

How we (don't) follow rules

Pietro Guarnieri

Second WINIR Conference
Rio de Janeiro, 10-13 September 2015

Contents

1	Introduction: the ontology (epistemology) of normativity	3
I	No representation without interpretation	8
2	Can right and duty be reduced to rule?	8
3	What does it rule the rules?	14
4	The paradox of commitment	22
II	No normativity without alternatives	30
5	Incompleteness of deontology	30
6	Interactive intentionality	39
7	How we don't follow rules	46
8	Conclusions: on rule-formation and change	52
	References	55

Abstract

The recent proposal of unification of the “equilibrium-based account” and of the “rule-based account” under the conception of “rules-in-equilibrium” by Hindriks and Guala (2014) reactivated the discussion about rules’ functioning within institutions. The debate remains open and it has witnessed the resistance by Searle (2015) to the attempt to reduce “constitutive rules” to “regulative rules”. This paper adopts Searle’s social ontology perspective in order to discuss it through some epistemological considerations about the rule-following problem. The discussion goes in the direction of a clarification of the modalities of norm formation and change, by working on the hypothesis that normative structure of institutional ontology has to be considered as intrinsically dynamic.

One possible way for showing this is to reflect on the psychological conditions of rule-following. I refer here to the (linguistic) modalities through which the subject experiences (feels, conceptualizes) the decision-making process while facing the problem whether to follow or not to follow a rule. But there is a logical (and hence ontological) asymmetry between the two cases, which makes the latter more relevant for the purpose of this paper. If we are interested in “desire-independent reasons for action” (hence, not considering the desire-driven, self-interested, unilateral deviation from the rule) the decision not to follow the given, socially accepted rule, reveals the grammatical structure of the motivations towards rule-change.

This account of rule-following from the perspective of rule-change provides the possibility of describing some distinctive features of interactive processes of rule formation. These are taken as given by the constitutive-rule approach, which, confined to deontic logic, cannot specify the purposive structure of agency and the actual content of power relations. Moreover, from this perspective, it becomes available a description of normativity, which, avoiding the dualism between commitment and enforcement, evaporates the alternative between rule and equilibrium, overcoming the problem of their unification.

1 Introduction: the ontology (epistemology) of normativity

It is the mark of a practice that being taught how to engage in it involves being instructed in the rules which define it [...]. Those engaged in a practice recognize the rules as defining it. The rules cannot be taken as simply describing how those engaged in the practice in fact behave: it is not simply that they act as if they were obeying the rules.

J. Rawls, *Two concepts of rules*, 1955

A spider conducts operations that resemble those of a weaver, and a bee puts to shame many an architect in the construction of her cells. But what distinguishes the worst of architects from the best of bees is this, that the architect raises his structure in imagination before he erects it in reality.

K. Marx, *Capital*, Vol. I, Ch. 7, 1867

The question “why do we follow rules?” is investigated here by addressing the parallel issue concerning “how don’t we follow rules?”. This alternative perspective entails two methodological assumptions which will be presented in this introduction: a) the rule-following is better understood when we are in the condition of comprehending cases in which rules are not followed; b) dealing with normativity, the explanation of “why” is better provided as a description of “how”. Before illustrating the reasons sustaining the opportunity of these assumptions, I introduce the focus and the purposes of the paper.

The focus is suggested by the reactivation of the debate about rules’ functioning within institutions (Hodgson, 2006; Aoki, 2007; Greif, 2006; Greif and Kingston, 2011) and, in particular, by the recent proposal of unification of the “rule-based account”¹ and of the “equilibrium-based account”² by Hindriks and Guala (2014). The dualism rule/equilibrium reflects underlying dichotomies at the level of the psychological presuppositions which are made in order to discuss the features of decision-making implied respectively by the two types of account. These presuppositions will be the object of investigation, together with the various declinations of the notion of *representation*, which is typically called for by the attempts of advancing a unified view of how rules coordinate interactions and shape institutions.

¹North (1990, 2005); Ostrom (2005); Hodgson (2006).

²Lewis (1969); Schotter (1981); Sugden (1986); Young (1998).

