Editorial Manager(tm) for The Review of Philosophy and Psychology
Manuscript Draft

Manuscript Number: ROPP36

Title: Explaining Social Reality: Are the Notions of Collective Intentionality and Background Apt for the Job?

Article Type: Joint Action: What is Shared?

Keywords: Collective Intentionality; Background; Social Facts; John Searle; Social Ontology

Corresponding Author: Mr. Boaz Miller, PhD Candidate

Corresponding Author's Institution: University of Toronto

First Author: Boaz Miller, PhD Candidate

Order of Authors: Boaz Miller, PhD Candidate

Abstract: John Searle's theory of the construction of social reality has three major aims. The first is to provide an ontology of social reality, namely the 'building blocks' from which social facts are constructed and the logical relations between them. The second is to explain how social facts and social institutions come into being. The third is to explain how come social facts persist, for example, how come governments manage to maintain their coercive power, and how come paper money bills keep their status as legal tender. For the second and third aims, Searle invokes the notions of collective intentionality and the background. While Searle's ontological project may be successful, I argue that his explanatory project, namely the project that tries to address the second and third aims, is inadequate. Searle's arguments for the existence of collective intentionality are unsuccessful as they manages neither to show that collective intentionality is irreducible to individual intentionality nor that it can perform the explanatory work he attributes to it. Similarly, the notion of the background collapses into behaviourism, and hence is incapable of explaining the persistence of social facts. I sketch a more fruitful direction an explanatory project compatible with Searle's ontology may take.

Suggested Reviewers:

Title Page

Author:
Boaz Miller

Title:
Explaining Social Reality: Are the Notions of Collective Intentionality and Background Apt for the Job?

Affiliation and Address:
Institute for the History and Philosophy of Science and Technology
Room 316, Victoria College
91 Charles Street West
Toronto, Ontario M5S 1K7
CANADA

E-mail: boaz.miller@utoronto.ca
Fax: +972-3-547-5187
Tel.: +972-54-747-3423

Acknowledgments:
I thank Mark Solovey for his helpful advice and Meital Pinto for her support and inspiration.

Explaining Social Reality

Are the Notions of Collective Intentionality and Background Apt for the Job?

Abstract John Searle's theory of the construction of social reality has three major aims. The first is to provide an ontology of social reality, namely the 'building blocks' from which social facts are constructed and the logical relations between them. The second is to explain how social facts and social institutions come into being. The third is to explain how come social facts persist, for example, how come governments manage to maintain their coercive power, and how come paper money bills keep their status as legal tender. For the second and third aims, Searle invokes the notions of collective intentionality and the background. While Searle's ontological project may be successful, I argue that his explanatory project, namely the project that tries to address the second and third aims, is inadequate. Searle's arguments for the existence of collective intentionality are unsuccessful as they manages neither to show that collective intentionality is irreducible to individual intentionality nor that it can perform the explanatory work he attributes to it. Similarly, the notion of the background collapses into behaviourism, and hence is incapable of explaining the persistence of social facts. I sketch a more fruitful direction an explanatory project compatible with Searle's ontology may take.

Keywords Collective Intentionality; Background; Social Facts; John Searle; Social Ontology

1. Introduction

John Searle in his book *The Construction of Social Reality* (1995) discusses the ontology of social reality. His goal is to find out how social institutions are possible. Searle asks:

> How can there be an objective world of money, property, marriage, governments, elections, football games, cocktail parties and law courts in a world that consists entirely of physical particles in fields of force, and in which some of these particles are organized into systems that are conscious biological beasts, such as ourselves? (1995, xi-xi).

Searle's project may be broken down into three subprojects, each of which has a different aim. The first aim is to provide an ontology of social reality, namely the 'building blocks' from which social facts are constructed and the logical relations between them. For Searle, social ontology is an extension of the ontology of the physical world. The second aim is to explain, drawing on his social ontology, how social facts and social institutions come into being, namely the mechanisms and processes involved in the human construction of social facts and institutions, such as marriage and property. The third aim is to explain how come social facts and institutions persist, for example, how come governments manage to maintain their coercive

power, and how come paper money bills keep their status as legal tender. The continued existence of social institutions constitutes a genuine explanatory puzzle for Searle. On the one hand, social institutions are often very powerful and enduring. On the other hand, their existence seems fragile, as it seems to depend only on the continued mutual agreement between members of society that they should continue to exist. Even more puzzling is the continued existence of social institutions, such as oppressive regimes, that do not benefit and even harm the people whose collective agreement is required for them to continue to exist.

To address the latter two aims, Searle invokes the notions of collective intentionality and the notion of the background. While Searle's ontological project may be successful, I argue that his explanatory project, namely the project that tries to address the second and third aims, is inadequate. Searle's arguments for the existence of collective intentionality are unsuccessful as they manages neither to show that collective intentionality is irreducible to individual intentionality nor that it can perform the explanatory work he attributes to it. Similarly, the notion of the background collapses into behaviourism, and hence is incapable of explaining the persistence of social facts.

I sketch a more fruitful direction an explanatory project compatible with Searle's ontology may take. I argue that a framework with a notion of individual intentionality that assumes that people have and take each other to have similar experience of the physical world may overcome the problems with collective intentionality. Similarly, I argue that an understanding of Searle's notion of the background as representational, rather than dispositional system may solve the difficulties it faces.

