INSTITUTIONS AS EMERGENT PHENOMENA: REDEFINING DOWNWARD CAUSATION

Documents de travail GREDEG
GREDEG Working Papers Series

Nicolas Brisset

GREDEG WP No. 2016-30
http://www.gredeg.cnrs.fr/working-papers.html

Les opinions exprimées dans la série des Documents de travail GREDEG sont celles des auteurs et ne reflètent pas nécessairement celles de l’institution. Les documents n’ont pas été soumis à un rapport formel et sont donc inclus dans cette série pour obtenir des commentaires et encourager la discussion. Les droits sur les documents appartiennent aux auteurs.

The views expressed in the GREDEG Working Paper Series are those of the author(s) and do not necessarily reflect those of the institution. The Working Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate. Copyright belongs to the author(s).
Institutions as Emergent Phenomena: Redefining Downward Causation

Nicolas Brisset

Université Côte d’Azur, CNRS, GREDEG, France.
Nicolas.BRISSET@unice.fr
GREDEG Working Paper No. 2016-30

The concept of emergence is frequently used in the social sciences in order to characterize social institutions. Nevertheless, philosophy of mind argues that the idea of emergence is problematic because it encompasses the dubious notion of downward causation, i.e. the fact that an entity at a given ontological level might have a causal influence on lower level entities. This work shows that although it is problematic in some fields, emergence is an ontological feature of the social world. In order to justify this point of view and to show how institutions relate to individuals’ actions, we define an institution as an exogenous device, which enables us to show that the relationship between institution and individual actions is not only a causal one but also an intersubjective and a constitutive one.

JEL codes: B52, B41.
Keywords: Emergent Phenomena; Institution; Downward Causation; Convention.
1. INTRODUCTION

Though widely used in social sciences, the concept of emergence still has a very polysemous character. On the one hand, the “revolution” of agent-based modeling (ABM) supports an intuitive definition of emergent phenomena as macroscopic patterns arising from interactions of heterogeneous individual agents, where these patterns cannot be straightforwardly derived from the properties of these agents (microstructure) (Epstein 1999; Nigel and Terna 2000). On the other hand, some recent developments in institutional economics endorse a deeper version of emergence, closer to what is well known as the classical British emergentism (Lawson 1997; 2003). From this second perspective, an emergent property of a system cannot be derived in principle from the microstructure of this system. This perspective defends the possibility of an ontological gap, i.e. that emergent social phenomena bring novelties into the real world.

In the original and still-dominant perspective, institutions as emergent phenomena are ontologically reducible to the agents and their activities. In the emergentist’s one, this reduction is nonsense. The main difference between both these perspectives rests on the controversial notion of downward causation. Indeed, classical emergentism argues that there is an irreducible feedback from emergent institutions to social agents that prevents the analytical reduction of institutions to social agents.

This article focuses on the notion of downward causation. We argue that there is room for strong emergence in the social sciences. We defend a symmetry thesis regarding the epistemic problems faced by the social scientists and the social agents: not only is emergence a necessary concept for the social scientist to explain social phenomena, institutions as emergent phenomena are also necessary for the real social agents to make choices and to act. We make this claim on an internalist basis: our argument relies on the very tools of economic theory, especially game theory. Indeed, pushing Gintis’ (2009) suggestion, we contend that once we make explicit the epistemic requirements tacitly assumed by game theory to account for coordination in the social world, a concept of strong emergence is needed.

2. INSTITUTIONS AND ECONOMICS
Institutions are now central in economic theory. Although there is still no well established definition of this concept (but see Greif 2006; Hodgson 2006), economists and social scientists tend to characterize an institution as a set of three embedded and self-enforced elements: firstly, institutions are made of habits of thought (individual beliefs) that evolve and change through the interactions between the agent and his natural and social environment. Secondly, once habits have accumulated and have become widespread enough in a population, they develop into a set of informal rules, such as social norms or conventions. A habit becomes a norm or a convention once everyone (or a sufficient fraction of the population) expects others to conform to it and believes that others expect he will conform (Lewis 1969; Bicchieri 2006). Thirdly, a social norm or a convention can become a formal rule once an organization endorsed with some form of legitimacy begins to enforce it.

In this paper, we will consider concept of institution as an answer to the analytical limits of narrow methodological individualism defined as a way of explaining the social world in terms of individuals alone. Institutional economics claims that such an endeavor is nonsense since it is impossible to reduce social explanations to individuals without involving at least a minimum set of interactive relations between them, i.e. previous institutional rules that enable social interactions. This tradition could be traced to those we commonly called the old institutionalists such as J. R. Commons and Thorstein Veblen (Hodgson 2004; Hédoin 2013). Game theory provides a good illustration of this problem (Aoki 2001). Explaining social interactions from a game theoretical point of view requires an institutional background shared by each agent: payoffs, guarantee of contracts’ enforcement, constitutive rules of the game. Such rules are necessary: without these rules, how to explain that agents trust others enough to take the risk of establishing a contract with them? The origins of this institutional background can be found in other games related to different social fields: economic, political, communal, and so on. Following the ambition of methodological individualism, one should explain institutions (as well as the common knowledge of them) as the results of individuals’ interactions in a free-institutional state. Yet, such an endeavor leads to an infinite regress since one has to spell out where constitutive rules of these other games come from. Any attempt to make endogenous each rule of the game in an individual explanation is thus nonsense. To solve this problem, game theory must assume that the rules of the game are common knowledge, which supposes a previous institutional and irreducible background. Because of the infinite regress of the individualist’s quest for the explanation of the rules of the game, Hodgson (2007, 8) argues that such an inquiry has actually never been achieved and, as a consequence, that the narrow definition of methodological individualism has never been applied in practice. In a more general framework, the old theoretical project in economics to
find microfoundations for the macroeconomic, held by New Classical School, met strong criticism. Such criticism takes root in several arguments; the problem of the institutional rules’ reduction is only one of these (Epstein 2014).

