

Laboratoire REGARDS (EA 6292)  
Université de Reims Champagne-Ardenne

## Working paper n° 2-2017

### History, Analytic Narratives and the Rules-in-Equilibrium View of Institutions

Cyril Hédoïn \*

- \* Professeur des Universités Sciences économiques, REGARDS (Université de Reims Champagne-Ardenne)
- \* Full professor of economics - University of Reims Champagne-Ardenne, France
- \* Contact: [cyril.hedoin@univ-reims.fr](mailto:cyril.hedoin@univ-reims.fr)

**Abstract:** Analytic narratives are case studies of historical events and/or institutions that are formed by the combination of the narrative method characteristic of historical and historiographical works with analytic tools, especially game theory, traditionally used in economics and political science. The purpose of this paper is to give a philosophy-of-science view of the relevance of analytical narratives for institutional analysis. The main claim is that the AN methodology is especially appealing in the context of a non-behaviorist and non-individualist account of institutions. Such an account is fully compatible with the “rules-in-equilibrium” view of institutions. On this basis, two supporting claims are made: first, I argue that within analytical narrative game-theoretic models play a key role in the identification of *institutional mechanisms* as the explanans for economic phenomena, the latter being irreducible to so-called “micro-foundations”. Second, I claim that the “rules-in-equilibrium” view of institutions provides justification for the importance given to non-observables in the institutional analysis. Hence, institutional analysis building on analytical narrative typically emphasizes the role of *derived* (i.e. non-directly observed) intentional states (preferences, intentions, beliefs).

**Keywords:** Analytic narratives – Rules-in-equilibrium view of institutions – Institutional analysis – Game theory

Word count: 11 000

*Les working papers d'économie et gestion du laboratoire Regards sont édités après présentation en séminaire et validation par deux relecteurs internes, sous la responsabilité du conseil de laboratoire.*



Laboratoire d'Economie et  
Gestion REGARDS (EA 6292)  
Université de Reims  
Champagne-Ardenne  
UFR de sciences économiques,  
sociales et de gestion  
57B Rue Pierre Taittinger  
51096 Reims

Directeur : **Martino Nieddu**

Responsable de l'édition des  
working papers : **Romain  
Debref**

## 1. Introduction<sup>1</sup>

Analytic narratives (henceforth, ANs) are case studies of historical events and/or institutions that are formed by the combination of the narrative method characteristic of historical and historiographical works with analytic tools, especially game theory, traditionally used in economics and political science. The analytic narrative (AN) methodology can thus be defined as the articulation of a historical narrative with a formal or informal game-theoretic model to account for a specific economic or political phenomenon and its underlying institutional mechanisms. In principle, the AN methodology can be used to study a great diversity of historical contexts and issues, such as for instance significant battles in military history (e.g. Mongin 2017). AN practitioners have however essentially focused on issues related to economic and political institutions and their role in the development of economic exchanges (especially long distance trade) or the rise and fall of political regimes (Bates et al. 1998).

As a way to articulate history with tools and theories of social sciences, the AN methodology has been met with many criticisms attacking in particular its commitment to rational choice theory (Elster 2000) and its reliance on non-observable and non-quantifiable factors as well as its inability to produce testable results (Clark 2007). The purpose of this paper is not to rebut these critiques per se but to give a broader, philosophy-of-science view of the relevance of AN for institutional analysis. The main claim is that the AN methodology is especially appealing in the context of a non-behaviorist and non-individualist account of institutions. Such an account is fully compatible with the “rules-in-equilibrium” view of institutions recently promoted in social ontology and economics by Hindriks and Guala (2015). On this view, institutions are neither only rules playing the role of constraints nor patterns of behavior corresponding to a game-theoretic equilibrium. While institutions have a behavioral manifestation which can be analyzed in terms of game-theoretic equilibrium, this behavioral manifestation is the result of rules that shape individuals’ beliefs, intentions, preferences and reasoning processes (Hédoïn 2017).

On this basis, two supporting claims are made: first, I argue that within AN game-theoretic models play a key role in the identification of *institutional mechanisms* as the explanans for economic phenomena, the latter being irreducible to so-called “micro-foundations” (Hédoïn 2012). Quite the contrary, institutional mechanisms correspond to complex arrangements combining individual features (e.g. individual beliefs), collective features (e.g. common knowledge) and structural features (norms, values). Second, I claim that the “rules-in-equilibrium” view of institutions provides justification for the importance given to non-observables in the institutional analysis. Quantitative economic historians and behaviorist-minded economists generally emphasize the primacy of both choice concepts (for the formulation of theories) and choice data (for the testing of theories). However, institutional analysis building on AN typically emphasizes the role of *derived* (i.e. non-directly observed) intentional states (preferences, intentions, beliefs). The rules-in-equilibrium view emphasizes the importance of such derived intentional states as well as of the underlying reasoning process through which they are formed. As these intentional states cannot be derived without an extensive knowledge of the historical, cultural and socioeconomic context in which individuals evolved, this confirms that institutional accounts building on AN cannot consist in authentic “micro-reductions”.

The paper is organized as follows. Section 2 presents the principles of AN in the context of institutional analysis and briefly discusses several examples. Section 3 argues that as a method to identify and characterize institutional mechanisms, AN is committed to give a great importance to “structural” factors that cannot be properly given micro-foundations in the analysis. Section 4 explains and defends the role given to non-observables by any institutional analysis using the AN methodology. Section 5 shows that these features of AN both support and are justified by the rules-in-equilibrium view of institutions. Therefore, AN militates for an account of institutions that is both anti-individualistic and anti-behaviorists, which helps to answer the critiques mentioned above. Section 6 briefly concludes.

---

<sup>1</sup> This paper is based on a lecture given at the “Interdisciplinary Symposium”, University of Corsica (France), 06/07/2017. I thank the participants and particularly Richard Arena and Philippe Steiner, for their comments.

## 2. Analytic Narratives and Institutional Analysis: Principles and Examples

As a methodological approach articulating historical case studies with theoretical tools of social sciences, especially rational choice theory and game theory, the AN research program has been explicitly endorsed and applied by economists and political scientists over the last three decades. The collective book *Analytic Narratives* (Bates et al. 1998) by a group of political scientists (Bates, Levi, Weingast) and economists (Greif, Rosenthal) provides the first systemic exposition of the methodological principles of the AN approach as well as several case studies illustrating them. Avner Greif's book *Institutions and the Path to the Modern Economy* (Greif 2006) is the other most significant contribution to the AN approach in the context of institutional analysis with a detailed methodological discussion. This section describes the constitutive principles of the AN approach essentially on the basis of these two references. It also briefly discusses several applications.

