Infallibilism and Human Kinds

Francesco Guala

Philosophy of the Social Sciences 40 (2010), pp. 244-64.

Abstract

Infallibilism and apriorism are still influential in the philosophy of social science. Infallibilists about human kinds claim that there are features of institutional entities about which we cannot possibly be wrong. But infallibilism is not implied by the theory of collective intentionality that supposedly grounds it. Moreover, it fails to account for the mode of existence of important institutional kinds, including the paradigmatic example of money.

1. Introduction

Infallibilists about human kinds hold that there are features of institutional entities about which we cannot possibly be wrong. This claim is derived from a plausible ontological analysis of what it means for a kind to be of the "human" type.[1] In particular, the infallibilist claim is derived from the thesis that collective acceptance of a set of rules and conditions is required for an entity to belong to a human kind. Because these conditions are accepted by the relevant members of the community for which entity X is of kind K, it is impossible for the members to be ignorant of the conditions that make X an entity of kind K.

[1] Throughout the paper I will use the expression "human kinds" to refer to the categories we use to make sense of, classify, and organize our social environment. Whenever appropriate I will use a more refined terminology to distinguish between social, institutional, and artifactual kinds.

If infallibilism is correct, then much of what we can know about social reality can be known a priori. A competent member of a community needs only examine her own beliefs and linguistic practices about a human kind, in order to find out what that kind is. In this paper I examine critically the infallibilist position. To begin with, infallibilism clashes with explicit claims made by collective acceptance theorists. These theorists weaken the notion of acceptance considerably to account for the fact that we are often unaware of the conventionality of human kinds. Once so weakened, the notion of collective acceptance cannot support the substantial epistemological claims that infallibilists would like to make. A further worry is that some essential properties of human kinds cannot be known a priori. I will illustrate this claim using the example of money, a case that is frequently cited by collective acceptance theorists. Money, like many other institutions, requires projectibility, and the latter in turn requires knowledge of the mechanisms that sustain our classificatory practices. Since knowledge of such mechanisms is largely a posteriori, infallibilism does not hold for paradigmatic cases such as money.

The paper is organized as follows: sections 3-5 are introductory and lay out the main elements of the infallibilist position. Sections 6-8 discuss some amendments made to the theory of collective acceptance, and their implications for the infallibilist thesis. Sections 9-11 analyse two counterexamples, and section 12 concludes.

2. Realism about human kinds

History is full of puzzled philosophers trying to reconcile the relative lack of progress of the social sciences with our apparent direct acquaintance with social phenomena. Giambattista Vico is a famous precursor of this ambivalent attitude. In the *Scienza nuova* he writes of a "truth which is beyond any possible doubt":

> that the civil world itself has certainly been made by men, and that its principles therefore can, because they must, be rediscovered within the modifications of our own human mind. And this must give anyone who reflects upon it cause to marvel

> how the philosophers have all earnestly endeavoured to attain knowledge of the natural world which, since He made it, God alone knows, and have neglected to meditate upon this world of nations, or civil world, knowledge of which, since men had made it, they could certainly attain. [1744, 331]

Friedrich Hayek echoed Vico and his own teacher Mises when he claimed that in social science "we must be able to find all that we can understand in our own mind" [1943, 8]. Claims of this kind can be derived from a plausible ontological analysis of social reality. As Vico says, we have direct access to the fundamental features of society simply because social reality is *created* by us. Thus, according to David Bloor,

> Given a description of a material object or physical or biological process we can, in principle, study it more closely by examining more minutely or experimenting on it in more detail. This is because it has a material existence independent of our current descriptions and the current state of belief about it. A 'social object', by contrast, is constituted by the descriptions actors and participants give it. It has no existence independent of their beliefs and utterances about it; hence it cannot be described 'more closely' by, as it were, getting behind these descriptions. [1997, 35]

Observations of this sort raise some interesting issues concerning philosophical realism. Generally speaking, metaphysical realists believe in the existence of entities and properties that do not depend on our theories and beliefs about them. The meaning of "our" must be spelled out carefully though: in a sense, we are all realists about human kinds. We believe that money, the president of the United States, laws against smoking in public places have a reality that is somewhat independent of whether we *individually* believe that George Bush is president, that it is unlawful to smoke in pubs, or that this particular piece of paper is a one dollar bill. However, we do not fail to notice that each of these facts is dependent on the existence and intentions of a *collective* of people – that there would be no laws, presidents, or indeed money, if the relevant collective was suddenly wiped out by the almighty God.

Let us distinguish this (weaker) form of *social* realism, from traditional (stronger) *natural* realism of the Putnam-Boyd variety.[2] Some philosophers have argued that social realism has non-trivial epistemic consequences for the social sciences. The most important one, according to David Ruben, concerns the possibility or *error*:

> The essential point of [natural] realism […] is that it is always possible that our theories are wrong: error and mistake are always possible. Where there is this distinction between theory and theorized reality, our theory, or our beliefs about that reality more generally, may have *failed* to grasp adequately or comprehend what they are about. (Ruben [1989], 60)

In the social realm, in contrast, "widespread general classificatory belief that there are things of social kind s is sufficient for there being things of social kinds s". As a consequence, "consistent and widespread error regarding general classificatory beliefs about society is indistinguishable from reality" (Ruben [1989], 74).

