background, causal efficacy, reductive physicalism, social and institutional facts
A Non-Reductionist Solution
to the Problem of Social Causation

In the book *The Construction of Social Reality* (in short, *TCSR*), John Searle faces the problem of the causal efficacy of social and institutional facts. He claims to answer it by an appeal to the notion of the Background, which was originally introduced in the book *Intentionality: An Essay in the Philosophy of Mind* (in short, *Intentionality*). This paper aims to show how and to what extent does the notion of the Background solve the referred problem. It also argues that, besides the similarities with the reductionist response to the problem of mental causation, Searle's solution does not imply the reduction of social facts to physical ones.

The fundamental question of contemporary philosophy

At the beginning of *Making the Social World*, Searle states that contemporary philosophy faces a fundamental question that would have no historical precedence. He formulates it as follows:

*How, if at all, can we reconcile a certain conception of the world as described by physics, chemistry and the other basic sciences with what we know, or think we know, about ourselves as human beings?* (Searle 2010, p. 3)

This question presupposes an apparent distinction between two sorts or kinds of facts. On the one hand, there are the facts investigated by the natural sciences, which Searle calls “brute facts”. To make the discussion more tractable, I will follow Braddon-Mitchell and Frank

---

1 With the expression “causal efficacy”, I mean the general capacity of something to play some determinate role in a causal nexus. I do not identify the property of being causally efficacious with the property of being a sufficient cause.
Jackson and indiscriminately call them “physical facts” (1996, pp. 13f.)\(^2\). On the other hand, there are the facts that, according to common-sense and some philosophical views, would compose the human impregnated part of the world (or, simply, the human world). Among the components of the human world, we count social and institutional facts\(^3\), as well as our mental states\(^4\). Although this list may not be exhaustive, I will only be concerned in this paper with the components of the human world just specified.

**Components of the human world**

Following a traditional characterization, mental states are understood here as states (or events) that have intentionality or phenomenal quality as one of its properties. According to Searle, mental states can be both phenomenal and intentional, or posses just one of these distinctive properties (Searle 1996, p. 7).

Social and institutional facts are characterized in this paper by means of the theoretical apparatus employed by Searle to account for their nature. Three conceptual tools have been originally presented: collective intentionality, assignment of function and constitutive rules (Searle 1996, p. 13). Collective intentionality is the most fundamental among them. In *TCSR*,

---

\(^2\) Someone may claim that proponents of reductive physicalism would reject this definition of physical facts as being too tolerant, since it does not avoid any of the problems emphasized in the debate concerning the characterization of what is physical, e.g. the Hempel's dilemma. This claim is, however, harmless to the present investigation. As suggested by Andrew Melnyk (2003, p. 11), reductive physicalists could take some facts as physical in a broad sense, insofar as they are reducible to physical facts in the strict sense, i.e. facts which fall under a more rigid definition of physical. The tolerant definition proposed here can be understood as a characterization of what reductive physicalists would take as being non-controversially physical in a broad sense.

\(^3\) Even though Searle characterizes institutional facts as a subclass of social facts (1996, p. 26), I will usually refer to both classes as if they were independent.

\(^4\) Searle endorses the thesis that non-human animals have mental states and are able to constitute social facts. In *TCSR*, he describes “the fact that those hyenas are attacking a lion” as an example of a social intentional fact (1996, pp. 38 & 122). However, the mental states referred in “what we know or think we know about ourselves as human beings” can be easily (though not substantively) distinguished from the mental states of other kinds of minds. The distinction follows from the fact that we conceive our mental states to be entertained by humans.
Searle stipulates that the expression “social facts” will refer to all facts involving collective intentionality, and only to them (Searle 1996, pp. 26, 38 & 122).