At first sight, the very expression “analytic narratives” may look like an oxymoron. Narratives are traditionally associated to historiographic reasoning, especially as it is used by historians. Meanwhile, the term “analytic” refers to tools and theories seeking rather for a nomological kind of knowledge, i.e. laws and regularities. This quest for generalization corresponds to what is sometimes called nomothetic reasoning.<sup>2</sup> Therefore, the conjunction of both terms seems to imply contradictory philosophical and scientific commitments, especially regarding the kind of knowledge to be pursued. The contradiction is only apparent however. It is resolved through the delicate articulation of a narrative giving qualitative and (eventually) quantitative insights on an historical case study with a (formal or informal) game-theoretic model capturing the most significant features of the case.<sup>3</sup> The point of the game-theoretic model is to capture the relevant outcomes identified in the narrative in terms of an equilibrium:

“By modeling the processes that produced the outcomes, we seek to capture the essence of stories. Should we possess a valid representation of the story, then the equilibrium of the model should imply the outcome we describe – and seek to explain. Our use of rational choice theory and game theory transforms the narratives into analytic narratives. Our approach therefore occupies a complex middle ground between ideographic and nomothetic reasoning.” (Bates et al. 1998: 12)

The analytic part of an AN plays a twofold role in the explanation and in the exposition of the case under study. I will consider the latter first. The articulation of the narrative and the analytic parts highlights what Mongin (2016) identifies as a *discursive paradox* in the exposition of a case study: “The paradox of a narrative that is also analytic is compounded by the fact that game theory, a formalized discipline, often provides the analytic tools, whereas narratives are of necessity limited to natural languages” (Mongin 2016: 2, emphasis in original). The point is that by its very nature an AN combines two quite different expository modes, one relying on a formal language the other on natural language. According to Mongin (2016), the ANs developed in economics and political science adopt either one of two expositions schemes: an *alternation* scheme or a *analyzed narrative* scheme.<sup>4</sup> The former consists in a three-step process: a narrative presenting the salient features of the historical case under study is first developed and used to generate an explanatory hypothesis about the causes of the relevant outcome; a game-theoretic model is constructed with respect to the salient features of the case to capture the

---

<sup>2</sup> The historiographic/nomothetic distinction has been essentially developed by neo-Kantian philosophers, especially those belonging to the so-called “Heidelberg School” (Wilhelm Windelband, Heinrich Rickert), toward the end of the 19<sup>th</sup> century. Those philosophers (on which Max Weber's methodological writings largely build) have popularized the distinction between “natural” sciences and “cultural” or “historical” sciences by arguing that only the former were seeking for laws and generalization through nomothetic reasoning. Cultural sciences were characterized as sciences of the particular, thus seeking a quite different kind of knowledge.

<sup>3</sup> The analytic part of the AN does not need to consist in a game-theoretic model, at least in principle. Other kinds of rational choice models or even “behavioral” analytic accounts emphasizing the role of emotions or cognitive biases could also be used, as suggested by some proponents of the AN approach. Still, these possibilities remain essentially virtual, though there are some exceptions, e.g. Schiemann (2007).

<sup>4</sup> Mongin (2016) identifies a third scheme that he labels “local supplementation”. Here, the model is thought to bring an analytical complement to the historical narrative resulting from historians' work. Mongin cites his own study of the Waterloo battle (Mongin 2017) as an instance of this scheme of exposition. It seems that there are few if any works in political science or economics that can be characterized along this way.

explanatory hypothesis; a refined narrative is finally produced, bringing additional historical evidence and borrowing theoretical terms from the modeling part. The latter scheme rather consists in stating *separately* the theoretical hypothesis on the basis of the game-theoretic model and the historical narrative and then to proceed to a comparison of the model with the relevant facts.

The point here is not to determine if and how these schemes of exposition solve the discursive paradox identified above; Mongin (2016) makes some remarks about this issue. Both are used by political scientists and economists doing ANs. An instance of the alternation scheme is for instance Greif's study of the political system in late medieval Genoa (Greif 1998). A form of the analyzed narrative scheme is used by Acemoglu and Robinson (2006) in their discussion of the commitment problem between elites and citizens in the transition from dictatorship to democracy. There are many more examples (some of them are discussed below) and also cases where the exposition scheme actually oscillates between these two polar cases. Most of them share the same commitment to the middle ground between ideographic and nomothetic reasoning referred to in the quotation of Bates et al. (1998) above. In both types of exposition schemes, the last stage (the refined narrative or the comparison between the narrative and the model) has indeed the same twofold function. First, it provides a motivation to revise *both* the model and the original narrative. This can be due either to a significant discrepancy between the revised narrative and the original narrative (in the alternation scheme) or between the model and the narrative (in the analyzed narrative scheme), or more generally to a "surprising" finding that emerges from the whole AN process. Second, the last stage encourages a limited form of generalization beyond the specific case under study. Indeed, assuming that the model actually captures the mechanism responsible for the salient outcome highlighted by the narrative, several implications can generally be conjectured concerning *other* historical events sufficiently similar in whatever relevant respects with the original case. In this way, ANs are not purely ideographic but may also allow for a (restricted) form of nomothetic reasoning. The doing of comparative statics building on the model is perhaps the most significant illustration of this twofold function. Comparative statics (where one varies the value of some parameters to infer implications regarding the equilibrium outcomes) is indeed useful both to check the consistency of the model with the narrative and to explore implications that were not salient in the original narrative. As it appears, several significant ANs developed in economics make extensive use of comparative statics.

Whatever the exposition scheme retained, the AN approach thus broadly relies on a ternary inference process akin to the philosopher Charles Sanders Peirce "Abduction-Deduction-Induction" schemata. Abduction consists in the movement going from the consequences to their hypothetical causes. Abduction is thus a hypothesis-generating inference that, within the AN approach, is made through the original narrative combined with prior theoretical knowledge. Deduction is the logical inference through which the implications of the hypothesis about the causes of the phenomenon are derived. Within the AN approach, deduction is done through the game-theoretic model.<sup>5</sup> Finally, induction confronts the logical implications with facts. As suggested above, this role is played by the refined narrative (or the comparison stage in the analyzed narrative scheme). This has the interesting – and at first sight, perplexing – implications of making the *explanatory* (rather than the exposition) scheme circular. In particular, as the model may reveal features that were not transparent in the original narrative, the introduction of these factors in the refined narrative may change what is to be explained. In other words, ANs are prone to an interplay between the *explanans* and the *explanandum*.<sup>6</sup> As I explain in section 4, this is essential to understand the importance of non-observables in the AN approach.

In the context of institutional analysis, ANs will be used to identify the institutions explaining some significant socioeconomic outcomes. That means that in principle, institutions will figure as the *explanans* and the socioeconomic outcomes as the *explanandum*. This is actually more complicated as in most cases the institution becomes itself a part of the *explanandum*. This is more generally due to the

---

<sup>5</sup> It is worth noting that in this perspective, deduction cannot be done on the basis of the hypothesis generated through abduction solely. Auxiliary theoretical hypotheses (e.g. which is the relevant equilibrium concept to be used) are also required.

<sup>6</sup> Clark's (2007) critique against the tendency of the AN approach to conflate evidence with hypotheses may be due to this subtle interplay between the *explanans* and the *explanandum*. Obviously, Clark regards the AN's explanatory scheme as viciously circular, but as I explain below such a negative judgment is not warranted.

underlying view of institutions on which the AN approach builds and that I discuss in section 5. Table 1 below gives a list of ANs, each time with their corresponding *explanans* and *explanandum* as well as with their exposition scheme.

**Table 1: A list of ANs in economics**

<b>References</b>	<b><i>Explanandum</i></b>	<b><i>Explanans</i></b>	<b>Exposition scheme</b>
Acemoglu & Robinson (2006)	(Un)Ability to solve the commitment problem for not using political power in various non-democracy	Democratization where the elites allow the citizens to set the tax rate today and in the future	Analyzed narrative
Casari (2007)	Management of common resources (pasture and forest) in the Trentino region from 1200 to 1800	Private-order governance through a legal institution ( <i>Carte di Regola</i> )	Analyzed narrative
Greif (1993)	Existence of long distance trade in the Mediterranean area in the 11 <sup>th</sup> century	Reputation-based community enforcement mechanism (multilateral punishment strategy)	Analyzed narrative
Greif (1998)	Stability and instability of the political system in late medieval Genoa	(i) Mutually advantageous mutual deterrence equilibrium in the presence of an external threat; (ii) stabilizing effect of the <i>podestà</i> institution corresponding to another mutual deterrence equilibrium	Alternation
Greif (2001)	Existence of impersonal exchange in late medieval Europe without a legal system enforced by a state	Community responsibility system: communities' courts are incentivized to punish cheating merchants in their community	Analyzed narrative
Greif, Milgrom & Weingast (1994)	Development of trade relationships between merchants and rulers in late medieval Europe	Merchant guild organization with coordinating abilities implementing a Markov perfect equilibrium	Alternation
Leeson (2007)	Organization and coordination of 16 <sup>th</sup> and 17 <sup>th</sup> pirates' criminal activities	Pirate ship organization based on a check & balances system	Hybrid
Milgrom, North & Weingast (1990)	Impersonal exchanges in the 11 <sup>th</sup> century Champagne Fairs	Law merchant enforcement system	Alternation
Rubin (2011)	Historical divergence regarding interest restrictions in Christianity and Islam between 1000 and 1800	Extent of the political authorities' dependence on religious authorities for ensuring legitimacy	Alternation