3. Collective acceptance

Why are widespread classificatory beliefs only *sufficient* for the existence of social kinds? As Ruben points out, the existence of some social entities does not require explicit recognition. "Exploitation, alienation and many other social phenomena can exist undetected" ([1989], 74). It would be useful, then, to have some general criterion to identify the class of entities for which collective belief is necessary and sufficient for existence. This would give us a more precise characterization of the class of things for which traditional (natural) realism fails.

Amie Thomasson [2003] has proposed such a criterion, using John Searle's [1995, 25-29] notion *institutional fact*. The simplicity of Searle's theory makes it particularly suitable

[2] Cf. Putnam [1975], Boyd [1991]. This way of splitting the doctrine of realism is consistent with the "localist" approach to realism advocated by Mäki [2008] and others.

for introductory purposes, so I will sketch out its main elements in this section and leave further complications for later. Searle's distinction between institutional and "merely" social facts can be cashed out in terms of *collective intentionality* and *collective acceptance*.

For a fact to be social, in Searle's terminology, it must involve a state of *collective intentionality*, where the members of a group perform an activity or assign a function in the "we-mode".[3] For example: if all[4] the members of the Corleone family jointly believe and desire that a member of the Provenzano family should be disposed of, when they plan and execute the killing they are performing a social action, and the ensuing feud itself is a social fact. Beliefs about social entities are thus *performative*, because the existence of social kinds (and the derivative facts) depends on a human activity of maintenance of such kinds and facts.

Some social entities have specific functions. When we say that this is a chair, we collectively believe that this object can (and typically should) be used to sit on. An *artifact*, in Searle's system, is created by the attribution of "agentive functions" to a physical object. We shall speak of "ordinary artifacts" when dealing with objects whose function can be performed purely in virtue of their physical properties. Many interesting social artifacts like money or totem poles, however, perform their function in virtue of characteristics that have been attributed to them by the relevant community and that go well beyond their physical features. (This is why a piece of paper, metal, or a shell can all be used as money in different contexts.) In cases like these, a "status function" has been imposed on a physical object, leading to the creation of "institutional entities" and "institutional facts". Beliefs about institutions are not only performative, but also

[3] The expression is due to Raimo Tuomela [2002a].

[4] I am using universal quantification for simplicity here. In fact this is not quite correct: the quantifier should be weakened to account for the more realistic case in which only a sufficient majority of individuals hold the relevant intentions.

reflexive, in the sense that the reference of an institutional kind is determined by the referential activity itself. [5]

An *institutional fact* exists in virtue of one or more *constitutive rules* of the form "X counts as K in C", where X is a pre-institutional entity[6] (defined according to some independent criterion), K is an institutional kind, and C is a set of conditions that X has to satisfy in order to belong to K.[7] For example: Vito Corleone is a convicted criminal because he has been sentenced to ten years in prison by a judge at the end of a fair trial. Clearly even the most common institutional facts are constituted by layers of other social and institutional facts that support one another, sometimes in fairly complex ways. But the basic structure, as laid out by Searle, is remarkably simple and accounts for much of this complexity in a powerful and elegant fashion.

4. Collective acceptance and infallibilism

According to the collective acceptance approach, for a token entity X to be of kind K it is both necessary and sufficient (i) that possession of some properties C be collectively accepted as sufficient for being K, and (ii) that C obtain. In a single formula:

(CA) X is K $\leftrightarrow$ [CA(X is K if C) & C].

It is useful for analytical purposes to break the (CA) principle in two parts. The first part states that collective acceptance is *necessary* for social kindness:

(CA1) X is K $\rightarrow$ [CA(X is K if C) & C].

[5] Collective intentionality, reflexivity and performativity are standard features of most social ontologies, but a good general discussion can be found in Tuomela [2002a, Ch. 8].

[6] More precisely, X must be pre-institutional *with respect to K*.

[7] Searle is not always clear on this point, and sometimes seems to treat C merely as a domain condition (as in: "notes issued by the Bank of England count as money *in the UK*"). In this paper I will follow other philosophers in taking C to define substantial conditions of kind-membership.

The second part states that collective acceptance is jointly *sufficient*, with the realization of C, for making X an institutional entity of type K.

(CA2) [CA(X is K if C) & C] → X is K.

Each part of the (CA) principle is important, depending on the philosophical point one is trying to make. The sufficiency part of the principle (CA2) has the advantage of distinguishing sharply human from natural kinds. For example, satisfying a set of commonly accepted conditions for being water (being a transparent, odourless and drinkable liquid, say) is not enough for something to be genuine water, if the conditions do not capture water's true nature (being H_2O).

The necessity part (CA1) in contrast has the advantage of allowing the derivation of an epistemic claim (everybody accepts that X is K) from a purely ontological claim (X is K). It is therefore especially relevant for the infallibilist project. Thomasson argues that the following realist theses do not hold in the case of institutional kinds:

> *Extensionality:* "there is a kind with natural boundaries that determine the extension of the term independently of anyone's concept(s) regarding the kind".
> *Error principle:* "since these boundaries are not determined by human beliefs about those boundaries, any beliefs (or principles accepted) regarding the nature of Ks could turn out to be massively wrong".
> *Ignorance principle:* "for all conditions determining the nature of the kind K, it is possible that these remain unknown to everyone" (Thomasson [2003], 583).