It is not entirely clear what Searle understands under the notion of collective intentionality. At one point, he characterizes it as mental states, such as beliefs, desires, etc. that are shared among humans or animals (Searle 1996, p. 23). At another point, collective intentionality is presented as “a biological primitive phenomenon” that could be realized inside the brain of a single person or an animal (Searle 1996, p. 25). These two claims are incompatible. If a collective intentional state can be realized by a single person (or a single brain), then it does not need to be shared. In “Collective Intentions and Actions” (2000), Searle explicitly opposes the claim that collective intentional states must be shared. In the mentioned paper, he emphasizes the view “that all intentionality, whether collective or individual, could be had by a brain in a vat or by a set of brains in vats” (Searle 2000, p. 96).

There are, nonetheless, claims that one could uncontroversially make about Searle’s account of collective intentionality. One is that, in his view, collective intentionality is instantiated only by intentional states in the we-mode (Searle 1996, p. 26). Intentional states in the we-mode do not present only to the person (or animal) who entertains them as their subject. Instead, a “we”, i.e., a group in which that person (or animal) presumably takes part, is presented as the subject of such intentional states.

The fact that collective intentionality is instantiated by intentional states in the we-mode is prima facie compatible with both positions of the debate concerning the role of sharedness in the constitution of collective intentionality. On the one hand, one could state that collective intentionality occurs only when intentional states in the we-mode are shared. On the other, one could state that a single (and, thus, not shared) instance of a mental state in

---

5 Another aspect that is uncontroversially recognized in Searle's account of collective intentionality is the thesis of the irreducibility of intentional states in the we-mode to intentional states in the I-mode.
the we-mode would be sufficient. However, an analysis shows that the latter position is unacceptable.

Searle defends that intentional states represent, by means of their content and their direction of fit, their own conditions of satisfaction (Cf. Searle 1983, p. 13; 2010, p. 29). Thus, he says, “a belief represents its truth conditions, a desire represents its fulfillment conditions, an intention represents its carrying out conditions” (Searle 2010, p. 29). But do intentional states in the we-mode have some specific condition of satisfaction that intentional states in the I-mode lack? Consider the intentional state expressed in “we believe that we are walking together”. Nothing prevents that this intentional state might be instantiated at some place and time by an individual with whom no one is intentionally walking. There are different possible scenarios in which that person might instantiate such state. She may think she is accompanied by someone, when, in fact, she is walking alone. She may be side by side with an acquaintance that, for any reason (Alzheimer or the works of a Malin génie), might forget that he is with her, etc. It does not matter how this single instantiation might occur, what matters is that the respective intentional state is false in all possible scenarios. One of the conditions represented by the intentional state is not satisfied in any of these cases. The truth condition involved in this particular intentional state would have been satisfied if the intentional state expressed in “we believe that we are walking together” were shared by the other person(s) to which the term “we” also refers. Given that an instantiation of an intentional state in the we-mode always attributes the same state to others, then any true intentional state in the we-mode must be shared among the individuals that are referred to as their subjects. The conclusion that intentional states in the we-mode are falsified, if they are not shared, is a strong reason to restrict the notion of collective intentionality to the cases in which such intentional states are shared among the individuals referred as their subjects. I will
employ the expression “collective intentional state” to refer to intentional states in the we-mode that are shared in this way.

It is important to highlight that social facts are not just correlated to collective intentionality, but are also constituted by it. The collective intentionality condition for the constitution of social facts determines two of their distinctive features. They are intentionality-relative, i.e., they do not exist independently of instantiations of intentional states, and they are objective, in the sense that the collective intentional state that constitutes them must be shared among individuals.

The conceptual tools that let us comprehend the specific nature of social objects and institutional facts, namely, the notions of assignment of functions and constitutive rules, also represent conditions for their constitution.