Table 1 covers a relatively heterogeneous set of ANs. Some of them use the alternate scheme of exposition and others the analyzed narrative scheme, with at least one (Leeson) using an exposition scheme not fitting with either. While most rely on a formal game-theoretic model, two (Casari and Leeson) use the concepts and the logic of rational choice theory only informally. Finally, they greatly differ regarding the temporal scope of the historical case under study. While this is not explicitly indicated in the table, in some cases the institutions that figure in the *explanans* is actually also part of the *explanandum* as these ANs are interested in their endogenous evolution.<sup>7</sup>

I cannot discuss all these ANs and so I will only briefly comment on three of them as they are especially representative. The study of the merchant law in the 11<sup>th</sup> century Champagne Fairs of Milgrom et al. (1990) is a prototypical AN – though it is criticized by Greif (2006: 315-8) for not paying sufficient attention to the historical context and relying on an inadequate microanalytic model. The authors start with a broad historical survey of the development of long-distance trade in late medieval Europe and describe the Law Merchant institution in the context of the Champagne Fairs. This institution provides the motivation for the developments of not one but several game-theoretic models seeking to show how it could have solved the classical commitment problem that affects impersonal exchanges. A first model shows that a simple reputation-based mechanism would not have been sufficient under plausible assumptions about the information distribution among traders. A second model shows that the Law Merchant system was able to bypass this information problem thanks to a specific player, the “judge”, collecting information about the traders’ past behavior. It is then shown that a path of play where the judge honestly collects the information, each trader queries the judge about the other merchant’s past behavior, each merchant cooperates if the other has no outstanding judgment and defects otherwise is part of a (symmetric) sequential equilibrium. The model is finally put in contrast with historical evidence to refine the original narrative and to proceed to a mild generalization regarding the role of the state in the further development of impersonal exchanges.

Greif’s (1993) study of the Maghribi traders community is also a very representative example. Quite similarly to Milgrom et al., Greif’s starting point is the development of long-distance trade in a context where no state authority was able to enforce contracts – in this case the Mediterranean area in the 11<sup>th</sup> century. Maghribi traders had to employ “agents” to transport commodities to other locations, most of the time across the seas. A classical principal-agent problem arises as Maghribi traders were obviously unable to monitor agents; as a result, agents were themselves unable to commit to honor the terms of the contract. Greif argues that Maghribi traders were able to solve this problem through a community-based enforcement mechanism corresponding to a game-theoretic multilateral punishment strategy. The game-theoretic model is a variation of an efficiency-wage account where it is shown that on the equilibrium path, agents are paid some wage  $w^*$  and never cheat traders, traders who are cheated never employ again the agent and cheating agents are never employed again by any trader in the community.

A last example is Rubin’s (2011) account of interest restrictions in Islam and Christianity. Compared with Milgrom et al.’s and Greif’s studies, Rubin’s AN has a far greater historical scope. Moreover, the players in his game-theoretic model are not individual traders but organizations with less well-defined borders (“political” and “religious” authorities). Still, Rubin’s AN closely follows the alternate exposition scheme: first, historical considerations related to interest restrictions are presented with some details and an hypothesis regarding the role of legitimacy relationships between political and religious authorities is formed. A game-theoretic model is then built to establish that under some assumptions (i) a greater dependence of the political authority on the religious authority for legitimacy entails stronger interest regulations and (ii) dependence endogenously changes as a function of the political and religious authorities’ choices. Finally, a refined narrative establishes that the dynamic of the model fits with the historical events, allowing for mild generalizations regarding the role of religious authorities in economic development.

---

<sup>7</sup> Not all the works referred in the table are self-consciously characterized as ANs by their authors. Indeed, for some of them this characterization might be disputed, especially for Acemoglu & Robinson (2006) and Casari (2007). Nonetheless, as they combine more or less detailed historical narrative with either a formal game-theoretic model or an informal use of rational choice theory, they broadly conform to the general methodological principles highlighted above.

### 3. Analytical Narratives, Micro-Explanations and Institutional Mechanisms

As I note above, the AN approach has been criticized by various authors on diverse and sometimes conflicting grounds. The role of “micro-foundations” in the explanatory endeavor of the AN approach is one of the main objects of such criticisms. Indeed, according to Elster (2000), the fact that the whole explanatory strategy of an AN depends on the assumption that agents are instrumentally rational contributes to making the relevance of the whole approach doubtful. The problem does not lie in the fact of looking for micro-foundations in the institutional analysis per se, but rather that the micro-foundations are deemed to be the “wrong” ones. Elster (2000: 692-3) also criticizes the tendency of ANs to ascribe to collective agents (clans, the elite...) intentional states and rationality properties, in contradiction with the principles of methodological individualism. Here, the objection seems to be that ANs is not sufficiently micro-founded. In a similar fashion, Clark (2007) regrets that the AN approach’s original aim to provide institutional analysis with new micro-foundations has been lost, at least in Greif’s (2006) subsequent works.<sup>8</sup> Quite the contrary, according to Clark (2007: 736, emphasis in original)

“Some will be shocked by, and perhaps hostile to, the path Greif has taken. Were economists of a more literary bent, the word *apostasy* would be on their lips. In a search for generality, Greif concludes that such a set of limited rational actor assumptions is not constraining enough to describe real-world institutions.”

The lack of proper micro-foundations in institutional analysis has – still according to Clark (2007: 736) – the unwelcome implications of introducing “into the explanatory apparatus almost all the entire sociological zoo of ill-defined and unmeasurable constructs”, with the effect that “we lose all explanatory power”. Without seeking to provide a defense of the AN approach against these critiques, I shall argue in this section for an interpretation of the aim of ANs that explains and in some way justify the absence of full micro-foundations. My point basically consists in the claim that developing micro-explanations of socioeconomic outcomes (arguably, what ANs are trying to do) is not the same as seeking for micro-reductionism. This point is clearer once we acknowledge that the explanatory strategy of most ANs consists in identifying *institutional mechanisms*.

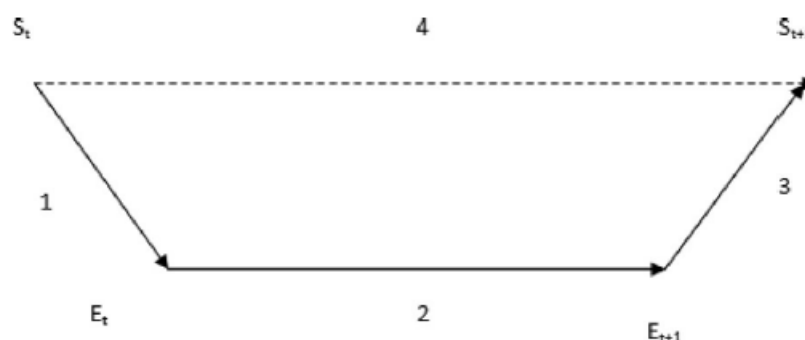
The nature of micro-explanations, especially in relation with methodological individualism, has generated a huge amount of discussions in the philosophy of social sciences. A good starting point is the well-known Coleman’s (1994) diagram which provides a graphical representation of any “good” explanatory scheme in the social sciences (Fig. 1):<sup>9</sup>

**Fig. 1: Coleman’s Diagram**

---

<sup>8</sup> Clark is referring to Greif’s 1993 article on the Maghribi traders in identifying this original aim: “The 1993 article seemed to point to new micro foundations for institutions that would ground them in individual maximizing behavior” (Clarke 2007: 735).