Giving up the extension principle is a direct consequence of the Searlean definition of institutional kinds. The boundaries of an institutional kind *must* depend on people's acceptance of a constitutive rule for it to be an institutional kind at all.[8] The error and

[8] Notice that the boundaries themselves may remain unknown to everyone; what cannot be unknown, according to the infallibilist, are the criteria that *determine* such boundaries. Once the criteria have been

ignorance principles seem to follow rather unproblematically, once you go down this route:

> Our acceptance of a set of conditions C as sufficient for being K is constitutive of what conditions suffice for being K, so what conditions there are is constituted by what conditions we accept. As a result, we could not turn out to be mistaken – our acceptance of the set of conditions C declaratively establishes the conditions for being K rather than attempting to describe preexisting and independent conditions for being K. So the Error Principle fails: any conditions we accept as sufficient for the existence of Ks must be free from error [...]. (Thomasson [2003], 588-9)

The failure of the Error principle is not universal however. There are features of social and institutional kinds that may remain opaque to observers and even sometimes to the relevant social actors. This is where the social sciences have important contributions to make. According to Thomasson ([2003], 606), these contributions fall in three categories: (1) the category of social facts that are known only to a small sub-community of insiders within society (a mafia family is a good case in point). (2) The category of phenomena that occur as unintended consequences of a series of collectively accepted social facts (e.g. inflation, racism do not require for their existence the collective acceptance that we are racists or that prices on average are 5% higher than last year). (3) The category of causal relations that hold among social entities and kinds. All other aspects of institutional entities and facts are transparent and are grist for the mill of the infallibilist thesis.[9]

---

accepted, the boundaries are automatically drawn, so to speak, pretty much in the same way as they are determined by the essential properties of a natural kind according to the causal theory of meaning. The main difference is that in the natural realm the boundaries are drawn by nature, whereas in the social realm they are drawn by us.

[9] Ordinary artifacts are also transparent, but I will largely ignore them in this paper: all the arguments concerning infallibilism can be formulated with the case of institutional kinds in mind, and I will stick to them for simplicity throughout the discussion.

5. Infallibilism and empirical discovery

For the sake of clarity, it is important to distinguish infallibilism from the (spurious) claim that we cannot be wrong about *specific instances* of social classification. Infallibilists recognize that the satisfaction of conditions C is an empirical matter, and hence that individual acts of classification can go wrong in various ways. Our natural inductive propensities can make us believe, for example, that X possesses a certain property C in virtue of the fact that it shares another property C* with an entity X* that does have C. We may infer, for example, that Giuseppe Corleone is a mobster because he hangs out with Mafiosi, which is usually a sign of belonging to Cosa Nostra. But such inference would be fallible: when we engage in analogical reasoning of this kind, we can certainly mistakenly classify X as K, where in fact it does not qualify as such.

Another way to put it is that infallibilists do not claim that *specific* issues of institutional classification can be resolved a priori. Whether a token X counts as K or not is a contingent fact, for the possession of C is itself a matter of fact to be established empirically. Infallibilists rather claim that the proposition "X is K if it has properties C" is true a priori. That a piece of paper counts as genuine money in virtue of the fact that it has been issued by the Bank of England is a piece of information that does not require empirical research in the same way as the fact that water is H_2O. It is a fact that holds by mere stipulation (or collective agreement) by the members of the relevant community. Infallibilism is a thesis about our knowledge of *general* facts concerning the *nature* of social kinds. It concerns our knowledge of *what it is*, or what it *means* for something to be a thing of type K, and does not preclude errors in identifying token entities as K-things.

Infallibilists thus challenge a fundamental plank of realism about scientific kinds. According to the causal theory of reference, the nature of scientific kinds is a matter of empirical discovery rather than conceptual stipulation. The process of discovery typically begins with the identification of a *sample* or paradigm – a set of entities that are prima facie interesting and similar enough to warrant their inclusion in a single category.

Empirical investigation then reveals that a set of "essential" properties or mechanisms explains these similarities, as well as the existence of other (accidental) properties and behaviours. Some members of the initial sample may drop out, if investigation reveals that they do not share the essential properties, while others may be added if they do. This way, our knowledge of scientific kinds is progressively refined and modified as the evidence accumulates.

This story is rejected by infallibilists about human kinds. What criteria apply to membership in K, according to the (CA) principle, is a matter of convention rather than empirical discovery. There is no deeper fact of the matter, regarding what K is, than our collective decision to classify in K all the entities that satisfy C. Infallibilists' challenge to the causal theory, if successful, would have deep implications regarding the methods of investigation that are appropriate for social ontology. If infallibilism is correct, then general facts concerning social reality can be known a priori. A competent member of a community needs only examine her own beliefs and linguistic practices, in order to find out what it means for something to be K. This would distinguish sharply social ontological from natural ontological investigations. While natural classification would fall in the domain of science, armchair conceptual analysis would maintain an important role in the realm of the social.