If economists need to introduce institutions in their analysis and models, few of them handle the question of institutions as ontological entities. The previous reasoning is epistemological; from an ontological point of view, institutions are no longer a concept, but a social fact distinguished from the individual level. Such a position contradicts ontological individualism and the idea that the social world is only made of individual agents. Hodgson (2000a; 2004) and Lawson (1997; 2003) contribute to this question by defining institutions as sui generis entities through the concept of emergence. Lawson defines emergence as follows:

“A stratum of reality can be said to be emergent, or as possessing emergent powers, if there is sense in which it

(i) has arisen out of a lower stratum, being formed by principles operative at the lower level;
(ii) remains dependent on the lower stratum for its existence; but
(iii) contains causal powers of its own which are irreducible to those operating at the lower level and (perhaps) capable of acting back on the lower level.”

(Lawson 2003, 44)

As we wrote in the introduction, such a definition is in sharp contrast with the reductionist one of agent-based modeling. This definition supports a particular view of the social realm, structured in different strata dependent on each other but sufficiently distinct for having irreducible causal effect on each other. Hence, institutions, as emergent properties, cannot be reduced to human practices and have a particular causal power on these. The concept of emergence appeared with classical British emergentism during the last quarter of the nineteenth century. George Lewes (1874-1875), distinguishes resultant and emergent effects:

“There are two classes of effects markedly distinguishable as resultant and emergent. Thus, although each effect is the resultant of its components, the product of its factors, we cannot always trace the steps of the process, as to see in the product the mode of operation of each factor. In this latter case, I propose to call the effect an emergent.” (Lewes 1874-5, 412)

If such an idea has been relegated for decades to the level of metaphysical curiosity, the 1970’s renewal of the mind-brain debates made it up-to-date. Especially with Davidson’s works about anomalous monism (Davidson 1980) and
the Putnam’s and Fodor’s defenses of Multiple Realizability (Putnam 1975; Fodor 1982). Although there are great differences between the so-called non-reductive physicalism and the classical British emergentism (Kim 1993, chap 14), Jaegwon Kim identifies three core common ideas required for emergence to happen: supervenience, irreducibility and downward causation.

**a. Supervenience**

Kim (2006, 550) defines supervenience as follows: “If property M emerges from properties \( N_1, \ldots, N_n \), then M supervenes on \( N_1, \ldots, N_n \). That is to say, systems that are alike in respect of basal conditions, \( N_1, \ldots, N_n \) must be alike in respect of their emergent properties.” It indicates that the higher-level property M cannot differ in systems that have the same set of lower level properties. This is Lawson’s (ii) point: there is an ontological dependence between the different strata of the social world: the sum of agents constitutes the institutions.

**b. Irreducibility**

The second condition for emergence is that an emergent property is not ontologically reducible to its microfoundations (the lower level). That is to say that even if it is ontologically constituted by its microfoundations, an emergent property cannot be explained from them because of an ontological gap between the microfoundations and the emergent property. Emergence is thus a *brute and unexplainable fact*: it is impossible to know on what kind of dependence the supervenience relation involved in emergence is founded (Kim 2006, 556). C. D. Broad, another major advocate of classical emergentism, summarizes:

“...the characteristic behaviour of the whole could not, even in theory, be deduced from the most complete knowledge of the behaviour of its components, taken separately or in other combinations, and of their proportions and arrangements in this whole. This alternative [...] is what I understand by the ‘Theory of Emergence’” (Broad 1925, 59)

This definition of irreducibility suffers from a lack of precision. What is the nature of the relationship between the macrostructure’s properties and its microfoundations? Stephan (2002) gives a useful definition of irreducibility divided in two characteristics. First, a systemic property M can be called irreducible if it cannot be deduced from the arrangement of its microfoundations and the properties they have in isolation. Second, a systemic property M can be said to be irreducible if the parts’ behaviors on which M supervenes are themselves irreducible: \( N_1, \ldots, N_n \) behavior cannot be deductively explained.
without taking M’s behavior into consideration (Stephan 2002, 88-89). As an
ontological statement, this second characteristic is essential because it leads to
another one, which embodies the brute aspect of emergence: there seems to exist,
in that case of irreducibility, a “downward causation” from the macro-system to its
parts.

c. Downward causation

The notion of downward causation is subject to sceptical claims in several
fields. Hulswit (2006) shows that the use of that concept entertains some
confusion since it is not even clear of what we mean by “causing”. In the case of
social sciences, what is caused in downward causation is the behavior of the
agents in a social interaction. We will explore the notion of causation later. For
the present step of our argumentation, we reduce this causal power to the fact
that institutions make a difference in social interactions. Finally, it is possible to
link up downward causation to irreducibility: the behavior of the microstructure
cannot be reduced to its own causal power and the causal power of the emergent
property cannot be reduced to the causal power of its microfoundations.
Downward causation is for Hodgson (2000b) the essence of institutional economic
theory. Given this statement, and in accordance with the three conditions for
emergence we pointed out before, institutions are both exogenous (not reducible)
to each agent, and endogenous to the system (given the principle of supervenience).