<sup>9</sup> Coleman’s diagram corresponds to an account of explanation in the social sciences that has been endorsed and generalized by proponents of “analytical sociology” [(Hedström and Bearman 2011a); (Hedström 2005); (Hedström and Swedberg 1998a)]. As far as I know, there is no cross reference between the analytical sociology and the AN literatures.



Coleman's diagram depicts four relationships between four vertices, two of them corresponding to social situations ( $S_t$  and  $S_{t+1}$ ) and the other two to individual states ( $E_t$  and  $E_{t+1}$ ). According to the underlying account of explanation within the social sciences, the latter aim at accounting for the "macro relationship" (relation 4) between a given social situation at time  $t$  ( $S_t$ ) and another social situation at time  $t+1$  ( $S_{t+1}$ ). More exactly, the point is to provide a causal explanation of the relationship between a socioeconomic outcome characterizing  $S_t$  (say, the degree of competition within a market) and a socioeconomic outcome characterizing  $S_{t+1}$  (say, the rate of innovation or the level of profits). This relationship however is purely statistical. A proper explanation of relation 4 must thus correspond to a "micro-explanation" accounting for three, "micro-founded" relations: a macro-to-micro relation (1), a micro-to-micro relation (2) and a micro-to-macro relation (3). Each of these relations corresponds to a kind of *social mechanism* which can be basically characterized as referring to "a constellation of entities and activities that are organized such that they regularly bring about a particular type of outcome... we explain an observed outcome by referring to the mechanism by which such outcomes are regularly brought about" (Hedström and Bearman 2011b: 5). Following Hedström and Swedberg (1998b), we can distinguish different kinds of social mechanisms according to their location in Coleman's diagram: *situational mechanisms* correspond to the way structural features of social situations (e.g. networks of interactions) systematically affect agents' beliefs, preferences and choices; *action-formation mechanisms* refer to decision-making operations through which agents make choices on the basis of their beliefs, preferences and other intentional states; *transformational mechanisms* account for the way agents' choices aggregate to generate behavioral patterns characterizing social situations and which are causally responsible for observable socioeconomic outcomes. A proper micro-explanation of a socioeconomic outcome thus consists in the concatenation of these three forms of social mechanisms, connecting the influence of the social situation the agents are embedded in with the collective consequences of the agents' choices through decision-making principles.

As explained in the preceding section, ANs take institutions to figure as the relevant *explanans*. Therefore, while the socioeconomic outcomes to be explained correspond to the social situation  $S_{t+1}$ , any AN conforming to the canons of micro-explanation must identify the relevant institution through the three kinds of social mechanisms just identified. I therefore suggest that the ultimate purpose of ANs is to identify *institutional mechanisms* defined as a *set of entities and activities through which institutional elements such as rules, norms or conventions influence agents' choices and their aggregation*. As I will explain below (see section 5), this characterization of institutional mechanisms is especially appealing in the context of the "rules-in-equilibrium" view of institutions. At this point, it is thus sufficient to acknowledge that the game-theoretic part of ANs is specifically designed to model institutional mechanisms and thus the concatenation of the three kinds of social mechanisms distinguished by analytical sociologists. Consider for instance the game-theoretic models in two of the ANs taken as illustrations above, Milgrom et al.'s (1990) study of the Law Merchant institution in 11<sup>th</sup> century Champagne Fairs and Rubin's (2011) analysis of interest restrictions in Christianity and Islam. Milgrom et al.'s second game-theoretic model (I ignore the first as it is not intended to capture the Merchant Law system) identifies two types of players: the traders and the judge (or law merchant). The traders' lack of information over others' past behavior is directly determined by the social situation in which they are embedded and that the model is intended to capture. This, as well as the strategies available to the judge and the traders (especially the possibility for the judge to register traders' behavior and for traders to query the judge), implicitly capture the effects of a situational mechanism. The action-formation mechanism is formalized in the model by a classical rationality principle according to which



the players choose their best strategies given their preferences and their expectations about others' future behavior.<sup>10</sup> Finally, the transformational mechanism is directly captured on the basis of the adoption of a solution concept, which in this case is the sequential equilibrium one. Indeed, by assuming that any acceptable strategy profile must correspond to a sequential equilibrium, the range of possible behavioral patterns is automatically restricted. The aggregation then leads to a particular sequence of plays satisfying the condition required by the solution concept.

Rubin's model similarly captures the three kinds of social mechanisms at once. As in the preceding case, the game-theoretic model includes a situational mechanism by determining the set of relevant parameters as well as the strategy space. In particular, Rubin's model has two key parameters  $\gamma$  and  $\alpha$  reflecting respectively the exogenous degree to which political authority derives legitimacy by conforming to the regulation of the religious authority and the exogenous degree of the religious authority's resistance to change. As it happens, the  $\gamma$  parameter is made endogenous in Rubin's model, but this endogenous property is also derived from an implicit situational mechanism affecting the players' preferences as a function of past plays.<sup>11</sup> The action-formation mechanism is captured by a standard best-reply principle where each player is assumed to play her best strategy given the other's strategy choice.<sup>12</sup> Finally, the transformation mechanism is directly derived through the combination of the Nash equilibrium solution concept with the endogenous nature of the  $\gamma$  parameter.

Both Milgrom et al.'s and Rubin's ANs thus satisfy the minimal requirements for an appropriate micro-explanation in the social sciences, at least according to the standards settled by Coleman's diagram. Now, this does not directly answer the questions of whether these explanations are properly "micro-founded" and whether they succeed or not in achieving an adequate micro-reduction. Consider first the issue of *micro-foundations*. In macroeconomics, a model is generally regarded as appropriately micro-founded if and only if agents' choices (households and firms) are directly derived from traditional optimization under constraints. In this perspective, both Rubin's and Milgrom et al.'s models are micro-founded, as well as for all models in ANs that are game-theoretic and/or rational choice-based. Note however that the issue of micro-foundations could be framed differently by asking whether or not the model's underlying behavioral assumptions are "realistic" or relevant. For instance, players in Milgrom et al.'s model are assumed to revise their beliefs along Bayesian lines.<sup>13</sup> Experimental studies in behavioral economics suggest that individuals actually are scarcely Bayesian. Alternative micro-foundations, capturing for instance biased belief updating due to focal points or framing effects, might be thought as more relevant to account for specific historical cases (Schiemann 2007). Whatever one's position on this specific issue, the point is that the determination of micro-foundations is directly related to the kind of action-formation mechanisms (relation 2 in Coleman's diagram) that operate in the decision-making process. Therefore, the fact of favoring micro-explanations in the institutional analysis does not commit one to specific micro-foundations, though it indeed implies that the analysis must be micro-founded.

Consider now the issue of *micro-reduction*. What this term is precisely intended to mean is not necessarily clear. Clark (2007: 736) seems to suggest that an institutional analysis successfully achieving micro-reductionism would account for institutions in terms of "a game theory equilibrium of purely self-interested rational individuals interacting with common knowledge". Such characterization of micro-reductionism obviously corresponds to a form of methodological individualism, defined as "the doctrine that all social phenomena (their structure and their change) are in principle explicable only in terms of

---

<sup>10</sup> As I will explain in the next section, the game-theoretic model and the whole AN are also used to *infer* the agents' intentional states and thus their preferences and beliefs. There is nothing viciously circular about this once it is recognized that the most usual way to account for individuals' behavior in the social sciences is to ascribe them intentional states.