6. Defending infallibilism

The infallibilist thesis is interesting, provocative, and has non-trivial consequences concerning the scope and methods of social ontology. But *pace* Thomasson, the main proponents of the collective acceptance theory do *not* endorse it, and in some cases make explicit pronouncements *against* it. Searle, for example, writes that

> The process of creation of institutional facts may proceed without the participants being conscious that it is happening according to this form. […] In the very evolution of the institution [of, say, money] the participants need not be consciously aware of the form of the collective intentionality by which they are

> imposing functions on objects. In the course of consciously buying, selling, exchanging, etc., they may simply evolve institutional facts. Furthermore, in extreme cases they may accept the imposition of function only because of some related theory, which may not even be true. They may believe that it is money only if it is "backed by gold" or that it is marriage only if it is sanctified by God or that so and so is the king only because he is divinely authorized. [1995, 47-8]

Two different claims are made here, which must be kept distinct. In any case of institutional construction, the relevant actors may be ignorant of:

> (1) The *institutional conditions* C in virtue of which "X counts as K".
>
> (2) The *collective acceptance condition* itself (CA) (i.e. the fact that "We accept that X counts as K if C" as a matter of convention).

In the first part of the paragraph, Searle recognizes that the second type of ignorance is fairly common. This is not, however, a threat to infallibilism. Infallibilism strictly speaking applies only to the *argument* of the CA function in (CA2). As a consequence, it does not require knowledge of the (CA) condition itself, or awareness of the conventional nature of institutional kinds.[10] Infallibilists are only committed to the claim that the relevant actors are aware of the conditions C that make X an instance of K. *Why* C matters may remain unknown to everyone – we may all mistake an institutional kind for a natural one, for example, and erroneously believe that C is a natural, instead of a conventional condition.[11]

---

[10] To put it more formally: infallibilism requires that CA(X counts as K in C), not that CA[CA(X counts as Y in C)].

[11] In fact there is extensive evidence in social psychology that we tend to endow institutional kinds with characteristics that are typical of natural kinds. We tend, for example, to posit the existence of essential properties shared by all members of a kind, to exclude that one individual can belong to more than one social category at the same time, to seek for superficial patterns of association in a population of individuals and to interpret them as evidence of "groupness". Finally, we tend to believe that kinds and their boundaries are more rigid and less historically persistent than they actually are (see e.g. Campbell

But while the first part of Searle's paragraph is consistent with infallibilism, the second one is not. Here Searle says explicitly that we can be wrong about the very conditions (C) that define the nature of K. In a similar vein, Tuomela states that collective intentionality theory does not "require anything […] about the participating agents understanding of anything" [2002b, 421]. Infallibilists like Thomasson are pushing collective acceptance theory well beyond its original boundaries.

The caution of Searle and Tuomela is motivated in part by their desire to hedge collective acceptance theory from cheap counterexamples. Clearly many social facts are not *consciously* accepted as such – at least by most of us, most of the time – so the very notion of collective acceptance must be formulated in such a way as to account for this fact. Collective acceptance must be turned into a technical concept that is somewhat weaker than its everyday counterpart. Such a strategy is relatively harmless in the context of ontological analysis – where our ultimate goal is to define the nature of social kindness, quite independently of its epistemological consequences. It may be problematic, however, for the infallibilist project: if we use ontological analysis in the service of epistemology, the ontological notions we employ must be rich enough to allow the derivation of *interesting* epistemic claims.

We need some preliminary criterion then to demarcate interesting from uninteresting versions of the infallibilist thesis. The following two conditions seem appropriate: first, an acceptable version of infallibilism about human kinds should not be so weak as to trivialize violations of the Error principle. In particular, it should preserve an interesting sense for the claim that no mistakes and learning regarding social kinds can take place. Secondly, in rejecting the Error principle we must be careful not to give up the "core" features of human kinds. Among these features, collective acceptance, reflexivity, and

---

[1958], Sigel et al. [1967], Yzerbit et al. [2001]; Rothbart and Taylor [1992] provide a survey and general discussion). The "biologization" of social concepts is a well-known manifestation of such biases, which has been documented extensively in the cases of gender, race, and mental illness (see Hacking [2002] for historical examples and philosophical analysis).

performativity are of primary importance. But *projectibility* will also play an important role in the arguments that follow. A kind is projectible, in Goodman's sense, if it can be used to predict with a reasonable degree of accuracy the properties and behaviours of individual entities that will be encountered in the future. Given the role played by human kinds in coordinating social behaviour, projectibility is clearly an important feature from the actors' as well as from the observer's point of view. [12]

7. Expert knowledge

Armed with these preliminary criteria, let us examine the technical notion of collective acceptance, as understood by the main proponents of the theory. Collective acceptance theorists to begin with do not require that *every* member of a social group explicitly accepts the conditions C that make X an instance of K. Lay people for example do not know exactly what conditions must be satisfied for a certain individual (Mary Elizabeth Windsor) to be the Queen of England. Not only most people have no idea what the relevant conditions exactly are in cases like this; as Searle points out, they probably have *false* beliefs about them. (They tend to identify "blue blood", lineage, and similar features as necessary and sufficient – whereas in fact none of them is).

