

Gräbner, Claudius; Ghorbani, Amineh

Working Paper

Defining institutions - A review and a synthesis

ICAE Working Paper Series, No. 89

Provided in Cooperation with:

Johannes Kepler University Linz, Institute for Comprehensive Analysis of the Economy (ICAE)

Suggested Citation: Gräbner, Claudius; Ghorbani, Amineh (2019) : Defining institutions - A review and a synthesis, ICAE Working Paper Series, No. 89, Johannes Kepler University Linz, Institute for Comprehensive Analysis of the Economy (ICAE), Linz

This Version is available at:

<https://hdl.handle.net/10419/193628>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Defining institutions – A review and a synthesis

Claudius Gräbner and Amineh Ghorbani

Defining institutions – A review and a synthesis

Claudius Gräbner and Amineh Ghorbani*

Despite being a key term in institutional economics, the term 'institution' seems to be used in various different ways. To what extent is this problematic, given that a shared understanding of key terms among scholars is a necessary condition for scientific progress?

We review prominent definitions of the term ‘institution’ and identify a set of shared elements of these definitions, which can be summarized as institutions being “codifiable systems of social structures (in particular norms and rules) that lead to inclinations for people to act in specific ways”. This suggests that the shared understanding of ‘institutions’ is sufficiently concrete.

Keywords: Definitions, institutions, epistemology

JEL classification code: B25, B4, B52

1. Introduction

The important role of institutions in economic systems is now widely recognized. They are said to determine not only growth (e.g. Rodrik et al. 2004), but also the distribution of income, wealth, and power (e.g. Bowles and Gintis 2002, Veblen 1899), to name just a few examples. Given the prominence of institutional economics, it seems surprising at first that authors tend to use the term ‘institution’ in quite different ways. For example, Calvert (1995) defines institutions as a certain kind of equilibria in coordination games, Aoki as cognitive media (Aoki 2011), and Hodgson (2006), building upon North’s (1990) seminal contribution, considers

* Claudius Gräbner (corresponding author): Johannes Kepler University Linz, Austria; Amineh Ghorbani: Technical University of Delft.

An earlier version of this paper has been presented at the Annual Conference of the *World Interdisciplinary Network for Institutional Research* (WINIR) in Utrecht in September 2017, and the Annual Conference of the *European Association for Evolutionary Political Economy* (EAEPE) in October 2017 in Budapest. We thank all the participants for their helpful feedback. Furthermore, we are particularly grateful to Wolfram Elsner, Svenja Flechtner, Geoffrey Hodgson, Christian Kimmich, Rolf Kunneke, Alex Lascaux, Pavel Pelikan, Paolo Ramazzotti, Alice Nicole Sindzingre, Antoon Spithoven, William Waller and Ulrich Witt for their extensive feedback and helpful discussions. All remaining errors are our own. CG acknowledges funding from the Austrian National Bank (OeNB) under grant number 17383 (*Jubiläumsfonds*).

institutions as “rules that structure social interaction” and rejects the adequacy of the equilibrium approach when defining institutions. Thus, there does not seem to be a generic and commonly accepted definition of ‘institutions’ (for a similar diagnosis of the heterogeneity of definitions in institutional economics see also Mayhew 2008 or Scott 2014).¹

At first sight this seems to be highly problematic: Scientific research is always a social process, and communication across scholars is important. For this communication to function properly, scholars must establish shared meanings of central concepts. If two scholars talk about ‘institutions’ but refer to entirely different phenomena, they are unlikely to be successful as a scientific community.

What is needed are shared *taxonomic definitions*. Taxonomic definitions are definitions that delineate certain phenomena and assign names to them. They are not concerned with the functioning of the phenomenon, but seek “to minimize controversy over the kind of objects under investigation” (Hodgson 2018). They are, thus, a pre-condition for further analysis and theorizing.

At the same time, it is asked too much that all scholars make use of the same taxonomic definition of an ‘institution’. For every analysis, the degree of concreteness of the main terms might differ: a case study of the role of institutions for wealth distribution in Sub-Saharan tribes requires a more concrete definition of institutions than a general model of long-term economic growth. Thus, while it is always important to provide explicit definitions about the terms used in an investigation, the level of concreteness of definitions can differ. What is important is that the more concrete definitions are sufficiently similar to each other such that the scientific community can collaborate in crafting better theories of how institutions work: they should be consistent with a more higher-level definition of institutions in general. Thus, taxonomic definitions of institutions must be sufficiently broad to allow for different forms of analysis, and sufficiently narrow to effectively delineate institutions from other phenomena that are not the prime interest of institutional economists. Without such minimum of shared understanding, effective research is not feasible.

¹ Well-aligned terminology is a pre-condition for effective communication, which is a pre-condition for well-functioning science. There is a whole body of literature on the philosophy of communication and collective understanding that backs this statement, on top of the practical comments of institutional scholars cited above (for more thorough discussions see e.g. McCloskey, 2001; van Bouwel, 2009; Jaccard and Jacoby, 2010, pp.16-18).

Thus, having a certain degree of shared meaning is important, and this shared meaning should express itself in commonly shared taxonomic definitions. The present paper seeks to discuss this state of affairs with regard to the term ‘institution’. Based on a – necessarily selective – review of definitions used in economics and related fields, we assess the degree of heterogeneity of definitions of institutions currently in use. While this review is of interest for institutional economists *per se*, it also helps to show that the authors considered here indeed have converged on a certain set of shared ideas of what an institution is. This might be surprising, given the wide scope of the review when it comes to disciplines and economic paradigms. This suggests that the current state of affairs in institutional economics is rather satisfactory. Nevertheless, as we argue in the conclusion of this paper, institutional research could benefit from a more explicit use of taxonomic definitions of institutions in the applied literature.

The rest of the paper is structured as follows: In Section 2, we make certain clarifications with regard to the nature of definitions, the demands they have to meet, and how definitions in models relate to definitions outside models. Section 3 reviews the literature and discusses a selection of existing definitions. Section 4 identifies the shared elements of these definitions and assesses their heterogeneity. Section 5 consolidates our findings and concludes.

2 *On definitions of definitions*

There are different kinds of definitions. Taxonomic definitions are a particular kind of nomological definitions, i.e. definitions that clarify the use of a word (in the present case: the word “institution”). In contrast to other types of nomological definitions, they list certain properties that a phenomena must have in order to be given a certain name. Thus, a taxonomic definition of an institution lists properties that phenomena must have in order to be called an ‘institution’. For example, defining institutions as “rules that structure social interactions” means to say that whenever we find a rule that structures a social interaction, we can call this rule an institution. Good taxonomic definitions in the social sciences refer to the way a word has been used in the scientific community so far, and do not make up entirely new meanings. One says that these definitions do not stipulate too much new meaning. This distinguishes them from most definitions used in mathematics, which tend to stipulate a lot, i.e. they establish a meaning of a term by defining it from scratch (e.g. by defining an Abelian group as a set and an operation that satisfy five axioms). Contrary to such *stipulative* definitions in mathematics, whose major goal is to allow for very precise analysis and logical derivation, the primary goal of *taxonomic* definitions is to establish *shared meanings* about phenomena within a scientific

community, such that the phenomena in the real world can be studied jointly (see Robinson 1962 and Hodgson 2018 for a more extensive discussion of the concept of a ‘definition’). In (institutional) economics, both kinds of definitions are relevant.

To see why we must appreciate that definitions can refer to phenomena in reality *directly* or *indirectly*, depending on the research practice of the scientists involved. Many economists, for example, use models to study real phenomena. This means they study reality *indirectly* (Weisberg, 2007): they build models, study the models, relate them to reality and study their target system through the lens of their model. Economists using game theory to study social interactions take this approach and there is an extensive literature in the philosophy of science dedicated to the question of what models are, how and when they represent a target, and how they become epistemologically meaningful (see e.g. the contributions in Magnani and Berolotti 2017 for further references, and Frigg and Nguyen 2017 for a survey on how models represent their targets). Consequently, the definitions used by these scientists – such as Calvert’s definition of institutions as equilibria in coordination games - often follow the language in which their models are formulated.