Considered to be obvious by several economists, this definition has been
strongly attacked by many philosophers. Indeed, we know since the so-called
supervenience argument of Kim (1999; 2000) that emergence is a problematic
concept. The critical question is: If an emergent, M, emerges from basal condition
P, why can’t P displace M as a cause of any putative effect of M? After all, without
the presence of lower-level conditions, higher-level properties could not even be
there. So how could these higher-level properties causally influence the conditions
from which they arise (Kim 1999, 25)? To argue that a social structure M has
caused a set of individual actions P* is knotty: if we define M such that it
supervenes upon a set of actions P, it follows that we only need P to explain P*.
Figure 1

Given this reasoning, it is clear that emergence should be removed from our social ontology: the concept of emergence would only be a ploy used in order to understand what we do not understand yet. In other words, emergence would just be an epistemological statement. As Hempel and Oppenheim wrote:

“...emergence of a characteristic is not an ontological trait inherent in some phenomena; rather it is indicative of the scope of our knowledge at a given time; thus it has no absolute, but a relative character; and what is emergent with respect to the theories available today may lose its emergent status tomorrow” (Hempel and Oppenheim 1948, 150-1).

Nevertheless, we claim that for the effort of a rough adaptation, institution could be call emergent properties. With that goal in mind, we will re-write the previous definition of emergence in terms of supervenience, irreducibility and downward causation by re-defining each of these terms.

3. A DEFINITION OF CAUSALITY

The core notion of emergence is the one of downward causation. Before targeting this notion, we have to set a clear definition of causation starting from Woodward’s Manipulability Theory of Causation. For Woodward, X is a direct cause of Y with respect to a set V of variables if manipulation of X has an impact on Y when other relevant variables are fixed. Woodward then defines the deeper notion of contributing cause (Woodward 2003, 59). To say that X to be a contributing cause of Y with respect to variable set V is to say that:

(i) There is a directed path from X to Y such that each link in this path is a
...direct causal relationship;

(ii) If there is only one path $P$ from $X$ to $Y$ or if the only alternative path from $X$ to $Y$ besides $P$ contains no intermediate variables (i.e., is indirect);

To be complete, it is necessary to define the idea of intervention. A manipulation $I$ of $X$ is an intervention with the respect to $Y$ iff:

1. $I$ causes $X$
2. $I$ acts as a switch for all other variables that cause $X$. That is, certain values of $I$ are such that when $I$ attains those values, $X$ ceases to depend on the values of other variables that cause $X$ and instead depends only on the value taken by $I$.
3. Any directed path from $I$ to $Y$ goes through $X$. That is, $I$ does not directly cause $Y$ and is not a cause of any causes of $Y$ that are distinct from $X$ except, of course, for those causes of $Y$, if any, that are built into the $I$-$X$-$Y$ connection itself; that is, except for (a) any causes of $Y$ that are effects of $X$ (i.e., variables that are causally between $X$ and $Y$) and (b) any causes of $Y$ that are between $I$ and $X$ and have no effect on $Y$ independently of $X$.
4. $I$ is (statistically) independent of any variable $Z$ that is on a directed path that does not go through $X$.

(Woodward 2003, 98)

From this definition, Kim's criticism of emergence can be translated as follows: in order to show that, in the figure 1, $M$ could be a cause of $P^*$, one has to show that an intervention $I$ on $M$ can respect the condition 11 to 14. Marcellesi (2010) shows that if 11 to 13 are, it is not the case for 14. Given the fact that there is only one way from $I$ to $M$, 11 and 12 are respected. Plus, the unique way from $I$ to $P^*$ go through $M$, 13 is also respected. Nevertheless, given the fact that $M$ supervenes upon $P$, $I$ cannot be independent of $P$, which is on the direct path that does not go through $M$.

4. A DEFINITION OF INSTITUTION

Even if supervenience and downward causation seem to be contradictory statements (which would imply that emergence is nothing but an analytical tool and not an ontological statement), we want to go further in emphasizing that both epistemological and ontological considerations about emergence in social sciences overlap. If emergence is obviously problematic in philosophy of mind, where we
deal with physical microfoundations, basic units of the social world (social actors) have the specificity of making expectations about the evolution of the social system they belong to. As we stated, an emergent property in the social world is thus a fact (an institution) taken as given (exogenous) by the agents, but in fine depending on them. It is, for instance, clear in Avner Greif’s work:

“...institutionalized rules and beliefs are man-made yet exogenous to each individual whose behavior they influence. They are exogenous to each individual in the sense that they are commonly known rules and beliefs in situation in which behavior is not technologically determined.” (Greif 2006, 34)