The purposes of the paper come from the idea of integrating some epistemological³ considerations about the rule-following problem within the framework of social ontology analysis. The adoption of an epistemological perspective lets me problematize the *interpretation* of rules and to specify modalities of rule *formation* as dependent not only on preexisting structures of normativity, but also on the intentional⁴ modalities of subjects, capable of introducing elements of *novelty* and *change*. The analysis of the epistemic (linguistic) conditions of rule-following is hence aimed at showing that rule-change is constitutive of normativity (and not a contingent occurrence, secondary to rule-maintenance), and that social ontology can be considered as intrinsically dynamic.

This last point lets me discuss briefly my first methodological assumption. There is a logical (and hence ontological) asymmetry between the decision whether to follow or not to follow a given rule, which makes the latter negative case (henceforth, *negative rule-following*) more interesting for the purposes of this paper. The intentional decision to deviate from rule for the purpose of rule change, in some crucial cases, escapes both the logics of *commitment* and of *enforcement*, with which rule-following decisions and motivations are traditionally conceptualized. Commitment to rules can in fact only be referred to established duties and rights, and hence depends on the (exclusively) positive acceptance of a given normative structure. As it will be clarified, commitment presupposes *recognition* so that, strictly logically speaking, you cannot be committed and deviate from the rule at the same time.

On the other hand, the logic of enforcement frames the problem by interpreting the benefit of rule-following only in terms of the consequences of the chosen action, as reflected by *extrinsic payoffs*. By defining “punishments” and “rewards” in this way, the enforcement perspective binds rule-following behaviors to utility maximization, and this does not allow the consideration of cases of negative rule-following in which the motivations towards deviation from rules do not depend on the production of any direct (individual) or indirect (group) advantage. In any case, utility maximization act as an already established and internalized rule expressed in individuals’ preferences and reflected in behavioral regularities, which are necessarily compliant with the given systems of material payoffs as objectively interpreted.

The asymmetry of rule-following is revealed by the fact that, beside the motivational factors sustaining established intrinsic duties and rights and

³“Epistemological” in the sense of the philosophy and theory of knowledge.

⁴“Intentionality” is commonly understood as “the power of minds to be about, to represent, or to stand for, things, properties and states of affairs”(Jacob, 2014).

given equilibria in extrinsic incentives, there are reasons for institutional agency through norm deviation which do not presuppose neither (only) commitment nor (only) enforcement. The hypothesis of this category of negative motivations and behaviors is hence thought to play the theoretical role of forcing the distinction intrinsic/extrinsic, and the dichotomy commitment/enforcement, while calling into question the individualistic assumptions about the *subject* of rule-following and the relation between subjective agency and normative structure of institutions itself (Davis, 2003; Hodgson, 2007). The analysis of the type of intentionality sustaining the decision not to follow a given rule, while revealing the grammatical structure of the (intrinsic) motivations towards rule-change, as we will see, necessarily expresses the social conditions of its possibility.

This analysis is developed by making reference to the elaboration by John Searle (1995, 2005, 2010) of the conception of institutions as *collective recognition of status function and deontic powers*. But on one crucial point my epistemological analysis diverges from his foundation of social ontology on collective intentionality. Being interested in the cases in which intentionality is directed toward deviation from the rule, I do not take deontological motivations and functions as given and I study the collective intentional conditions which make possible the *purposive choice* of one alternative end (or, in general, alternative ends), with respect to that end dictated (or simply assumed) by the given deontic provision.

The clarification of the intentional modality of *purposiveness* – defined as the possibility of the choice among alternative ends – is fundamental for my attempt to re-conceptualize rule-formation and rule-following on the ground of modalities of rule-change. My second methodological assumption comes to the aid of this attempt. The possibility of interpreting negative rule-following depends on the capacity of characterizing decision-making in terms of the relations between the subject and alternative ends, and between the subject and the multiplicity of subjects in which those ends are defined, shared, and pursued. This consideration suggests the adoption of a *relational methodology*⁵ which leads me to specify the notion of *interactive intentionality* as the peculiar modality through which *new* ends are advanced as an alternative deontology, reorganizing, at the same time, the structure of practice relations

⁵As it will be clarified, relational methodology is hypothesized as alternative both to methodological individualism and methodological holism or collectivism, which I interpret as “object-centered”, i.e engaged, in both cases, in pointing at features, capabilities, and activities of the individual or of the totality of the collectivity as a unified object. However, by supposing relational methodology I do not embrace ontological relativism, but I only sustain the opportunity of looking at institutional reality *also* from the perspective of the description of relations, and, in particular, of purposive relations.

among subjects and between subjects and their institutional reality.

