Please cite the final version, which appeared in Synthese.

Developing the Incentivized Action View of Institutional Reality

JP Smit, Filip Buekens and Stan du Plessis

Abstract

Contemporary discussion concerning institutions focus on, and mostly accept, the Searlean view that institutional objects, i.e. money, borders and the like, exist in virtue of the fact that we collectively represent them as existing. A dissenting note has been sounded by Smit et. al. (2011), who proposed the *incentivised action* view of institutional objects. On the incentivised action view, understanding a specific institution is a matter of understanding the specific actions that are associated with the institution and how we are incentivized to perform these actions. In this paper we develop the incentivised action view by extending it to institutions like property, promises and complex financial organisations like companies. We also highlight exactly how the incentivised action view differs from the Searlean view, discuss the method appropriate to such study and discuss some of the virtues of the incentivised action view.

1. Searle - The collective acceptance of status functions

John Searle (1995, 2010) has claimed that it is constitutive of the existence of an institution that we *collectively* accept that a (type of) object has a certain *status-function.* The problem that Searle tries to solve arises from the fact that certain objects in the world, typically money, borders, driver's licences and the like, perform social roles that cannot be understood purely in virtue of the physical properties of these objects. Since money is, in a sense, just pieces of paper, borders are lines that need not be intrinsically difficult to cross, and driver's licences are typically just bits of laminated paper, the question arises: how do these objects manage to do what they do, i.e. serve as a medium of exchange, serve as national boundaries and certify one's ability to drive?

Searle's ingenious answer to this question is that these objects 'do what they do' in virtue of the fact that we collectively regard them as having certain status-functions. In this way we collectively regard certain bits of paper as being a medium of exchange, we collectively regard certain lines as being borders, and we collectively regard the laminated bits of paper as certifying that we are legally allowed to drive. This means that such objects can, for the most part, be thought of as having the logical structure *X* counts as *Y* in *C*. The *X*-term is the non-institutional characterisation of the object, the *Y*-term is the institutional characterisation of the object, and the *C* term refers to the context in which the object counts as having such social properties. Consider, for example, a piece of paper currency. It would seem to fit Searle's formula as it is a *piece of paper* that counts as *currency* in contexts where it can be used to *effect trade*.

Searle views institutional facts as constituting a "huge, invisible ontology" (2005:1) that cannot be fully described without employing terminology that is irreducibly institutional. We think that this *irreducibility claim* is too strong. One way of seeing the problem that we have with Searle's irreducibility-thesis is to ask how one would go about defining a notion like money. In the Searlean theory, something is money if we collectively regard it as being money. Such a definition does not get us very far, as it does not tell us what we are regarding something *as* if we regard it as money. Searle acknowledges this, and states that we may try to define it in terms of other notions, but that we will not be able to define it in terms of notions that are not, themselves, institutional (see Searle, 1995: 32, compare Ruben 1997). This traps us in a sort of definitional circle where even ostensive definitions would seem to be of no help, as all parties agree that no physical property of dollar bills can serve to capture the social role of money. Searle denies that there is anything inherently problematic about being caught in such a circle (1995: 52). We do not share his confidence in this matter.

In contrast to the above, the *incentivised action* account of institutional reality, proposed in Smit et. al. (2011), avoids such a definitional circle. On the incentivised action view it is claimed that institutional facts can be individuated, and understood completely, in terms of certain types of *actions,* namely the ones that we are incentivized to perform when confronted with some institutional object. The basic

idea is that, where an institutional fact exists, this is always a matter of *some people being incentivized, not merely by the brute nature of the relevant object, but by human agency (punishments, rewards) or moral belief, to perform some specifiable action or set of actions.* If institutional reality can be captured in this manner, then, provided we can characterize the relevant actions in non-institutional terms, we can similarly characterise the institutional facts in such terms. Hence we can reject Searle's idea of the irreducibility of institutional reality and the very idea of objects that exist in virtue of being collectively represented as existing.

In this paper we develop the incentivised action view by, firstly, defining the notion of ownership and the notion of a promise in terms of incentivised action. This is done in order to show that the incentivised action view can handle radically different types of institutions and as these notions are useful in achieving our second objective, namely showing that the incentivised action view can, in principle, explain complex financial entities like limited liability companies, etc. We then highlight the 'bottom-up' nature of the incentivised action view, discuss the method appropriate to such study, and some of the virtues of thinking of institutional reality in terms of incentivised action. We start with a brief explanation of the incentivised action view.

2. The incentivized action view of institutional facts

2.1 Institutional objects and artefacts

In Smit et. al. (2011) the notion of traffic light is defined in terms of the fact that we are incentivised to coordinate our driving behaviour (typically by performing actions like stopping on red, going on green, etc.) with a certain set of lights (2011: 5); the notion of a border is defined as a line, whether concrete or abstract, that divides two areas that we are differently incentivised towards (2011: 11); and the notion of money is defined as an object, either abstract or concrete, that we are incentivised to acquire for exchange (2011: 11 – 20). One aspect of the above definitions should immediately be noted, namely that, while the above definitions require that people be somehow incentivised to perform the relevant actions, *the source and nature of the incentivization are not individuating facts.* In other words, the above definitions allow for traffic lights, borders and money to come about in a variety of ways and for a variety of reasons. This means that, while all three of the above institutions are

currently typically brought about by grand-scale institutions like states, this is not definitionally required, and so the ontological nature of these objects can be characterised without requiring a prior analysis of institutions like the state.

The guiding idea behind the incentivised action view is that the strategy of defining institutional objects in terms of incentivised action(s) can be generalised to all institutional objects. Hence the key claim behind the incentivised action view is that, where institutional objects are concerned, it will *always* turn out that, due to human action or moral belief, such 'objects' give some relevant subjects a reason to act in various ways and that, without the things providing such reasons, we would not say that the relevant institutional object exists. *This is not a trivial claim.* Borders that no-one pays any reason to heed will eventually disappear off the map *qua* borders, bits of paper that no-one has any reason to trade will eventually not be considered money anymore, and so on, but this is not true of most ordinary objects; rocks that no-one throws will forever be rocks and numbers that are never used in computation will forever be numbers. We take it that this is the crucial clue concerning the nature of institutions, as it suggests that we can individuate institutional facts by the associated actions[1].