<sup>11</sup> Formally speaking,  $\gamma$  is best described as a "quasi-parameter", i.e. a parameter whose value is endogenously determined but which is taken as given by players. Crucially, it is assumed that the players *do not account* for the parameter's value change when choosing their optimal action.

<sup>12</sup> Rubin's model does not formalize beliefs explicitly. It is simply assumed that each player plays her best-reply given other's strategy choice, thus implementing a Nash equilibrium.

<sup>13</sup> Though this is not made explicit in their article, this is implied by the authors' use of the sequential equilibrium solution concept.

individuals – their properties, goals and beliefs” (Elster 1982: 453). Whether or not the AN approach can be seen as an instance of methodological individualism is not the most interesting question here. More interesting is asking what the kinds of social mechanisms entering into the explanation are and whether or not they could be *in principle* given a pure micro-reductionist interpretation. Clark’s critique of Greif’s ANs already indicates that the answer is probably negative. This presumption is reinforced by looking at our two examples discussed above. Milgrom et al.’s model faces the traditional problem of equilibria multiplicity. While their analysis singles out one sequential equilibrium, a range of different belief profiles sustain different strategy profiles and thus multiple sequential equilibria exist. Which equilibrium is actually selected therefore depends on the way players form their beliefs. Milgrom et al.’s model is silent about this – it only establishes the “cooperative” equilibrium existence, not whether and how it is selected. The narrative could in principle provides hints about the mechanisms responsible for the agents’ beliefs but few if any elements are offered.<sup>14</sup> Beyond the fact that Milgrom et al.’s micro-explanation is thus not fully satisfactory, what is relevant is that *even if it had been it would not have achieved complete micro-reduction*. Belief formation is indeed typically accounted for by invoking structural features, such as the cultural or religious context. Greif’s (1994) comparative study of Maghribis and Genoese traders’ “cultural beliefs” is a representative example.

A similar lesson is taught by Rubin’s model. The role played by non-reductive factors is emphasized by the equilibrium analysis done through comparative statics. In particular, the exploration of the model establishes that the equilibrium values of the players’ strategies (level of regulation of interest policies) change as a function of the values of parameters  $\gamma$  (measuring the dependence of the political authority on the religious authority for legitimacy) and  $\alpha$  (measuring the religious authority’s resistance to change). In the same way, the endogenous evolution of the  $\gamma$  parameter depends on other parameters whose value is taken as exogenous. The point is thus that the analysis not only fails but moreover does not try to account for the value of the parameters or the functional relationship between the quasi-parameter and parameters in a reductive or “individualistic” way. Contrary to Milgrom et al.’s study, Rubin’s AN does provide however a justification for these modeling choices through the historical narrative. In conclusion, it appears that these ANs as well as all the ANs listed in table 1 develop micro-explanations that rely on non-individualistic features. I would argue that this is not a deficiency of the analysis, but rather an unavoidable implication of the kinds of mechanisms that are singled out (institutional mechanisms) and the necessarily limited scope of the case studies.

#### 4. Analytic Narratives and the Role of Non-Observables in Institutional Analysis

A different but related feature of ANs is their reliance on non-observable elements, i.e. elements that enter into the social mechanisms but that cannot be directly observed through the available data on which the analysis relies. As these elements are part of the social mechanisms (and thus of institutional mechanisms), they will play a key role in the explanatory endeavor. This dependence on non-observables may be regarded as problematical, at least from a certain standpoint. Elster (2000) is especially concerned by the fact that explanation within ANs relies on the imputation of intentional states:

“Much of applied rational choice theory is a combination of just-so stories and functionalist explanation. One constructs a model in which the observed behavior of the agent maximizes their interests as suitably defined, and one assumes that the fit between behavior and interest explains the behavior... Unless one can demonstrate an intention (first version) or a causal feedback loop from the consequences of the behavior to the behavior (general version), the coincidence of behavior and interest may be only that – a coincidence.” (Elster 2000: 693)

Elster’s point is that unless one has direct evidence for the agents’ intentional states, rational choice explanations essentially consist in imputing goals and beliefs such that the observed behavior can be

---

<sup>14</sup> This is indeed one of Greif’s (2006: 317) complaints against Milgrom et al.’s account: many other models and thus institutional mechanisms could have accounted for the same behavioral pattern. Without a proper account of the way agents form their beliefs and how the relevant equilibrium is selected, there is no reason to favor Milgrom et al.’s explanatory hypothesis over the other ones.

explained by its consequences through their effects on the agents' intentional states. Clark (2007) exhibits a somewhat different but related kind of worry about Greif's (2006) use of the AN methodology. As we have seen in the preceding section, multiplicity of equilibria implies that equilibrium selection will depend on features such as beliefs and norms that do not transpire into the observable elements of institutions. This leads to a confusion between hypothesis (i.e. such or such elements could account for the fact that this equilibrium has been selected) and firm results that should be tested in light of the available evidence.<sup>15</sup>

The general critique regarding the (misplaced) importance of non-observables (especially intentional states) can be interpreted as resulting from a more or less strong endorsement of behaviorism and revealed-preference theory. In its most extreme version, the revealed-preference view banishes from economics all kinds of data that is not choice-based.<sup>16</sup> From this standpoint, the reliance on non-observable intentional states (i.e. not directly inferred from observed behavior) is of course deemed to be problematical. What would be an institutional analysis in a revealed-preference/behaviorist perspective then? First, it would identify institutions to formal rules (laws, constitutions) and ultimately to observed behavioral patterns. Second, it would seek to account for these behavioral patterns in terms of incentives and opportunity costs that are derived for instance from publicly known market prices and other observable and formal institutional devices. Third, it would provide a quantitative estimate of the relationship between formal rules, the resulting behavioral patterns and their consequences in terms of some relevant economic variables (growth, wealth inequalities...). Such kind of institutional analysis actually already exists and has generated interesting results regarding the importance of institutions for economic development.

Obviously, this is not the methodological road taken by the AN approach. Rather than trying to evaluate which approach is the best one, I will rather point out several considerations justifying the importance given to non-observables in the institutional analysis, especially concerning intentional states. A first point to be made is that in spite of the rhetoric of some revealed-preference theorists, it is actually hard if not impossible to completely avoid any reference to non-choice data in economic and institutional analysis. While it is true that the institutional analysis can be in principle restricted to econometric exercises looking for statistical relationships between formal institutional variables (e.g. property rights protection as estimated by the rate of expropriation) and economic variables (e.g. investment rate), this will not be sufficient to account for the *causal* relation between institutions and outcomes. Here, it is necessary to engage in (micro-) explanation on the basis of a combination of theoretical and empirical insights. I would argue that at this stage it is rarely if ever possible to do without postulating elements that cannot be directly observed and measured. This is true even if one conducts the analysis in terms of incentives and opportunity costs. While the latter may sometimes be directly inferred from market prices, they most of the time also depend on assumptions about the kinds of motivations (e.g. extrinsic or intrinsic) that underlie individuals' behavior. Moreover, in many cases, the institutional analysis will be done in a context where no market data are available. In this case, there is no other possibility but to *impute* motivations to agents and to make inferences about the relevant incentives that may account for their behavior. Recent results in behavioral and institutional economics that emphasize the variety of motivations and the complexity of their interactions responsible for agents' behavior only strengthen this point (e.g. Bowles 2016).