In this respect, the description of the “how” of these specific relational modalities appears worth assuming. In general, it avoids the doubling of *explanans* and *explanandum* characteristic of the explanation of “why”, which, as we will see, is often responsible for *problems of foundation* (circular argument and infinite regress) when applied to social ontology. In particular, the duality of the object-centered relation between individual’s choice and the object of choice – typical of the standard rationality assumptions, even when interaction is (game-theoretically) taken into account – prevents the possibility of a proper understanding of the (negative) rule-following choice, because, while reducing it to an instrumental, objective relation, it lacks to represent the structure of multiple, various, and evolving relations involved by the modes of purposiveness, which represent the condition of possibility of individual-and-collective actions within institutional normative networks.

The work is organized around two general problems, which divide the paper in two parts, and which can be emphatically stated as follows:

1. No representation without interpretation
2. No normativity without alternatives

In the first part, Hindriks and Guala’s attempt of *unification via reduction* will be criticized (starting from section 2) by focusing on Aoki’s and Searle’s theories of rules (to which Hindriks and Guala make reference in order to advance their “rules-in-equilibrium conception”). Aoki’s “rules-as-cognitive-media” (Aoki, 2007, 2011) and Searle’s “constitutive rules” (Searle, 1995, 2005, 2010) have in common a conception of rule as *representation*. But they are based on the alternative psychological notions of *shared beliefs* and *collective intentionality*, which respectively entails a view of institutions as self-enforcing conventions (based on common knowledge) and as commitment concerning “status functions” (based on *collective recognition/declarations*). I discuss and compare these two conceptions in order to show how, even if rules are explicitly taken into account by both of them, they make the dualism equilibrium-rule reappear in the forms of the alternative between rules as coordination devices (given individual’s preferences and payoffs maximization) and rules as bearers of deontology (given the pre-recognition of structures of duties and rights).

The *problem of interpretation* consists in the fact that if rule is a representation, than it has a meaning, the semantics of which has to be theoretically determined. Aoki’s problem of interpretation depends on the circumstance

that rules have equilibria (or part of them) as object of representation, but equilibria depends on rules to be established. This circularity leaves semantics undetermined in so far as the correspondence to an experienced past equilibrium (in many cases) is not a sufficient condition for a rule (or for other institutional representation or societal artifacts) to have an unequivocal meaning (section 3). This problem is solved within Searle's social ontology which, by making reference to intentionality, can set deontological meanings in the shared intentional contents, collectively recognized, as status attributes, characteristic of institutional entities.

However, Searle's treatment of the relation between language and deontology entails a problem, which I deal with in terms of a *paradox of commitment* (in section 4). Also in this case we have a problem of circularity, because language, in Searle's conception, entails commitment, but also depends on the commitment towards linguistic conventions. This would be untenable, unless we accept Searle's assumption of intrinsic aprioristic (mental) intentional states. From my point of view, this assumption could be unnecessary, at least for the discussion of social ontology. Actually, when applied to institutional reality, this foundational strategy ends up producing under-determination, revealed by the incapacity of Searle's deontology to deal with its own origin and on its dependence on actual institutional practices.

I discuss this sort of under-determination under the title "incompleteness of deontology" (section 5). The problem I want to highlight derives from the circumstance that Searle's approach, while catching how institutional reality stems from the collective recognition of deontological rules, lacks a comprehension of what does make deontology be what it is: that specific constitutive rule against many possible others. This problem introduces to the second part of the paper. By stating "no normativity without alternatives", I investigate the circumstance that rules acquire their capacity to specify their normative content only by comparison with alternative possible specifications of their prescriptive or constitutive directionality. Rules specify their content as an implicit or explicit identification of purpose, which, as for Searlian "status function" or "deontic power", has to be collectively recognized and practiced. My point is that this recognition can work only on the ground of an identification and selection of that end as one among multiple alternative others.

The problem of the origin of deontology and of its embeddedness in practices can hence be solved by explicitly thematizing the choice of an end among alternatives. This is what I try to do by making the hypothesis of interactive intentionality as the linguistic interactive modality of ends selection (section 6). Interactive intentionality reveals an intentional structure capable of accounting for both subjective commitment and inter-subjective enforcement

behind motivations in collective actions. This makes it the proper analytical tool for treating negative rule-following and rule-change (section 7) as the condition of possibility of rule-formation. Rule-following actually depends on the possibility of its negative case, which is the possibility of interactive selection of and of individuals' commitment to an alternative, not given end. Under the light of this type of decision-making, an integration of the rule-based and of the equilibrium-based approaches, of the logics of commitment and enforcement can be provided by considering them as complementary aspects of the unified structure of purposiveness which involves subjects and ends within institutions.