An intuitive way of explaining the incentivised action view is by contrasting institutional objects with another class of objects that also feature a close link between the nature of the object and the existence of a reason for action, namely ordinary artefacts like chairs, houses and hamburgers. Chairs are things that we generally have a reason to sit on as they offer us a good resting place, houses are things that we generally have a reason to live in as they offer shelter and hamburgers are things that we generally have a reason to eat as they offer nourishment. In the case of such artefacts, the relevant 'incentivization' is accomplished in virtue of physical properties of the objects, delicately coupled with and attuned to certain basic facts about human psychology and physiology.

[1] Searle has, upon occasion, made remarks that similarly indicate a close connection between institutional objects and actions. Consider: "Social objects are always... constituted by social acts; and, in a sense, *the object is just the continuous possibility of the activity*" (1994: 36). Or: "What we think of as social *objects... are in fact placeholders for patterns of activities"* (1995: 57). We think that a deep understanding of these points should lead one to reject the collective acceptance view in favour of the incentivised action view.

Institutional objects differ from ordinary artefacts in that the latter do not provide a reason for action solely in virtue of our psychological properties, our physiological properties and the intrinsic properties of the object. A line drawn in the sand, while well-suited to being a border in ways that a block of gold, or lights on top of a pole are not, does not, even in conjunction with our physiology and psychology, serve to give us a reason to not cross it. The incentivization done by the line is incomplete; the full incentivization *needs to be brought about* by some human action, or moral belief, that serves to incentivize us not to cross the line in the sand. This can be done by some credible threat about what will happen if the line is crossed, or due to the belief that crossing the line without permission is immoral, or some mixture of moral and prudential motivation. Once this is done the incentivization is sufficient for the line to become a border. But, and this is the important point, while the incentivization of the incentivized *is brought about* by an incentivizer independently of the nature of the object, this is the *only* relevant difference between an artefact and an institutional object. It would be profoundly odd to talk of the reason for action that derives from the nature of hamburgers and chairs as showing that these objects are constitutively imbued with 'deontic powers'. On the incentivised action view it is equally odd to say, just because the incentivization in question is partly brought about by an incentivizer, that money and borders are constitutively imbued with deontic powers.

2.2 Rules of the game and clarifications

We use the term 'institutional fact' as it is used by Searle, i.e. as a technical term to refer to phenomena that are essentially similar to these paradigm cases, i.e. that pre-theoretically fit the *'X* counts as *Y* in *C'* characterization. Note that Searle does not use 'institutional fact' as a catch-all category to cover all objects that are not obviously physical, or that involve human practice, or that the ordinary term 'institution' applies to. For example, Searle treats an abstract line[2] like the sequence of extensionless points between Colorado and Utah as a brute fact (2003:308), treats a piece of music as a non-institutional fact (2003: 305) and does not view science, religion or education as institutions (2005: 19; 2010: 92). Hence we take it that our

[2] Any account of institutional reality must presuppose notions like abstract objects; it would be very odd to deny the theorist use of numbers when explaining social security numbers, winning lottery numbers, etc.

account can avail itself of things like abstract objects, ordinary artefacts, actions, beliefs and other types of things that do not obviously fit the 'X counts as Y in C' characterisation and that he has never claimed to be institutional in the relevant sense. Note that such things include moral norms and beliefs; Searle has never indicated that either a norm like 'thou shalt not kill', or the belief that one should not kill, are institutional in the relevant sense. Simply put, we feel free to avail ourselves of things that philosophers are committed to accounting for *anyway*, in virtue of non-institutional reality, i.e. things like abstract objects, artefacts, actions, abilities, beliefs and so on.

We do not take a stand on the nature of collective intentionality. Searle claims that the sense of collective acceptance used in his theory is not something that can be reduced to a matter of a group of individuals who individually accept that some object is to count as having a certain status-function (2010: 45-48). We doubt this claim, but will not focus on it, beyond giving examples of cases where institutional facts are created without any collective acceptance[3]. Our problem is with Searle's strong irreducibility-thesis, not with collective intentionality.

Note that we do not in any sense deny that institutional facts and objects exist. This is emphatically not the case; everyone agrees that money, borders and limited liability companies are real. We simply claim that for institutional reality to exist just *is* for ordinary bits of non-institutional reality, individuated by a non-intrinsic property of such bits of non-institutional reality, namely some incentivized action relevant to it, to exist. In this way the same mountain range can be a border at one time and not be a border at another time, without this being due to any physical change intrinsic to the mountain range, but due to the relevant incentives no longer obtaining. The border, however, *qua* institutional object, is no more than the mountain range, provided that the relevant incentives obtain. Note that it is not a peculiarity of institutional facts that non-intrinsic properties may be individuating properties. Consider footprints[4]; if I see a footprint in the sand and scratch out an exact replica next to it, the replica is still not a footprint. It is essential to a footprint that it must be the mark that has been left by a foot descending into the sand.

[3] For a non-Searlean view that is mostly congenial to our way of looking at things, see Tieffenbach (2010).

[4] This example adapted from Stalnaker (1989: 289).

Footprints are individuated by their causal histories, despite such causal histories not being intrinsic properties of footprints.

Note that our talk about incentives is not supposed to indicate any allegiance to the idea that people are rational actors in the economic sense, or that we exhaustively calculate the pros and cons of doing something before acting. Also note that we are not cynics who think that all apparently altruistic action is actually based on self-interest or anything of the sort. Rather our talk of incentives reflects the fact that human action is motivated, i.e. based on *reasons*. These reasons come in many kinds; people may act for reasons that are selfish, altruistic, self-regarding or other-regarding, moral or prudential, and so on. When we say that someone is incentivized to perform an action we merely mean that there is. For that person, some reason for action, whatever this may be.