In section 2, I present Searle's ontology and logic of social reality. In section 3, I critically discuss his notion of collective intentionality and his arguments in favour of it. I argue

that Searle fails to show that collective intentionality exists, and even if it exists, he fails to show how and when it amounts for the construction of social facts. In section 4, I discuss his notion of the non-intentional background on which intentionality interacts with the social world. I argue that while he manages to show what functions the background performs, he fails to explain how it performs them, specifically how the background is responsible for the continued existence of social facts.

2. The Ontology and Logic of Social Reality

Searle's ontology of social reality makes explicit metaphysical assumptions. Searle assumes that there exists a mind-independent reality governed by laws of nature. In addition, he assumes that this is the only reality that exists. Therefore, under Searle's account, there are brute physical facts, which would exist if there were no human beings around. Human beings are part of the physical world. Through natural selection, they have developed consciousness which entails intentionality, namely the capacity to represent other things (Searle 1995, 5-7).

Searle distinguishes between two types of features of things in the world: intrinsic and observer-related. Intrinsic features of things exist independently of our representation of them. For example, a stone has a mass. Observer-related features exist relatively to the intentionality of observers. For example, that a stone is a paperweight is observer-related. Observer-related features are ontologically subjective – their existence depends on the existence of subjects that believe in them. Some observer-related features are also epistemologically objective – it is not just somebody's opinion that a certain object is a screwdriver, but rather this claim is objectively ascertainable (Searle 1995, 9-13).

A particular kind of observer-related feature is a function. Searle stresses that under his account, all functions are never intrinsic to their object.[1] Within functions, Searle distinguishes agentive from non-agentive functions. Non-agentive functions are relative to a teleological framework, but are not functions for an agent. For example, the function of the heart is to pump blood. Searle argues that having the function of pumping blood is not an intrinsic feature of the heart. It is relative to a system in which a goal, such as sustaining life, is defined. On the other hand, pumping blood is not a function for an agent; the heart pumps blood independently of any agent's intention. By contrast, for an object to be a paperweight or a screwdriver is an agentive function. It is a screwdriver for *someone*. When agents assign a function to an object in the world, it is often because of its intrinsic features, such as the shape of an object that makes it suitable to serve as a screwdriver (Searle 1995, 13-20).

Within agentive functions, Searle identifies a particular subgroup, whose agentive function is to symbolize, represent, stand for something, etc. He calls these functions 'status functions'. Status functions are the basic building blocks of social reality. (1995, 23) The basic formula for creating an institutional fact is 'X counts as Y (in context C)'. Searle calls a rule of this form 'a constitutive rule' (Searle 1995, 28). Status functions can be assigned to physical objects, e.g. 'this fence is a border between two states'. They can be assigned to persons, e.g. 'this person is a Canadian citizen'. They can also be assigned to other status functions, and iteratively, e.g. 'this Canadian citizen is a parliament member; this parliament member is the prime minister', etc. When specifying the context in which a status function is assigned, other

[1] Searle also defends this position against alternative positions according to which functions are, at least sometimes, intrinsic to their objects (1995, 16-17).

social or institutional facts[2] are mentioned, so an intricate web of relationships between status functions is formed. This web is social reality (Searle 1995, 79-84).

Searle stresses that physical facts are logically prior to social facts. When we iteratively impose status functions, we will find at the bottom of the pile physical objects. Something physical in the world must bear the status eventually. Otherwise, we will have endless regress (Searle 1995, 55-56).

Status functions are used to empower people. For example, if someone is the Persistent of the United States, he or she can veto a law passed by the Congress. If a physical object is involved, status functions empower people beyond the power of intrinsic properties of the object involved. For example, having paper money bills empowers people to buy things, in a way which is not a consequence of the intrinsic properties of the bills. By contrast, this is not the case when dealing with causal agentive functions: a screwdriver empowers its user in a way closely related to its intrinsic prosperities, such as toughness and shape (Searle 1995, 94-97). (I will challenge this distinction in section 3.)

Searle identifies four kinds of powers which status functions give to people. First, they give symbolic powers. They enable people to represent things in the world, using other things, such as words. Second, there are deontic powers, which create rights and obligations. Third, there are honours and dishonours, which give people status for its own sake, without rights or obligations. Last, there are procedural steps on the way to honour or power, e.g. being nominated as the Democratic candidate for president is a step in becoming the president. Searle sees deontic powers as the most important category and argues that all other powers are reducible to them (Searle 99-110).

[2] The difference, according to Searle, between social and institutional facts is the degree of their explicit codification. For example, there is usually no explicit codification of friendship or a tea party, but there is an explicit codification of who counts as the President of the United States (Searle 1995, 87-90).

Last, Searle's logical-ontological analysis of the social world has a hierarchical structure, at the bottom of which lies the physical world. This hierarchical structure is represented in Figure 1 (taken from Searle 1995, 121):