It is because agents must take institutions as given that they are qualified as exogenous and that they are emergent with respect to the three characteristics of emergence we give above. This idea is known in economics since J.M. Keynes. Keynes (1937; 1973) emphasizes the necessary conventional aspect of investment practices, which is due to the market’s fundamental uncertainty. Indeed, if the occurrence of an event cannot be ascribed a probability, “In practice we have tacitly agreed, as rule, to fall back on what is, in truth, a convention. The essence of this convention [given our definition of institutions, convention, institution and social structure are necessarily embedded] – though it does not, of course, work out quite so simply – lies in assuming that the existing state of affairs will continue indefinitely, except in so far as we have specific reasons to expect a change.” (Keynes 1973, 152). Individuals accept some conventions as given since they are confronted to incomputable situations. Our argument follows R. Koppl and B. Rosser’s work on computable economics (Koppl and Rosser 2002; Rosser 2004; Koppl 2008a; 2008b) based on self-referential phenomena in the social world. In mathematics, a function is said to be non-computable if it cannot be calculated via a mechanical procedure (an algorithm). In economics, that kind of phenomena is frequent since some decision problems lead to infinite recursive processes. For instance, if ego has to conjecture about alter’s actions which are based on conjecture about ego actions. For Koppl and Rosser (2002), self-reference leads to non-decidable problems and shows the limits of our knowledge in the social world. They illustrate this idea by the famous Morgenstern’s Holmes-Moriarty game (Koppl and Rosser 2002, 343), where the decision of Holmes to leave the train to Dover or Canterbury depends on Moriarty’s decision:
In figure 2, there is no Nash equilibrium in pure strategies, but one in mixed strategies where Holmes and Moriarty assign a probability weight of \( \frac{1}{2} \) to each pure strategy. However, the concept of equilibrium in mixed-strategies is problematic from the perspective of computability because it is one of the properties of this equilibrium that if one believes that others will play their mixed-strategy, then one has no reason to also play their mixed-strategy; indeed, any pure strategy which has a positive probability of being played at the mixed-strategies equilibrium will give the same utility. Binmore (1987) goes further by arguing that even games with (multiple) equilibria in pure strategies can be non-computable. This makes, on the one hand, the analyst unable to understand players’ choices and, on the other hand, impossible for the players to rationally make a choice. We fall in an infinite regress of ‘I expect that you expect that I expect’ and so ad infinitum, therefore no certain decision can be made. In this game, best-reply strategies are not computable even for perfectly rational agents (Binmore 1987).

In the social world, if an actor has to make a decision depending on expectations about other’s expectations, he cannot do it by a pure act of knowledge about others, but by taking an exogenous feature as given. As a consequence, he gives to this feature a certain independence from his personal expectations. This exogenous feature is an emergent property. From this point of view, infinite regress as we find it in the Holmes-Moriarty problem is not only an epistemological problem for the economist who wants to discover why Holmes finally decided to get off the train at Dover or Canterbury, but an epistemological problem (defined as a problem of knowledge) for Holmes himself, who cannot decide uniquely from his expectations about Moriarty’s expectations. Here, the epistemological and ontological sides of emergence overlap because the epistemic limitations of agents - isomorphic to the epistemological limitations of the economists - is a characteristic of the social ontology. There is a symmetry between the epistemological problems of economists (the limit of methodological individualism) and the epistemic problems of agents (forming expectations about the social world): economists and agents cannot know were an institution come from and how it emerged. This symmetry leads to support our view of the ontological features of the social world: institutions are necessarily independent
from individual agents and can be seen as emergent properties (we will later detail the causal link involved between agents and institutional structure). This is why we call our argument for emergence an argument of symmetry between agents and economists. Finally, by defining downward causation as a specific solution to self-referential problems for the social agents, we claim that downward and same-level causation can be seen as independent from each other. Thus, taking Kim’s argument, we will show that a set of microfoundations $P^*$ can be the result of two causal forces: firstly, from microfoundations $P$, secondly from the institution $M$.

The main conclusion of this argumentation is that some institutions are emergent properties that solve self-referential problems the social actors are confronted with.

For simplicity, we describe an interaction involving only two players, though any $n$-players game would do as well. Say that two individuals, Nicolas and Cyril, arrive simultaneously at a crossroads from perpendicular directions. This situation can be described by the following 'stop-and-go game':

|        | Cyril
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Go</td>
</tr>
<tr>
<td>Go</td>
<td>0; 0</td>
</tr>
</tbody>
</table>
| Nicolas
| Stop  | 1 ; 3 | 1 ; 1 |

Both players have two pure strategies available, to go or to stop. The matrix describes the consequences of any strategy combination in terms of utilities for each player. The situation is problematic from each player’s point of view because neither Nicolas nor Cyril has a clear-cut strategy: Nicolas’ best choice depends on what Cyril decides and the same is true for Cyril. However, contrary to the Sherlock Holmes/Moriarty game of section 3, this game has fixed-point solutions even if players only play pure strategies. Indeed, if Nicolas decides to go, Cyril prefers to stop and Nicolas does not regret his choice. Similarly, if Cyril decides to go, Nicolas would prefer to stop and Cyril would be happy with his initial choice. Therefore, there are two Nash equilibria in pure strategies, namely (Stop; Go) and (Go; Stop). There is also a mixed-strategy Nash equilibrium where each player decides to go with a probability of $2/3$. The interesting point is that this simple interaction clearly involves an intersubjective dimension. Clearly, what Nicolas will do depends on what he expects (or believes) Cyril will do; but what Cyril will do depends on his beliefs of what Nicolas will do, meaning that Nicolas’ beliefs
are about Cyril’s beliefs about Nicolas’s beliefs, and so on. It is a well-known result in game theory that for players to be expected to play a Nash equilibrium, they must have common knowledge of three features (Aumann and Brandenburger 1995; Vanderschraaf 2011): 1) of the structure of the game, 2) of each player’s instrumental rationality and 3) of their mutual beliefs. This third condition is required when there is more than one Nash equilibrium because the contrary leaves players unable to form convergent expectations. On the contrary, common knowledge of mutual beliefs together with common knowledge of rationality implies that players will form common priors on the way the game will be played and so we can expect them to coordinate on one of the Nash equilibriums.