Yet models are not the only vehicle to study reality: others study reality *directly* without the intermediation of models (Weisberg, 2007). Definitions used by this group of scientists do not employ the language of a particular form of formal modelling. Rather, they provide a direct description of the phenomenon in the real world to which their investigation refers. Greif’s definition of institutions as systems of particular rules and norms is an example.

The two kinds of definitions are not in competition but often work together: whenever one uses a model to derive a statement about a real economics system, one must define concepts twice. First, one has to define concepts *within the model*. For example, one defines an institution within a game theoretic model as a Nash equilibrium, i.e. a situation in which no player has an incentive to change her strategy. This is a *stipulative definition*, as frequently used in mathematics. Second, one has to define those phenomena in the target system of the model, to which the Nash equilibrium shall refer (one speaks of ‘keying up’ the concept in the model, with the phenomenon of interest in the target system, see Frigg and Nguyen 2017: 93-95). This means, one must list properties that delineate real-world phenomena that are considered institutions (and to which the model applies) from those that do not (and to which the model does not apply). In other words, one must refer to *taxonomic* definitions of institutions. Thus,

the (stipulative) definition of institution in a particular model should refer to and complement, but not replace a taxonomic definition. Greif and Laitin (2004) express a similar view when they “define an institution not as an equilibrium but in a way that distinguishes between the object of study – institutions - and the analytical tools used to study them.” (p. 635). Keeping this distinction in mind is important to appreciate, and to classify the various definitions used in the literature.

To summarize: We can meaningfully assign the term “institution” to particular real-world phenomena by listing the necessary and sufficient conditions that must be met by these phenomena to be called ‘institution’. Such *taxonomic* definitions of an institution should provide a list of characteristics identifying the real-world phenomena to which word ‘institution’ shall be assigned. To conduct a model-based analysis of ‘institutions’, such definitions can be refined in the language of a particular modelling framework to study their functioning more analytically. This involves *stipulative* definitions in terms of the modelling framework used. However, to ‘key up’ the results from the model to the system under investigation a reference to the model-independent taxonomic definition is necessary. The main focus of this paper will be on *non-model-based* definitions, i.e. taxonomic definitions of institutions, whose main goal is to establish a shared understanding across institutional economists.

3 *A review of existing definitions of institutions*

Table 1 summarizes our review of selected definitions. In this section we explain these definitions in a concise way and refer to the original publications for details. The shared elements are discussed in section 4. The selection of the definitions surveyed is based on our perception of which definitions have been particularly influential at the present time and a consideration of the citation counts for the scholars discussed here. While citations are, at best, an approximate measure for academic influence, and our own views as well as our research communities’ do affect our selection, it is beyond doubt that all scholars discussed below have made influential contributions to institutional research, and their definitions deserve particular attention. Moreover, after considering definitions from economics, we also consider definitions

Table 1
Existing definitions of institutions
Non-model-based definitions

<i>An institution is...</i>	<i>Complete definition</i>	<i>Origin</i>
... a system of constitutive rules	Institutions are systems of constitutive rules, i.e. rules of the form “X counts as Y in C”.	Searle (2005)
... a normative action system	Institutions as collectively constructed normative action systems, which are constituted by norms and social practices.	Tuomela (2002; 2003; 2013)
... a structure of conventions and norms that regulate joint activities	Structures of conventions and norms that regulate systems of joint and individual actions directed at an overarching collective end.	Miller (2004)
... a system of rules	Institutions as durable systems of established and embedded social rules that structure social interactions.	Hodgson (2006)
... a structured process of interaction	Systems or structured processes of interaction (collecting together rules, relations, and positions as well as habits and other practices) that are relatively enduring and recognized as such.	Lawson (1997)
...the set of the rules-in-use	The set of rules actually used by a set of (actors) to organize repetitive activities that produce outcomes affecting those (actors) and potentially affecting others	Ostrom et al. (1992)
... a system of rules, beliefs, norms, and organizations	Institutions are systems of social (i.e. man-made and nonphysical) factors that are exogenous to each individual whose behaviour they influence. They consist of rules, norms, beliefs and (potentially) organizations and “provide individuals with the cognitive, coordinative, normative, and informational micro-foundations of behaviour as they enable, guide, and motivate them to follow specific behaviour”	Greif (2006a)
... a cognitive medium	Institutions in their deep structure are commonly cognized, salient patterns of the ways in which societal games are recursively played and expected to be played. Institutions in their substantive forms are social artefacts that ensure the societal games are in equilibrium.	Aoki (2001; 2007; 2011)
<i>Model-based definitions</i>		
<i>Definition</i>	<i>Origin</i>	
Institutions as game forms	Hurwicz (1996)	
Institutions as equilibria	Calvert (1995), Aldashev et al. (2012), Elsner (2012)	
Institutions as correlative devices / equilibria	Gintis (2009), Hedoin (2012)	

from other social sciences and philosophy (see 3.7 – 3.10) to see whether a shared understanding of institutions emerges even across disciplinary boundaries.²

3.1 Institutions as rules of the game

The conception of institutions as rules dates back to Max Weber and Talcott Parsons but was most famously articulated by North (1990). He considers institutions to be “the rules of the game in a society...[and] the purpose of the rules is to define the way the game is played” (North, 1990, p. 3). As we will see below, many have taken this definition as a starting point to develop very sophisticated (model-based and non-model based) theories of institutions as systems of particular rules. For this reason, we will not discuss the original formulation in more detail. Moreover, because of the conceptual criticism of Hodgson (2006) we also do not concern ourselves with his distinction between ‘formal’ and ‘informal’ rules.³ Rather we will scrutinize the approach of those who have taken North’s definition literally and interpreted the “rules of the game” as the constitution of a (game theoretic) game form.

In this case, the “rules of the game” constitute the strategy space and the outcome function of a game. Institutions are then defined as families of game forms that share the same structure, yet may differ in their exact numerical values (Hurwicz 1996).⁴

For such a definition to be useful, three clarifications are needed (Hurwicz 1996): firstly, the area of applicability, i.e. the category of situations and players to which the institution refers – already a hint towards the need for non-model based definitions. Secondly, institutions should not be defined as game forms proper, but as correspondences between environmental factors, which are taken as given and cannot be changed by the people involved, and game forms. This way, the institution refers only to those arrangements that are “human made”, and not those that are “given by nature” (Hurwicz 1996). Finally, only those game forms that are actually effective should fall under the umbrella of “institutions”. This is the same argument as that of (Ostrom 1992) who restricts her definition of institutions to the rules-*in-use*. As we will see below, *assuming* institutions to be effective has been one of the major criticisms of the institutions-as-

² Furthermore, particularly interesting contributions emerge whenever the styles of various disciplines are combined. Guala (2016) provides a nice example: the book was written by a philosopher who uses game theory extensively, and regularly refers to contributions from economics. Similar examples are provided by Binmore (2014) or Elster (2009).

³ North originally distinguished formal and informal rules. Hodgson (2006) criticizes the ambiguity of the term and that it is often not clear whether ‘formal’ means ‘legal’ (and ‘informal’ ‘illegal’ or ‘non-legal’), or ‘explicit’. It seems that the latter interpretation is the most frequent one, but it seems preferable to abandon this ambiguous use of the terms, and to be very explicit about the kind of rules and institutions one is talking about.

⁴ The game form is distinct from a game as such, since the latter also includes the set of players and their payoff functions.

rules approach since there is no reference to the motivation of the agents to actually consider the institution (see also Hurwicz 2008 or Hindriks and Guala 2015a). It becomes immediately clear that this definition of institutions as rules in the narrow sense is a typical *model-based* definition, and more stipulative than taxonomic. It is, thus, less useful in establishing a shared meaning of institutions within the institutional economics research community.