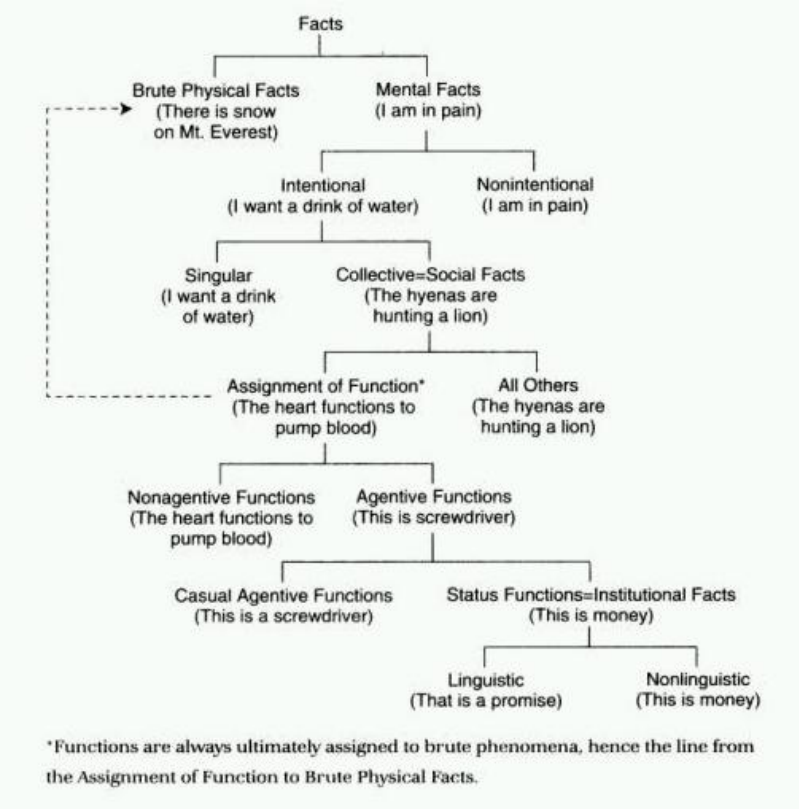

Figure 1: The Hierarchical Structure of Social Reality

In this section I have presented the ontology and logical structure of Searle's theory of social reality, with the exception of his notion of collective intentionality, which I will discuss in the next section. Various objections may be raised against parts of Searle logical and ontological analysis, but I will not pursue them in the context of this paper. Rather, within Searle's ontological framework, I will critically evaluate the process of *creating* social institutions and *maintaining* them, as Searle presents it. Two central notions are involved in these processes, collective intentionality and the background. In the next section I will first deal with collective intentionality.

3. Collective Intentionality and its Discontents

How are social facts created? To answer this question, Searle depicts two hypothetical examples. First, Searle asks us to imagine a village whose inhabitants have built a wall to mark some boundary. At first, people cannot cross this boundary because of the intrinsic features of the wall. As years pass, the wall wears down, but the territory in which it was build has acquired a *status* of a symbolic boundary. Because of the imposition of this status on the territory, people still treat this territory as a boundary and do not cross it (Searle 1995, 37-40).

A more complex example is the example of money. In a hypothetical historical genealogical narrative, first, commodity money such as gold was used because gold was considered valuable (this situation already assumes that the function of value had been imposed on gold). Then, the function of money, namely medium of exchange, is imposed on gold. Then, contract notes that promise to pay their holder a certain amount of gold are assigned the function of money. Last, gold itself is taken out of the picture, and the notes themselves become the medium of exchange (Searle 1995, 41-43).

The pivotal point in such processes is the moment in which things start functioning due to the status that is imposed on them rather than their intrinsic properties, e.g. when the wall wears downs. Usually, the imposition of the function is done collectively and implicitly. There is typically no one individual who assigns the status function, but rather a group of people, and there is no official moment at which one utters a constitutive statement, such as ‘let these paper notes be money’. (However, at later sages, official codification of existing social institution may be done.)

Two things seem quite mysterious in this process: First, how come a group of people collectively assigns a status function without ever implicitly agreeing to do that? Second, why do people continue to accept the status function after it is imposed?

A notion which is supposed to clear this mystery, especially with relation to the first question, is *collective intentionality*. Collective intentionality means that a group of people shares intentional states such as beliefs, desires and intentions. However, Searle's collective intentionality is also committed to methodological individualism. It posits no collective consciousness, Hegelian world spirit, or anything of this sort, i.e. no collective consciousness over and above the consciousness of the individual members of the collective. Collective intentionality is an individual mode of thought that takes the collective as its atomic unit. Collective intentionality is contrasted with the more familiar notion of individual intentionality, in which each individual has his or her own individual intentional states, which include beliefs about other individuals' intentions (Searle 1995, 23-25). The diagrams in Figure 2 illustrate this distinction. The upper one represents individual intentionality and the lower collective intentionality (Searle 1995, 26):

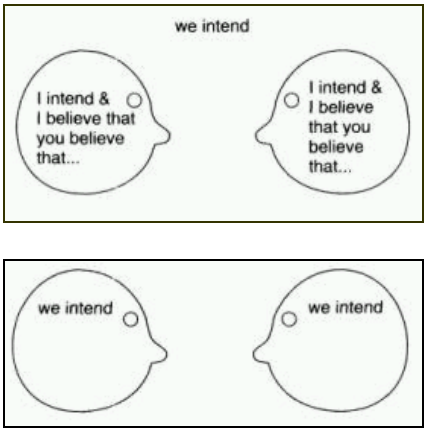

Figure 2: Individual and Collective Intentionality

To explain the difference between collective and individual intentionality, Searle asks us to consider two cases. In the first case, a group of people sits in a park, when it suddenly starts to rain. To take cover from the rain, all the people in the park run to a shelter which is located at the centre of the park. In the second case, a group of ballet dancers in the park converges into a central location in park as part of a ballet dance. While when observed externally, the two cases look the same, they are quite different in terms of intentionality. In the first case, each person has the intention 'I am running from the rain'. In the second case, each person has the intention 'we are converging into a central location' (Searle 2002 [1990], 92).

