Bruno Celano

Collective Intentionality, Self-referentiality, and False Beliefs: Some Issues Concerning Institutional Facts

Comment to John R. Searle "Social Ontology and the Philosophy of Society" (Analyse & Kritik 20, 143–158)

Abstract: J. R. Searle's general theory of social and institutional reality, as deployed in some of his recent work (*The Construction of Social Reality*, 1995; *Social Ontology and the Philosophy of Society*, 1998), raises many deep and interesting problems. Four issues are taken up here: (1) Searle's claim to the effect that collective intentionality is a primitive, irreducible form of intentionality; (2) his account of one of the most puzzling features of institutional concepts, their having a self-referential component; (3) the question as to the point, or points, of having institutions; (4) Searle's claim to the effect that false beliefs on the part of the members of the relevant community are compatible with the existence of related institutional facts. It is argued that, under all four respects, Searle's theory proves to be hardly satisfactory.

1. Introduction

The importance of J. R. Searle's views about institutional facts for an understanding of social, political and legal phenomena could hardly be overestimated. Since the beginning of the sixties, Searle's rather sketchy remarks about "constitutive" rules, as contrasted with "regulative" ones, and about institutions and institutional facts, as contrasted with "brute" facts (Searle 1964, 111–4; 1965, 41–2; 1969, 33–42, 50–3), have played a key role in such different fields as, e.g., metaethics and jurisprudence. Thirty years later, in *The Construction of Social Reality* (1995), Searle presents us with "a general theory of the ontology of social facts and social institutions" (1995, xii). His *Social Ontology and the Philosophy of Society* (1998) provides a restatement of the main tenets of the book, and further develops some of its main themes.

Searle's theory of social institutions is rich and—its astonishing neatness and apparent simplicity notwithstanding—highly sophisticated. I shall not, in this paper, address all of its core tenets, nor shall I attempt to pass judgment on the theory as a whole.[1] Rather, I shall focus my attention on four

---

[1] I have attempted a fully-fledged critical assessment of the theory in Celano 1998. Some of the arguments deployed in the latter work are taken up again, and further developed, in sections 3 and 4 below.

separate issues: (1) collective intentionality; (2) self-referentiality of institutional concepts; (3) the point, or points, of institutions; (4) false beliefs about institutional facts. I shall argue that: (1) collective intentionality, as understood by Searle, raises deeper problems than Searle himself grants; (2) self-referentiality cannot adequately be accounted for along the lines suggested by Searle; (3) Searle's theory does not provide a satisfactory answer to the question as to the point, or points, of having institutional facts and activities; (4) it is not to be taken for granted that, as Searle suggests, the existence of institutional facts is compatible with the members of the relevant community having false beliefs concerning such facts.

2. Collective Intentionality

Institutional facts are, in Searle's terminology, a subclass of social facts. The latter are defined as facts "involving collective intentionality" (1995, 26, 88–9). Collective intentionality—intentionality of the form 'We (collectively) intend (desire, believe, etc.)...' (or, 'We-intend', 'We-desire', 'We-believe', and the like)—is, according to Searle, "a primitive form of intentionality" ("a primitive form of mental life"). It cannot be reduced, Searle claims, to singular intentionality (intentionality of the form 'I-intend', 'I desire', 'I believe', and the like), not even to singular intentionality supplemented with mutual beliefs (1990, 401, 404, 407; 1995, 23–6; 1998, 149–150). "We-intentions—Searle writes—cannot be analyzed into sets of I-intentions, even I-intentions supplemented with beliefs, including mutual beliefs, about the intentions of other members of the group" (1990, 404).

According to Searle, claims to the effect that (apparently) collective intentionality, or 'We-intentionality', may be reduced to singular intentionality, or 'I-intentionality', rest on a mistaken assumption; the assumption, namely, that "we have to choose between reductionism, on the one hand, or a supermind floating over individual minds, on the other" (1995, 25; see also 1998, 149: "a kind of Hegelian Weltgeist that is floating around overhead"). This dilemma is, Searle claims, "a false one" (1995, 259; see also 1998, 150). Allowing collective intentionality as a primitive, irreducible form of intentionality does not, according to Searle, commit one to unplausible assumptions. Why? Because, Searle argues, "it is indeed the case that all my intentional life is inside my brain, and all your intentional life is inside your brain, and so on for everybody else. But it does not follow from that that all my mental life must be expressed in the form of a singular noun phrase referring to me. The form that my collective intentionality can take is simply 'we intend', 'we are doing so-and-so', and the like." (1995, 25–6)[2]