3. Ownership and exchange

3.1 Ownership defined

Smit et al. (2011) defined money as *an object that we are incentivized to acquire for exchange*, rather than for direct consumption. Note that we can be incentivized to acquire an object for exchange independently of whether this incentivization is due to the existence of other institutions like banks or shops. This shows that such institutions are not *constitutively* required for something to be money. The matter of 'being incentivized to acquire for exchange' is supposed to capture what fiat currency, gold currencies, tins of mackerel, shells, large stones, virtual currencies like bitcoins[5], cigarettes, etc. have in common in cases where they 'function as money'[6]. Under the appropriate circumstances the incentives that operate in a society can be such that a significant amount of people have reason to acquire object from a specific class for exchange, rather than consumption. If this institution is stable, i.e. enough people have reason to acquire the objects for exchange, and over a long enough time-period, we express this fact by saying that such objects are 'money'.

[5] A non-state issued, wholly electronic currency. (www.bitcoin.org).

[6] A further problem concerning money is the so-called problem of 'free-standing *Y*-terms', first raised by Barry Smith in his exchange with Searle (2003). The total number of notes and coins in circulation only amount to a fraction of the actual money in existence. This raises the question: where is the thing that Searle's *Y*-term is supposed to refer to? In Smit at al. (2011: 17-20), we defend the view that, where fiat money is concerned, the actual 'thing' traded is the abstract ability to acquire goods and services.

'Money', above, was defined in terms of 'exchange'. 'Exchange' is itself an institutional notion, and so the proponent of the incentivised action view owes a non-Searlean account of 'exchange'. As 'exchange' can be defined as 'a transfer of ownership', we will do so by defining the notion of 'ownership'. The general strategy, as explained above, is to identify institutional facts with what they give us reason to do, i.e. with a certain simple or complex incentivization structure. In general, and ignoring some details that depend on the vagaries of various legal systems[7], owning something means that the owner can use the object without third parties interfering with such use. Such unencumbered use comes about as third parties are *disincentivized* to interfere with the owner's use of the object. Hence we define ownership in the following way: *a subject, S, owns an object, o, if, and only if, third parties are disincentivized to interfere with S's use of o.* Practically speaking, this means that, when I say I own my pen, this means that I can write with it, use it to help balance an uneven table or chew on it, and that third parties are disincentivized to interfere with any of this. This incentivization is accomplished by moral beliefs about private property, the informal social norms governing such interference (unauthorised use, theft, etc.) and the police's willingness and ability to punish physical interference with how I use my pen.

Note that ownership can, in principle, be sustained by radically different types of incentives. This can be seen by constructing various 'Robinson Crusoe - style' cases in which some or other source of incentives is completely absent. Consider two people stranded on an island, where one is the selfish and powerful Jack. If Jack declares that he will use any object in any way he sees fit and that he recognises no constraint on his own action, then, provided Jack has the (brute) power to make good on such a claim (i.e. no-one can meaningfully disincentivize Jack from performing such actions) then he owns, in any conceivably relevant sense, everything on the island. Or consider the case of two people with sincerely held moral beliefs about private property who wash up on an island. If both are committed to the claim that, if anyone 'mixes their labour' with a natural object and in doing so creates an artefact, then they are entitled to unencumbered use of such an artefact, and this belief

[7] An incentivising agency may choose to protect only certain uses. We won't deal with this complication here.

regulates their actions, then both people ‘own’, in any conceivably relevant sense, everything that they create.

The above also indicates that ownership is a matter of the relevant incentives being present and does not *constitutively* or *conceptually* presuppose an institution like the state. This can also be seen by looking at the historical record. Economists have catalogued a variety of different, non-state institutions that have effectively created, in whole or in part, the incentives that constitute property rights. These include coalitions of Maghribi traders of the 11th century (Greif, 1993), the Bourse in Amsterdam in the 17th century (Greif, 1993), cattlemen’s associations and land clubs in the American West (Smith, 2003) and countless others. The Bourse in Amsterdam is especially interesting as many of the financial instruments that were traded on it, such as short sales, forward contracts and options were in legal limbo, *or actually illegal*, until formal institutions caught up to where the market had led. Despite this the degree of ownership that arose from social reputation and other informal sources of incentivization sustained an active market (Greif, 2005). Hence property can exist without being created by a state, which means, once again, that we can *define* the notion of property without first needing to define other institutions, most prominently the notion of a state.

3.2 Does our analysis presuppose other institutions?

It may well be thought that the problem that Searle addresses has merely been swept under the rug. Is it not so that, in most cases, the incentives sustaining ownership only came about because of another institution, i.e. in virtue of the institutions like the state imposing certain penalties on those who do interfere with one’s use of what you ‘own’? This may well be so, but as an objection to our account, it misses the point. Note that, independently of the merit of the objection, we have succeeded in *defining* ownership without reference to Searlean mechanisms or *irreducibly* institutional facts. While there must be some source of the relevant incentives, *facts about such a source are not individuating facts* and hence ownership can be characterised without mentioning the source of such incentivization. All that is *constitutively* required is that the relevant incentives are such that it becomes a good idea to not interfere with someone’s use of an object. It will typically be the case that some are more swayed by the relevant incentives than others. It will also typically

be the case that are relevantly disincentivized for purely for practical reasons, whereas some will be so incentivised partly in virtue of thinking that the incentivizer is morally justified in creating the relevant incentives. Others may think that not interfering with such use morally binding norm of behaviour. These differences are irrelevant to the nature of ownership; all that matters is that the relevant disincentivization is somehow brought about.