Once this is acknowledged, the reliance of ANs on non-observables becomes less problematic. A more interesting and decisive aspect (which here is more directly targeted at Elster's critique) is that the AN methodology is specifically designed to form inferences about such non-observables. Start by considering the precise kinds of non-observables that enter into most ANs:

- Informal institutional elements (norms, conventions).

---

<sup>15</sup> Clark (2007: 739) also regrets that the use of the AN approach to resolve the indeterminacy problem "closes institutions off from some other promising paths that might help narrow the range of possible equilibria and their dynamics". Among these other promising paths, Clark cites the economic efficiency constraint and the pressures emanating from gene-culture coevolution.

<sup>16</sup> For recent discussions on contemporary revealed-preference theory and behaviorism, see Dietrich and List (2016), Hands (2013) and Okasha (2016), among others.

- Intentional states (beliefs, desires, intentions).
- Reasoning modes (i.e. forms of inferences).

Informal institutional elements like norms and conventions are obviously central in most ANs since they most of the time play the role of *explanans*. They are non-observable as soon as we recognize that they cannot be reduced to behavioral patterns but also depend on specific intentional states and reasoning modes. As an illustration of this point, consider the simple coordination problem depicted by fig. 2 below:

**Fig. 2: A Coordination Game**

	A	B
A	1 ; 1	0 ; 0
B	0 ; 0	1 ; 1

The matrix depicts an interaction where players' interests are fully convergent. As a consequence, players' sole objective is to coordinate over the same strategy choice. This game has three Nash equilibria but principles of rational choice are unhelpful to decide which one will be selected. One function of conventions is specifically to permit coordination of expectations and actions and thus to solve the equilibrium selection problem [(Lewis 1969); (Sugden 2005)]. Now, suppose that we observe that members of the relevant population constantly play the A strategy, thus implementing one of the Nash equilibrium. In spite of the fact that coordination is achieved, on most accounts of conventions this would not be sufficient to infer that a convention is actually followed by the members of the population. For instance, Lewis's (1969) definition of conventions requires a common knowledge or belief among agents that everyone expects a given equilibrium to be implemented.<sup>17</sup> Were the coordination of players over the A strategy due to pure luck or randomness, it would arguably be mistaken to conclude that they are following a convention. To follow a convention thus implies to have specific intentional states, especially beliefs, beyond the fact of implementing a given behavioral pattern. The same is true for social norms, which generally depend on a conditional preference for conformity and appropriate beliefs regarding what others will do (Bicchieri 2005).

Therefore, unless one subscribes to a purely behaviorist view of institutions (which is hardly sustainable, as I argue above), at least some institutional elements have to be regarded as non-observable by nature. The point is of course that most of the intentional states that underlie conventions and norms are themselves hardly directly observables. This is especially the case for beliefs and intentions. The same is true for what I call "reasoning modes", i.e. how individuals make epistemic and practical inferences on the basis of the knowledge/belief of some event. Suppose one observes a given event (e.g. the traffic light has just turned green). On this basis, she will eventually make a series of inferences generating beliefs about what other drivers will do as well as inferences about what she should do, given her desires/preferences. Though hardly observable, reasoning modes are thus essential to account for the way agents form beliefs and intentions in strategic interactions. Most ANs build on the assumption that agents are Bayesian rational and thus form posterior beliefs on the basis of Bayes's rule, given their prior beliefs. Interestingly, even on this strong assumption it is still necessary to recognize that an individual can interpret the same event in different fashions and ascribe to it several possible meanings (Bates, de Figueiredo, and Weingast 1998). This is indeed directly related to the multiplicity of (perfect Bayesian or sequential) equilibria in incomplete information dynamic games. In particular, it is quite usual in such games to have at least one pooling equilibrium and at least one separating equilibrium. In the former, the other player's action at a given information set is not taken as any evidence for this

---

<sup>17</sup> This is only a rough characterization of Lewis's account of conventions but details are not relevant here. The point is that to follow a convention implies to have specific intentional states.

player's type by the agent observing this action. In the latter, the other player's action is quite the contrary used to update one's belief about the other's type. The same event (i.e. the other player's action) can thus be interpreted in two different ways and lead to radically different practical and epistemic inferences. The fact that agents are scarcely Bayesian rational in practice only strengthens this point: there are many ways individuals may update prior beliefs and form new ones and this will depend on how they reason in a given situation but also on their beliefs regarding how others are reasoning. Of course, all of this is hardly observable and the same behavioral pattern may hidden quite different intentional states formed on the basis of quite different reasoning modes.

Admittedly, there are cases where a more or less direct access to the actors' intentions or beliefs (for instance through historical archives or interviews) is possible. While this kind of information almost surely helps to feed the narrative, it cannot be a full substitute to the interpretative work consisting in inferring non-observables from what is observed. In part, this is due to the fact that agents might misreport their intentional states or even not be fully aware of them. The more general point is that an institutional analysis building on the AN methodology is committed to take the *intentional stance* both to account for the agents' behavior and to account for the institutional mechanisms that explain the relevant socioeconomic outcomes. Far from being, as suggested by Elster, a defect of "applied rational choice theory", the imputation of derived intentional states is an unavoidable part of the process of explaining and predicting agents' behavior, not only in the scientific context but also in our daily life. Taking the intentional stance consists in treating as a default assumption the agent as rational and then figuring out which intentional states (beliefs, desires) the agent ought to have given her rationality and the context in which she is embedded. On this basis, it becomes possible both to predict her behavior and to account for it. As eloquently put by Daniel Dennett (1987: 25-6), taking the intentional stance is not only natural, it is also the unique way to uncover "real patterns" underlying observed behavior:

"Our imagined Martians might be able to predict the future of the human race by Laplacean methods, but if they did not also see us as intentional systems, they did not also see us as intentional systems, they would be missing something perfectly objective: the *patterns* in human behavior that are describable from the intentional stance, and only from that stance, and that support generalizations and predictions."

The importance granted to non-observables in institutional analysis becomes unproblematic as soon as we recognize the unavoidability of taking the intentional stance.<sup>18</sup> Moreover, the whole methodology of AN seems to be designed to encourage and to facilitate taking this stance. Indeed, the use of the intentional stance works on the basis of some "triangulation" where (i) individuals are assumed to be minimally rational and (ii) a socioeconomic and cultural context is taken as given. Game theory and more generally rational choice theory provide the formal and conceptual apparatus to study the behavior of minimally rational agents. The historical narrative not only offers partial evidence for the individuals' intentional states, reasoning modes and degree of rationality, it also has the role of giving information about the context in which the agents make decision. In the perspective of the AN approach, the behavioral pattern that characterizes an institution is thus the result of a complex interaction between two general kinds of non-observable elements: on the one hand, individuals' minds constituted by intentional states and reasoning modes, on the other hand institutional elements such as norms and conventions. The AN approach thus offers a methodology to *infer* these elements in the context of a case study.

## 5. Analytic Narratives and the Rules-in-Equilibrium View of Institutions

The two preceding sections have established that ANs favor a non-individualist and non-behaviorist methodological approach within institutional analysis. In this section, I shall argue that the AN methodology also has conceptual and theoretical implications regarding the nature of institutions. In

---

<sup>18</sup> Dennett's account of the intentional stance can be given either an instrumentalist or a realist interpretation. On a realist reading, for an agent *i* to have a belief *B* is fully exhausted by the fact that *i*'s behavior is explainable and predictable by ascribing *B* to her on the basis of the intentional stance. This issue is of secondary importance here as my discussion stays at the methodological level and ignore the ontological one.

particular, it strongly supports what Hindriks and Guala (2015) have characterized as the *rules-in-equilibrium* view of institutions. I first briefly surveys the standard accounts of the nature of institutions and then provides argument for my claim that the AN approach favors the rules-in-equilibrium view. Finally, I draw some implications for the kinds of scientific contributions that the use of ANs is ultimately making to economics.