The notion of an assignment of function is entitled to explain how certain things in the world acquire functions. Typical examples of functional facts are artifacts and biological structures. Differently from mere constituents of causal mechanisms, a functional fact would be determined by “purposes, goals and values” (Searle, 1996, p. 19). It is also characterized by a specific normative feature, namely, that a functional fact possess a function not only when it works, but also when it is not employed, and even when it fails to perform the causal role indicated by its function (ibidem). According to Searle, functions are always intentionality-relative (Searle 1996, p 14). This implies that functional facts are not constituted by their physical components, but depend on intentional states to exist. Such intentional states have the form of an assignment or an imposition of function. The functional object constituted will be a social fact (or a social object, as I will call it for now on), if the assignment of function is performed by a collective intentional state.\(^6\)

According to the theory presented in *TCSR*, institutional facts are constituted by the establishment, i.e., the collective recognition, acceptance or acknowledgment of constitutive rules (Searle 1996 p. 40). Constitutive rules have the form “X counts as Y in C”, and constitute institutional facts by assigning status-functions (Y) to physical facts or to previously constituted institutional facts (X) (Searle 1996, pp. 40 & 80).

**The components of the physical world; the causal closure and the reductionist solution to the problem of mental causation**

Without expecting to offer an exhaustive list, I want to emphasize two distinguishing features of physical facts. According to the definition mentioned earlier, physical facts are the ones investigated by the natural sciences. However, attention to the actual natural sciences shows that this definition is problematic. A widely acknowledged aim of biology is to reveal the functions of biological structures. If being the subject of a natural science is the only condition for something to be a physical fact, then the functional facts investigated in some sub-disciplines of biology would be physical. But a functional fact is, in Searle's view, always intentionality-relative. Its existence is dependent on an intentional state that assigns a function to it. In a plausible interpretation of the fundamental question mentioned above, if a fact depends in this way on our intentionality, then it is a component of the human impregnated part of the world. In order to avoid the danger of classifying things both among the components of the physical world and the components of the human world, we can exclude any intentionality-relative fact from the set of physical facts. Searle does this by stipulating that physical facts, or, in his words, the features described in “our fundamental ontology”, are *intrinsic features of the world* (Searle 1996, p. 9). By doing so, he excludes all functional

---

7 Searle argues that the intrinsic features of the world are not restricted to the physical facts, but include all mental states as well (Searle 1996, p. 11). He contrasts these features with the intentionality-relative features of
facts from the set of physical facts. This does not mean that biological structures – and the causal mechanisms in which they take part – are not included among what is understood here as physical. I will soon refer to neural processes and structures as physical facts, and the possibility of such characterization is important for the purposes of this paper. Nonetheless, the functions that are imposed or assigned to these biological structures – with all the features that determine them, such as “purposes, goals and values” and the normative aspect explained above – are excluded from the physical world.

Based on scientific findings, it is claimed that physical facts have at least another shared and distinctive feature (Papineau 2009, pp. 57 & 60). They are components of a domain, “the physical world”, that is causally closed. As it is common in the current literature, I will use the expression ‘physical world’ to distinguish the set of physical facts.

The thesis of the causal closure of the physical world says that:

Every physical effect has an immediate sufficient physical cause, in so far as it has a sufficient physical cause at all (Papineau 2009, p. 59; Cf. also Kim 2005, p. 15).

Given the causal closure of the physical world, two similar problems emerge. It is clear that these problems stand against the reconciliation of the two conceptions of the world referred in the fundamental question presented above. One is the problem of the causal efficacy of social and institutional facts. This problem can be formulated as follows: How can social and institutional facts have causal efficacy in a causally closed physical world? The other, but very similar problem concerns mental causation and it is formulated as: How can mental states have causal efficacy in a causally closed physical world?

The problem of mental causation is one of the most discussed themes in the current philosophy of mind. Based on the thesis of the causal closure of the physical world, Jaegwon
Kim and others have formulated the most compelling argument in favor of reductive physicalism, the so-called causal argument. The thesis of the causal closure of the physical world implies that any physical fact or (supervenient) mental state that is allegedly caused by a prior mental state would have two independent sufficient causes (Papineau 2009, p. 61). There would be a sufficient physical cause, as the thesis of the causal closure of the physical world profess, and the alleged mental cause. But the second premise of the causal argument, namely, the metaphysical principle of causal exclusion, makes the claim of the two independent sufficient causes unacceptable. Kim characterizes this principle as follows:

If an event e has a sufficient cause c at t, no event at t distinct from c can be a cause of e (unless it is a genuine case of causal overdetermination) (Kim 2005, p. 17).