Note that we are not denying that the incentivised action view needs to eventually give some non-Searlean account of the state. The state (and countries, laws, etc.) are institutions, and so clearly we do. We cannot do all things at once and so leave it for future work; our current point is just that the nature of ownership can be characterised without reference to other institutions.

3.3 An objection concerning the ownership/possession distinction

Note that, when discussing issues like non-state sanctioned forms of ownership, we need to be careful to distinguish matters of substance from matters of terminology. Economists typically consider examples like the Bourse in 17th century Amsterdam to be cases of non-state sanctioned ownership of goods. Cannot, however, all such examples be rejected *a priori* by objecting that someone who 'owns' a forward contract on the Amsterdam Bourse in the 17th century does not *really* own the forward contract, but merely has some sort of possession of it? We think that such an objection misses the mark by confusing a narrow, legal sense of 'ownership' with a broader use of the term. In the legal sense of 'ownership', of course, the person who owns the forward contract does not own it. But, then again, by such a standard, no cigarettes or sea-shells have ever been money, no borders have ever come about by informal practice and it is a *priori* impossible for some small community to agree to put up a 'real' traffic light without involving the state. Note that, by such a standard, the study of institutional facts would be very easy; everything is what the law says it is. Our interest, however, is not in this narrow legal sense of these terms, but, as has already been explained, in *how it comes about that certain objects play a social role that is not fulfilled purely due their brute nature*. Facts concerning such objects are often tracked by what may be termed the 'broader' use of terms like 'ownership', 'money' and 'borders', used to refer to *de facto*, if not always *de jure*, states of affairs, and so our interest mainly lies in this broader sense of 'ownership',

'money', 'border', etc. Hence, the fact that some[8] of what we will call 'ownership', legally only counts as 'possession', is no more an objection to our view than the fact that, legally speaking, cigarettes have never been legal tender, or that the UN does not consider Somaliland[9] a country, etc. The fact that, legally, Microsoft is a 'person' has little relevance for philosophy of mind; in the same way the fact that, legally, no-one 'owns' illegal drugs has little relevance for the study of the broader notion of ownership[10].

With the notion of 'ownership' as third-party disincentivization in place we can now define the notion of a transfer of ownership. *Ownership is transferred when the subject, whose use of an object third parties are disincentivized to interfere with, is switched.* In other words, my pen becomes Bob's pen when Bob is the person whose use should not be interfered with, and I join the ranks of those who are disincentivized to interfere with Bob's use of the pen. This, then, leads to a definition of money that does not tempt us to postulate a new, autonomous ontological realm. Something is money in case I am incentivized to acquire it not for consumption, but in order to effect exchanges, where such exchanges are understood as bringing it about that the agent, whose use of an object others are disincentivized to interfere with, is switched by mutual consent.

4. Promises and agreements

We now turn our attention to a completely different kind of institution, namely promises. In accord with the incentivised actions view we take it that the way to answer such a question is to ask about the actions that are incentivized in virtue of the existence of a promise. Here the answer is simple: the action that is primarily associated with a promise is that of keeping or not keeping it. This is not, however, unique to promises. Consider the following two utterances, both made by Jean:

(1) I will go to the party tonight.

(2) I promise I will go to the party tonight.

[8] Namely where non-trivial, extra-legal third party disincentivization is involved.
[9] Legally Somaliland is considered an autonomous region of Somalia, but it is typically considered a *de facto* state.
[10] The difference arises in virtue of the fact that the interest of law is *regulative*, and its definitions implicitly amount to a commitment concerning where the state will and will not apply its incentivising muscle. Our interests are *descriptive*, and supposed to capture what actually exists, not what the state commits itself to doing.

Utterance (1) can be understood in two different ways. It can, firstly, be the mere expression of an intention that does not bring with it any specific commitment or obligation. This would be the case if Jean was merely mentioning this fact to a person who does not have any particular interest in whether Jean attends the party or not. Call this the *intention reading*. (1) can also, however, be understood as involving a commitment on the part of Jean to go to the party. Call this the *commitment reading*. While such an utterance can, under certain circumstances, amount to a full-blown promise, it need not. We generally recognise a level of commitment between merely expressing an intention to act in a certain way and promising to act in a certain way.

What is the difference between the intention reading of (1), the commitment reading of (1) and the full-blown promise in (2)? The difference lies in the seriousness of the consequences of not acting in accord with the expressed intention. If Jean merely uttered (1) with the intention reading in mind, then he will suffer no harm in virtue of not going to the party. On the commitment reading, however, Jean can expect to suffer adverse consequences if he does not go the party. The person he spoke to is likely to feel somewhat aggrieved, and Jean somewhat guilty, about the fact that he did not act as he said he would. These consequences are magnified if Jean had uttered (2). The person who Jean made the promise to is likely to feel extremely aggrieved if Jean does not go to the party, Jean may well feel very guilty, and, in future, when Jean makes a promise, he may well be met with rolled eyes.

The difference, in other words, between a mere expression of an intention, the expression of a commitment and a full-blown promise lies in the level of sanction that results from non-compliance. We believe that this fact holds the key to understanding the institution of promising. What makes some act a promise is that the promisor willingly and intentionally opens himself up to the possibility of such sanction. This allows us to characterise an explicit promise as *the expression of an intention to act in a certain way, coupled with an expressed willingness to be open to a relatively extreme sanction if the action is not accomplished.*