---

[2] See also Searle 1998, 150: "the puzzle is, assume that all intentionality is in the heads

It seems to me, however, that such a simple theoretical move (such a "trivial notational solution of the puzzle"; Searle 1998, 150) cannot by itself suffice in coping with the difficulties raised by the notion of collective intentionality, as understood by Searle. Taking collective intentionality to be a primitive phenomenon, Searle concedes, has a price (1998, 150). The price is however, higher than Searle himself suggests. It is not only that, once we have allowed collective intentionality as a primitive phenomenon, we also have to grant that "I can be mistaken about the very mental state that I have" (1998, 150). This is not, Searle maintains, an untenable suggestion, and I fully agree with him. The problem rests, however, with *the way* in which, on Searle's account, in cases of (purported) collective intentionality I may in fact turn out to be mistaken about my own mental states. On Searle's account, whether in such cases I am mistaken about the very mental state that I have—about, e.g., *my* 'We-intending' to do such-and-such—turns out to depend on what *somebody else's* mental state is (e.g., whether *I* have a collective intention to do *A* turns out to depend on whether *you* have a collective intention to do *A*). It does not depend, that is, only on what *my* mental states are, but also on what *yours* are.

Notice that, according to Searle's view of collective intentionality, the problem is not whether I may happen to be mistaken, in my 'We-intending' to do such-and-such, because of my falsely *believing* that you also 'We-intend' to do such-and-such. Were collective intentionality reducible to singular intentionality plus belief about other people intentions (were my 'We-intending' to do such-and-such to be understood as my 'I-intending' to do such-and-such plus my beliefs about your 'I-intending' to do such-and-such, etc.), there would be no difficulty in maintaining that I can be mistaken in my 'We-intending' to do such-and-such, and that whether, in such cases, I am indeed mistaken or not may depend on what your mental states in fact are. It might happen, trivially, that my belief that you also 'We-intend' to do such-and-such is false, and that this makes my 'We-intending' mistaken. Such a possibility is not, however, the relevant one here; it is not in this way that, according to Searle, I may happen to be mistaken about my 'We-intending' to do such-and-such.

of individual human and animal agents, how can it be the case that it's all in our individual brains, if it is irreducibly collective? And the answer is, that we can have intentionality in your brain and in my brain, which is in the form of the first person plural as much as we can have it in the form of the first person singular. (...) The irreducible form of the intentionality in my head, when we are doing something collectively is, 'we intend"'. Searle (1990, 404, 406; 1995, 24–5; 1998, 149–150) provides some arguments against reductionist hypotheses—specifically, against attempts to reduce collective intentions to singular intentions plus common knowledge, or mutual beliefs. I have scrutinized these arguments elsewhere (Celano 1998, 29–36; 1999); they cannot, I have argued, be taken to rule out the possibility of a conventionalist account of some subclasses of social belief-dependent facts (such as, e.g., the existence of a 'convention', in D. Lewis' sense of this word; see Lewis 1969).

Rather, the issue is whether, literally, I am mistaken about whether *I* in fact have the mental state I believe myself to have; and the answer to this question turns out to depend on whether *you* have a certain mental state or not. According to Searle's account of collective intentionality, in short, claiming that *X* may be mistaken about his own collective intentionality (about, i.e., his own having an intention of the form 'We-intend') because *Y* does not have the corresponding mental state is like claiming that *X* may be mistaken in his believing that he has four hands because it may happen that *Y* lacks one.

Thus, in Searle's theory collective intentionality mental states have to be conceived of as mental states being both *mine* and *yours*, at the same time and under the same respect; or, what in fact amounts to the same, as being neither exclusively mine nor exclusively yours, but *ours*. The very same mental states have to be thought of as belonging, at the same time and under the same respect, both to *me* and to *you*. Allowing such a possibility amounts, I think, to allowing that, in cases of collective intentionality, *I* and *you* are one and the same agent (or, one and the same subject, as the bearer of the relevant mental states). Thus, it seems to me, you cannot take collective intentionality to be a primitive phenomenon, in the way Searle suggests, without having to postulate a plural subject; a single agent, that is, conceived of as being made up of *you* and *me*, at the same time and under the same respect. It looks as if, after all, allowing collective intentionality as a primitive, irreducible form of intentionality does indeed commit one, *pace* Searle, to postulating some sort of super-individual mind, individual minds being parts, or organs, of such a composite mental entity.