In fact decisions concerning classificatory issues are often delegated to experts, who are believed to have privileged access to the properties that determine social kind attribution (cf. Rothbart and Taylor 1992). A satisfactory theory thus must account for the role of experts in determining the nature and extension of human kinds. Tuomela's distinction between "operative" and "non-operative" members of a social group is introduced for this purpose.

> It must be emphasized that in the case of developed societies there is division of labor also with respect to collective acceptance. Thus not all members of the

[12] This point is often overlooked by theories that focus exclusively on the synchronic aspects of social ontology. Major exceptions are Lewis [1969], Barnes [1983] and Sugden [1998]. I will come back to this important point below.

> community actually need to know all the details and may not even have heard of the [institution] in question [for that institution to exist]. [2002a, 200]
>
> […] in such collectives it is the operative members for decision who decide what will be money, for instance. In this realistic cases the other, non-operative members only need to tacitly accept what the operative members have decided. [2002b, 427]

This move preserves a principled distinction between correct and incorrect applications of a concept (or correct and incorrect attributions of kind-membership). The distinction between, say, a legitimate and an illegitimate sovereign may be entirely independent of lay people's beliefs about whether X is really the Queen or not, for it is even independent of whether they know (or agree about) the institutional conditions C that apply in this case. All they have to (tacitly) agree upon is that there is a procedure for resolving such matters, and perhaps that certain experts know how to resolve it. Once this is agreed, the "ontological consequences" of this fundamental collective acceptance cascade down, so to speak, to whatever logically follows from it.

In such cases knowledge of institutional kinds is to be interpreted as "social knowledge" or "expert knowledge", knowledge that is stored somewhere in society, perhaps readily accessible only to a minority of experts. This is not a far-fetched suggestion: we speak of knowledge in this sense when we say that "Great Britain knows how to build the atomic bomb", even though the vast majority of British citizens have no idea how it can be done; or when we say that "21st Century mathematicians know how to demonstrate Fermat's Theorem", even though perhaps only a handful of people are able to understand and reproduce the proof. Under this interpretation what particular individuals (even the majority of individuals) believe is rather irrelevant – they can be systematically wrong and yet infallibilism true, because *as a society* we cannot be wrong about institutional kinds.

8. Dispositions

Even experts' acceptance, however, should not always be taken literally. The conventionality of social entities is certainly something we do not pay much attention to, for most purposes, in our everyday lives. We have a tendency to treat social kinds in a naively realist fashion, and we engage with the conventional nature of social reality only occasionally, upon reflection. To account for this fact most collective acceptance theorists give up the requirement that collective acceptance must be constantly operative.

A large part of Searle's [1995] *Construction of Social Reality* is devoted to the concept of the "Background" – a set of mechanisms and dispositions that subconsciously and automatically support our social practices without requiring full cognitive engagement with the logical presuppositions of such practices. Similarly, Tuomela ([2002a], [2002b]) posits the existence of "virtual" mechanisms that *would* bring individuals' behaviour back in line, were certain disturbing factors to disrupt the regularity of a social practice:[13]

> The we-attitudes need to be respected in the various institutional activities undertaken in the institution, but, being dispositional states, they need not be made occurrent and reflected upon in normal circumstances but only in cases of institutional breakdown (or something analogous). [Tuomela 2002b, 426]

The dispositional view preserves an interesting meaning for the denial of the Error principle that is a central plank of the infallibilist position. Even though the relevant actors are not constantly entertaining the thought that X is K if it satisfies conditions C, they can nevertheless retrieve such thought in the appropriate circumstances.

These amendments to the notion of collective acceptance are still consistent with a non-vacuous infallibilist thesis. Even though not all members of a group know that X is K in virtue of having C, some of them – the "experts" – do. And even though perhaps no one

[13] Pettit [1996] gives an account of the explanatory role of such "virtual mechanisms" in biology and social science.

is always consciously aware of its conventional, institutional character, those who are in charge of determining and recognizing K-ness are able to form the appropriate intentions when required.

9. Witches

Searle [1995, 47] however mentions "extreme" cases, where the attribution of K depends on some false theory that is universally endorsed by the members of a community. Consider a classic case of social construction. Vernacular theories of witchcraft in the Medieval and Renaissance period tended to follow a standard narrative structure, centred on the so-called "pact with the devil".[14] For simplicity, let us assume that the pact is generally considered a sufficient condition (C) for being classified as a witch (W). Clearly, X is W if she fulfils the conditions that the members of her community take as sufficient for being W. But it surely would be bizarre to say that believers in witchcraft could not be wrong about W. Witchcraft practices were sustained by hugely mistaken beliefs concerning the nature of W, and the actors (including the relevant experts) lacked a disposition to retrieve the *real* conditions of W-ness upon disruption of that practice.[15]

One complication is that the pact with the devil was supposed to be a real but unobservable event. Because of its very nature, then, C had to be inferred from other signs and clues. Let us call the evidence that was used to build a witchcraft accusation (a confession, testimony, or mark) E. In reality, of course, the conditions for X to count as W were E: a suspect in a trial was classified as a witch if she satisfied the evidential requirements set by the jury. No further condition could be fulfilled, for it was impossible

[14] An old woman meets a man in black, who offers help and money (the money later usually turns into leaves or grass). There is intercourse, and the devil gives the woman a substance (typically powder) that can be used to harm other people. The victim sometimes tries to resist, but the devil forces her to carry on with her crimes (see e.g. Briggs [1996]).