3.2 *Institutions as integrated systems of rules that structure social interactions*

The definition of institutions as *established social rules that structure social interactions* was most precisely formulated by Hodgson (2006). A social rule is considered a “socially transmitted and customary normative injunction or immanently normative disposition that in circumstances *X* do *Y*” (Hodgson, 2006, p. 3). While Hodgson warns to define rules (and institutions) in terms of regularities in behaviour, a rule (and thus an institution) cannot be defined without any reference to its effects: a social rule that is not followed is no institution and “[i]gnored laws are not rules” (Hodgson, 2006, see also section 3.6 on Ostrom). Hodgson concludes that rules cause *dispositions* to engage in particular behaviour rather than actual behaviour. This way, there are similarities to the definition of Lawson (1997) who considers them as “systems or structured processes of interaction (collecting together rules, relations, and positions as well as habits and other practices) that are relatively enduring and recognized as such”.

While this is not part of his definition of an institution, Hodgson also elaborates on the functioning of institutions. As in the institutional theory of Miller (2004), *habit*, plays an essential role. The exact functioning has been operationalized by Hodgson and Knudson (2004) in an agent-based model in which agents need to decide to drive on the left or the right side of the street. This shows that Hodgson’s definition of institutions is broader than that of e.g. Searle, who does not consider traffic rules to be institutions because they are “only” regulative, and not constitutive rules.

Note that rules as defined by Hodgson do not refer exclusively to legal rules, but also to unwritten social conventions or unwritten injunctions to behave in a particular way. This way, many rules remain tacit and people follow them unconsciously. The only essential prerequisite for a normative injunction to become a rule is *potential codifiability*. The justification is the same as it was for Searle (2005). The need for codifiability becomes even more apparent if one

studies how rules actually evolve: since rules must be culturally transmitted via social learning because they cannot be encoded in our genes, they need to be codified so that they can be explained to other people.

3.3 *The set of rules-in-use: the ADICO grammar for social institutions*

The definition of ‘institutions’ as “the set of rules actually used by a set of (actors) to organize repetitive activities that produce outcomes affecting those (actors) and potentially affecting others” used by Ostrom (1992) is very similar to the one of Hodgson discussed above: both consider institutions as systems of social rules, thereby shifting the burden of definition to the concept of a social rule. Crawford and Ostrom (1995) developed the ADICO grammar, which clarifies the essential components of a social rule: *attributes*, *deontic*, *aim*, *conditions*, and *or-else statements* (see table 2).

Table 2

Overview of the ADICO grammar of Crawford and Ostrom (1995) used to describe social rules.

Attributes	The object(s) to which the rule applies to (e.g. all female people in Europe).
Deontic	The specification of the deontic content of the rule: <i>may</i> , <i>must</i> , or <i>must not</i> .
Aim	The action or the outcome to which the deontic refers to (e.g. driving a car in public)
Conditions	Specification of time, place, circumstances, and extent to which the deontic applied to the aim (e.g. “always” or “in large cities”)
Or else	Specification of the sanctions for non-compliance with the rule.

These definitions of rules and institutions are compatible with the concept of Hodgson(2006) - albeit being a little less explicit on the particular nature of the rules: Hodgson defines the effects of institutions more precisely as *inclinations* to do something, and links institutions explicitly with the concept of *habit*. As Hodgson Ostrom suggests a broader definition of institutions than Searle: ADICO is not exclusively directed at constitutive rules (which would be captured as a rule with the deontic *may*, see Crawford and Ostrom, 1995), but also at regulative rules, i.e. rules with the deontics *must* or *must not*.

3.4 *Institutions as systems of man-made nonphysical elements*

According to Greif’s ontological definition, institutions are systems of social (i.e. man-made and non-physical) factors that are exogenous to each individual whose behaviour they influence

(Greif, 2006). Greif considers these social factors as *institutional elements* because institutions are no monolithic entities, but systems composed of interrelated, yet distinct components. For Greif, the relevant institutional components are *rules*, *beliefs*, and *norms*, but also the manifestation of these in organizations. Conjointly, these components are expected to “provide individuals with the cognitive, coordinative, normative, and informational micro-foundations of behaviour” (Greif, 2006, p.14). Consequently, institutions are expected to generate behavioural regularities.

Contrary to other definitions we have encountered before, Greif demands more than only systems of rules. A rule alone, for example, does not cause any individuals to follow it and rules that do not influence behaviour “are not components of an institution” (p. 30).⁵ For rules to be followed, *motivation* is needed (p. 31). This is why Greif defines institutions as *systems* of social elements: one element alone is insufficient. Rules, however, do play an important role in the functioning of an institution: they represent behavioural instructions that help individuals to get a cognitive understanding of their situation and thereby facilitate their decision-making. The motivation to follow rules comes from norms (i.e. socially constructed behavioural standards) or beliefs (i.e. either cognitive models used to understand the decision environment or expectations about the behaviour of others).

The role of organizations in Greif’s account of social institutions is more complicated. On the one hand, they constitute the fourth institutional element considered by Greif: the task of courts (which help to sustain property rights) or driving associations (which help to enforce street rules) is “to produce and disseminate rules, to perpetuate beliefs and norms, and to influence the set of feasible behavioural beliefs” (Greif, 2006, p.37). On the other hand, organizations can be institutions themselves. In these cases, they constitute systems of rules, beliefs, and norms exogenous to each individual whose behaviour they influence, usually the members and stakeholders or the organization. This consideration of organizations is similar to Hodgson’s account of organizations according to which organizations are institutions involving criteria to determine their members, their sovereignty, and their internal hierarchy and responsibilities (Hodgson, 2006). Whether an organization is to be considered an institutional element or an institution itself depends on the situation under study.

⁵ By this he means actual behavior (and not as Hodgson, *potential* behavior), yet by providing the impeachment of the US president as an example for an institutional element he makes clear that the behavior does not need to transpire very frequently.

In Table 3, we illustrate Greif’s definition of an institution via the institution of traffic systems. The effect of the institution as a whole is a certain regularity in behaviour, usually that people drive according to the rules. From our viewpoint, it is worth mentioning that Greif and Laitin (2004) explicitly define institutions in a way “that distinguishes between the object of study—institutions—and the analytical tools used to study them” (p. 635), a practice that reflects our distinction between model-based and non-model based definitions discussed in section 2: “Institutions are not game-theoretic equilibria, games are not the basic unit of institutional analysis, and game theory does not provide us with a theory of institutions. Indeed, the key to advancing institutional analysis by using game theory is precisely to recognize the difference between game-theoretic equilibrium analysis and institutional analysis” (Greif, 2006, p.19).

Table 3

The institution of the traffic system as defined by Greif (2006, p. 38).

<i>Name of the institution</i>	<i>The traffic system</i>		
	<i>Rules</i>	<i>Beliefs & Norms</i>	<i>Organizations</i>
<i>Institutional elements</i>	Rules of the road	Beliefs about the meaning of traffic signs and symbols; belief that drivers and law enforcement officials behave in accordance of the rules	Traffic police, motor clubs, departments of motor vehicles
<i>Implication</i>	Regularity in the behaviour of people participating in official traffic		

3.5 *Institutions as cognitive media*

Aoki defines institutions within his institutions-as-cognitive-media-view (Aoki, 2011). Given his heavy reliance on game-theoretic language, his definition of institutions lies at the edge of model-based and non-model-based definitions. With regard to his definitions, he distinguishes between the more fundamental *institutions in deep structure* and the more concrete *institutions in substantive form*. To understand his definitions, some knowledge of his general theory of institutions is required, which is why we provide a very concise outline of the institutions-as-cognitive-media-view. Aoki (2011) gives a more extensive introduction.

Aoki defines institutions in deep structure (or the “societal rules”) as “commonly cognized, salient patterns of the ways in which societal games are recursively played and expected to be played.” This definition exemplifies his strong reliance on game theory. Nevertheless, the

definition is not model-based in the narrow sense: “Societal games” should be understood broadly as interactions between members of the society.

Aoki then turns to the more specific *institutions in substantive form*, which are essential for the institutions in deep structure to be self-sustaining. For this to happen, the “societal games” must be in equilibrium, and equilibrium usually requires what in game theory is referred to as “common knowledge” (Aumann, 1987). Institutions in substantive forms are social structures⁶ that ensure this.