Searle claims that collective intentionality is 'a biologically primitive phenomenon' (1995, 24), and 'the capacity for collective behavior is biologically innate' (1995, 37). In nature, an example of collective cooperative behaviour is a pack of hyenas which move together to kill an isolated lion. In the case of humans, so Searle claims, groups can collectively have intentional states in a similar manner (1995, 38). He claims that 'the forms of collective intentionality cannot be eliminated or reduced to something else' (1995, 37). Most importantly for Searle, collective intentionality is irreducible to individual intentionality.

If collective intentionality is indeed biologically innate and calls for no further analysis, then it can explain, on its own, how social facts emerge, namely how groups of people collectively impose status functions without ever explicitly agreeing to do so.

But is collective intentionality indeed irreducible to, or non-analyzable in terms of individual intentionality? An attempt to reduce collective intentionality to individual intentionality will adhere to the notion of *mutual belief*. For example, in the case of a collective intention, it will ascribe to each individual an intention to do her part in a collective act. It will

also ascribe to her the belief that her fellows intend to do their part in the collective act and believe that she intends so as well, etc. (Searle 2002 [1990], 93)

Searle's arguments that aim to show that collective intentionality is irreducible to individual intentionality are surprisingly weak. Searle provides two such arguments. According to the first argument, a reduction of collective intentionality to individual intentionality requires an agent to have too many beliefs than she can possibly have:

> My intentionality, if I am part of a collective, is, "I intend to do such and such," and, "I believe that you also have this intention." Furthermore, I have to believe that you believe that I believe that you have the intention, and this then generates a nonvicious regress of the form "I believe that you believe that I believe that you believe that I believe," and so on; on your part, you believe that I believe that you believe that I believe that you believe, and so on (Searle 1998, 119).

Searle argues while the infinite regress in such a case may not be vicious, he finds this framework cumbersome and implausible. He states: 'I don't think that my head is big enough to accommodate so many beliefs' (1998, 119). He finds collective intentionality to be a simpler and hence a preferable option.

Searle's argument, however, is not persuasive. Searle does not distinguish between having a belief and consciously entertaining a belief. Consider the following example. I go down the street and pass by my car. I know that it's white. I know that it's a Suzuki. I know that it has four wheels. I know that I can use my key to open it and drive it. I know that if I drive it south for half an hour or so, I will reach the big city, and so on. Moreover, I know that this is my car; I know that I know that this is my car; I know that I know that I know that this is my car; and so on. I know infinitely many things about my car, which means that I potentially have infinitely many beliefs about it. Yet, in normal circumstances I do not and arguably cannot consciously entertain all of my beliefs about my car at any given moment. Rather, the particular beliefs I entertain at a given moment are triggered by some relevant event. By the same token, when I

take part in collective intentionality, just as in the mundane car example, I do not need to consciously entertain all the beliefs and meta-beliefs Searle mentions. Hence, his objection fails.

Searle's second argument is that collective intentionality includes a component that is irreducible to individual intentionality. He asks us to imagine a group of businessmen who have all been indoctrinated in the same business school with Adam Smith's theory of the hidden hand. Each of them comes to believe that he can best help humanity by perusing his selfish interests without cooperating with others. Each of them also believes that his fellows believe the same. Searle argues that this situation satisfies the conditions of mutual belief a reductive analysis of collective intentionality in terms of individual intentionality arguably must meet. Yet, so he claims, there is no collective intentionality involved in this example. There is no *we-intention* such as 'we are helping humanity', only *I-intentions* such as 'I am doing my part in helping humanity by perusing my selfish interests' (Searle 2002 [1990], 94).

This argument hinges on the premise that the indoctrinated businessmen example is different from the ballet-dancers example in that it does not involve a collective intention. Searle supports this premise only by an appeal to intuition. I argue that this intuition, inasmuch as it is decisive, is misguided. Furthermore, Searle's reliance on an example which explicitly involves a group of people who all believe in the existence of a social 'hidden hand' mechanism is particularly problematic. Hence, Searle's argument according to which collective intentionality is irreducible to individual intentionality fails.

Let us, then, examine Searle's example more closely. Adam Smith's theory involves claims about *unintended consequences*. Under Smith's theory, the betterment of humanity is an unintended consequence of each individual's perusing of his or her selfish interests. It is one thing for a person to act in a way that has unintended consequences without him being aware of

them. It is another thing for a person to act while being aware of the unintended consequences of his actions. These are two different intentional states. When a person believes in Smith's theory, he does not believe that his actions *alone* cause the betterment of humanity. Rather, he believes that his actions causally interact with the actions of others who live and interact with him in society, even if he does not cooperate with these others, to improve the state of humanity. He takes part in a collective intention.

In the indoctrinated businessmen example, the businessmen already see themselves as members of a society in which social institutions exists, and their actions as part of interactions of commerce and labour with other members of society. They also believe that their actions, *which involve social interactions with others who are acting the same as them*, bring about a chain of events that will contribute to the general wealth, and that in a society in which the institutions of commerce and labour do not exist, they cannot bring about such a chain of events. If the ballet-dancers dance is an example of collective intentionality, so is the indoctrinated businessmen example.[3]

To see how this is so, consider the following example. Suppose that in Searle's park example, the shelter in the centre of the park, to which people run when it starts raining cannot occupy more than twenty people without collapsing. Suppose that when it starts raining, twenty one people run to this shelter. In this case, each of them has the intentional state 'I am running to the shelter', and none of them has the intentional state 'we are causing the shelter to collapse'. By contrast, suppose there is a sign in the park which states that the shelter cannot occupy more than twenty people without collapsing. Suppose further that everybody in the park has read the sign and counted the number of people in it. In this case, when running to the shelter, each

[3] If the businessmen acted in the same way, but without being aware ore believing in Smith's theory, the condition of their having mutual beliefs about each other's intentions would not be met, and would not be considered as a candidate case for collective intentionality.

person has the intentional state ‘I am running to the shelter’, *as well as* the intentional state ‘we are causing the shelter to collapse’. The second case is equivalent to the Searle’s alleged counterexample of the indoctrinated businessmen, where the sign that everybody has read, understood and believes in is parallel to Adam Smith’s theory.