[15] In fact awareness of the real nature of witchcraft institutions would likely contribute to their demise. That's why enlightenment attacks against superstition typically took the form of exposing alleged natural (or "supernatural") kinds for what they are – a social construction that does not capture any underlying reality.

for C (the pact with the devil) to have occurred. But the actors were unaware of this, and believed that X was W if C was satisfied. To say that "any conditions we accept as sufficient for the existence of K must be free from error" seems misleading at best.

Examples of this kind can be dealt with in three different ways. First, (a) one may deny that W is an institutional kind at all. Since infallibilism applies mainly to institutional kinds, it would be unaffected by cases like this. Alternatively, (b) one could accept that W is an institutional kind, thus saving infallibilism, but arguing that collective acceptance must be interpreted in a stronger non-cognitivist form than Searle's and Tuomela's. Or finally, (c) one could reject infallibilism about institutional kinds altogether.

Strategy (a) is not very promising in my view, for W ticks all the boxes of institutional kindness. Witchcraft had a number of social functions that could not be performed simply in virtue of the physical properties of the entities (the women) who were classified as witches. Moreover, the collective acceptance of X as W relied on the specification of a set of conditions that X had to satisfy to count as a "real" witch.

The second strategy (b) is to retreat to a strong non-cognitivist interpretation of collective acceptance: while believers in witchcraft did not explicitly accept the 'right' conditions for W-ness, one could say that they did so *in practice*. Thomasson [2003] notices that social life typically consists of a series of practices rather than conscious decisions based on explicit belief-desire deliberation. Accordingly she suggests that the infallibilist thesis should be translated in a strong non-cognitive mode:

> Some might argue that in fact we seldom have explicit cognitive awareness of the relevant principles for institutional kind membership, we just have the practice of accepting certain sorts of things and rejecting others as putative kind members. I have spoken of the acceptance of principles in order to make the logical relations clearer, but the basic points can be made in a less explicit cognitivist scheme. The result in that case would be that even if (on a realist view) certain kinds of massive error in treating entities as members of a certain kind are possible for natural kinds

> (e.g. treating whales as fish), the same is not true for practices involving institutional kinds (e.g. treating cowry shells as money). [2003, 590, note 12]

People may well be ignorant about the conditions that truly make X an entity of type K, and yet *treat* X exactly as a K-type entity. Infallibilism is turned into a thesis about *what we do* rather than about our explicit cognitive states: when the infallibilist says that we cannot fail to know that X is K, she means "knowing how" rather than "knowing that". "Collective acceptance" becomes a theoretical concept, a state that is imputed by a theorist who observes a certain regularity of behaviour.[16] Paraphrasing Dennett, we may say that the theorist takes the "*collective intentional stance*" toward a certain community and the behaviour of its members.

Does such a move take all the interesting content away from infallibilism? Infallibilism now boils down to the claim that the members of a community are competent speakers, as far as social categories are concerned. This effectively collapses the semantic and the epistemic challenges to realism: there is a boundary to K because *de facto* we label some entities, but not others, as K – even though we may not know why. If you speak a language correctly, then of course you know (tacitly) what is to be named K and what is not – where "correct" is just what people in the community happen to call K. Obviously this non-cognitivist version of infallibilism does not have any particular implications for the social sciences, which pursue knowledge in the sense of "knowing that", rather than "knowing how". But also from the viewpoint of the actors, it drastically reduces the significance of the Error principle.

Which leaves us with strategy (c). One can, of course, interpret acceptance in a strong non-cognitivist fashion – as Searle [1995], for example, does. But at the same time one must refrain from deriving any substantial epistemic claim from the collective acceptance theory of social ontology. The collective acceptance doctrine, in Tuomela's words, has no implications regarding "the participating agents understanding of anything". This in turn

[16] Notice that the theoretical concept does not have to be a mere fiction: it may well have a referent, namely the behaviour of the social actors.

implies that the infallibilist project should be abandoned. The best way to account for the existence of institutions, like witchcraft, that rely on massively mistaken beliefs about their central categories, is to recognize that there are institutional facts about which we can be massively wrong.

10. Money

The case of witchcraft is in some ways peculiar, for it involves the blatant naturalization (or, we should say, "super-naturalization") of a human kind. It instantiates, however, many typical I shall examine a paradigmatic example of social construction, which plays a prominent role in current debates in this field. This will help introducing another important issue – the projectibility of institutional kinds – that sits uneasily with the infallibilist stance.

Institutions help us coordinating. To fulfil this crucial function (perhaps *the* function they have evolved for) they require a certain amount of stability. Projectibility is therefore important both for external observers (for the predictions and generalizations of social science) and for the actors (who seek coordination). In spite of all this, theorizing in the collective acceptance tradition has a remarkable synchronic bias. Once a concern for the temporal dimension is introduced, our outlook on some philosophical issues – including infallibilism – changes quite dramatically.