They represent a social cognitive medium in linguistic form, which helps agents to summarize information about the beliefs and behaviours of other agents. The idea of institutions as means to summarize information required for societal coordination goes back to Hayek (although Aumann did not explicitly refer to him). Hayek studied extensively the cognitive underpinnings of how people classify information about their environment and become capable to act under more complex circumstances (Hayek, 1952). This reasoning later became the foundation for the game theoretic concept of “common knowledge” (Aoki, 2011). Aoki refers to the classical work of Lewis (1969) on the emergence of shared beliefs to identify the conditions under which the assumption of common knowledge is adequate. Yet the essential prerequisite for common knowledge to emerge is a shared public proposition from which agents can infer the current state of the societal game they play, and – using the assumption of symmetric reasoning⁷ - the beliefs of other agents. This way, the public proposition forms the “basis for common knowledge” (Lewis 1969) and summarizes information about the current state of the societal game. Aoki now defines these public propositions as the *institutions in substantive form*. Examples include written property rights such as “One must not harvest the land of someone else”, employment contracts a la “Whoever does not work with the required effort will be fired” or norms of conduct in a given community such as “Whoever extracts too much of our common pool resource will be ostracized from community interactions” (see Aoki 2001 for more examples). As these examples show, the substantive forms of institutions usually have the syntactic form of *regulative* or *descriptive* rules and they tell agents how they should play the

⁶ Originally, Aoki speaks of *social artefacts*. Yet, this term is rarely used in the social sciences, and much less understood than the term *social structures*, which will be used instead in this paper and which usually denotes social relations in their broadest sense. It is likely that *social structures* are what Aoki had in mind.

⁷ According to the assumption of symmetric reasoning, the public proposition indicates to every agent that every other agent also believes in the proposition and derives the same inferences from it. This assumption is strong but may be justified on the ground of recent results in the cognitive sciences according to which human beings are very good at pattern recognition (Aoki, 2011). Since institutions represent a pattern of play that is commonly perceived by all the agents, agents can reasonably be assumed to perceive and process these patterns. Further justifications refer to shared culture, which usually encapsulates past experience that is shared by the majority of the members of society.

societal game they face. Here, Aoki (2011) directly refers to Searle and makes clear that for constitutive rules to work, people must have a collective understanding about what is meant by the Y term in the “X counts as Y in C”. Therefore, he argues, there must be more primitive rules such as descriptive rules into which status functions can be nested. In this sense, Aoki considers his definition of an institution to be more generic than the one of Searle.

The *substantive form* of an institution is a social structure in linguistic form on which the agents agree. It is created by all agents together and is neither reducible to agents’ individual attributes nor can it be created by a single agent. This social structure creates an identical belief in the minds of the individuals. It summarizes the shared beliefs of the agents and induces the partial convergence of individual behavioural beliefs. Once these beliefs are such, the states of strategic play are reproduced and the beliefs are verified. This way, the substantive forms of institutions are external to individual agents and play an essential role for equilibrium correspondence between physical states of the game and the mental beliefs of the agents. Practically, it serves as a resource into which agents can off-load some computational burden because they do not need to compute the state of play and the beliefs of the other agents. In this sense, an institution is the substantive form of common knowledge in the game theoretic sense and under certain conditions the strategic choices of the agents are consistent with their mutual beliefs – and a stable state of the societal game emerges.

While the stability of the situation and the institution stem from the theoretical assertion that agents maximize their utility, Aoki makes clear that proactive agents do not always make conscious and optimizing choices: once an institution in substantive form emerges as an equilibrium, following it can also become a habit for the agents, or even an internalized norm according to which a deviation from the equilibrium behaviour would be considered normatively wrong.

3.6 *Institutions as equilibria*

The approach of defining institutions in term of equilibria in games originates from the desire to remedy two shortcomings of the institutions-as-rules approach (see 3.1). Firstly, institutions should be treated as endogenous elements of a theory, not as exogenous constraints. Secondly, researchers wanted to consider the role of motivation to follow an institution in their theory. If these two drawbacks were remedied, one could explain the stability of institutions, and it is this

stability and self-reinforcement that make adherents of the equilibrium approach to claim that “equilibrium is part of the ontological nature of institutions” (Aoki, 2015, p.487).

The origins of the approach are usually associated with Schotter (1981), who proposed to model institutions as equilibria in the framework of non-cooperative game theory. This approach has been refined by Calvert (1995) who advocated the now prominent definition of an institution as the strategy profiles satisfying the conditions for a Nash equilibrium. In its pure form, such a definition considers institutions as a full derivative of individual preferences: “there is, strictly speaking, no separate animal that we can identify as an institution. There is only rational behaviour, conditioned on expectations about the behaviour and reactions of others... ‘institution’ is just a name we give to certain parts of certain kinds of equilibria” (Calvert, 1995).⁸

Other advocates of the equilibrium approach have identified institutions with the means that ensure the convergence to a Nash equilibrium (e.g. Aldashev *et al.* 2012). These authors refer to institutions as something external to the game that selects a particular equilibrium by ensuring the coordination of agent’s expectations, by creating a certain commitment of agents to behave in a certain way, or by providing other agents the ability to enforce a particular behaviour of others. This makes this notion of an institution necessarily incomplete since the authors must specify what exactly the equilibrium selection device is. A way out of this is to look for a *correlated equilibrium*, rather than a Nash equilibrium. To get the gist of the idea, consider a standard normal form game and extend it by letting a *choreographer* (or a *correlating device*) issue a directive to all players on which strategy they should play. If following this directive represents mutual best response for all players, the resulting equilibrium is called a correlated equilibrium. Institutions can then be defined as correlating devices that lead to a correlated equilibrium (Aumann, 1987; Gintis, 2009; Hèdoin, 2012). But in all cases, we face a similar problem as in the strict definition of institution as game forms in 3.1: defining an institution as a Nash equilibrium (or a correlative device) is, strictly speaking, possible only within a game theoretic model. It is not a taxonomic definition as such, and to link a model with real world cases, an additional taxonomic definition of an institution is required.

⁸ We will not deal with the criticism of such a strong interpretation. It has been shown elsewhere that even such individualistic approaches to define institutions always need to refer to some macro social structures (Arrow, 1994; Hèdoin, 2012). See Tuomela (2002) for a concise summary of the main philosophical criticisms of the approach.

3.7 *Institutions as constitutive rules*

The next three definitions come more from the philosophical literature, but might also be of interest to economists, because these definitions highlight very different – but also important and complementary – aspects of institutions. This first definition of this kind comes from John Searle, who links institutions to collective intentionality:⁹ collective intentionality allows people to develop a language and – using this language – to collectively impose functions on objects. These objects do not get these functions because of their physical appearance, but because people collectively accept these objects to have a certain status. A piece of paper, for example, is not money because of its physical appearance, but because it has been collectively accepted that the paper counts as (or “has the status of”) money. Thus, it can fulfil the function of a mean for exchange. It has this function only because its status as money is collectively accepted. Searle calls such functions “status functions” and they are allocated to objects via collectively accepted *constitutive rules*. Constitutive rules come in the form of “X counts as Y in C” and institutions are – according to Searle – systems of constitutive rules.

To reach this definition, Searle elaborates on the following ideas:¹⁰

1. The distinction of institutions and institutional facts.
2. Institutions require collective intentionality and the institution of language
3. Institutions are normative and work through the creation of power relations

The distinction of institutions and institutional facts For Searle, institutional facts are status functions, i.e. functions of objects or people that they cannot perform due to their biological capabilities or physical appearance, but because they are collectively accepted to perform this function. Status functions are imposed on objects via *constitutive rules*, i.e. rules that *constitute* new forms of behaviour. The rules of chess, for example, constitute the game of chess because the game ceases to exist once the rules of chess cease to exist.¹¹ Constitutive rules usually come in the form of “X counts as Y in context C”. An institution is ultimately defined as a system of constitutive rules. They are, in this sense, systems of collectively accepted rules that enable the creation and maintenance of status functions.