By taking as premises both the thesis of the causal closure of the physical world and the metaphysical principle just mentioned, proponents of reductive physicalism argue that mental states must be reducible to physical facts, in order to be able to cause other physical facts and other mental states.

**Searle’s partial solution to the problem of social causation.**

A survey of Searle’s writings reveals that he has referred to a Background for different reasons (Cf. Searle 1983, pp. 145-152; Searle 1996, pp. 132-137 & Searle 2010, 155-160). In *TCSR*, however, the most important reason for appealing to a notion of the Background is to

---

8 In *Mind: A Brief Introduction*, Searle explicitly acknowledges the causal closure of the physical world (p. 136), and sketches a non-reductionist response to the problem of mental causation (pp. 146f). This response may be useful to understand how the Background explains the causality of social and institutional facts, but I do not think it will help us to understand their irreducibility. Searle’s argument for the irreducibility of mental states is based on the fact that they have a “first-person ontology”, while physical facts have a “third-person ontology” (Searle 2004, p. 147).

9 M. Ratcliffe claims that Searle “regards the Background as a necessary (though not sufficient) condition for almost all human experience, agency and language” (2004, p. 151).
answer the question of how institutional and, I want to claim, also other kinds of social facts, have causal efficacy.

At the beginning of the book, Searle states:

... to explain the causal functioning of institutional structures, we will introduce (...) the Background of capacities that humans have for coping with their environment. (Searle 1996, p. 13).

Before investigating the nature of the Background, it is important to consider a restriction indicated by Searle himself to the thesis that the Background would explain the causal efficacy of social and institutional facts. According to Searle, the notion of the Background is entitled to explain:

how we can relate to rule structures such as language, money, marriage, and so on, in cases where we do not know the rules and are not following them either consciously or unconsciously (Searle 1996, p. 129).10

In the passage just quoted, two ways of relating to institutional facts are acknowledged.

1. In some cases, we relate to institutional facts by following constitutive rules. For example, when someone learns the rules of chess.

2. In the majority of cases in which we relate to institutional facts, however, we do not actually follow a constitutive rules – neither consciously, nor unconsciously. We just act in conformity with them. For example, when a professional chess player plays chess.

I claim that there are analogous ways of relating with social objects (or artifacts):

1’. In some cases, we relate to social objects by assigning them functions in a theoretical way. In these cases, we entertain the thought that the social object in question has a certain function that characterizes it. For example, when someone learns how to use the most unusual tools of her Swiss army knife.

---

10 The emphasis is mine.
2’. In the majority of cases in which we relate to social objects, however, we do not entertain any thoughts about their characteristic function. We assign them a function in a strictly practical way. For example, when I put my glasses after waking up.

Based on these distinctions, a formulation of the claim involved in Searle’s solution to the problem of the causal efficacy of social facts would be as follows:

Searle’s solution: The notion of the Background explains the causal efficacy of social objects and institutional facts in the cases when we relate to (2.) institutional facts just by acting in conformity with their constitutive rules; and when we relate to (2’.) social objects by assigning them functions in a strictly practical way.

The physical Background

Searle appealed to a so-called hypothesis of the Background for different reasons, and – so it seems – has conceived different things under this title. I will attempt to present a notion of the Background that is compatible with the claims he made in *Intentionality* and *TCSR*.