To clarify the basic idea, imagine a society of English speakers who do not use the word ‘promise’ or any simple equivalent. They do, however, have a practice they call ‘inviting being socially sanctioned’ and say things like “I will go to the party. If I

don't, you may think ill of me". Now consider the case of Robert, who says something like this to his friend Eugene. The vital thing to notice is that Robert, in virtue of having uttered this phrase, has an increased reason to go to the party. Firstly, not going to the party after having invited the potential sanction is more likely to cause Eugene to think ill of him than if he had not explicitly invited the sanction. This is so, partly because Robert has no plausible way of denying that he knew Eugene will be relying, to a significant degree, on him going to the party. If Robert does not go he implicitly signals to Eugene that he does not much care about any problems that may arise for Eugene in virtue of his inaction. Robert also signals to Eugene that Eugene's opinion of Robert does not much matter to Robert, and hence Robert's not going constitutes a type of insult. The second reason that Robert has to go to the party is that not going will decrease the chances that Eugene will, in future, attach a higher credence to the likelihood of Robert performing some action in virtue of the fact that Robert explicitly invited some sanction for non-compliance. Thirdly, Eugene is likely to tell others about Robert's actions, and hence third parties are also less likely to attach higher credence to the likelihood of Robert performing some action in virtue of the fact that Robert explicitly opened himself up to the possibility of social sanction. Now further assume that this society generally holds the moral belief that one should perform those actions where one explicitly invites sanction. Given this fact, Robert may well feel guilty if he does not go, Eugene and third parties may well attach moral blame to Robert, etc.

Given all the negative consequences that may well accrue to Robert, one may ask why he would ever explicitly open himself up to the possibility of social sanction. It is, however, precisely because non-compliance may lead to such consequences, that the practice has a point. If Robert wishes to induce others to trust him, then the act of opening himself up to extreme sanction, should he fail, gives Robert an increased reason to perform the relevant action. This means that others, i.e. Eugene and third parties, *have an increased reason to trust Robert*. In this way the act of inviting social sanction allows Robert to gain trust, and hence to undertake projects that depend on being trusted by others.

We see no reason to deny that what happens in the above hypothetical society *are* acts of promising, even though the way such promises are phrased may sound a bit

strange. The act of promising just is, in our view, a practice that allows an agent to significantly increase the incentive he has to perform some action. If the person breaks the promise then the fact that he has violated a moral norm of trying to fulfil intentionally created expectations, the disdain of his peers and his decreased capacity to have his future promises accepted by others all conspire to punish his lapse. Promises are a form of self-incentivization where we try to gain trust by primarily putting up some part of our social reputation as collateral.

A promise is nothing more than the expressed intention to act in a certain way, where the person making the promise has willingly incurred a strong reputational incentive to perform the action[11]. The same goes for agreements. If A and B agree to perform certain actions, as opposed to merely state their intentions to perform them, then this is a case of both parties willingly opening themselves up to a certain degree of sanction if they fail to perform the acts. This incentivizes both parties to perform the acts in question, and hence helps to make their commitments credible[12]. The sanctioning that I willingly commit myself to can, of course, be a matter of degree. For our purposes we can distinguish between 'agreements', where the sanctioning is relatively mild, and a 'mutual promise'[13], where the relevant sanction is mutually understood to be much harsher. The sanctions themselves are not mysterious, but will arise in virtue of some mix of moral and prudential motivation.

5. People who lack the propositional attitudes required by Searle's theory can form complex, interlocking institutions

5. 1. Companies, dividends, etc.

We have now defined the notions of ownership and the related notions of promises and agreements. Ownership, in modern society, typically takes the form of a heavily regulated, complex institution where most of the incentivization is done by the state, while promises take the form of simple, informal agreements; we hope to have shown that the incentivised action view can characterise the essence of such seemingly

[11] This puts us at odds with Searle's controversial claim (1964) that, in some sense, the institution of promising suffices to make it true that one should keep one's promises. We do not think that it is the job of a theory of promises to fully account for the normative status of promises - or, at least, no more than it is the job of a geological theory to account for the beauty of mountains.
[12] More on how incurring a cost in making a commitment can serve to make the commitment credible can be found in Green (2007), who explores ideas proposed in Zahavi et al. (1997).
[13] A legal contract is just an agreement, or mutual promise, with an added incentivization mechanism, namely legal enforcement.

radically different institutions and bring them under a common rubric. Another reason for choosing these two institutions to analyse is that they are both *basic* in the sense that, once they are in place, they can serve to generate further institutional entities. Below we will make a case for the claim that the incentivised action view, by availing itself of notions like promises and ownership, can account for complex financial objects like companies, employers, dividends and so on. Note that our efforts will, for now, fall far short for defining such institutions. All we wish to show is that such institutions can, in principle, come into being in recognisable form without any of the Searlean machinery (collective acceptance; propositional attitudes that cannot be cast in non-institutional terms) being involved.

Consider ten people who independently and simultaneously wash up on a desert island and who all speak a version of English that contains no institutional fact terms. Due to the practical gains to be had from sticking together and their belief in the edifying effects of human society they live in the same place, talk, interact, and so forth. (None of this necessarily presupposes the existence of an institutional fact, *as Searle uses the term 'institution'.*) Stipulate that the institution of private property develops, meaning that there are situations where third parties are disincentivized to interfere with someone's use of an object. Some are disincentivized in virtue of believing that there are conditions under which such interference is morally wrong, some are disincentivized in virtue of a norm whereby those who interfere with other's use of relevant objects are likely to suffer such interference themselves, some are disincentivized in virtue of the fact that the 'owner', i.e. the party not so disincentivized, can punish them for such interference or has friends who will, and so on. Further stipulate that there are shells that are used as currency, i.e. that the people on the island are incentivized to acquire them for exchange and not for some other specific use, with exchange being defined as being a voluntary change in the object of third party disincentivization, as explained above. Now suppose that Alex and Bob make an 'agreement', as defined above, according to which Bob will use Alex's bucket to carry water from a river to where our ten unlucky souls have set up camp. Bob will receive SS10 (ten sea-shells) for his efforts and Alex then sell the water for SS20, making a profit. Alex, in other words, has the means of production (capital, i.e. the bucket) and Bob is a paid labourer.