⁹ In social philosophy, many prominent definitions are built on (or in demarcation of) the concept of *collective intentionality*. It refers to the (alleged?) ability of humans to refer themselves to states of the world collectively. See Searle (2010, chapters 2 and 3) for a concise introduction.

¹⁰ An introduction to Searle’s theory can be found in Searle (2010). For a critical introduction see Miller (2004, p.182 ff).

¹¹ Constitutive rules differ from regulative rules, which just regulate existing behavior: traffic rules, for example, regulate behavior, which exists independent of the existence of the (regulative) traffic rule.

Institutions require collective intentionality and language Just as imposing the function of cutting on a stone requires individual intentionality, imposing *status* functions on objects requires *collective* intentionality. Searle insists that “[a]ll intentionality, whether individual or social, must exist in individual heads” (Searle, 2010). The individual *intentions* in collective action only refer to actions that the individual herself can engage in, but for collective intentionality, these intentions must be accompanied by a *belief* that the other members of the collective are contributing their part such that the collective goal can be achieved. An important question that many economists have tried to answer is: what conditions are necessary for a belief to be stable? This question does not, however, play a central role in Searle’s theory of institutions. He rather assumes stability (and provides a nice example of how other disciplines can benefit from results in economics).

Aside from collective intentionality, imposing status functions requires some form of codification: otherwise it would be impossible to assign a function to something that does not have this function by virtue of its physical or biological appearance. One could give a hammer to someone, and this person would be able to use it without any manual or explanation. This does not work in the case of money: for paper to function as a mean for exchange, there need to be a collectively accepted representation of the paper as money, and a declaration of the status function of money. Without codification, such a representation and declaration – and in consequence the institution of money – cannot exist.¹²

Institutions are normative and work through the creation of power relations The channel through which institutions affect individuals is the creation of deontic powers, such as rights or obligations. These powers originate from the status that is assigned to a person or object. For instance, once a person has the status of a police officer and other people recognize this, the person has, in virtue of the status of a police officer, certain rights, duties, and obligations.¹³ Searle is not very clear on how these deontic powers affect the behaviour of human beings: he claims that they create *desire-independent* reasons for action, but it is not very clear what is meant by this (see also the critical remarks by Hindriks and Guala 2015b).

¹² Searle is not really interested in why people agree on constitutive rules, how they are committed to follow them – a central question for many economists at least since North and Weingast (1989) – and how they persist. In this way he does not refer at all to the functioning of institutions.

¹³ For critical remarks on the link between the accepted status and deontic powers see Miller (2004).

3.8 Institutions as collectively constructed normative action systems

In contrast to Searle, Raimo Tuomela does not rely on the (in the philosophical literature controversial¹⁴) speech act theory. Fundamental to his definition is the distinction between the *we-mode* and the *I-mode*. An individual who acts in the *we-mode* makes his decisions not on an individualistic base, but as a group member. When acting in the *we-mode*, people are able to engage in joint social practices and confer a particular status on these practices by collectively accepting a system of norms. An institution is then defined as the resulting system of a social practice and the associated norms, or, more precisely as *collectively constituted (i.e. constructed) normative action systems*.

The two ingredients for institutions in this sense are (i) social practices and (ii) a system of social norms. Tuomela's definition is thus twofold, referring to institutions both in terms of the induced behaviour as well as the social facts and norms underlying this behaviour. To appreciate this definition, we need to introduce some elements of the theory of sociality proposed by Tuomela. For him, a social practice and the imposition of a particular status on a practice via a system of social norms require people to act in the *we-mode*. Thus, we need to clarify (i) the distinction between the *we-mode* and the *I-mode* and (ii) the concept of a collective social action.

Table 4

The different modes of reasoning in Tuomela's social theory.

I-mode			We mode
Plain	Pro-group		
Promoting the individual interests (even in group contexts)	Promoting the group interests		Acting in the we-form. This way, the “artificial” group agent is the primary actor, not the members. ¹⁵
Individuals are committed only to themselves to further their interests	Individuals are committed to themselves to further the group interests. ¹⁶		Individuals are committed to the group to further group interests.

Distinguishing the we- and I-mode Central to Tuomela's approach is the distinction between the *we-mode* and the *I-mode*, as summarized in table 4.

¹⁴ For a critical discussion see e.g. Grewendorf and Meggle (2002).

¹⁵ Translated into classical game theory, a group agent can select the outcomes or the games, and not, as individuals, their strategy. This is because the group agent can select the strategies to be played by all agents.

¹⁶ In a classical game theoretic setting, this would be the case when the individual preference relation would take into account the goals of the group.

Acting in the I-mode can happen in two ways. In the plain I-mode, the individual follows his own (“selfish”) interests. In the pro-group I-mode it also acts according to his interests, which take into account the well-being of other group members. This captures the idea of other-regarding preferences, which are well-known in the economics literature: if individuals have other-regarding preferences they make their choices individually (i.e. based on their own utility function), but take into account the utility of others.¹⁷ Acting in the we-mode, by contrast, means acting as a group member, i.e. to commit one’s interests fully to the group. Although Tuomela is more interested in the consequences rather than the sources of these commitments, analogies to the economic discussion of commitment are obvious – be it on the small scale when it comes to team reasoning (Sugden, 2000) or on the societal scale when it comes to stability of political constitutions (North and Weingast, 1989) or the occurrence of political transitions (Acemoglu and Robinson, 2001). The disciplinary integrability of Tuomela’s work is also evidenced by the game theoretic work inspired by his theory (e.g. Hakli et al. 2010).

A central claim of Tuomela is that thinking in the we-mode is not only irreducible to thinking in the I-mode (see also Gräbner and Kapeller, 2017), but that the aggregated choice of a group of agents with group-oriented preferences may differ from the choice made by a group agent constructed by the same agents. Hakli et al. (2010) use the game theoretic model of Bacharach (1999) to formally study situations, which are perceived as situations of group decision making.¹⁸ Bacharach proved that the set of equilibria in such situations depends on whether agents reason in a pro-group I-mode or we-mode. One should note that in terms of preferences, the situation in which agents reason in the I-mode does not differ from the situation in which the agents reason in the we-mode: in both cases group members would like to maximize group welfare. Yet, the two cases differ with respect to the mechanisms with which the decisions in the group are made.¹⁹

What is a collective social action? A collective social action is an action performed for a social reason, e.g. a shared we-attitude a group has agreed upon via collective acceptance. A shared we-attitude entails that all agents in the group have this attitude, all agents believe that the other

¹⁷ Most economic theories analyze individual decision making in the I-mode because individual preferences are taken as a starting point.

¹⁸ See chapter 8 of Tuomela (2010) for a more philosophical argument in favor of the irreducibility of the we-mode to the I-mode.

¹⁹ An obvious question is: when and why do individuals act in the we-mode? Working out a game theoretic model and deriving the conditions under which agents are willing to commit themselves to a group would ignore Tuomela’s claim that it is not possible to reduce the I-mode to the we-mode. For empirical studies see Colman et al. (2008), Krueger (2008), Sugden (2008) and their interpretation in Hakli et al. (2010).

group members have this attitude, and believe that all others have them, which again is believed by all agents (etc.).

After having clarified his basic terminology, Tuomela's definition of social institutions becomes clearer:

1. Individuals construct a group in the we-mode via collective acceptance.
2. The individuals engage in collective social action. The reasons for their behaviour come from "shared we-attitudes" of the group.
3. If the collective social action gets repeated, we can speak of a social practice.
4. The group confers a special status to the social practice (or an item that plays an important role during this activity) by collectively accepting a system of social norms in the we-mode that confers the status on the social practice.
5. The result is a social institution, i.e. a social practice and a corresponding system of norms.

Note that – as in Searle's approach – normativity is an indispensable element: firstly because of the norms required for the construction of institutions, secondly because the agents need to act in the we-mode, i.e. as group members, which logically entails that they can act either right or wrong (Tuomela, 2002).