This example is reminiscent of the methodological individualism problem: there must be ‘something’ that makes this coordination possible. Game theory can help to account for this ‘something’ through the concept of correlated equilibrium (Aumann 1987; Gintis 2009; Vanderschraaf 1995; Hédoin 2012). Basically, a correlated strategy is a strategy that is made contingent on an external signal. It is of the type ‘if context A occurs, then play R’. A correlated equilibrium is simply a Nash equilibrium of correlated strategies. The crucial difference is that in the preceding example, players were deemed to form their beliefs independently, while correlated strategy is due to the collective following of an external rule. This is the heart of the David Lewis concept of convention, i.e. a behavioral rule deduced by agents from a state of affairs A: if context A occurs, I know I have to play R. The notion of convention presupposes such an epistemic frame in the sense that one has to identify the different possible strategies (R or R’) in a specific context. A convention is therefore at the same time a frame of the decision-making problem, and a solution to it. In other words, a convention is at the same time a constitutive epistemic device and a regulative one, what is in fine a correlated equilibrium.

5. RE-DEFINING EMERGENCE

1 This is the Harsanyi-Aumann doctrine (Hargreaves-Heap and Varoufakis, 2004). Loosely, it states that two persons having common knowledge of their (Bayesian) rationality and of their posterior beliefs cannot agree to disagree (i.e. cannot have different prior beliefs). If they have different posterior beliefs, then it necessarily means that at least one of them has some private information. Each individual will then use Bayesian updating until beliefs converge. The fact that this process is fictitious allows one to claim that the two persons must have the same priors.

2 Neither Binmore (1987) nor Koppl and Rosser (2002) seek the solution to the problem of non-computability in the existence of institutions. Binmore (1987), using the computer metaphor, explains that humans-as-machines can make choice thanks to a ‘guessing algorithm’ which has been selected through an evolutionary process. Koppl (2008a) use the problem of non-computability as an argument against planning and argue that the market process is efficient precisely because it does not try to solve a non-computable problem. Note that both Binmore and Koppl ultimately use an evolutionary argument to escape from non-computability.
Let us set up a definition of social institutions as emergent phenomena. We have to re-define the three elements of emergent property’s definition: supervenience, irreducibility and downward causation.

5.1. Supervenience

As we defined above, in the philosophy of mind, systems that have the same basal conditions must be alike in respect of their emergent properties. This definition of supervenience does not fit with our characterization of institutions as an external feature that helps people solve multi-equilibria indeterminacy. Given this last characteristic, contrary to the standard definition of supervenience, in the social world, some systems that are alike in respect to basal conditions can have different emergent effects. We must give a broader definition of social supervenience. In respect to John Searle’s theory of the social world (Searle 1995; 2010), social facts are ontologically subjective, that is to say that they depend on the representations of society’s agents. Taking this only characteristic into consideration, we define social supervenience by the fact that if agents would not exist, social facts, i.e institutions, would not exist either.

5.2. Irreducibility

The main conclusion we draw from our definition of institution as multi-equilibria solving devices is that it is impossible to trace back the origins of the institutions in pure individual agents’ characteristics. This implies that it is necessary to take institutions as given in all social models. Our definition is the same than the one we took for the standard definition of emergence. If this conclusion is at first glance an epistemological conclusion, i.e. from the theorist’s point of view, this is either a conclusion that can be drawn from the agent’s point of view: in each game we studied the social scientist as well as the agents can not conclude which equilibrium will be chosen (for scientist) or what behavior to adopt (for the agent).

5.3. Downward causation

Taking Woodward’s definition of causation and the criticism of downward causation we set in part 3, downward causation seems to be nonsense. Nevertheless, given our broad definition of supervenience coming from the idea of multi-equilibria, P does not determine M by itself. Thus, an intervention I upon M is independent from P because there is no clear way from P to M given the fact that people cannot solve multi-equilibra problems. Let us consider that P is a set of agents defined as perfectly rational agents and P* a set of the same
agents when they are coordinated on a coherent strategy. As we set in the previous part, agents in $P^*$ must consider an external rule, $M$, as given in order to reach the coordinating strategy. We know from the previous part that an institution that allows the coordination process cannot itself be deduced from an individualistic process by agents in $P^*$, that is to say that the way from $P$ to $M$ cannot be known for the agents in $P^*$ because of incomputability problems. Thus, $P$ cannot be a cause of $P^*$ in that kind of situation, because the only element that would be added in order for $P$ to cause a coordination strategy is a macro element $M$.

6. INSTITUTION, CAUSALITY AND CONSTITUTIVE RELATIONSHIP

One cannot reduce the difference institutions make to causality. Yet, it seems to be meaningless to speak of human agency without referring to institutions. Hence, the relationship between institutions and individual actions is not only causal, it is also constitutive (Hédoin 2012; Brisset 2015). The idea of constitutive relationships, which for example has been endorsed in social sciences by Vromen (2010) in the case of organizational routines, applies to part/whole relationships. This implies a first qualitative difference with causal relations: constitutive relationships are synchronic. The synchronic dimension is explained by the fact that a whole is defined by its parts, and that a property of parts is a property of the whole. Hence, if we change one property of either the whole or its parts, we logically change the property of both the whole and the parts. An important corollary is that constitutive relationships are symmetric while causal relationships are asymmetric. Causation implies a one-directional relation between the cause and the effect; constitutive relationships are bi-directional or more precisely non-directional. Hence, it is meaningless to separate – even counterfactually – both ends of a constitutive relationship because then we lose the very nature of this relation.