In *Intentionality*, when explaining how the Background works, he says: “... the repeated experiences create physical capacities, presumably realized as neural pathways” (Searle 1983, p. 150); and that “practice enables the body to take over and the rules recede into the Background” (*ibidem*). These passages seem to imply that the components of the Background are physical. But subsequently, Searle makes a very cautious claim. He says that the things composing the Background (i.e., skills, stances, preintentional assumptions and presuppositions, practices, and habits) “are realized in human brains and bodies” (Searle 1996, p. 154). The claim that they are realized in human brains and bodies suggests a characterization of the nature of the Background that differs from the one based on the other
quoted passages. The components of the Background should, according to this last claim, not be conceived as identical with physical facts, but as being realized by them.

In TCSR, however, Searle states that:

*It is important to see that when we talk about the Background we are talking about a certain category of neurophysiological causation. Because we do not know how these structures function at a neurophysiological level, we are forced to describe them in a much higher level.* (Searle 1996, p. 129).

In this passage, Searle clearly elucidates that (at least in TCSR) when he uses terms such as “preintentional capacities” to describe the components of the Background, he is actually talking about neural processes and structure, but in a different level of description. Based on this passage, I propose an interpretation, according to which the Background is a set of neural states or events, that means, a set of physical facts. I will call it the physical Background.

A passage already quoted reveals an important aspect of the physical Background. It says that “... practice enables the body to take over and the rules recede into the Background” (Searle 1983, p. 150). The structure in which the physical components of the Background are related to each other is determined by practice or, in other words, by learning processes. The causal roles that the physical components of the Background are expected to perform in the cases described in (2.) and (2’.) are, in fact, programed responses to social objects and institutional facts.

**How does Searle's Solution work, and to what extent?**

It is time to consider how the notion of a programmed physical Background might solve the problem of the causal efficacy of social facts. We must remember, however, that the notion of a programmed physical Background is not entitled to explain all kinds of cases in which we
may relate to a social object or an institutional fact. Only the cases described in (2.) and (2'), i.e., the cases in which we relate to social objects and institutional facts in a practical way, are expected to be explained by an appeal to the notion of the Background. I claim that two aspects of interactions of this kind (or three, in the case of social and institutional objects) justify taking social objects and institutional facts to be causally efficacious. These aspects are the following:

i. The physical components of the Background participate in the causal nexus in question. They participate as components of the sufficient physical cause of the caused event.

ii. In cases of interaction with social and institutional objects (e.g., a 5 Euros note), the physical components of these objects will also take part in the causal nexus. Together with some physical components of the Background, they participate as components of the sufficient physical cause of the caused event.

iii. The components of the physical Background perform a role in the causal nexus only because the Background was programmed to respond in that way to the social or institutional facts to which the subject is related.

If these aspects are actualized, the social or institutional fact to which we are related in a practical way can be rightfully characterized as causally efficacious.

The accusation that Searle’s account implies the reducibility of social to physical facts

Proponents of reductive physicalism have described the explanation of a token (or instance) of some higher-level type through the mention of a simultaneous token (or instance) of some lower-level type as a reduction or, more precisely, as a reductive explanation (Cf. Melnyk 2003, p. 82; Kim 2008, pp. 94 & 96). Examples of reductive explanations are the identity statements between instances of mental states and physical facts that, according to reductive
physicalism, must be true if the instances of the mental states are causes of any event. The definition of a reductive explanation compels us to recognize in Searle's solution a large a set of reductive explanations. Given that it refers to components of the Background (i.e., to certain physical facts) in order to explain the causal efficacy of social facts, it seems that Searle's solution consists in reductive explanations of the practical relations we maintain to these facts.

In the last part of this paper, I will investigate whether, according to Searle's account, social objects and institutional facts are reducible to physical components of the Background, or to these in conjunction with their own physical components.

The constitution of social and institutional facts, and their irreducibility to the components of the Background that make them causally efficacious

As argued above, social and institutional facts are constituted if some determinate conditions are satisfied. The specific features that take part in the constitution of social and institutional facts determine their distinctive features. I will argue that, because of some of these distinctive features, social objects and institutional facts cannot be reduced to physical components of the Background, or to these in conjunction with their own physical components.