3.9 Institutions as structures of conventions and norms that regulate joint activity

Contrary to Searle and Tuomela, Miller (2004) rejects the idea of primitive we-intentions and refers only to individualistic notions. He also denies that any form of collective acceptance is required for institutions such as money: if people use money as a mean for exchange, the only thing they need to agree upon is the function of money. The same, he argues, applies to the use of language. In many instances, "collective acceptance collapses into actual or realized substantive functionality" (Miller, 2004, p.191).

He defines institutions as "structures of conventions and norms that regulate various species of generic joint activity" (p. 192), the latter being defined as a system of many joint and individual actions directed at an overarching collective end. A collective end is an individual end that can be achieved only by the action of a group. The behavioural building blocks of institutions – conventions – are also understood individualistically, i.e. as joint habits, in the sense of repeated action in given circumstances (Miller, 2004).

Rather than by collective intentionality, an institution is constituted by its functionality to achieve a collective aim. If it does not serve its purpose, it ceases to be an institution: “any given institution is to be understood in terms of the collective end or ends to which their activities are and/or ought to be directed” (p. 198).

This illustrates an important difference to Searle’s and Miller’s account of institutions: for Searle, a person who holds an official certificate giving her the title of a surgeon, is a surgeon. It does not matter whether the person actually has the skills to perform surgery. Miller, on the other hand, claims that a person who cannot perform surgery is not a surgeon – notwithstanding any certificate she might hold for whatever reasons. In this respect, Miller’s approach is compatible with functionalist approaches pursued by many original institutional economists and, more recently, the theory of Guala (2016), which aligns well with his individualistic and functionalist approach.

Miller claims that any institution is necessarily normative: either the end of an institution itself is a moral good, or the end is a mean to achieve a certain moral good. The moral goods serve as the justification of all the norms that govern the activity of the members within the institution. Consider the example of the police officer: according to Miller the deontic powers of the police officer are justified with the end the police officer serves, usually social stability and the effectiveness of law. If the police officer stops serving these ends, the deontic power of the institutions will cease to exist.

3.10 Institutions outside social theory, economics and philosophy

The preceding elaborations have focused on contributions from influential economists and philosophers. While we believe that our selection is reasonable given the high impact of all those authors in their respective field, there are many other approaches to define institutions that are worth mentioning. Due to space limitations we restrain ourselves to a very concise summary of definitions from other social sciences, in particular sociology, anthropology, psychology and political science, well aware that such separation into disciplines is - fortunately - getting out-dated. As shown below, these definitions mostly align well to the definitions discussed above.

In political sciences one might roughly distinguish between a *historical* and a *rational choice* institutionalism (Scott, 2014, p.38). The former bears much similarity to evolutionary institutionalism in economics, and institutions usually understood as the “rules and procedures

that structure social conduct” (Thelen and Setinmo 1992), thereby being very similar to the institutionalist definitions as put forth by Lawson or Hodgson discussed above. While rational choice political scientists differ from historical institutionalists in terms of methods, epistemological convictions, and theories on how institutions emerge, they also view institutions mainly as governance and rule systems (Scott, 2014).

Cognitive scientists and anthropologists have a slightly different perspective on institutions, and usually define them in relation to the functioning of human decision making. Consequently, they are often considered “mental models”, “collective representations” or “behavioural scripts”, but also “values, ideas, and other symbolic meaningful systems” that help individuals to make decisions in the face of complex and uncertain decision problems (see e.g. Scott 2014, p. 45ff for an overview). While such view of institutions is rather narrow and specialized, it has similarities to the (broader) definitions of Greif, Aoki, and Hodgson, and also stresses the importance of codifiability of institutions.

Finally, institutions are traditionally prominent in sociology, although many different approaches to their investigation exist. Yet, when it comes to the *definition* of institutions we again find many similarities both among different approaches, as well as to the definitions discussed above: phenomenologists traditionally take a more structuralist view on institutions by considering them, for example, as “symbolic systems that are ‘experienced as possessing a reality of their own, a reality that confronts the individual as an external and coercive fact’” (Scott, 2014, p.48). While most previous definitions are more agnostic with regard to how institutions are perceived by individuals, the idea of institutions as ‘symbolic systems’ that impact upon individuals is consistent with what has been presented so far.

In all, while different schools and disciplines differ with regard to the preferred methods of how to study institutions, different views of how institutions emerge and evolve, as well as competing views on their relationship to individuals, the *definitions* of institutions seem to align rather well with the definition from economics and philosophy. As the next section will demonstrate this is no coincidence.

4. *Aligning the definitions: identifying shared meanings*

We now come back to our question posed in the beginning: how different is the term ‘institutions’ used in scientific practice? Do we see a shared meaning of the term across institutional economists and other social scientists? To submit an answer to this question we now search for common and divergent elements in the just reviewed definitions (see table 5). Some aspects of institutions are widely agreed upon: Firstly, all definitions summarized above stressed that institutions are *human-made* and that they cannot exist outside the social sphere: there are no institutions in a one-person world.

Secondly, institutions must be *codifiable*, i.e. it must be possible to describe them in verbal or written form: since institutions are social constructs, it must be possible to communicate them from one person to another.

Thirdly, institutions are usually considered *systems* of social structures. This means that institutions have various components that play distinctive, yet related roles. Except Aoki, who remains silent on this, all non-model-based definitions reviewed above (and also Hurwicz’ operationalization of institutions as game forms) consider institutions explicitly as systems or sets of social structures.

Fourthly, while researchers also refer to other social structures (e.g., conventions), rules and norms were mentioned most frequently as the essential components of institutions. This makes sense since rules and norms fulfil two important roles for the functioning of institutions: Rules provide recipes to behave, thus lowering complexity of the decision problem faced by agents – a feature that is regularly highlighted by cognitive scientists working on institutions (see section 3.10). This is spelled out most explicitly by Crawford and Ostrom (1995), Hodgson (2006), and Aoki (2001), but is implicit in all definitions except the one of John Searle, who focuses on constitutive rules alone. Norms provide motivation to follow an institution by relating them to values. It is one of the essential contributions of the (model-based) institutions-as-equilibria approach to stress that one should also pay attention to the sources for the motivation to follow rules and norms, and to relate this with the self-stabilizing nature of institutions. This is a nice example of how models of institutions can constructively inform non-model based definitions.

Table 5

Common elements of the definitions reviewed in section 3.

	Pro	Neutral	Contra	Comments
<i>Institutions are human-made, i.e. social structures</i>	Everybody.	-	-	<i>Clear unanimity.</i>
<i>Institutions are always codifiable</i>	Everybody.	-	-	<i>Follows from the previous aspect and the fact that institutions are passed on among humans.</i>
<i>Institutions consist of components, each serving different functions</i>	All except...	Aoki	-	<i>Components are usually social structures as well; sometimes institutions impose functions on physical objects, which are, however, interchangeable.</i>
<i>Institutions provide receipts of how to behave</i>	All except...	Searle	-	<i>Usually done via rules as one component of institutions.</i>
<i>Institutions provide motivations to behave in a particular way</i>	Greif, Aoki	The rest	-	<i>Usually done via norms as one component of institutions; highlighted by the IAE approach.</i>
<i>Institutions manifest themselves in behavioral regularities</i>	Tuomela	Miller, Ostrom, Greif, Aoki	Hodgson, Searle, Lawson	<i>Unclear why this should be necessarily the case; better talk about inclinations rather than actual behavior.</i>
<i>Institutions exist because of collective intentionality</i>	Searle, Tuomela	Aoki, Hodgson, Lawson, Ostrom, Greif, Aoki	Miller	<i>Claimed only by particular school of thought in philosophy; not useful for a general definition.</i>
<i>Institutions constitute new forms of behavior</i>	Searle	-	The rest.	<i>Unclear why this should be the case, c.f. ADICO grammar.</i>

Less unanimity exists when it comes to functioning of institutions and their effect on individuals.