Even if collective intentionality existed as a primitive and irreducible phenomenon as Searle argues, under Searle’s characterization of it, it cannot account for the existence of social facts. A constraint on collective intentionality which Searle imposes is that it must be independent of the fact of whether or not one gets things right about what is actually occurring. Another way to put this is that collective intentionality must be consistent with the claim that all intentionality could be had by a brain in a vat (Searle 2002 [1990], 96).

This constraint makes collective intentionality alone, even if we assume that it exists in an irreducible manner, insufficient for explaining the creation of social facts. Margaret Gilbert gives the example of a young couple, a guy and a girl. Both the guy and the girl believe that they are going to get married. They both have a *we-intention* to get married. Each of them privately tells his or her patents that they are going to get married. However, as Gilbert argues, if the parents learn that the couple has actually not discussed the possibility of getting married, they would judge their children’s assertions as inaccurate (Gilbert 2007, 42). In this case, the existence of collective intentionally alone, though it is not mistaken, is insufficient for creating the social fact of engagement. In other words, Searle does not provide a criterion for distinguishing a mere shared belief or intention from a social fact.

In an opposite case, the appearance of cooperative behaviour can somehow occur while each individual has a different we-intention in mind. The problem in this case is that each individual may have a unique separate causal history for forming a belief or an intention.

Searle's account does not explain how it comes to be that different people form the same intention (Gilbert 2007, 36).

Let us sum up the objections up to this point. First, intentionality is ultimately individual and therefore the creation of social facts must be explained in individualistic terms. It thus remains a mystery how, within the framework of Searle's social ontology, a group of individuals can collectively impose a status function without ever explicitly forming an agreement to do so. Second, even if we assume that collective intentionality irreducibly exists, it is insufficient for creating social facts, as there are cases in which collective intentionality arguably exists but not a corresponding social fact.

In my view, the lesson from these criticisms is that the construction of social reality cannot be explained purely in mental terms. Searle's 'brain-in-a-vat' constraint on intentionality stems from his rejection of anti-individualism in philosophy of mind. Within the context of his social ontology, however, it seems extraneous and redundant. As you recall from section 2, Searle's social ontology presupposes metaphysical realism. It presupposes that individuals live in a world, in which there are brute physical facts and physical objects, some of which are other human beings who have the capacity of intentionality. The 'brain-in-a-vat' scenario seems incompatible with these assumptions, as the world the brain in the vat perceives is not the real world, and other persons it thinks are out there may not be real.

In the remainder of this section I sketch an alternative account of how people are capable of collectively imposing a status function without forming an explicit agreement to do so. My account, which is largely consonant with and in the spirit of Searle's social ontology, is twofold. First, in my view, the fact, which is an explicit assumption in Searle's social ontology, that all people in society live in the same world, have and take each other to have similar cognitive

mechanisms and similar experience of the world, holds at least some of the key to explaining how groups can collectively impose status functions without explicitly agreeing to do so. Put differently, the shared world and the shared experience of it constitute an anchor for the collective imposition of a status function, which may, in some cases, substitute explicit verbal deliberation among members of a group. Because every individual assumes that other individuals have a similar experience to hers, some things need not be explicitly stated and said.

Second, Searle makes a clear-cut distinction between status functions, such as prime minister, and causal agentive functions, such as a screwdriver. While causal agentive functions are assigned to objects because of their intrinsic features, status functions are purely conventional – they depend only on social acceptance (Searle 1995, 124). By contrast, I suggest seeing status functions and causal agentive functions as lying on a continuum. I suggest that an object's or a person's properties always play *some* role in the assignment of a function. Take the example of money. As Searle notes, gold was not chosen entirely arbitrarily to serve as a medium of exchange. It has some properties such as being rare and shiny that make it lucrative to people. Even when gold was taken out of the picture, paper money bills have some properties that make them suitable for having the social status of legal tender, such as being lightweight, durable, easy to carry around, and hard to forge. The assignment of the status function of legal tender to paper money bills is therefore not entirely conventional. When it comes to the assignment of status function to people, things are similar. For example, it is not a matter of pure convention who can become a lawyer. A person needs certain abilities and knowledge, as well as official authorization, where the latter is an indication for the former.

How do these two assumptions help better explain the process of creating social facts? I suggest that since at first, functions are assigned to things due mainly to their intrinsic features,

the mystery around the collective assignment of them somewhat dissolves. People share similar cognitive abilities and recognize the same intrinsic features of objects in the world. Moreover, they take each other to have similar experience. *This is why, at least at first, different individuals may come to share the same intentional state – as an individual, each of them recognizes that it makes sense to impose a certain status function on some physical object due to its intrinsic properties, and assumes others recognize this as well.* Over the course of time, I suggest, newer generations become used to seeing the status assigned to different objects and persons as part of their world without questioning it, and may assign more status functions, gradually making them more conventional and less causal.