3. Self-Referentiality

Institutional facts have, Searle claims, a puzzling feature; they only obtain if they are believed to obtain. Where the fact that *p* is an institutional fact, 'It is the case that *p*' entails '*p* is believed to be the case'; or, where *P* is an institutional property, *x* is *P* only if it is believed to be *P* (so, e.g., "it is only money, because we believe it to be money", 1998, 144; see also 1995, 1, 3, 13, 32–4, 52–3, 63, 69, 76, 88–9, 96, 191–4, and Lagerspetz 1995, 6, 14). Institutional facts are, in short, belief-dependent.

It follows, Searle explains, that there is a kind of self-referentiality in institutional concepts (1995, 32–4; 1998, 145; see also Lagerspetz 1995, 6, 23, 209). Where *P* is an institutional concept, part of the definition of *P* will have to be the clause 'believed to be *P*'; claiming that *x* is *P* amounts to claiming that *x* has certain features *and* it is believed to be *P*.[3]

---

[3] Searle 1995, 32: it seems "that the concept of money, the very definition of the word 'money', is self-referential, because in order that a type of thing should satisfy the definition,

The self-referentiality of institutional concepts, in turn, seems to lead us in an infinite regress, or some kind of vicious circularity (1995, 33, 52; 1998, 145). So, for instance, "if part of the content of the claim that something is money is the claim that it is believed to be money, then what is the content of that belief?" (Searle 1995, 33). Where P is an institutional property, the belief that x is P will have to be, in part, the belief that x is believed to be P; it will, thus, have to be, in part, the belief that x is believed to be something that is believed to be P, and so on (Searle 1995, 33; see also 1998, 145).[4]

The issue is, thus, whether there can be self-referentiality without vicious circularity or infinite regress (Searle 1998, 155). This question should, according to Searle, be answered in the affirmative; the self-refentiality of institutional concepts does not, appearances notwithstanding, lead us in an infinite regress, or in vicious circularity. Searles's "resolution of the paradox" runs as follows (1995, 52; see also 1998, 155).

Let us consider an institutional word, such as, e.g., the word 'money'. The word 'money', Searle claims, "marks one node in a whole network of practices" (owning, buying, selling, earning, etc.); "as long as the object is regarded as having that role in the practices—he argues—, we do not actually need the word 'money' in the definition of money, so there is no circularity or infinite regress. The word 'money' functions as a placeholder for the linguistic articulation of all these practices". True, Searle allows, "to explain the concept [of money] we do need other institutional concepts such as 'buying', 'selling' and 'owing', and thus we avoided the vicious circularity only by expanding the circle by including other institutional concepts". There is, however, nothing wrong with this strategy; "we are not trying to reduce the concept 'money' to noninstitutional concepts" (1995, 52–3).

In other words. A satisfactory definition of the concept of money will have to include the clause 'believed to be money' (something is money only if it is believed to be money). This, however, does not, according to Searle, involve any vicious circularity, or infinite regress. Why? Because, he suggests, we may specify the content of the relevant belief—believing that something is money—by resorting to further institutional concepts: something is money only if it is believed to be a medium of exchange, a repository of value, payment for debts, etc.; i.e., we may, in the required belief-clause, replace the word 'money' with further institutional expressions, such as 'medium of exchange', 'repository of value', 'payment for debts', etc. Thus, "we can cash out the description in terms of the set of practices in which the phenomenon is embedded" (1995,

---

in order that it should fall under the concept of money, it must be believed to be, or used as, or regarded as, etc., satisfying the definition".

[4] Searle 1995, 33: "if the content of the belief that something is money contains in part the belief that it is money, then the belief that something is money is in part the belief that it is believed to be money, and there is, in turn, no way to explain the content of that belief without repeating the same feature over and over again".

53). Vicious circularity and infinite regress are avoided "not by our eliminating the circle but by widening it" (1995, 105); by resorting, namely, to further institutional expressions.