At least, almost all authors – some implicitly, some explicitly – stressed that institutions need to have some impact if they are said to exist. This means that the function of an institution should be part of its taxonomic definition: a rule that nobody follows should not be considered an institution. But most of the authors are not clear as to whether institutions necessarily manifest themselves in *actual* behaviour. We believe there are good reasons not to relate institutions with actual behaviour, but with *inclinations* to behaviour (Veblen, 1909; Hodgson, 2006). The important point is that institutions need to have some effect on individual cognition: it might well happen that two institutions suggest an individual to behave in mutually exclusive

ways. If, after a long-lasting decision-making process, the individual behaves in line with what has been suggested by institution *A*, this does not mean that institution *B* has had no impact on the reasoning of the individual, let alone did not exist.

Not surprisingly, there are also instances where the existing definitions are contradictory. The concept of collective intentionality, for example, is a specialized, yet in the philosophical literature still contested concept. Few scientists outside philosophy are familiar with it. We believe that in certain applications one may refer to collective intentionality to substantiate the theoretical underpinning of one's institutional theory, but it is not a concept generally (and necessarily) employed by institutionalist scholars.

Similarly, there is also no agreement on as to whether institutions are necessarily constitutive rules. This view is particular to John Searle's approach, but not representative for the discourse as a whole. As the ADICO grammar suggests, rules, an important component of institutions, can come in a variety of forms that affect social interactions. Of course, one could reserve the name "institutions" to constitutive rules as suggested by Searle, but then one would need other names for social phenomena that are systems of regulative rules. Moreover, the speech act theory underlying this account of institutions is quite controversial, even within philosophy. All this considered, we do not see many reasons for and many against the reservation of the term 'institutions' only for constitutive rules.

But aside from these two aspects, the definitions are surprisingly consistent. Does this mean that a universal definition of institutions is likely to emerge? On an applied level, we are sceptical as to whether this is possible, and even desirable: research practices of institutional economists (and social scientists) differ depending on their paradigmatic origins, their methodological orientation, and the level of abstraction that they use to study institutions. Using the same definition for a very specific case study and a general model of human decision-making is not instrumental. Only a certain degree of shared understanding on a general level is important for a research community to be functional.

Interestingly, it is even possible to use shared elements across the definitions above, to formulate very general definition of *institutions as codifiable systems of social structures (in particular norms and rules) that lead to inclinations to act in specific ways*.

This ‘meta-definition’, composed out of the shared elements identified above, consists of three main parts: the first part states that institutions are codifiable systems of social structures, the second part specifies the content of institutions, and the third part relates to its functioning. Although very general, such an idea of institutions has some implications and certain characteristics follow from it:

Institutions constrain and enable people to act and coordinate in complex and uncertain interaction arenas. Many institutionalist theorists refer to habits, heuristics, and similar cognitive mechanisms to explain how institutions induce informed beliefs about the state of the world and the behaviour that can be expected from others. Consequently, institutions enable people to act when they are not able to perceive their environment completely and derive optimal decisions from this environment. Thus, by stabilizing the expectations of individuals about the behaviour of others they are not only constraints of human behaviour, but also *enabler* of behaviours and coordinated interaction (Hodgson, 2006; Elsner, 2012).

Institutions are normative and create power. There has been some disagreement in the literature to what extent institutions are normative (Veblen, 1899; Hodgson, 2006; Searle, 2010; Hindriks and Guala, 2015a; Aoki, 2015). We believe that the arguments that normativity plays an important role are convincing. If we question the motivation of people to follow institutions, we will ultimately reach social values. Additionally, since institutions always serve specific functions, normativity enters via the justification of the function (which always refers to values in some sense). The fact that institutions may create power relationships seems to be rather uncontroversial. Since institutions structure the interaction of people, they enable and constrain them in certain ways, and the way they do this affects the capability space of the people involved. Of course, the normativity of institutions does not – *a priori* – tell us something about the desirability of institutions.

Institutions have some self-stabilizing effects. If there were no self-stabilizing aspects, institutions would disappear. It is an important contribution of the institutions-as-equilibria approach to stress this aspect: analytically, it is not enough to simply state that a given institution exists, but one has to work out why it does. To do so, one must refer to the sources of its self-stabilization, which might work directly via the provision of motivations or commitment to follow the institution, or indirectly via establishing enforcement and sanctioning mechanisms involving third parties. There is nothing that prevents these self-stabilizing effects

to occur for dysfunctional institutions that deprive many people affected by them (one might think of the institutions of slavery, for example). The question of how such dysfunctional institutions can be changed is obviously an important area of ongoing research (see e.g. Aldashev et al., 2012; Guala, 2016 ch. 14).

5 Conclusion

The vantage point of our undertaking was the recognition that every research community requires a shared understanding of its key terms in the form of sufficiently clear taxonomic definitions. We asked whether such shared understanding is present in institutional economics, and related disciplines. By surveying influential definitions of institutions, we found that most of these definitions have a shared core, which itself can be formulated as a very general definition of institutions as *codifiable systems of social structures (in particular norms and rules) that lead to inclinations for people to act in specific ways*. This indicates that institutional economists seem to have developed a shared understanding of their key term, and that this shared understanding even reaches beyond the discipline of economics as such.

We hope that our work can be useful for future research in several respects: First, while our review of existing definitions is unavoidably incomplete, it provides a summary of influential definitions from within and outside economics, which might constitute a useful reference for institutional economists.

Second, our discussion of various forms of definitions, particularly the distinction between model-based and non-model-based definitions, might contribute to a better taxonomy of existing definitions of institutions, and hopefully facilitates the discourse about the meaning of the term.

Third, identifying the shared elements of existing definitions contributes to more clarity with regard to *the* key term of institutional economics. As noted by Ostrom (1986), “[n]o scientific field can advance far if the participants do not share a common understanding of key terms in their field” (p. 4). While we clarified that - given the different purposes definitions must serve - a single, unanimous definition is neither feasible nor desirable, a certain high-level agreement on taxonomic definitions among researchers is a pre-condition for successful science. Our work suggests that – although explicit discussions of definitions are rare – institutional economists not only have a sufficiently clear shared understanding of the key term ‘institution’, but that this shared understanding also expands beyond the realm of economics, and is compatible with

influential work in other social sciences and philosophy. At the same time, definitions from outside economics highlight different aspects of institutions that might be interesting to consider in future research.

Another avenue for future research that can be suggested by this work – although more for the field of history of economics thought – is to study the relationship between model-based and non-model-based inquiry of institutions. As we have seen from the example of the “institutions-as-equilibria” approach, model-based reasoning can have significant impact on the way people consider institutions in general. Yet the precise relationship between model-based and non-model-based inquiry of institutions is not fully understood and deserves further attention.

Finally, we want to stress that we do not expect all institutional economists to use the general umbrella definition that emerges from the shared meanings of the scholars surveyed in this paper. It is way too general for concrete inquiries, and mainly serves the purpose of highlighting the shared understanding that has emerged among institutional economists. Yet we believe that research on institutions can benefit from scholars defining explicitly what they refer to by ‘institutions’ in their work. While doing our survey we found many influential papers that do not define institutions explicitly. Doing so, however, would not only facilitate the synthesis of existing research by others and to advance our understanding of economic institutions as a research community, it would also help to sustain the shared understanding of the concept of ‘institutions’ that has emerged within institutional economics until today.