This last point underlines a second key difference between causal and constitutive relationships. While the former is amenable to counterfactual analysis, the latter is not. That is to say, if the relation between $A$ and $B$ is causal, we could ask “what if not-$A$?”. However, if the relation between $A$ and $B$ is constitutive, then the last question is meaningless because $B$ necessarily implies $A$. One cannot imagine a counterfactual world where there would be $B$ and not $A$. Therefore, a constitutive relationship between two sets of entities or two sets of properties $A$ and $B$ can be defined as a relationship where the occurrence of $A$ (resp. $B$) logically entails the occurrence of $B$ (resp. $A$). Hence, to say that institutions are constitutive of individual agencies means that the existence of institutions entails the existence of human agency (this is a specific instance of
supervenience) while human agency logically entails the existence of institutions. To use the influential Aristotelian Four Causes Theory, institutions are some material causes of agencies’ behaviors, and not some efficient causes. Indeed, institutions are to a certain extent the material social agencies are made of, because of the intrinsic computational problems of social relationships. So it cannot be taken apart from these agencies: they necessarily take institutions as given, as external devices. Our argument support the same sort of understanding of emergence we find in the works of Lawson or Hodgson, but try to provide an original ontological reasoning about it. In order to justify his vision of social ontology, i.e made of emergent social structures not reducible to agents, Lawson uses, following Bhaskar and Critical Realism, a transcendental argument:

(1) Actions are generally successful.
(2) These actions are based on beliefs about what make them feasible.
(3) Such beliefs cannot be made of regularities per se but have to be based on conditions that make for it.
(4) Therefore, it must be social structures (what we call institutions) that render acts feasible.

To sum up, human actions would be impossible without social institutions taking the form of non-observable deep regularities of the structures of the social world that involve regularities of beliefs. My own argument consists in the same kind of transcendental argument, which I based on the argument of computability: individuals have no choice but accepting institutions as ways of proceeding that co-ordinate social interactions. I am agree with Lawson (2013) when he defends that it makes no sense to define causation in terms of an emergent entity having a causal effect by way of causally impacting its own components. If there is causality, it comes from organizational structure. He provides the example of house, which is at the same time a set of different components, and an organizational structure. If it is clear that the house as en entity cannot causally impact its owns parts, the organizational structure can do it. Following this analogy, causality from the higher level to the lower one is formal causality. I think the analogy of the house hide the fact that the basic components of the institution, individuals, have no choice but to consider the structure of institutions as given to be what they are what they are, social agents. In contrast, a brick does not need the house to be a brick. That is why I consider institutional structure as material causes as well as formal causes of individuals behaviors.
7. INSTITUTION, CONTRACT AND SOCIAL MARKING: TWO KINDS OF DOWNWARD CAUSATION.

Emergence and downward causation are a necessary tool, ontologically founded. This foundation, as we saw, rests on the necessity for agents to take as given some social devices that indicate which strategy to play in specific coordination contexts. We will not evoke two different cases of subjective downward causation. One concerning the possibility of promise and contract; another one concerning the resistance towards market organization. In both cases, the concept of downward causation is a direct opposition to the extreme, and in an actually scarce, constructivism stance. That is to say the defense of a pure subjective and malleable social reality: conventional representations shared in a population are at the same time necessary to interact with autrui and very restrictive where one wants to set a new social organization.

7.1. Institution, promise and contract

Beyond the simple institutional devices presented in section 4, one of the most fundamental institutions concerning economic life is the respect of the set of rules (we can call contract) that frames the exchanges. If this question runs throughout the entire history of economic thought, David Hume (1740) clearly drafted it: the guarantee that the contractual rule is an essential point for a mutually beneficial exchange to take place. Nevertheless, no one can logically deduce such a guarantee from purely individualistic reasoning: how to know if alter will respect his promise to give me what I want against what he wants since it is impossible to predict and control the future state of mind (the desire or not to fulfill the contract) of our contractor? Above all when one introduces a temporal gap between the contract and the exchange.

"But'tis certain we can naturally no more change our own sentiments, than the motions of the heavens, nor by a single act of our will, that is, by a promise, render any action agreeable nor disagreeable, moral or immoral" (Hume 1740, 332)

So to participate in an exchange, I need to be embedded in a set of common rules, some institutions that structure social interaction, in this case, the guarantee that alter will respect his promise. Such an institutional structure leads to a general trust that is more than a simple "lubricant of a social system" (Arrow 1974, 23), but a necessary condition of the existence of the economic system. So
agents need to take as a priori a given set of rules that guarantees the respect of the contract and that cannot be deduced only from others’ behaviors since we are in the kind of intersubjective dimension (and computational problem) we evoked in section 4: agent $\beta$’s participation to exchange depends on his expectation of agent $\mu$ to respect the contract, that depends on his expectation about propensity of $\beta$ to respect the contract, that depends on $\beta$’s expectation about propensity of $\mu$ to respect the contract, and so ad infinitum. And the same question concerns the potential authority that is in charge of enforcing the contract: who controls the controller? Something has to make coordination possible: an institution, a correlated equilibrium à la Aumann. Since such equilibrium cannot be deduced from some expectations about others’ behaviors, agents must take the “trust equilibrium” as given. As pointed out above, the classical methodological individualism is logically a dead-end since the conditions ensuring that individuals participate in the exchange (in a game) are not limited to their own personal characteristics. These conditions also include social intersubjective characteristics (some common priors). Following our argument of symmetry between economists and agents, this limit of the so-called methodological individualism is also supported by the individuals who must consider as given some institutions as coordinating devices, such as the "trust equilibrium" that ensures that individuals participate in the economic life. The relation between agents and this “trust equilibrium” is an ontological and constitutive one since without it there is simply no social life.