The notions of agreement, property, money and trade have already been defined in terms of actions and incentives and all can understand their situation in terms of being incentivized to act in a certain way. In other words, all know that they have a reason to act in accordance with professed intentions, that they have a reason not to interfere with how relevant individuals use certain objects, that the person whose use should not be interfered with can change, and so on. Now let us suppose that Alex spent his sea-shells quickly, whether on food stuffs gathered by others or alcohol brought onto the island, and can no longer pay Bob. He now makes an 'agreement', as defined already, with Carol to the effect that Carol will provide SS10 so that he can pay Bob to collect water. In return Carol will get SS1 from Alex every day when there is at least SS1 profit. The agreement further stipulates that, should Alex be unable or unwilling to continue the deal, Carol will take ownership of the bucket, but will not be able to claim anything else. Stipulate that the relevant beliefs, expectations and power-relations are such that all have an incentive, whether moral or practical, to keep their agreements.

In the above case we have certain familiar financial objects in their most basic form. Alex has a certain financial liability which is limited to the value of his invested capital, i.e. the value of his bucket. Hence Alex owns a limited liability company, Bob is an employee of this company and Carol is an investor in Alex's company who receives a fixed, daily dividend. The balance sheet of Alex's company comprises the following: equity invested by Alex and Carol to the value of the bucket and the working capital, assets to the value of the bucket and the cash used for working capital. There are no liabilities on the balance sheet. All of this can be described in terms of people being incentivized to act in a certain way. Alex is incentivized to tell Bob what to do in a way that other people are not and incentivized to hand Bob SS10 every time he hands him a filled bucket, after which Alex is disincentivized from interfering with Bob's use of the SS10. Alex is incentivized by his agreement with Carol to give the water to whoever hands him SS20, after which he is disincentivized to interfere with the person's use of the water and the person disincentivized to interfere with his use of the SS20, Alex is incentivized to hand Carol SS1 every day and disincentivized to interfere with what she does with it, and so on.

To encounter a limited liability company is no more or less than to encounter actual people locked into a certain stable pattern of incentivization. Interestingly enough, when Searle's initial account of corporations was challenged, he responded by saying that "[a] corporation is just a placeholder for a set of actual power relations among actual people" (2010: 22). We think that this idea is on the right track. But we also think that following such ideas to their logical conclusion leads to the view that we can characterise corporations in terms of the relations of incentivized action that obtain between the relevant people, and so dispense with the Searlean apparatus of collective intentionality, irreducibly institutional facts that have the structure of declaratives, etc.

In the above case we have described multiple, interlocking institutions - i.e. an agreement, employer, employee, investor, limited liability company, dividends – in terms of actions and incentives. This should make it plain that, once basic notions like ownership and agreement are in place and defined in non-institutional terms, we can recreate financial facts of any desired complexity on our hypothetical island.

5.2 A note on method

The method employed thus far primarily consists in the presentation of various Robinson Crusoe – style thought-experiments[14]. We follow this time-honoured method for the same reason that economists often do, namely that it allows us to abstract away various real-world complications that do not affect the essential issue at hand. The most basic advantage this has is that it allows us to deal the fact that institutional facts are interlinked. In this way, I may withdraw *money* from the *bank* in order to allow me to *pay* for a *visa* that allows me *legal entrance* into the *United Kingdom*. There exist an enormous number of types of institutional facts, many of which constitutively depend on the existence of other institutional facts. Our mortality prevents us from analysing them all; our aim is merely to give some substance to our claim that doing so is possible in principle. For this reason it is often convenient to stipulate that several people show up on an island, that there exists no institutional facts (in the technical sense used here) on the island or between the relevant people, and then to describe the process whereby such facts are created

[14] Searle himself has defended the use of thought-experiments in clarifying the nature of institutional facts (Searle, 2005: 20-21).

without any appeal to (irreducibly) institutional facts that consist in the collective acceptance of statements of the form *X* counts as *Y* in *C*.

All that our examples require is that the relevant incentives do not spring from an (Searlean) institutional source, i.e. from a pre-existing irreducibly 'institutional' fact. Given that, as has been explained before, Searle's use of 'institution' is much narrower than the ordinary one, this is not as difficult a demand to meet as one might think. Searle, for example, does not view "the fact that today is 24 September 2004" or the Christian calendar in general, as institutional, as these do not create deontic powers. The same goes for the fact that someone is an 'underachiever', or practices like science, religion and education, (2005: 18, 2010: 92), as he uses the term 'institution'. Nor does he view moral rules, norms and beliefs as such facts. Searle *only* talks of institutions where certain status-functions, i.e. where certain kinds of rights, duties and obligations, enter the picture (2005: 18). Where Searle would see someone with a duty which can only be characterized in irreducibly institutional terms, we only see someone who is incentivized to perform certain actions, with these actions being expressible in non-institutional terms.

Note that making our argument is not an empirical enterprise, but a theoretical-conceptual one. One may object by saying that we cherry-pick our examples so as to be amenable to our theory, i.e. that Searle may also come with hypothetical examples that vindicate his views. We do not think that this is so, for while specific cases have illustrative value, they do not serve to decide the theoretical issue. To illustrate this claim, consider a case as non-cherry picked, and favourable to Searle as conceivable, i.e. a society of Searlean literalists that only creates institutional facts by getting together and chanting things like 'we count sea-shells as money". Such chanting *will have no effect* unless the people involved understand it as making it come about that at least some of them will exchange sea-shells for goods, which means that all now have an incentive to acquire such sea-shells for exchange. Hence we still insist that the sea-shells are money purely in virtue of the fact that the professed intentions served to incentivize everyone to acquire the sea-shells for exchange. In other words, our analysis still applies unchanged. Searle, presumably, would similarly interpret our hypothetical cases as cases where it comes about that a society collectively accepts

some sort of 'X counts as Y in C' type of fact. We do not dispute that he could do so, but claim that our account is preferable as it does not complicate ontology.