The case of money has become an undisputed classic in social ontology. Money is usually cited as a paradigmatic social entity endowed with functions that are completely independent of its physical characteristics. (Coins, shells, cigarettes, furs, can all be used as money in different societies and in different contexts.) Unfortunately the case of money tends to be discussed in a highly idealized fashion by philosophers – abstracting from many details that obfuscate its nature. But it is partly by ignoring these details that theses like infallibilism are erroneously derived from the collective acceptance model of social reality.

Economists subscribe to the principle that "money is what money does". What counts as money does not depend merely on the collective acceptance of some things as money, but on the causal properties of whatever entities perform money-like functions, regardless of whether they are classified as belonging to the same folk categories or not (cheques, debit cards, bank accounts are all money, in economists' sense).[17] What these properties are, exactly, is far from obvious however, and certainly not a matter of arbitrary stipulation. Kind membership is largely a process of discovery, rather than stipulation.

Money performs several functions at once. Monetary theories usually begin by identifying three principal functions of money, as (1) medium of exchange, (2) store of value, and (3) unit of accounting. Economic theory identifies (1) as the core function of money: a currency, in order to be money, must first and foremost be used as a medium of exchange. The second function (store of value), however, is strictly linked with it: "if an asset were not a store of value, then it would not be used as a medium of exchange" (Dornbusch and Fischer [1994, 374]). The reason is simple: exchanges take place *in time*. Selling a good now with the aim of purchasing something else in the future makes sense only if the revenue from the first trade does not lose its value during the time it takes to effect the second one.

Being a store of value, then, is an important precondition for X (a currency) to be M (money). But what sort of condition is it? It would certainly be inappropriate to include it in the constitutive rule "X counts as M in C". The "store of value" condition is not *logically* or *conceptually* connected with money-hood. The connection is *causal*, and not a matter of arbitrary stipulation like the relation between being money and being issued by the Bank of England. It is a condition that holds a posteriori, in virtue of the way the world is.

This helps explaining some of the folk theories that are often associated with money. In most societies throughout history, the legitimacy of a currency was backed up by belief in

[17] This is one reason why measuring and controlling the quantity of money circulating in an economy is such a difficult and controversial scientific task, by the way.

the currency's linkage with some underlying goods. The statement "I promise to pay the bearer on demand the sum of 10 pounds" is still written on Sterling notes, and signed by the Chief Cashier of the Bank of England. It is in fact the relic of an age when currency was backed up by solid commodities rather than merely by collective beliefs or expectations.[18] To say that a commodity interpretation plays a crucial role in sustaining the institution of money, to be sure, does not mean that beliefs in the commodity-value of money must be true. Even in the commodity age the value of money did not always track the value of the underlying goods (e.g. silver, or gold).[19] The commodification of money nevertheless provides intuitive anchoring and hence projective stability to what could otherwise be dangerously perceived as a "mere" convention. Collective acceptance is too slender a foundation for an all-important institution such as money. At the same time, the commodity interpretation facilitates the representation of a system of institutions and mechanisms that are too complicated to be captured by any simple, cognitively manageable model (cf. Mäki [2004], 16).

Although in the post-gold standard era money officially lacks such an anchoring, folk theories of money still inform lay people's understanding of this fundamental and ubiquitous institution. The value of a given currency is generally perceived as linked with the soundness of a whole nation's economic and political system. The appraisal of this soundness, in turn, is partly delegated to experts who apply tests and criteria derived from scientific economic theories. These theories, however, can be wrong. Indeed, like all scientific theories they probably *are* wrong, to some extent. Knowledge of the nature of money is partly causal, a posteriori, and fallible.

[18] This age – the "commodity age" – accounts for the overwhelming majority of monetary history (Britain abandoned the gold standard only in 1931).

[19] Historians of debasement have shown that currencies could preserve their value even when the link with the commodity was eroded (provided it was not *too* blatantly eroded – notice the importance of cognition here).

11. Time

The example of money is not an odd curiosity. Its implications can be easily generalized to many other institutional kinds. Of course if we all believe that X is K at time t, then X *is* K at t. But this sort of synchronic analysis is of limited use, because most social institutions are supposed to endure in time. Institutions must coordinate behaviour over relatively long periods, so we are rarely interested in X being K at t only. But it is precisely on diachronic matters that we are most likely to go wrong, because grasping the causal mechanisms that govern social change (and change in beliefs) can be rather difficult.

Suppose we all believe that a certain bank is sound, to use a classic Mertonian example. Although that perhaps makes it sound *now*, it cannot make the bank sound in the future, as if by magic.[20] But soundness is intrinsically a *diachronic* concept, for it incorporates the idea that an institution can be trusted – and trust surely is future-oriented. The same applies to concepts such as "police", "government", or "money" that involve the exercise of some sort of power. All power-related concepts are future-oriented, and in order to be so oriented must rely on mechanisms that guarantee certain consequences for certain classes of acts or events.

These mechanisms must be correctly understood for these concepts to be correctly applied. And there is no guarantee that we possess enough knowledge to do it properly. Indeed, what these mechanisms are is largely a matter of discovery. Because knowledge of institutions is a posteriori, we can all be wrong regarding the nature of institutional kinds.