According to Searle's account, the most general condition for the constitution of social facts is that they must be correlated to collective intentionality. The condition that places the role of collective intentionality reveals that social facts are intentionality relative. But the constitution through a collective intentionational state imposes yet another feature on social facts. They are objective in the sense that their existence as social facts does not follow from a singular instance of an intentional state in the we-mode, such as my belief that we are walking
together, in a case when no one is intentionally accompanying me. If a fact is truly social, the
tentional state in the we-mode that constitutes it must be shared among the individuals that
are referred as subjects of the intentional state.

Social objects have yet another constitutive condition, namely, the role of assignment
of functions. In conjunction with the collective intentionality condition, a social object must
have a function assigned to it by instantiations of an intentional state in the we-mode among
the individuals referred as the subjects of the intentional state. In contrast to more simple
social facts, an intentional state in the we-mode that assigns a function to a social object
usually refers to an indeterminate set of subjects. Consider, for example, the intentional state
expressed by the sentence “we believe that scissors have the function of cutting”. The
continued existence (or maintenance) of a social object is guaranteed as long as there are
individuals instantiating the intentional state in the we-mode that assigns a function to it. This
feature is compatible with the possibility that a social object might have a continued existence
in spite of changes, in the course of time, in the set of individuals that collectively assign a
function to it.

Institutional facts are also intentionality-relative and objective in the sense just
elucidated. As in the case of social objects, the constitution of institutional facts is distinct
from that of more simple social facts by demanding a specific condition. According to
Searle's position in TCSR, an institutional fact must have a status-function assigned to it by
constitutive rules. Considered in conjunction with the collective intentionality condition, the
assignment of a status-function that constitutes an institutional fact must be realized by a
group of individuals. According to Searle, this occurs through a collective acceptance or
recognition of a constitutive rule of the form “X counts as Y in context C” (Searle 1996, p.
28). Similar to the case of social objects, institutional facts have a continued existence as long
as there is a collective recognition of their constitutive rules among a (usually indeterminate)
set of individuals. This feature is compatible with the possibility that an institutional fact might maintain their status-function in spite of changes, in the course of time, in the set of individuals that recognize its constitutive rule.

The irreducibility of social objects and institutional facts to the components of the Background that make them causally efficacious – or to these in conjunction with their own physical components – is based on the fact that the latter are not sufficient to constitute the first. The components of the Background can, at most, constitute single instantiations of collective intentional states in one individual. Thus, it enables practical assignments of functions to social objects or actions in conformity with the constitutive rules of institutional facts. But this is not enough for constituting social objects or institutional facts. As I have argued, in the cases of social objects and institutional facts, a (usually indeterminate) set of individuals, that may even change in the course of time, must instantiate intentional states in the we-mode in order guarantee their constitution.

**Conclusion**

To conclude, I would like to summarize the main theses that I have argued for in this paper. Firstly, I have defended the claim that the notion of a physical Background solves (at least partially) the problem of the causal efficacy of social facts in the cases in which we relate to them in a practical way. Secondly, I have shown that the physical components of the Background that secure the causal efficacy of a social object or an institutional fact are not sufficient for their constitution. Thus, an account was given about how the notion of a programmed physical Background solves, at least partially, the problem of the causal efficacy of social and institutional facts, without implying their reducibility to the physical facts.
Acknowledgments

I would like to express my gratitude to Prof. Pirmin Stekeler-Weithofer (University of Leipzig) and Prof. Neil Roughley (University of Duisburg-Essen), as well as to the participants of their research seminars. I would also like to thank DAAD for the financial support. I am grateful to the participants of the Conference on Collective Intentionality VII: Perspectives in Social Ontology in Basel and the section Metaphysik/ Ontologie of the XXII. Deutscher Kongress für Philosophie: Welt der Gründe in Munich for stimulating discussions.

References