Note that, even if our hypothetical Searlean literalists only act in virtue of fully believing the Searlean theory of institutional facts, their beliefs and behaviour *still* do not constitute an objection to our theory. Our theory is not about the beliefs in the mind of people, but about the actions that people are incentivized to perform, and even action based on *systematically* false beliefs can still incentivize. For instance, if a bookseller only accepts dollars due to the belief that they are backed by gold, I still, given that a choice of currency is a kind of coordination game, have an incentive to give the bookseller dollars in order to acquire books from him. In the same way, if some people are only stopping at traffic lights due to a belief that the existence of a traffic light is an irreducibly institutional fact, or the police only punishes those who run red lights due to their belief that such objects are irreducibly institutional, such action still gives someone free of such beliefs an incentive to treat traffic lights in the appropriate way. Hence, we need not deny that people have Searlean *beliefs* about institutions; we only need to deny that the irreducibly institutional reality that such beliefs are supposed to be *about* exists in a way that can only be explained in institutional terms. Our view concerns ontology, not epistemology; all we are committed to concerning beliefs is, as indicated earlier, that the actions appropriate to legitimate institutions can, in principle, be cast in some non-irreducibly institutional form. We take it that the non-irreducibly institutional construal of paradigm institutional objects (traffic lights, borders, money, ownership, exchange, promises, agreements), and our illustration that people without propositional attitudes that feature non-irreducibly institutional concepts, can form complex, interlocking institutions (employers, employees, investors, limited liability companies, dividends), amounts to at least a *prima facie* case for this contention.

6. Bottom-up ontology and bottom-up genealogy: the order of explanation

The explanations given by the incentivised actions view are naturalistic, 'bottom-up' explanations. In fact, it is bottom-up in two ways, both ontologically and genealogically. To explain this point, consider the analogy between our view of institutions and another theory that is similarly 'bottom-up', namely the Gricean view of the development of natural language conventions, due to Grice (1957).

The Gricean view of language concerns the emergence of signals with a non-natural meaning, i.e. the kind of meaning a sign has, not because it is a natural symptom of something else (as in 'clouds mean rain'), but because the speaker *intended* it to signify something. Grice's key insight was that an action or utterance means non-naturally that *P* only if it was produced with the intention that the audience comes to believe that *P* by recognizing that very intention – i.e. by recognizing what the speaker intended him to believe. In other words, a sign non-naturally means 'There is a lion' if it was produced with the intention of letting the hearer come to believe 'There is a lion' by recognising the speaker's intention to cause the hearer believe 'There is a lion'.

Grice identifies the above mechanism as the essence of communicative action. From this, then, emerges the conventional system we know as natural language. Grice views natural languages as systems of stabilized shared habits which could supplant the Gricean mechanism of recognition of intentions. In other words, a signal comes to mean something 'timelessly' and 'conventionally' when the meaning it acquired on an initial occasion of use, i.e. a meaning it has in virtue of the Gricean mechanism has sufficiently 'stabilized'. On such a view the signal and its stable meaning are the *sediment* of a practice of communication that is prior to a system of conventions (Blackburn, 1984), and not some ontologically novel entity. In this way Grice presents a 'bottom up' account of how conventional meaning could emerge out of a practice where people meant to communicate something and their intentions were correctly recognized by their intended audience.

Our explanation of institutions is a similarly 'bottom up' explanation. Take, for example, a case where tins of mackerel become used as currency[15]. It can start with some individual realising that such tins are both generally popular and durable. This then leads the person to acquire them with the intention of exchanging, rather than consuming them. Other individuals, either independently, or by picking up the idea from others, then realise that the incentives operative in the local economy are such that they have good reason to act similarly. At some point this behaviour becomes so

[15] As happened in the case of US prisons. See Scheck (2008).

widespread that that we start calling the tins of mackerel ‘money’. But, as was the case with the Gricean story about the emergence of natural language, the tins of mackerel ‘being money’ is no more than the sediment of a simple practice that has stabilised. *The ‘mackerel-institution’, genealogically, comes about in virtue of a series of individuals being incentivized to act in a certain way, and is, ontologically, no more than the fact that such incentives are widespread and stable within a community.*

Due to the above we hold that there is an important sense in which Searle gets the ‘order of explanation’ the wrong way round. It is not the case, as Searle claims, that institutional facts are constituted by some sort of widespread social acceptance, however this is understood, and that such widespread acceptance gives us something fundamentally new. Rather ‘widespread acceptance’, i.e. recognition that a pattern of incentives obtains, is simply the end of a continuum that starts with a point where one person recognizes the existence of an incentive to act in a certain way. Hence we hold that the origin and nature of institutional facts are to be understood in terms of individuals who recognise that certain incentives obtain, just as, on the Gricean view, stable natural languages emerge from individual cases of intention-recognition. There is merely a difference of degree between the case where such incentives are fleeting and recognition rare, and the case where the incentives are stable[16] and their recognition universal.

7. Further virtues of the incentivised action view

The biggest gap in the incentivised action view at present, as we see it, is that it still owes an account of politically created entities like nations, governments and codified law. In future work we plan to try and analyse such institutions; hopefully this will serve to bridge the gap between the seemingly far-fetched nature of ‘desert island’ style examples and our common experience of institutional reality. We can, however,

[16] Above we portray institutional facts as no more than the more stable end of a continuum of which the other end is more loose incentive structures. An anonymous referee has pointed out that being at the end of such a continuum does not necessarily amount to being ontologically of the same kind as other points on the continuum. We agree but, in the absence of a positive argument in favour of viewing such a stabilised set of incentives as ontologically distinct from incentives in general, will keep our ontology sparse.

only do so much in one paper; here we will end by highlighting some of the virtues of the incentivised action view that have not been addressed thus far.

The incentivised action view allows for easy vertical integration with other theories. Our theory of promising as self-incentivization, for instance, fits seamlessly into game theory and mainstream economic theory in general. Game theory is antecedently committed to the existence of agents with various beliefs, courses of action and incentives to perform such actions. The incentivised action view attempts to show that all talk of money, borders, promises and the like can be accommodated into game theory as simply being yet more talk about actions and incentives[17].