The account I have presented here, of course, is a sketch, which requires further development that exceeds the scope of this paper. Of particular philosophical interest, in my view, is how strong the commitment to metaphysical realism needs to be to explain the formation of social facts of different kinds (cf. Hacking 1999, 80-84). Moreover, I do not believe that such work can be done solely from the armchair. A robust theory of they formation of social facts must be strongly supported by relevant historical, sociological and anthropological evidence. My aim here iss to argue that when the mysterious notion of collective intentionality is taken out of Searle's theory and when Searle's realist metaphysical assumptions are allowed to play the explanatory role Searle has originally denied them of, it becomes possible to explain, using Searle's social ontology, how social facts may come into being.

In this section I have removed some of the mystery around Searle's account of the process of creating social facts by collectively imposing status functions. But once social are created, how and why are they maintained? When a social fact becomes dependent mostly on convention, why do people continue to accept it, especially in cases when it seems to conflict

with their interests and goals? Searle offers a partial answer to this problem with his notion of the background. In the next section I will scrutinize this notion and the explanatory role Searle ascribes to it, and propose a more fruitful direction an explanation of the persistence of social facts, which is compatible with Searle's social ontology, may take.

4. The Explanatory Failure of the Background

It is clear from Searle's account that social facts depend on their collective acceptance in order to remain facts. But what explains their continued existence? As the collapse of the Soviet Union, for example, has shown, they can also collapse (Searle 1995, 91-92). What, then, explains their continued existences in some cases and their collapse in other cases? A common explanation attributes continued existence or collapse of social institutions to brute force or lack of it. Searle rejects this explanation:

> The temptation in all these cases is to think that in the end it all depends on who has the most armed might, that brute facts will always prevail over institutional facts. But that is not really true. The guns are ineffectual except to those who are prepared to use them in cooperation with others and in structures, however informal, with recognized lines of authority and command. And all of that requires collective intentionality and institutional facts (Searle 1995, 117).

Searle further observes:

> …the unfortunate person with a gun is likely to be among the least powerful and the most exposed to danger. The real power resides with the person who sits at a desk and makes noise through his or her mouth and marks on paper. Such people typically have no weapons other than, at most, a ceremonial pistol and a sword for dress occasion (Searle 1995, 118).

The use of brute force, then, depends on the existence of a social system, rather than vice versa. Searle also doubts common explanations such as social contract theories, which try to attribute the existence of society to one major cause. Searle notes that in cases in which social systems are on the verge of collapsing, such as the 1992 Los Angeles Riots, those who are in

charge of enforcing state power, such as the police, tend to overlook the real rioters, whom they fear, and pick on innocent people (Searle 1995, 90-91). He says:

> We cannot assume that the Leviathan will come to our aid in a genuine crisis; on the contrary, we are in the state of nature all the time, but the state of nature is precisely one in which people do in fact accept systems of constitutive rules, at least nearly all the time. (Searle 1995, 91)

This reference to a 'state of nature' is not accidental. Searle views innate biological abilities to socially interact as necessary for the creation of social reality. For him, biology is as essential to political theory as the features it has traditionally tended to address, such as the social contract (Searle 1995, 86).

This reference to biology leads us to Searle's notion of the background. Searle states that it is 'the biologically primitive sense of the other person as a candidate for shared intentionality'. Searle notes that the existence of the background is a necessary condition for the existence of collective intentionality and hence social reality (Searle 2002 [1990], 105). Searle says that the background is 'the set of non-intentional or preintentional capacities that enable intentional states to function' (Searle 1995, 129). Searle emphasizes that his talk about the background as enabling intentional states does not refer to logical conditions of possibility, but to 'neurophysiologic structures that function causally in the production of certain sorts of intentional phenomena' (Searle 1995, 120).

What does the background do? Searle identifies seven main functions:

1. *It enables linguistic interpretation*. The background does not only contain the meaning of words, but also the pragmatics involved in their use. For example, from the background one knows that in order to cut a cake she needs a knife, but in order to cut the lawn she needs a lawn-mower.

2. *It enables perceptual interpretation*. It allows people to see, for example, certain objects *as* chairs.
3. *It constructs consciousness*. This seems to refer to Searle's reference to seeing the other as an agent like yourself.
4. *It enables temporal interpretation of processes and narratives.* This seems as the application of (1) and (2) over time. It enables the interpretation of processes that take time, such as getting married.
5. *It gives a set of motivational dispositions.* These dispositions are our reasons for actions.
6. *It facilitates certain kinds of readiness.* It tells one what to except in certain situations. For example, the existence of other skiers in a ski resorts (as opposed, let us say, to giraffes).
7. *It disposes people to certain kinds of behaviour.* For example, to laugh when hearing a joke and cry in a sad movie (Searle 1995, 132-137).

How does the notion of the background explain the continued existence of social institutions? As we have seen, according to Searle, the structure of human social institutions is a structure of constitutive rules. While he takes these rules to be codifiable, he notes that they are often not codified. Searle invokes the notion of the background to explain why people continue to follow these rules long after they are constituted. Searle states that background provides a person with 'a certain sort of knowledge about how the world works […] a certain set of abilities for coping with the world' (1995, 131). While earlier generations construct social facts and the rules of behaviour associated with them using collective intentionality, successive generations inherent these rules from previous generations as part of the background, typically without ever questioning them. While the moment of the creation of a social fact involves a collective imposition of a status function, the subsequent use of the entities in question *need not contain the*

intentionality of the original imposition. For example, while the first generation imposes a function on a screwdriver, the next generations take screwdrivers for granted. The same applies to social institutions. Their existence *becomes accepted as part of the background*. This is why social facts and social institutions remain in existence, often long after they were created and long after the people who originally created them using collective intentionality passed away (Searle 1995, 126-129).