I do not, however, find such a purported resolution of the paradox satisfactory. Once the self-referentiality of institutional concepts is allowed, one cannot, I think, avoid being caught in an infinite regress by simply "expanding the circle" in the way suggested by Searle. By expanding the circle, it seems to me, the relevant regress is, step by step, reproduced; each step in the explanation, in fact, merely generates a further instance of the very same regress. Let me explain.

Institutional concepts have, let us assume, a self-referential component (in the sense explained above). This means that, where P is an institutional property, for any possible specification of what it is, for a certain kind of thing x, to be P, the following clause will have to be added: '...and x is believed to be something having the very features that have just been specified'. That is, for any further step in the explanation of any institutional concepts themselves involved in the explanation of a given institutional concept, we shall have to add, to the list of the necessary, sufficient (contributory, or whatever) conditions for the relevant institutional features $P_1 ... P_n$ to apply to x, the clause '...and x is believed to have these very same features'. The issue is not, thus, whether we necessarily have to use, in the *definiens* of any given institutional word, the *definiendum* itself. Such a kind of circularity—a merely terminological one—may, indeed, be avoided by 'expanding the circle' in the way suggested by Searle. The problem is, rather, that—if, as Searle himself assumes, the relevant words or phrases, involved in the definition, have themselves to be institutional ones—for any given specification of what it is, for x, to be P, we shall have to add a clause to the effect that x is also *believed* to be those things which we have just taken to be the very same thing as being P (i.e., to have those features which we regard as being the very same property as P). Whenever we (non-circularly) define a word, the *definiens* has to be regarded as in some sense expressing what the *definiendum* itself expresses. The problem is, however, that, where institutional concepts are concerned, being believed to express what the *definiendum* itself expresses has to be regarded as part of the *definiens* itself. It makes no difference, it seems to me, whether we say that the new institutional properties we resort to in explaining our initial institutional property themselves generate further instances of the same kind of infinite regress, or we say that we are simply stuck with further and further steps of one and the same regress. What matters is that we are not simply moving through semantic equivalences of some sort; we are, rather, compelled to assume that being believed to be an equivalent is part of the content of each member of the relevant equivalences.

Let us consider, once again, the concept of money. True, in defining the

word 'money' we may explain the content of the required belief to the effect that something is money by resorting to further institutional words or phrases ('believed to be a medium of exchange, repository of value, etc.'), thereby avoiding using the very same word, 'money', in the *definiens* of 'money'. The problem is, however, that for any given specification of what it is, for x, to be money, we shall have to add a clause to the effect that x is also *believed* to be those things which we have just taken to be the very same thing as being money (i.e., that x is also believed to have those features which we have just taken to sum up to the very same property as being money). This is something more than merely requiring that any purported *definiens* of 'money' should be regarded as in some sense expressing what 'money' itself expresses. The problem is that where money (or any other institutional concept) is concerned, being believed to express what the word 'money' expresses has to be regarded as part of the *definiens* itself. It makes no difference whether we say that the phrases 'medium of exchange', 'repository of value', etc., themselves generate further instances of the same kind of regress, or we say that we are proceeding through subsequent steps of one and the same regress. What matters is that we are not simply moving through a chain of subsequent definitions. We are, rather, compelled to assume that being believed to be equivalent to 'money' (or to 'medium of exchange, repository of value, etc.') is part of the very content of 'medium of exchange, repository of value, etc.' (or of whatever further institutional words or phrases the latter phrase is, in turn, explained in terms of), and so on.

In short. Given an institutional concept C, part of the content of C will have to be the clause 'believed to fall under C'. What is the content of that belief? The belief under consideration will have to be, in part, a belief to the effect that the sort of thing falling under C also falls under concepts $C_1 \ldots C_n$ (e.g., that x, which is money, is a medium of exchange, repository of value, etc.). We will also have to add, however, a belief to the effect that it is believed that the thing falls under $C_1 \ldots C_n$ (otherwise, $C_1 \ldots C_n$ would not, as Searle requires, qualify as institutional concepts). Now, what is the content of the latter belief? Let us assume that whatever falls under $C_1 \ldots C_n$ also falls under $C_a \ldots C_x$. If—as Searle postulates—$C_a \ldots C_x$ are themselves to be taken as institutional concepts, we will have to add, in answering the latter question (in specifying, i.e., the content of the belief that the thing falls under $C_1 \ldots C_n$), the clause 'believed to fall under $C_a \ldots C_x$', and so on. 'Expanding the circle' does not, in short, remove the relevant regress; it merely conceals, or postpones, it.[5]