7.2. Downward causation and the institutional rejection of the market

The Nobel Prize winner Alvin Roth (2007) points out how difficult it is to set a market mechanism when one is confronted with what he calls repugnant market. A market is said to be repugnant when the existence of a supply and a demand is not sufficient to allow the creation of a market: ‘even where there may be willing suppliers and demanders of certain transactions, aversion to those transactions by others may constrain or even prevent the transactions’ (Roth 2007, 40). This is not a new idea. The repugnance towards the market treatment of some kinds of objects that belong to other institutional fields is reminiscent of the famous Polanyi’s concept of fictious commodity, which is a commodity that cannot be exchanged in a market system regarding its social statute (Polanyi 1944). Examples of these kinds of commodities are well-known: blood, eggs, sex with a prostitute, organs and so on. The point is not that these commodities cannot be exchanged but that there are some beliefs in the social world that indicate them as unmarketable: only some other distribution structures will be accepted.
Roth especially emphasizes the specific case of the *inter vivo* kidney transplantation. Confronted with the increasing waiting list, economists’ first reaction is to propose the market as the ideal solution. Following the standard economic paradigm, decentralized exchange would reach the social utility maximum (Becker and Elias 2007). Nevertheless, as soon as a market has been proposed to solve the problem of *inter vivo* kidney transplantation, strong criticism led to the ban of bio-markets in the USA (Radcliffe, et. al, 1998). Two kinds of reasons for this rejection can be identified. First, Frey and Oberholzer (1997) show that market remuneration of organs can make the supply decrease because, they say, the market trade of organs reduces the *intrinsic motivation*, that is, the motivation linked to the act itself, independently from any income. This kind of reason refers to the efficiency of the market.

Some other studies (Cohen 2003) show that the implementation of a bio-market can reduce the liberty of choice of a poor population because the organs become a last resort collateral for credit. Such a situation is judged as unacceptable by a great majority of citizens since it is considered as an inadmissible domination from a certain part of the population of the world on another one. This second kind of reason refers to the immorality of the market.

Both cases take root in repugnance toward the market. In the first one, people do not accept to give organs for cash because an organ is simply not considered as a commodity in the sense that it is not socially accepted to be paid for it. The second case also relies on the same mechanism. Considering the fact that, as supported by Becker and Elias, there is no objective reason to distinguish between a classical market relation and the market for organs since all market relations could be interpreted as a domination relation, there is a social device which indicates that an organ is not a commodity. This is what Viviana Zelizer (1985) calls social marking. Social marking is a cognitive phenomenon of classification: for a specific object, a market can be set up if and only if this object is marked as a good. The repugnance is a clear manifestation of institutional. With respect to Lewis’ definition of convention, social marking can be seen as a framing device, for instance: “an organ is a good”, indicating a way to behave: “a good is an object that can be traded in a market”. As a result, social marking works as a conventional environment toward which market design has to be coherent.

So, our first conclusion is that the repugnance towards the market is the effect of a specific institutional device (an organ is not a commodity) that indicates how people have to behave: do not use the market as a social device for organ trading. This is a case of downward causation from an institutional emergent characteristic towards agents’ behaviors. If not, sanction can take place, such as legal sanction. But this is not the only way the social institution of the *repugnant market* impacts the social world. This repugnancy constitutes a limit for market logic at large (Satz 2008). For instance, the *market design*, that is the fact for an economist to
implement some markets, as the market for organs proposed by Becker and Elias (2007) or Kaserman and Barnett (1991), is also limited by conventions that can also take the form of a simple tacit resistance as a formal law system. In this way, Alvin Roth (2007) explains how he (Roth, Sönmez and Ünver 2004; 2005) designs a matching market, the Kidney Exchange Clearinghouse, in order to bypass this conventional limit. In this case, the market can be substituted for a trade system between some pairs of incompatible people. It is no more an exchange connecting one person who needs an organ for money and another who needs money for an organ. However, this is a gift-giving system between pairs of incompatible people: one who needs and one who gives a kidney, which is exchanged with another pair (Steiner 2010). Such a system permits to get out of a market system that was not accepted, and to build a structure that fits into the institutional nexus.

This example indicates that an institutional device defined as a correlated equilibrium is at the same time a coordination device and a framing device: it shapes the way agents see the world and the different strategies they have. In the previous section, we show how an institution such as trust is necessary to contract. In the case of the bio market, the question concerns the type of contract is socially acceptable or not. Here again, it is a coordination problem: how to classify and how to behave in different cases, with different kinds of objects?