Two dominant and essentially conflicting views of institutions prevail in the modern literature on institutional economics. The *institutions-as-rules* view is particularly developed by North (1990) and Ostrom (1990). On this view, institutions are the “rules of the game”, setting constraints upon individuals’ choices. The *institutions-as-equilibria* view has been essentially entertained in game-theoretic accounts of institutions. Schotter (1981) provides the first systematic theory of institutions based on a game-theoretic framework. Subsequent works such as Sugden (1986) and Young (1998) also build on the idea that institutions correspond to game-theoretic equilibria. While interesting in their own rights, both views have obvious limitations. The main difficulty associated to the institutions-as-rules view is that it lacks a convincing account of the conditions under which rules are indeed constraining people’s choices. This is evident with formal rules which fail to be followed on a regular basis. Moreover, the fact that rules act as constraints depends on whether or not they can be enforced and how. Postulating that some kind of third party will act as an external enforcer is generally insufficient as such assumption begs the question of why this third party is indeed enforcing rules. The institutions-as-equilibria is able to deal with this latter problem. On this view, all the agents’ strategies, including the third party’s strategy, should combine to form an equilibrium in some game. By the very definition of the equilibrium concept, this implies then that each player is incentivized to play her strategy given the strategy choice of other players. Institutions then correspond to self-enforcing behavioral patterns (i.e. strategy profiles) that reproduce through time due to the fact that they correspond to game-theoretic equilibria. I have already alluded to the main limitations of the institutions-as-equilibria view in the two preceding sections. First, by reducing institutions to equilibria and the corresponding behavioral patterns, this view fails to notice what is specific to *human* institutions. Basically, as evolutionary biologists have recognized long ago, equilibrium play is far from being human specific as most animal populations seem to implement some kind or another of equilibrium. If institutions are nothing but game-theoretic equilibria, then ants or birds have institutions too. Second, this view essentially ignores the underlying reasoning process and more generally the reasons that lead players to implement an equilibrium. While this view guarantees that an institution only exists if the players are incentivized to implement it, it does not state *why* the players are incentivized. These are more than mere philosophical babblings: the kinds of motivations and reasons that underlie an institution may affect the way people will respond to a change in incentives and thus the endogenous dynamics of the institution.<sup>19</sup>

The *rules-as-equilibrium* view of institutions is intended as a response to the defects of the two standard accounts. As characterized by Hindriks and Guala (2015) and Guala (2016), this view analyzes institutions as being *both* equilibria and rules, depending on the perspective taken. Following Vanderschraaf (1998) and Gintis (2009), respectively for conventions and social norms, Hindriks and Guala (2015) identify institutions to correlated equilibria in games. Informally, institutions then correspond to correlating devices (“choreographers” in Gintis’s words) sending signals indicating to each player what she should do given some exogenous state of the world.<sup>20</sup> Another way to put the matter is that an institution corresponds to a correlated distribution of the strategy profiles such that the players are incentivized to follow the signal they have received given what they believe about the signal received by the other players. The so-called hawk-dove game provides a simple illustration (fig. 3):

---

<sup>19</sup> The possibility of a “crowding-out effect” between intrinsic and extrinsic motivations is the most obvious example, though not the only one. See Bowles (2016).

<sup>20</sup> More formally, a correlated equilibrium is a system of linear inequalities such that each player maximizes her expected utility given her beliefs conditional on the strategy she is playing. As established by Aumann (1987), common knowledge of Bayesian rationality and a common prior over the outcome space are jointly sufficient conditions for a correlated equilibrium to be implemented in a game.

**Fig. 3: The Hawk-Dove Game**

	<b>H</b>	<b>D</b>
<b>H</b>	0 ; 0	5 ; 1
<b>D</b>	1 ; 5	3 ; 3

Suppose that two individuals are competing for an asset (resources, a territory) and have the choice between fighting (H) or negotiating (D). The two Nash equilibria in pure strategies are payoff-asymmetric while the mixed-strategy Nash equilibrium is Pareto-inferior. This game has however an infinite number of correlated equilibria with some of them being on the Pareto frontier. Figure 4 below gives an example of a Pareto-efficient correlated distribution corresponding to a correlated equilibrium:

**Fig. 4: A Pareto-Efficient Correlated Distribution**

	<b>H</b>	<b>D</b>
<b>H</b>	0	1/2
<b>D</b>	1/2	0

This correlated distribution indicates that the strategy profiles (H, D) and (D, H) are each played half of the time. A quick look is sufficient to see that this distribution is sustainable as an equilibrium and that the resulting allocation of the asset is Pareto-optimal.<sup>21</sup> As suggested by Hindriks and Guala (2015), this correlated equilibrium can be interpreted as corresponding to a proto-institution of property. Indeed, among the many ways the correlated equilibrium can be implemented, we may imagine that players' choices are correlated on an exogenous state of the world on the basis of the following simple rule:

If first arrived, then play H;

If not first arrived, then play D.

Here, the fact of being or not the first arrived on the territory is the exogenous state of the world. Clearly, by the very fact of corresponding to an equilibrium, if the other player follows this rule then one is also incentivized to follow it. This example then shows how an institution can be both a rule and an equilibrium at the same time. From an external observer point of view (e.g. a game theorist), this proto-institution of property corresponds to an equilibrium with an associated behavioral pattern. But from the *players' perspective*, the institution rather corresponds to a rule followed on the basis of (conditional) beliefs which are formed through an inductive inference. This inductive inference may itself be sustained by the fact that each player believes or knows that the other follows the rule, that the other believes or knows that one is following the rule, and so on. The point is that the whole existence of the institution,

---

<sup>21</sup> Suppose the row player receives the indication to play H. Then, at the equilibrium she conditionally believes that column will play D and therefore she has no interest to deviate. The same logic applies if row receives the indication to play D. As the same is true for column, this distribution is a correlated equilibrium. On average, each player will then gain a payoff of 3, which is obviously Pareto-optimal.

including its behavioral manifestation, is sustained by a more or less complex form of reasoning entertained by the players.

I would argue that such a view of the nature of institutions conforms to the non-behaviorist and non-individualistic account favored by the AN approach. On the one hand, it is clearly non-behaviorist since it does not take the nature of institutions to be exhausted by the corresponding behavioral pattern. On the other hand, the very fact that any institution depends on an underlying rule (or set of rules), as well as on the corresponding intentional states and reasoning modes indicates that non-observables are at the core of its functioning. As suggested in the previous section, the need to take the intentional stance then implies to include in the micro-explanation features that cannot be reduced to the individual level. I would also suggest that though the correlated equilibrium solution concept provides an intuitive way to apprehend the rules-in-equilibrium view, the latter is not dependent on the former. In fact, this view seems also compatible with a variety of other equilibrium solution concepts, including those relevant for sequential games such as subgame perfect and perfect Bayesian equilibria. The latter for instance can be used to capture the fact that agents' choices depend on how they interpret a given piece of information, as pointed out by Bates, de Figueiredo, and Weingast (1998). How events are interpreted is constitutive of a rule and more generally of an institutional mechanism. Perhaps more controversially, the rules-in-equilibrium can also be combined with the self-confirming equilibrium solution concept to capture institutional mechanisms that depend on the fact that agents' hold false beliefs out-of-the-equilibrium-path. The self-confirming equilibrium solution imposes that each player correctly predicts others' choices and plays her best-response accordingly but sets no restriction regarding the players' beliefs at information sets that do not are on the path of play. Hence, it can account for institutions grounded on partially false or eccentric beliefs.<sup>22</sup>

Ultimately, on the rules-in-equilibrium view, institutions can be seen as practices where individuals follow a rule or a set of rules. Rules shape individuals' beliefs, intentions and reasoning modes and thus help to explain their behavior. On this view, to follow a rule is nothing but to have the specific intentional states and reasoning modes that result in the corresponding behavior (Hédoïn 2016). Moreover, to be followed rules must be self-enforcing. This depends on the cultural, geographical and technological context, as well as on the individuals' preferences and beliefs. Therefore, the identification of institutional mechanisms through the AN methodology consists in inferring, through the combination of a historical narrative and a game-theoretic model, the rules that are followed (or have been followed) by individuals in a given context.