---

[5] I have suggested elsewhere (Celano 1998, 36–41; 1999) that, in some cases at least, the self-referentiality of social concepts should be accounted for, along conventionalist lines (see e.g. Lewis 1969; Lagerspetz 1995; Celano 1995; Den Hartogh 1998), as stemming from a common knowledge, or mutual beliefs, structure in the intentionality which is constitutive of such concepts. Where the fact that p is a conventional fact, part of a (true) belief to the

4. The Point of the Game

Searle does resort to the 'expanding the circle' strategy in dealing with a further issue as well. Here, too, the strategy raises, I think, serious difficulties.

Institutional facts are, in the framework of Searle's theory, a matter of—conventional, non-physical—power (see Searle 1995, 95–6; 1998, 157). Basically, the collective intentionality which is constitutive of institutional facts has, Searle claims, the form:

(1) We accept (S has power (S does A)).

Such is, according to Searle "the primitive structure of the collective intentionality imposed on the X term, where X couns as Y in C" (1995, 104). So, e.g., the "underlying form" of 'X, this piece of paper, counts as Y, a five dollar bill', will be in part:

(2) We accept (S, the bearer of X, is enabled (S buys with X up to the value of five dollars)).

But, it might be asked, what kinds of acts do people have the power to perform, when institutions are in place (when, for instance, one has a five dollar bill)? In order for our account of institutional reality to be a satisfactory one, it might be thought, such acts will have to be conceived of as being, at bottom, 'brute'—i.e., not-institutional—ones; they will, namely, have to be thought of as being acts which can exist, and can be performed, independently of any institution (independently, that is, of any constitutive rules). Otherwise, it will be objected, there will remain something, in the logical structure of institutional reality, which instances of (1) do not prove capable of capturing. Suppose that an instance of (1) happens to hold. The kind of act, A, that S has, by hypothesis, the power to perform may be either an institutional one (such as., e.g., 'buying up to the value of five dollars' in (2) above), or a 'brute' one (such as, e.g., gaining physical control

---

effect that it is the case that p will have to be the belief that each member of the relevant community believes that each member of the relevant community believes (and so on, *ad infinitum*) that it is the case that p. Such a common knowledge structure is, I suggest, a good model for the very kind of self-referentiality Searle ascribes to institutional concepts (provided, of course, that one does admit the possibility of chains of interlocking beliefs of infinite levels, as required by the notion of common knowledge. I cannot go into this problem here; see Heal 1978, Bacharach 1992; and, for an hint as to a possible argument against this possibility, Searle 1998, 150). Or, in other words, assuming that the collective intentionality which is constitutive of institutional facts includes common knowledge would provide a plausible explanation of the self-referentiality of institutional concepts. As hinted above (sect. 2), however, Searle emphatically rejects purported accounts of collective intentionality in terms of singular intentionality plus common knowledge. Such an explanation of the self-referentiality of (some) social concepts is, thus, precluded to him.

of something). But, it might be argued, in case the former hypothesis holds, it must be possible, on pains of an infinite progress, further to analyze the relevant intentional structure, until a (conventional) power to perform a brute act appears. Our analysis will only be complete, it seems plausible to hold, when we will have reached an instance of (1) where A, the kind of act S has, by hypothesis, the power to perform, is a brute act. Otherwise, we will have to go *ad infinitum*.[6]

According to Searle, however, this line of reasoning is, in fact, misleading. He admits both possibilities, that A may stand for a brute or for an institutional kind of act (see 1995, 104–6). Furthermore, he explicitly takes into account the objection I have just suggested (one cannot, in explaining institutional powers, go on *ad infinitum* adding institutional concepts). His reply draws, once again, on the purported possibility of avoiding vicious circularity by 'expanding the circle' itself: "don't be worried—he writes—by the apparent circularity of using institutional notions such as 'buy' or 'value of five dollars' in the account of the intentional content, since these notions (...) cash out not by our eliminating the circle but by widening it" (1995, 105).

Once again, however, I find the 'expanding the circle' strategy troublesome. This time, it seems to me, the strategy turns out to be unsatisfactory under two respects. Both respects are related to a question of basic importance— a question which, however, Searle addresses only in passing. The question, namely, what is the point of having institutions (institutional facts and activities)?