8. Conclusion

The goal of this article was to show that a new definition of emergence that is founded on a characteristic of the social world, that is the necessity of coordination, permits to oppose the traditional criticism against this notion. If individuals are epistemically obliged to take as given some coordination devices, exactly as the economist is, so these devices can be seen as emergent phenomena: they supervene upon agents, not logically reducible to them, and there is downward causation since peoples cannot reduce them the agents around them. So emergent phenomenon are, in social sciences, not epiphenomenal. Our demonstration permit to clarify what can of ontological commitment can be made in social sciences. One could say, for instance, that there is something real behind the concept of institution defined as a correlated equilibrium or as a convention à la Lewis.
REFERENCES


2016-01  Christian Longhi, Marcello M. Mariani & Sylvie Rochhia  
*Sharing and Tourism: The Rise of New Markets in Transport*

2016-02  Nobuyuki Hanaki, Eizo Akiyama & Ryuichiro Ishikawa  
*A Methodological Note on Eliciting Price Forecasts in Asset Market Experiments*

2016-03  Frédéric Marty  
*Les droits et libertés fondamentaux à l'épreuve de l'efficacité économique : une application à la politique de la concurrence*

2016-04  Alessandra Colombelli, Jackie Krafft & Marco Vivarelli  
*Entrepreneurship and Innovation: New Entries, Survival, Growth*

2016-05  Nobuyuki Hanaki, Angela Sutan & Marc Willinger  
*The Strategic Environment Effect in Beauty Contest Games*

2016-06  Michael Assous, Muriel Dal Pont Legrand & Harald Hagemann  
*Business Cycles and Growth*

2016-07  Nobuyuki Hanaki, Emily Tanimura & Nicolaas Vriend  
*The Principle of Minimum Differentiation Revisited: Return of the Median Voter*

2016-08  Charles Ayoubi, Michele Pezzoni & Fabiana Visentin  
*At the Origins of Learning: Absorbing Knowledge Flows from Within or Outside the Team?*

2016-09  Dino Borie  
*Error in Measurement Theory*

2016-10  Dino Borie  
*Expected Multi-Utility Representations by “Simplex” with Applications*

2016-11  Dino Borie  
*Additively Separable Preferences Without the Completeness Axiom: An Algebraic Approach*

2016-12  Frédéric Marty  
*Professions réglementées du droit et aiguillon concurrentiel: réflexions sur la loi du 6 août 2015 pour la croissance, l’activité et l’égalité des chances économiques*

2016-13  Agnès Festré & Pierre Garrouste  
*Hayek on Expectations: The Interplay between two Complex Systems*

2016-14  Dino Borie & Dorian Jullien  
*Weakening Description Invariance to Deal with Framing Effects: An Axiomatic Approach*

2016-15  Pascal Seppecher, Isabelle Salle & Dany Lang  
*Is the Market Really a Good Teacher? Market Selection, Collective Adaptation and Financial Instability*

2016-16  Muriel Dal Pont Legrand & Harald Hagemann  
*Business Cycles, Growth and Economic Policy: Schumpeter and the Great Depression*

2016-17  Dorian Jullien  
*All Frames Created Equal are Not Identical: On the Structure of Kahneman and Tversky’s Framing Effects*

2016-18  Dorian Jullien  
*À la recherche de la signification perdue du paradoxe d’Allais*
<table>
<thead>
<tr>
<th>Year</th>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-19</td>
<td>Tom Juille &amp; Dorian Jullien</td>
<td><em>Narrativity from the Perspectives of Economics and Philosophy: Davis, Ross, Multipleselves Models... and Behavioral Economics</em></td>
</tr>
<tr>
<td>2016-20</td>
<td>Dorian Jullien</td>
<td><em>Under Uncertainty, Over Time and Regarding Other People: Rationality in 3D</em></td>
</tr>
<tr>
<td>2016-21</td>
<td>Cyril Hédoin &amp; Lauren Larrouy</td>
<td><em>Game Theory, Institutions and the Schelling-Bacharach Principle: Toward an Empirical Social Ontology</em></td>
</tr>
<tr>
<td>2016-22</td>
<td>Stéphane Gonzalez &amp; Aymeric Lardon</td>
<td><em>Optimal Deterrence of Cooperation</em></td>
</tr>
<tr>
<td>2016-23</td>
<td>Ismaël Rafai, Sébastien Duchêne, Eric Guerci, Ariane Lambert-Mogiliansky &amp; Fabien Mathy</td>
<td><em>Ethnic and Racial Disparities in Children’s Education: Comparative Evidence from Ethiopia, India, Peru and Vietnam</em></td>
</tr>
<tr>
<td>2016-24</td>
<td>Nobuyuki Hanaki, Alan Kirman &amp; Paul Pezanis-Christou</td>
<td><em>Counter Intuitive Learning: An Exploratory Study</em></td>
</tr>
<tr>
<td>2016-25</td>
<td>Mohamed Arouri, Adel Ben Youssef &amp; Cuong Nguyen</td>
<td><em>Ethnic and Racial Disparities in Children’s Education: Comparative Evidence from Ethiopia, India, Peru and Vietnam</em></td>
</tr>
<tr>
<td>2016-26</td>
<td>Amel Ben Abdesslem &amp; Raphaël Chiappini</td>
<td><em>Cluster Policy and Firm Performance: A Case Study of the French Optic/Photonic Industry</em></td>
</tr>
<tr>
<td>2016-27</td>
<td>Patrice Bougette, Hugues Bouthinon-Dumas &amp; Frédéric Marty</td>
<td><em>Vers une prévisibilité des sanctions concurrentielles ? La procédure de transaction à la Macron</em></td>
</tr>
<tr>
<td>2016-28</td>
<td>Aymeric Lardon</td>
<td><em>Jeux de coalitions et oligopoles</em></td>
</tr>
<tr>
<td>2016-29</td>
<td>Gérard Mondello &amp; Evens Salès</td>
<td><em>Tort Law under Oligopolistic Competition</em></td>
</tr>
<tr>
<td>2016-30</td>
<td>Nicolas Brisset</td>
<td><em>Institutions as Emergent Phenomena: Redefining Downward Causation</em></td>
</tr>
</tbody>
</table>