[20] I say "perhaps" because should the bank fail a month later, there is an intuitive sense in which we should retrospectively admit that it was not sound earlier – when we all believed that it was. Our judgment of soundness was *wrong*, regardless of what was commonly believed back then. Since these questions are, however, more terminological than substantial, I will not put too much weight on them.

12. Conclusion

Social ontology relies heavily on conceptual analysis, often at the expense of empirical research. Many historical, sociological, and philosophical factors explain this state of affairs. One is that social science is generally considered a weak source of knowledge. As a consequence philosophers have for a long time treated social ontology as a subfield of the philosophy of language or the philosophy of mind, where important insights can be obtained independently of (sometimes in spite of) scientific research on its subject matter. Infallibilism in many ways continues this tradition. By rejecting the causal theory of reference, infallibilists broaden drastically the range of facts that can be known purely a priori.

The failure of this project is a reminder that scientific methods of investigation remain our main and best source of information regarding what there is – both in the natural and in the social realm. Overall, I suspect that infallibilists like Thomasson are simply making bad use of a good philosophical theory. The theory of collective acceptance is rich of insights concerning the structure of society. It is, however, best interpreted as an ideal type or idealized model like those found in science. While such models capture important aspects of reality, they rarely provide an entirely accurate description of entities in the real world. This does not mean that ideal types are useless. Quite the contrary, by observing the deviations of concrete phenomena from their idealized representations we can learn a lot about the way the world is. Theoretical analysis with empirical research is our best recipe for success, in social ontology as in social science.

References:

Barnes, S.B. [1983] “Social Life as Bootstrapped Induction”, *Sociology* 17: 524-545.

Bloor, D. [1997] *Wittgenstein, Rules and Institutions*. London: Routledge.

Boyd, R. [1991] “Realism, Anti-foundationalism, and the Enthusiasm for Natural Kinds”, *Philosophical Studies* 61: 127-148.

Briggs, R. [1996] *Witches and Neighbors: The Social and Cultural Context of European Witchcraft*. London: Penguin.

Campbell, D.T. [1958] “Common Fate, Similarity, and Other Indices of the Status of Aggregates of Persons as Social Entities”, *Behavioral Sciences* 3: 14-25.

Dornbusch, R. and Fischer, S. [1994] *Macroeconomics.* New York: McGraw-Hill, 6th edition.

Gilbert, M. [1989] *On Social Facts*. Princeton: Princeton University Press.

Hacking, I. [2002] *Historical Ontology*. Cambridge, Mass.: Harvard University Press.

Hayek, F.A. [1943] “The Facts of the Social Sciences”, *Ethics* 54: 1-13.

Lewis, D. [1969] *Convention: A Philosophical Study*. Oxford: Blackwell.

Mäki, U. [2004] “Reflections on the Ontology of Money”, unpublished paper, University of Helsinki.

Mäki, U. [2008] "Putnam’s Realisms: A View from the Social Sciences", in *Approaching Truth: Essays in Honour of Ikka Niiniluoto,* edited by S. Pihlström, P. Raatikainen and M. Sintonen. London: College Publications.

Pettit, P. [1996] “Functional Explanation and Virtual Selection”, *British Journal for the Philosophy of Science*, 47: 291-302.

Putnam, H. [1975] “The Meaning of ‘Meaning’”, in *Mind, Language and Reality. Philosophical Papers, Vol. 2*. Cambridge: Cambridge University Press.

Rothbart, M. and M. Taylor [1992] “Category Labels and Social Reality: Do We View Social Categories as Natural Kinds?”, in *Language, Interaction and Social Cognition*, edited by G.R. Semin and K. Fiedler. London: Sage.

Ruben, D. [1989] “Realism in the Social Sciences”, in *Dismantling Truth*, edited by H. Lawson and L. Appignanesi. London: Weidenfeld and Nicolson, pp. 58-75.

Searle, J. [1995] *The Construction of Social Reality*. London: Penguin.

Sigel, I.E., E. Saltz, and W. Roskind [1967] “Variables Determining Concept Conservation in Children”, *Journal of Experimental Social Psychology* 74: 471-475.

Sugden, R. [1998] “The Role of Inductive Reasoning in the Evolution of Conventions”, *Law and Philosophy* 17: 377-410.

Tajfel, H. [1970] “Experiments in Intergroup Discrimination”, *Scientific American* 223: 96-102.

Thomasson, A. [2003] “Realism and Human Kinds”, *Philosophy and Phenomenological Research* 68: 580-609.

Tuomela, R. [2002a] *The Philosophy of Social Practices*. Cambridge: Cambridge University Press.

Tuomela, R. [2002b] “Reply to Critics” in G. Meggle (ed.) *Social Facts & Collective Intentionality*, Frankfurt: Hänsel-Hohenhausen AG, pp. 419-436.

Vico, G. [1744] *La scienza nuova*. Roma: Laterza, 1974.

Yzerbyt, V., O. Corneille, and C. Estrada [2001] “The Interplay of Subjective Essentialism and Entitativity in the Formation of Stereotypes”, *Personality and Social Psychology Review* 5: 141-155.