As Searle observes, on its own, this explanation is incomplete. A complete explanation must also explain how the background causally interacts with people's behaviour, namely how it causes them to follow the set of rules it encapsulates. I argue that it is here where his explanation fails, and this is why he does not ultimately manage to satisfactory explain the persistence and continued existence of social facts and social institutions.

Searle identifies two paradigms of causal explanation of human behaviour in current social science, both of which he rejects, and he presents the background as an alternative to both. The first, which he associates with behaviourism, aims to explain human behaviour in terms of 'brute physical causation' (Searle 1995, 139). Searle rejects it as it fails to account for the complexity of social behaviour and leaves no room for intentional states. The second is the paradigm of mental causation, which describes behaviour as a set of rational procedures, and regards reason for actions as causes. The problem with it, according to Searle, is that people often do not act rationally, and they also do not seem to consciously apply rules all the time. While their behaviour may be rule-described, it does not seem to be rule-governed (Searle 1995, 137-141).

Searle suggests an alternative causal mechanism, based on the notion of the background. According to his explanation, when people encounter rule-describable behaviour around them,

they develop dispositions to act in certain ways with respect to certain intentional states. Such dispositions are skills and abilities that are *functionally equivalent* to the system of social rules, but they do not contain any *representations or internalization of rules*. Therefore, in their behaviour, people do not apply rules consciously or unconsciously. In addition, these dispositions are not constituted by intentionality – rather, intentionality is preconditioned on them (Searle 1995, 142).

Searle claims that he does not argue that people follow rules unconsciously or that their behaviour is caused by an undifferentiated mechanism that happens to look as if it were rule structured. Rather, the mechanism has evolved precisely that it will be sensitive to the rules, but in itself it is not a system of rules. This is what, in his view, distinguishes his theory from behaviourism, which he rejects (Searle 1995, 146).

I argue that Searle's theory nevertheless collapses into behaviourism. Thus, it provides an insufficient causal explanation for human behaviour, and hence for the persistence of social institutions. It is true that Searle's explanation is not put in purely mechanical terms of stimulus and response. Nor is the background a mere set of rules. Nevertheless, it is still mechanical in spirit. As you recall, Searle situates his social ontology within the larger ontology of the natural world. For him, human beings are merely complex physical systems which have developed consciousness. Extending the ontology of the natural world to encompass social reality and include intentional states enables the expansion of the behaviouristic programme to include intentional states in its system of stimulus and responses. Despite his contention otherwise, this is what Searle's programme is doing.

This becomes apparent when we examine Searle's explanation of how dispositions to follow rules evolve in the first place. As he notes, when first created, rules are usually not

codified, but then they are incorporated into the background. How does this happen? Searle answers this question with a hypothetical thought experiment. He asks us to imagine a tribe where children just grew up playing baseball. They never learn the rules as codified rules, but are '*rewarded and criticized'* (Searle 1995, 144; emphasis added) for doing the right or wrong thing. In this way, the children develop the necessary dispositions for playing baseball (Searle 1995, 144-145).

It is exactly at this crucial moment, the moment at which social institutions and rules become part of the background, that Searle's framework collapses into (one would even say a simplistic) form of behaviourism. This, in turn, makes his explanation question-begging. The notions of reward and criticism which Searle uses strongly resonate with behaviourism, and its corresponding notions of stimulus and response. As such, they cannot provide a satisfactory explanation. Most rewards or punishments are in themselves social statuses. Rewarding someone is the imposition of a certain status function on her. In order to react to such rewards, i.e. to enjoy them, one needs to have already developed the necessary dispositions. Therefore, notions such as reward presuppose the existence of a background and cannot be used to explain how the background is formed.

Moreover, in many cases, certain status functions are imposed on certain physical feelings, such the status of sin that is imposed in Christian society on the pleasure that accompanies sex or masturbation. In such cases, the imposition of a social status makes some physical feelings and sensations feel bad, even thought arguably, from a purely physiological perspective, they actually feel very good. Similarly, in some circumstances, such as Shia Muslim religious rituals of self flagellation in the Day of the Ashura, people may cause themselves extreme pain, for which they are socially rewarded. It is thus even questionable whether

explaining the formation of the background by reference to so-called pure physical sensations of pleasure and pain can get us out of the problem.

As in the previous section, I would like now to briefly point out an alternative route an explanatory framework of the causal relations between the background and human behaviour that is largely consonant with Searle's framework may take. Searle considers and rejects two paradigms of explanations in social science, behaviourism and mental causation. He does not consider the interpretivist paradigm, which, in my view, may serve as a better theoretical framework to accommodate his notion of the background.

In fact, the interpretivist paradigm in social science features notions that bear much resemblance to Searle's notion of the background, two of which are Charles Taylor's notion of intersubjective meanings and his notion of moral space. The intersubjective meanings are a system of distinctions, by which certain behaviours and practices are judged as appropriate or inappropriate (or more appropriate or less appropriate) in a given society. Taylor, who develops this notion partly drawing on Searle's concept of a constitutive rule, states that the intersubjective meanings are 'the background to social action' (1985, 36). He argues that social institutions such as voting or negotiations, which involve certain forms of rule-governed behaviour, depend on the prior existence of relevant intersubjective meanings in society (Taylor 1985, 34-37). Similarly, his notion of moral space refers to 'the background, explicit or implicit, for our moral judgments, intuitions or reactions' (Taylor 1992, 26), namely a system of distinctions by which forms of the good life are judged (Taylor 1992, 25-52). Similarly, Clifford Geertz defines culture as a system of public symbols that provide people with patterns for action that serve as a background for rule-governed behaviour such as rituals (1973, 90-123).