The incentivised action view helps to prevent confusion by making certain matters much clearer than they otherwise are. These are not the type of confusions that economists are prone to, but they are common among broader society. Take, for example, the matter of private property. On our view, the whole idea of owning something amounts to third parties being disincentivized to interfere with the use that an agent, namely 'the owner', makes of a certain object. This immediately avoids the mistake of thinking of ownership as a relation between a person and a thing, but makes it plain that private ownership is a relation between an individual, a thing, and society. It also makes it plain that ownership is not something that can only be created by government decree, as well as making it plain that, if the government lacks the necessary power, a government decree is not even sufficient for the creation of private property. Furthermore, it shows that our habit of talking of ownership as something that either exists or does not is dangerously misleading. The question of whether one owns something *in a purely legal sense* may be an absolute matter. But the law is only one way of making ownership come about, and actual ownership is a matter of third party disincentivization. Such disincentivization is a matter of degree. This means that whether someone owns something is, in an important sense, similarly a matter of degree. Simply put, there is a substantial difference between the degree to which a Danish farmer owns his farm and the degree to which a Zimbabwean farmer owns his farm. It is this sense of ownership, i.e. *actual*

[17] Alternatively, consider the Marxist claim that we mystify the social by reifying social relations. Our project can be described as the dereification of institutional facts.

ownership, that explains what we see in society and that should be the concern of sociological and economic theory.

This leads us to a related point, which can be clearly seen once we recognise institutions for what they are. When we talk of the political/economic system that is in place in a certain country, we typically categorise them in terms of the system that codified law tries to incentivize. In this way we will call countries capitalist if the legal system is capitalist, call them socialist if the legal system is socialist, etc. But, on the incentivised action view, it is plain that the system that prevails in a country is fundamentally a matter of the actual incentives that do obtain, and not the incentives that the government tries to create[18]. When we categorise countries, or, for that matter, companies or any institution, by codified law we are effectively attempting to capture facts about a country, namely incentive sets, in terms of one mechanism of bringing such incentives about. Simply put, the 'natural kinds' at play when studying economics, sociology and history are the stable sets of incentives that actually obtain, and they need not correspond closely to codified law; the patterns are the *de facto* classifier, not the self-descriptions.

A further vice that often results from identifying an institution with the legal incentives that try to co-create it, is to think that institutions are the kinds of things that can easily be created or changed by simple decree. On our view it is plain that, whereas legal decrees obviously play a role, they can do no more than cause a change in an already existent set of incentives, among which they may well have to 'compete' for priority and on whose prior existence they depend[19]. While institutions are frequently created by law and decree, they can also be created and maintained by social norms and moral motivation. The study of these matters is both famously complicated and highly variable across both cultures and time; it is our hope that the (still relatively undeveloped) incentivised action view will, once brought to full fruition, provide a clear and unified framework for thinking about such matters.

---

[18] An interesting example is of how these thing should be done is Cukierman et al. (1992), who tried to measure the *de facto*, as opposed to *de jure*, independence of central banks.

[19] Or, as Hodgson (2006: 18) puts it: "[F]ormal institutions... *always* depend on nonlegal rules and inexplicit norms in order to operate. If laws or declarations are neither customary nor embodied in individual dispositions, then - "formal" or not - they have insignificant effects. They are mere declarations or proclamations, rather than effective social rules."

References

Blackburn, S. 1984. *Spreading the Word.* Oxford: Oxford University Press.

Cukierman, A, Webb, S and Neyapti, B. 1992. Measuring the Independence of Central Banks and Its Effect on Policy Outcomes. *The World Bank Economic Review*. 4: 353-398.

Green, M. 2007. *Self Expression.* Oxford: Oxford University Press.

Greif, A. 1993. Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition. *American Economic Review*, 83: 525-48.

Greif, A. 2005. Commitment, Coercion, and Markets: The Nature and Dynamics of Institutions Supporting Exchange. In Mènard, C. and Shirley M. (eds.) *Handbook for New Institutional Economics*. Rotterdam: Springer.

Grice, P. 1957. Meaning. *Philosophical Review* 66: 177-188.

Hodgson, G.M. 2006. What are Institutions? *Journal of Economic Issues.* 40: 1 – 25

Ruben, David-Hillel. 1997. John Searle's The Construction of Social Reality. *Philosophy and Phenomenological Research* 57:443-447

Scheck, J. 2008. Mackerel Economics in Prison Leads to Appreciation for Oily Fillets. *The Wall Street Journal*. New York. 2 October 2008.

Searle, J. R. 2010. *Making the Social World: The Structure of Human Civilisation.* Oxford: Oxford University Press.

Searle, J. R. 2005. What is an Institution? *Journal of Institutional Economics.* 1: 1 – 22.

Searle, J. R. 1995. *The Construction of Social Reality.* London: Penguin Books.

Searle, J. R. 1964. How to Derive 'Ought' From 'Is'. *Philosophical Review* 73: 43-58.

Smith, B. and Searle, J. R. 2003. The Construction of Social Reality: An Exchange. *American Journal of Economics and Sociology*, 62: 285 - 309.

Smit, J. P, Buekens, F. and du Plessis, S. 2011. What is Money? An Alternative to Searle's Institutional Facts. *Economics and Philosophy*, 27: 1-22.

Smith, V. 2003. Constructivist and Ecological Rationality in Economics. *American Economic Review.* 1993: 465 – 507.

Stalnaker, R. 1989. On What's In the Head. *Philosophical Perspectives*, 3: 287 – 316.

Tieffenbach, E. 2010. Searle and Menger on Money. *Philosophy of the Social Sciences.* 40: 191-212.

Zahavi, A. and Zahavi, A. 1997. *The Handicap Principle: A Missing Piece of Darwin's Puzzle*. Oxford: Oxford University Press.