First, the hypothesis that institutional powers may be powers to perform institutional acts, and so on *ad infinitum*—so that such powers could not, in principle, be reduced to institutional powers to perform brute acts—seems to me equivalent to the hypothesis that an institution may have no point whatever apart from the very activity of participating in the institution itself (and, perhaps, the pleasure that could be gained thereof). That, in other words, there may be no point, in participating in an institutional activity, other than the attribution (modification, and extinction) of institutional qualifications and statuses (such statuses being "valued—or disvalued—for their own sake"; 1995, 101), and the various steps, or stages, leading thereto ("procedural" conventional powers; 1995, 102, 109). Now, this hypothesis holds true, I think, of a fairly definite kind of institutions and institutional activities, which may be termed 'games' (or, of what regarding an institutional activity as 'playing

---

[6] It is perhaps worth emphasizing that the issue addressed in the text is *not* whether (Searlian) institutional powers may, or should, prove to be reducible to brute (physical) powers. My argument is by no means intended to suggest this possibility (I am not, namely, suggesting that institutional reality should be reduced to brute facts). The issue is rather, whether the kinds of acts which people have institutional powers to perform have to be conceived of as being, at bottom, brute ones.

a game' is).[7] But, is Searle suggesting that all institutions are to be understood along these lines? That, namely, all human institutions are, at bottom, games? Adopting the 'expanding the circle' strategy, and the denial that, in the last resort, institutional powers should be conceived of as powers to perform brute acts, would draw us towards this conclusion. I think it would be more to the point to regard the distinction at hand—the distinction, namely between, on the one hand, those institutions which are, indeed, a matter of conventional powers to perform brute acts, and, on the other hand, those institutions participation in which has no point other than participation itself (i.e., 'games')—as a crucial difference between two different sorts of social phenomena.[8]

Second, Searle does in fact provide, in his 1998 lecture, a sketchy, tentative answer to the question as to the possible point, or points, of having institutions. "The whole point of the institutional"—he suggests—"is often to structure the brute" ("a structure whose point is not just to empower other institutional structures, but to control brute reality"; 1998, 155; see also 1998, 156-157). This suggestion, it seems to me, points in just the opposite direction as the 'expanding the circle' strategy does. Where the point of the institutional is to structure the brute, it may plausibly be held, the acts institutional powers are powers to perform will have to be, at bottom, brute acts.[9] There is, thus, a tension—an oscillation—in Searle's views. What explains the oscillation is, I think, Searle's substantial disregarding of the question as to the possible point, or points, of having institutions. In the quoted passage, what does, precisely, the phrase "structuring the brute" mean? Searle does not, in fact, provide any definite answer to this question.

The tension stems, I think, from the contrast between two opposite, plausible views about the rationale for institutions. What is the point of having institutional facts and activities? One answer to this question is, because it's fun—or, because it's beautiful—to have them. The weight of this answer does not have to be underestimated. Participating in an institution for its own sake is what playing a game amounts to (i.e., creating and maintaining new kinds of facts and activities just for the sake of it); and games are, no doubt, impor-

---

[7] In the framework of the theory presented in *The Construction of Social Reality*, the hypothesis considered in the text amounts to the possibility that all institutional powers might turn out to be either "symbolic", or "procedural", or, finally "honorific" powers (1995, 99-102). On this hypothesis, there would be—contrary to what Searle himself claims—no categorical, or "terminal", institutional powers other than the honorific (and, perhaps, the symbolic) ones (see 1995, 105, 109-10). This is, I submit, precisely what happens where institutions are created and maintained for their own sake; where, namely, games are played.

[8] This distinction may, of course, be regarded as a distinction between two different attitudes people may take towards the very same institutional set up (an individual player may well play chess with a purpose other than gaining the status of winner).

[9] So. e.g., the point of buying a plane ticket is to make a physical movement of my body possible (1998, 155).

tant (a 'playing a game' attitude is, in social life, pervasive). This first answer is, I submit, the answer which the 'expanding the circle' strategy implies. A different answer is, however, that the point of institutions is to structure the brute (institutional facts as facts which "make the brute facts possible"; 1998, 155). That, namely, institutions purport to control the exercise of physical powers, by setting up norms for the performance of brute acts, or by making the achievement of desired, non-institutional ends possible.