## 6. Conclusion

The AN approach provides an original methodology for doing institutional analysis. It is grounded on the articulation of ideographic and nomothetic reasoning through the combination of a historical narrative with a game-theoretic model. It aims at explaining salient socioeconomic outcomes by identifying the underlying institutional mechanisms. The main claim I have made in this paper is that this methodological approach favors a rules-in-equilibrium view of institutions. This is due to the fact that the study of institutions through ANs leads to develop micro-explanations that are neither individualist nor behaviorist. While the AN approach remains limited in its ability to from general theoretical and historical accounts of institutions, it nevertheless seems relevant to develop "middle-range" theories of institutional mechanisms shedding light on significant historical episodes.

## References

- Acemoglu, Daron, and James A. Robinson. 2006. *Economic Origins of Dictatorship and Democracy*. Cambridge University Press.
- Aumann, Robert J. 1987. "Correlated Equilibrium as an Expression of Bayesian Rationality." *Econometrica* 55 (1): 1–18.
- Bates, Robert H. 1998. *Analytic Narratives*. Princeton University Press.

---

<sup>22</sup> The self-equilibrium solution concept is developed by Fudenberg and Levine (1993). For an example of an AN using this solution concept, see de Figueiredo, Rakove, and Weingast (2006).



- Bates, Robert H., Rui J. P. de Figueiredo, and Barry R. Weingast. 1998. "The Politics of Interpretation: Rationality, Culture, and Transition." *Politics & Society* 26 (4): 603–42.
- Bicchieri, Cristina. 2005. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.
- Bowles, Samuel. 2016. *The Moral Economy: Why Good Incentives Are No Substitute for Good Citizens*. Yale University Press.
- Casari, Marco. 2007. "Emergence of Endogenous Legal Institutions: Property Rights and Community Governance in the Italian Alps." *The Journal of Economic History* 67 (1): 191–226.
- Clark, Gregory. 2007. "A Review of Avner Greif's Institutions and the Path to the Modern Economy: Lessons from Medieval Trade." *Journal of Economic Literature* 45 (3): 725–41.
- Coleman, James Samuel. 1994. *Foundations of Social Theory*. Harvard University Press.
- Dennett, Daniel Clement. 1987. *The Intentional Stance*. MIT Press.
- Dietrich, Franz, and Christian List. 2016. "Mentalism Versus Behaviourism in Economics: A Philosophy-of-Science Perspective." *Economics and Philosophy* 32 (2): 249–81.
- Elster, Jon. 1982. "The Case for Methodological Individualism." *Theory and Society* 11 (4): 453–82.
- . 2000. "Rational Choice History: A Case of Excessive Ambition." *The American Political Science Review* 94 (3): 685–95.
- Figueiredo, Rui J. P. de, Jack Rakove, and Barry R. Weingast. 2006. "Rationality, Inaccurate Mental Models, and Self-Confirming Equilibrium: A New Understanding of the American Revolution." *Journal of Theoretical Politics* 18 (4): 384–415.
- Fudenberg, Drew, and David K. Levine. 1993. "Self-Confirming Equilibrium." *Econometrica* 61 (3): 523–45.
- Gintis, Herbert. 2009. *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton University Press.
- Greif, Avner. 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition." *The American Economic Review* 83 (3): 525–48.
- . 1994. "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies." *Journal of Political Economy* 102 (5): 912–50.
- . 1998. "Self-Enforcing Political Systems and Economic Growth: Late Medieval Genoa." In R.H. Bates, A. Greif, M. Levi, J-L. Rosenthal and Barry Weingast, *Analytic Narratives*, 23-63, Princeton University Press.
- . 2001. "Impersonal Exchange and the Origins of Markets: From the Community Responsibility System to Individual Legal Responsibility in Pre-Modern Europe". In M. Aoki and Y. Hayami (eds.), *Communities and Markets in Economic Development*, 3-41, Oxford University Press.
- . 2006. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge University Press.
- Greif, Avner, Paul Milgrom, and Barry R. Weingast. 1994. "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild." *Journal of Political Economy* 102 (4): 745–76.
- Guala, Francesco. 2016. *Understanding Institutions: The Science and Philosophy of Living Together*. Princeton University Press.
- Hands, D. Wade. 2013. "Foundations of Contemporary Revealed Preference Theory." *Erkenntnis* 78 (5): 1081–1108.
- Hédoïn, Cyril. 2012. "Linking Institutions to Economic Performance: The Role of Macro-Structures in Micro-Explanations." *Journal of Institutional Economics* 8 (03): 327–49.
- . 2017. "Institutions, Rule-Following and Game Theory." *Economics and Philosophy* 33(1): 43-72.
- Hedström, Peter. 2005. *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge University Press.
- Hedström, Peter, and Peter Bearman. 2011a. *The Oxford Handbook of Analytical Sociology*. Oxford University Press.
- . 2011b. "What Is Analytical Sociology All About? An Introductory Essay." In *The Oxford Handbook of Analytical Sociology*, 3–24. Oxford University Press.
- Hedström, Peter, and Richard Swedberg. 1998a. *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge University Press.

- . 1998b. “Social Mechanisms: An Introductory Essay.” In *Social Mechanisms: An Analytical Approach to Social Theory*, 1–31. Cambridge University Press.
- Hindriks, Frank, and Francesco Guala. 2015. “Institutions, Rules, and Equilibria: A Unified Theory.” *Journal of Institutional Economics* 11 (03): 459–480.
- Leeson, Peter T. 2007. “An-arrgh-chy: The Law and Economics of Pirate Organization.” *Journal of Political Economy* 115 (6): 1049–94.
- Lewis, David K. 1969. *Convention: A Philosophical Study*. John Wiley and Sons.
- Milgrom, Paul R., Douglass C. North, and Barry R. Weingast. 1990. “The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs.” *Economics & Politics* 2 (1): 1–23.
- Mongin, Philippe. 2016. “What Are Analytic Narratives?” [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2796567](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2796567).
- . 2017. “A Game-Theoretic Analysis of the Waterloo Campaign and Some Comments on the Analytic Narrative Project.” *Cliometrica*, April, 1–30.
- North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge University Press.
- Okasha, Samir. 2016. “On The Interpretation of Decision Theory.” *Economics and Philosophy* 32 (3): 409–33.
- Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press.
- Rubin, Jared. 2011. “Institutions, the Rise of Commerce and the Persistence of Laws: Interest Restrictions in Islam and Christianity.” *The Economic Journal* 121 (557): 1310–39.
- Schiemann, John W. 2007. “Bizarre Beliefs and Rational Choices: A Behavioral Approach to Analytic Narratives.” *The Journal of Politics* 69 (2): 511–24.
- Schotter, Andrew. 2008. *The Economic Theory of Social Institutions*. Cambridge University Press.
- Sugden, Robert. 2005. *The Economics of Rights, Cooperation and Welfare*. 2nd ed. Palgrave Macmillan.
- Vanderschraaf, P. 1998. “Knowledge, Equilibrium and Convention.” *Erkenntnis* 49 (3): 337–69.
- Young, H. Peyton. 2001. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press.