References

- Acemoglu, D. and Robinson, J.A., 2001. A Theory of Political Transitions. *American Economic Review*, 91(4), pp.938–963.
- Aldashev, G., Chaara, I., Platteau, J.-P. and Wahhaj, Z., 2012. Formal Law as a Magnet to Reform Custom. *Economic Development and Cultural Change*, 60(4), pp.795–828.
- Aoki, Masahiko. 2001. *Comparative Institutional Analysis*. Cambridge, MA, London, UK: The MIT Press.
- Aoki, M., 2011. Institutions as cognitive media between strategic interactions and individual beliefs. *Journal of Economic Behavior and Organization*, 79(1-2), pp.20–34.
- Aoki, M., 2015. Why is the equilibrium notion essential for a unified institutional theory? A friendly remark on the article by Hindriks and Guala. *Journal of Institutional Economics*, 11(3), pp.485–488.
- Arrow, K.J., 1994. Methodological individualism and social knowledge. *American Economic Review*, 48, pp.1–9.
- Aumann, R.J., 1987. Correlated Equilibrium as an Expression of Bayesian Rationality. *Econometrica*, 55(1), pp.1–18.
- Bacharach, M., 1999. Interactive team reasoning: A contribution to the theory of co-operation. *Research in Economics*, 53(2), pp.117–147.
- Binmore, K., 2014. Bargaining and fairness. *Proceedings of the National Academy of Sciences*, 111(Supplement_3), pp.10785–10788.
- Bowles, S. and Gintis, H., 2002. The Inheritance of Inequality. *Journal of Economic Perspectives*, 16(3), pp.3–30.
- Calvert, R., 1995. Rational Actors, Equilibrium, and Social Institutions. In: J. Knight and I. Sened, eds., *Explaining Social Institutions*. University of Michigan Press.
- Colman, Andrew M, Briony D Pulford, and Jo Rose. 2008. Collective Rationality in Interactive Decisions: Evidence for Team Reasoning. *Acta Psychologica* 128 (2), pp. 387–397.
- Crawford, S.E.S. and Ostrom, E., 1995. A Grammar of Institutions. *American Political Science Review*, 89(3), pp.582–600.
- Elsner, W., 2012. The Theory of Institutional Change Revisited: The Institutional Dichotomy, Its Dynamic, and Its Policy Implications in a More Formal Analysis. 46(1), pp.1–44.
- Elster, J., 2009. *Reason and Rationality*. Princeton, NJ: Princeton University Press.
- Frigg, R. and Nguyen, J., 2017. Models and Representation. In: L. Magnani and T. Bertolotti, eds., *Springer Handbook of Model-Based Science*. Dordrecht, Heidelberg, London and New York, pp.49–102.
- Gintis, H., 2009. *The Bounds of Reason. Game Theory and the Unification of the Behavioral Sciences*. Princeton and Oxford: Princeton University Press, pp.1–305.

- Gräbner, C. and Kapeller, J., 2017. The Micro-Macro Link in Heterodox Economics. In: T.-H. Jo, L. Chester and C. DiIppoliti, eds., *The Routledge Handbook of Heterodox Economics*. London, UK, New York, NY, pp.145–159.
- Gräbner, C., 2018. How to Relate Models to Reality? An Epistemological Framework for the Validation and Verification of Computational Models. *Journal of Artificial Societies and Simulation*, 21(3).
- Greif, A. and Laitin, D.D., 2004. A Theory of Endogenous Institutional Change. *American Political Science Review*, 98(4), pp.633–652.
- Greif, A., 2006. *Institutions and the Path to the Modern Economy*. Cambridge, UK: Cambridge University Press.
- Guala, F., 2016. *Understanding Institutions*. Princeton, IL: Princeton University Press.
- Hakli, R., Miller, K. and Tuomela, R., 2010. Two Kinds of We-Reasoning. *Economics and Philosophy*, 26(3), pp.291–320.
- Hayek, von, F.A., 1952. *The Sensory Order*. Chicago, IL: University of Chicago Press.
- Hédoin, C., 2012. Linking institutions to economic performance: the role of macro-structures in micro-explanations. *Journal of Institutional Economics*, 8(03), pp.327–349.
- Hindriks, F. and Guala, F., 2015a. Institutions, rules, and equilibria: a unified theory. *Journal of Institutional Economics*.
- Hindriks, F. and Guala, F., 2015b. Understanding institutions: replies to Aoki, Binmore, Hodgson, Searle, Smith, and Sugden. *Journal of Institutional Economics*, 11(3), pp.515–522.
- Hodgson, G. and Knudsen, T., 2004. The complex evolution of a simple traffic convention: the functions and implications of habit. *Journal of Economic Behavior and Organization*, 54(1), pp.19–47.
- Hodgson, G., 2006. What are institutions?, *Journal of Economic Issues*, 40(1), pp.1–25.
- Hodgson, G., 2015a. *Conceptualizing Capitalism. Institutions, Evolution, Future*. Chicago and London: University of Chicago Press.
- Hodgson, G., 2015b. On defining institutions: rules versus equilibria. *Journal of Institutional Economics*, 11(3), pp.497–505.
- Hodgson, G. (2018). Taxonomic definitions in social science, with firms, markets and institutions as case studies. *Journal of Institutional Economics*, 1-18. doi:10.1017/S1744137418000334
- Hurwicz, Leonid. 1996. “Institutions as Families of Game Forms.” *The Japanese Economic Review* 47 (2): 113–32. doi:10.1111/j.1468-5876.1996.tb00038.x.
- Hurwicz, L., 2008. But Who Will Guard the Guardians? *American Economic Review*, 98(3), pp.577–585.
- Jaccard, J. and Jacoby, J., 2010. *Theory Construction and Model-Building Skills*. New York, London: The Guilford Press.
- Krueger, Joachim I. 2008. Methodological Individualism in Experimental Games: Not So Easily Dismissed. *Acta Psychologica* 128 (2), pp. 398–401.
- Lewis, D., 1969. *Convention*. Cambridge, MA: Harvard University Press.

- Magnani, L. and Bertolotti, T. eds., 2017. *Springer Handbook of Model-Based Science*. Dordrecht, Springer.
- Mäki, U., 1992. On the method of isolation in economics. *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 26, pp.319–354.
- Mayhew, A., 2008. Institutions, Culture and Values. In *The Elgar Companion to Social Economics*, edited by John B Davis and Wilfred Dolfsma, 28–43. Cheltenham, UK and Northampton, MA: Edward Elgar.
- McCloskey, D.N., 2001. The genealogy of postmodernism: an economist's guide. In: S. Cullenberg, J. Amariglio and D.F. Ruccio, eds., *Postmodernism, Economics, and Knowledge*. London, pp.102–128.
- Miller, S., 2004. *Social Action. A Teleological Account*. Cambridge, UK: Cambridge University Press.
- North, D., 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
- North, D.C. and Weingast, B.R., 1989. Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England. *The Journal of Economic History*, 49(4), pp.803–832.
- Ostrom, E., 1986. An agenda for the study of institutions. *Public Choice*, 48(1), pp.3–25.
- Ostrom, E. 1992. *Crafting institutions for self-governing irrigation systems*. San Francisco: ICS Press.
- Robinson, R., 1962. *Definition*. Oxford, UK: Clarendon Press.
- Rodrik, D., Subramanian, A. and Trebbi, F., 2004. Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development. *Journal of Economic Growth*, 9, pp.131–165.
- Scott, W.R., 2014. *Institutions and Organization*. Los Angeles: SAGE.
- Searle, J.R., 2005. What is an institution? *Journal of Institutional Economics*, 1(1), pp.1–22.
- Searle, J.R., 2010. *Making the Social World*. New York: Oxford University Press.
- Sudgen, R., 2008. Nash Equilibrium, Team Reasoning and Cognitive Hierarchy Theory. *Acta Psychologica*, 128(2), pp.402–404.
- Sugden, R., 2000. Team Preferences. *Economics & Philosophy*, 16(2), pp.175–204.
- Tuomela, R., 2002. *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge, UK: Cambridge University Press.
- Tuomela, R., 2013. *Social Ontology*. New York: Oxford University Press.
- van Bouwel, J., 2009. The problem with(out) consensus: The scientific consensus, deliberative democracy and agonistic pluralism. In: J. van Bouwel, ed., *The Social Sciences and democracy*. Chippenham and Eastbourne: Palgrave, pp.121–142.
- Veblen, T., 1899. The Barbarian Status of Women. *American Journal of Sociology*, 4(4), pp.503–514.
- Veblen, T., 1909. The Limitations of Marginal Utility. *Journal of Political Economy*, 17(9), pp.620–636.
- Weisberg, M., 2007. Who is a Modeler? *The British Journal for the Philosophy of Science*, 58(2), pp.207–233.