While such notions bear resemblance to Searle's notion of the background, they also differ from it in two important aspects. First, while Searle, as I have argued, talks about the background in quasi-behaviouristic language as conditioning people to have certain behavioural dispositions, Taylor and Geertz talk about their respective notions primarily as schematic mental systems of *representation* of social reality, which are both public and private. Accordingly, their function is parallel to the function of maps and our spatial mental representations of physical reality. They do not directly cause action, but rather guide and orient our behaviour in space, be it physical or social. Taylor stresses the similarities between orientation in physical space and orientation in moral space, and argues that both are necessary for human agency (Taylor 1992, 29-31; 47-48). Similarly, Geertz talks about culture as a system of *models* which are both representations *of* the social world, and models *for* action (Geertz 1973, 93-94). In my view, this explanation of the role of the background in regulating human behaviour holds more promise than Searle's explanation.

This conception of the background as a representation of social reality has a second explanatory merit. As we have seen, Searle poses not only the persistence of social institutions but also their occasional collapse, such as in the fall of the Soviet Union or the 1992 Los Angeles Riots, as an explanatory challenge to his theory. After posing this challenge, however, he does not pursue it further. It seems that his conception of the background is incapable of explaining how such collapse may occur. The source of the difficulty, in my view, is that Searle describes the background as something that is *inherited* from previous generations, and once it is inherited, it is *taken for granted and not questioned* by successive generations. Under this picture, it is hard to understand how social institutions can ever collapse.

The representational conception of the background as a system of models may better face this explanatory challenge. Models of social reality, like any other representational models, may be more or less accurate in their representation of their target systems.[4] Note that according the interpretivist paradigm, the models of social reality in the background not only represent reality, but also guide action. I tentatively suggest that social institutions collapse when a critical mass of people in a society no longer deems these models as sufficiently accurate representations of social reality or as sufficiently useful in guiding or orienting action. Put differently, when only few individuals do not deem the background models as accurate representations or as useful guides for action, social facts remain intact, but when a sufficient number of people thinks so, people may 'break the rules' and social collapse may occur, sometimes in an escalating manner.

I present this only as a conjecture, which requires further development and support by empirical evidence. I believe that phrasing the problem in such terms opens up a fruitful research programme for the social sciences, which highlights the potential of Searle's theory to enhancing our understanding of social reality.

5. Conclusion

When trying to characterize the exact nature of the contribution of Searle's theory of social reality, namely the question Searle aims to answer, Ian Hacking writes:

> The question is not about psychology (what are the psychological preconditions for such human institutions to evolve). It is not about cognition (what must be known, or what cognitive states or abilities must exist, in order for there to be human institutions). It is not about sociology (what social practices or institutions must exists in order for there to be others). It is, for lack of a better name, about logic. The task is to state a set of preconditions, itself explicable in terms of Searle's 'basic ontology', such that, if they are satisfied, one can in quite direct way see how institutions and other social facts are possible (Hacking 1997, 89).

 See Hughes (1997) for a detailed discussion.

Searle may indeed provide us with a plausible ontology and logic of social reality. However, as I have shown, he fails to provide an explanation of how the different elements of his ontology initially come into being or how they maintain their existence. I have argued that Searle fails to show the existence of collective intentionality in a manner that is irreducible to individual intentionality, and how it is able to construct social facts. Similarly, his notion of the background provides a plausible model of the requirements for intentionality, but he fails to show how it casually interacts with human behaviour.

Nevertheless, in my view, Searle's theory provides the social sciences with a powerful ontology and vocabulary for studying social reality. I have sketched two ways in which his social ontology may be amended to overcome its current explanatory deficiencies, turning it into a more plausible and potentially more fruitful theoretical framework for social scientists and philosophers who are interested in studying the formation and construction of social facts.

References

Gilbert, Margaret. 2007. Searle and collective intentions. In: Intentional acts and institutional facts: Essays on John Searle's social ontology, ed. Savas L. Tsohatzidis, 31-49. Dordrecht: Springer.

Geertz, Clifford. 1973. The interpretation of cultures: Selected essays. New York: Basic Books.

Hacking, Ian. 1997. Searle, reality and the social. History of the human sciences 10(4): 83-92.

Hacking, Ian. 1999. The social construction of what? Cambridge, MA: Harvard University Press.

Hughes, R. I. G. 1997. Models and representation. Philosophy of science 64: S325-S336.

Searle, John R. 1995. The construction of social reality. New York: Free Press.

Searle, John R. 1998. Mind, language and society: Philosophy in the real world. New York: Basic Books.

Searle, John R. 2002 [1990]. Collective intentions and actions. In: Consciousness and language, 90-105. Cambridge, Cambridge University Press.

Taylor, Charles. 1985. Interpretation and the science of man. In: Philosophy and the human sciences, 15-57. Cambridge: Cambridge University Press.

Taylor, Charles. 1992. Sources of the self: The making of the modern identity. Cambridge, MA: Harvard University Press.