We do not have to choose between these two answers. Both of them may hold true of different institutions, or in different situations, or for different people. They ground, however, a crucial distinction among two different kinds of social phenomena. No account of the structure of institutional facts disregarding this distinction may, I think, qualify as a satisfactory theory of institutional reality.

5. False Beliefs

Let us now consider a further claim of Searle's. On the one hand, Searle maintains, institutional facts are belief-dependent; they only obtain if they are believed to obtain (see above, section 3). On the other hand, he also claims, it may happen, and it often does happen, that members of the relevant community (i.e., participants in the institution) have false beliefs about institutional facts (i.e., about the institution itself). So, for instance, the existence of money, or X and Y being married, or X's being the king, are, all of them, institutional—thus, belief-dependent—facts; they only obtain if people believe them to obtain. It may happen, however, that people "believe that it is money only if it is 'backed by gold' or that it is a marriage only if it is sanctified by God or that so and so is the king only because he is divinely authorized" (Searle 1995, 47). Thus, according to Searle, in order for an institutional fact, p, to be the case it is not a necessary condition that p be recognized as such—namely, as a belief-dependent fact; "as long as people continue to recognize the X as having the Y status function, the institutional fact is created and maintained. They do not in addition have to recognize that they are so recognizing, and they may hold all sorts of other false beliefs about what they are doing and why they are doing it". "The participants—Searle claims—may accept the imposition of function only because of some related theory, which may not even be true" (1995, 47–8; see also 93, 96, 127, and 1998, 157).

According to Searle, then, belief-dependence of institutional facts is fully compatible with people having false beliefs about them (he even claims that "often the institution functions best when people hold false beliefs about it"; 1998, 157). This is, it seems to me, a mistake.

Institutional facts are belief-dependent. Thus, e.g., when the concept of

king is taken to be an institutional concept, being a king entails being believed
to be a king (x is a king only if, and because, he is believed to be a king).
Suppose now that members of the relevant community claim X to be their
king, and that, when asked how it is so, their answer is that X is their king
"only because he is divinely inspired" (Searle 1998, 158; see also 1995, 47, 96).
Suppose, further, that, if asked whether the truth of their belief to the effect
that X is their king does depend on their believing X to be their king, they
would sharply deny this to be the case—they would, perhaps, consider such
a thought to be a dreadful offence against God's majesty—, claiming that
what makes their belief true is a fact obtaining independently of any belief
they may happen to have thereabout. These people are, in short, explicitly
denying that the fact that X is their king is a belief-dependent fact. The
question is, is X's being their king an institutional fact? The answer, it seems
to me, has to be in the negative, for the following reason.

Belief-dependence entails, Searle rightly claims, self-referentiality. When
the concept of king is taken to be an institutional concept—let us name the
relevant concept '$king_i$'—X's being a $king_i$ entails X's being believed to be
a $king_i$. Thus, X may be held to be a $king_i$ only if he is believed to be
someone who is believed to be a $king_i$ (only if he is believed to be someone
who is a $king_i$ in virtue of his being believed to be a $king_i$). The people in
our imaginary example, however, do not believe X to be their king in virtue
of his being believed, by them, to be their king (they do not believe that it
is a necessary condition for X to be their king that he should be believed, by
them, to be their king). They are, moreover, sharply denying this to be the
case; to their lights, it would simply be blasphemy to claim that X is their
king in virtue of his being believed, by them, to be their king (X is their king,
according to them, in virtue of God's will, and this is what makes their belief,
that X is their king, true). Thus, they do not believe X to be their $king_i$;
and this entails that X is not their $king_i$. Believing someone to be the $king_i$
requires believing him to be someone who is believed to be a $king_i$ (someone
who is a $king_i$ in virtue of his being believed to be a $king_i$); and—such is the
import of the belief-dependence of institutional facts—being, in fact, a $king_i$
requires being believed to be a $king_i$. In our imagined example, therefore, X is
not the people's $king_i$. There are, in fact, two different concepts involved here:
the concept of $king_i$ (*ex hypothesi*, an institutional concept) and the concept
of king. In order for the fact under consideration to be an institutional fact,
the former concept has to be applicable. In our imagined example this is not,
however, the case. Thus, in our imagined example X cannot be said to be
the people's $king_i$; he is, rather, (believed by them to be) their king.

Let us generalize. When P is an institutional property, it is a necessary
condition for x to be P that x should be believed to be P. But, if it is
a necessary condition, for x to be P, that x should be believed to be P,

then it is a necessary condition, for x to be P, that x should be believed to be something such that it is a necessary condition, for it to be P, that it should be believed to be P. x would not, otherwise, be believed to be P; the property actually ascribed to x would in fact be different from its being P (the applicable concept would be a different one). Thus, when x is believed to be P, but it is at the same time explicitly denied that x may be P only if it is believed to be P (that, i.e., it is a necessary condition for x to be P that it should be believed to be P), P does not qualify as an institutional concept (x's being P is no institutional fact). I conclude, therefore, that, *pace* Searle, institutional facts being belief-dependent is not compatible with people having false beliefs about them.[10]

Bibliography

Bacharach, M. (1992), The Acquisition of Common Knowledge, in: C. Bicchieri/M. L. Dalla Chiara (eds.), *Knowledge, Belief, and Strategic Interaction*, Cambridge, 285–315

Celano, B. (1995), Consuetudini, convenzioni, in: P. Comanducci/R. Guastini (eds.), *Analisi e diritto 1995. Ricerche di giurisprudenza analitica*, Turin, 35–87

— (1998), Fatti istituzionali: la teoria di J. R. Searle, in: P. Comanducci/R. Guastini (eds.), *Analisi e diritto 1997. Ricerche di giurisprudenza analitica*, Turin, 19–54

— (1999), Fatti istituzionali e fatti convenzionali, in: *Filosofia e questioni pubbliche*, forthcoming

---

[10] What I am claiming, it should be stressed, is not that the relevant, true beliefs (e.g., the belief that X is the king in virtue of his being believed to be the king) should be consciously held by the people participating in the institution, nor that they should be in fact held by them, either consciously or unconsciously. As far as the former issue is concerned, Searte himself clearly distinguishes (see 1995, 47, 127; 1998, 157) between the hypothesis that the participants may unconsciously believe the fact that p to be an institutional—thus, a belief-dependent—fact and their having false beliefs about p. False, as well as true beliefs may be held both consciously and unconsciously. The latter issue—whether, namely, the relevant beliefs should in fact be held by the participants, either consciously or unconsciously—touches upon the possibility that intentional structures may play a role in the explanation of behaviour though not being part, neither consciously nor unconsciously, of the individuals' intentional makeup. This possibility is dealt with by Searle under the rubric of the "Background" of capacities, abilities, tendencies and dispositions (see 1995, 5, 127–46). I cannot, here, dwell on Searle's "thesis of the Background"; so I shall simply state my point in a somewhat dogmatic form. What the argument provided in the present section shows is that either the relevant true beliefs are taken to be part of the "specific structures of intentionality" (Searle 1995, 142) to which participants in the institution are, *ex hypothesi*, assumed to be sensitive in the form of background capacities, abilities, tendencies and dispositions, or the claim that instutional reality is belief-dependent poses a serious problem for the thesis of the Background. To put it in a nutshell, that X might be regarded to be the king out of mere training or habit, without reflection, does not conflict with the argument provided in the text.

Den Hartogh, G. (1998), A Conventionalist Theory of Obligation, in: *Law and Philosophy 17*, 351–76

Heal, J. (1978), Common Knowledge, in: *The Philosophical Quarterly 28*, 116–31

Lagerspetz, E. (1995), *The Opposite Mirrors. An Essay on the Conventionalist Theory of Institutions*, Dordrecht

Lewis, D. (1969), *Convention. A Philosophical Study*, Oxford

Searle, J. R. (1964), How to Derive 'Ought' from 'Is', repr. in: P. Foot (ed.), *Theories of Ethics*, Oxford 1967, 101–114

— (1965), What Is a Speech Act?, repr. in: J. R. Searle (ed.), *The Philosophy of Language*, Oxford 1971, 39–53

— (1969), *Speech Acts. An Essay in the Philosophy of Language*, Cambridge

— (1990), Collective Intentions and Actions, in: P. R. Cohen/J. Morgan/M. E. Pollack (eds.), *Intentions in Communication*, Cambridge (Mass.), 401–415

— (1995), *The Construction of Social Reality*, London

— (1998), Social Ontology and the Philosophy of Society, in: *Analyse & Kritik 20*, *143–158*