Part I

No representation without interpretation

2 Can right and duty be reduced to rule?

The opposition between *rule-based* and *equilibrium-based* approaches to institutions can be interpreted as the opposition between two alternative conceptions of the notion of “rule” and of the psychology of rule-following. On one hand, the rule-based account views rule as the top-down imposition of a constrain on behavioral choices, exerted through the influence on mutual expectations and systems of commitment. On the other hand, in the equilibrium-based perspective, rule is the result of the bottom-up procedure for which rationality and motivations produce stable and self-enforcing patterns of inter-subjective configurations, implying the conflation of the notion of rules with that of behavioral regularity (Aoki, 2007; Greif and Kingston, 2011).

By focusing on the epistemology of rule-following, my aim is to demonstrate that the two modalities of rule-formation can be considered as the two complementary aspects of the same unitary, evolutionary phenomenon. Acknowledging the complementary nature of this duality is the only way for analyzing the relations between the two aspects and hence for viewing them as attributes of the same institutional substance. This is an alternative methodological strategy with respect to the “unification via reduction” proposed by Hindriks and Guala (2014). This is discussed analytically in this first two sections exactly for the purpose of demonstrating how any attempt

of reduction of one aspect on the other can nothing but make the dualism and the complementarity reappear at another level or in another aspect of the theory. In this respect, the unification attempts respectively advanced by Aoki's and by Searle's conceptions of rules – used by Guala and Hindriks for identifying the two theoretical elements to be reduced and unified – are useful references for showing that conventions and deontologies are the two inseparable sides of the practice of normativity in human institutions.

The proposal of unification by Hindriks and Guala is a two steps procedure. From the point of view of the structure of the argumentation⁶, the reduction of “constitutive rules” to “regulative rules” is a pre-condition for the unification of the rule-based and the equilibrium-based approaches to institutions. If it wasn't possible to reduce constitutive rules to regulative rules, unification couldn't be obtained because the notion of equilibrium, as assumed by the authors, can only be applied to regulative rules, reinterpreted as the standard game-theoretical behavioral strategies. My discussion follows this logical order, for the purpose of, firstly, providing, in this section, arguments in favor of the irreducibility of constitutive rules, and secondly, preparing the comparison of Hindriks and Guala's rules-in-equilibrium solution with Aoki's *dualist* conception of rules, which will be developed in section 3.

To begin with, it is worth recalling Searle's definitions⁷, to which Hindriks and Guala's first step of reduction is addressed. The notion of “constitutive rules” makes references to propositions of the form “X counts as Y in C”. This formulation accounts for the attribution, made within the context C, of the function Y to the object, person, or state of affair X. Through constitutive rules, the person, object, or state of affair is provided with a *status*, such that the object or person can perform the attributed function, not in virtue of its physical structure, but only thanks to its publicly recognized and accepted status. Typical examples are money, private property, and positions of political leadership. In each of these cases, the object, state, or person (the piece of paper, the land, the man/woman) performs a function (for example, paying bills, organizing production, legislating) only in virtue of the collective recognition of the status it plays within the institutional context.

The possession of a *status function* is hence considered distinctive of in-

⁶The logical structure of their argument does not correspond to the narrative of the paper which, firstly, discusses from a general perspective the rule-based and the equilibrium-based approach, suggesting how to integrate them into the rules-in-equilibrium conception, and then, secondly, adjusts in order to incorporating, through elimination, Searlian constitutive rules.

⁷These definitions, firstly elaborated in Searle 1995, are basically taken from Searle 2005; see also Searle 2010.

stitutional reality. Performing it, in fact, consists in having and realizing *deontic powers*. In Searle's perspective, institutions and constitutive rules do not have the (only) purpose of constraining, but of creating new potentialities as power relationships (between persons, and between persons and objects). They create power by establishing structures of deontologies – i.e. systems of rights, duties, obligations, authorizations, permissions, empowerments, etc. – which make institutions function by providing subjects with *desire-independent reasons for action*. The ontological intrinsic implications of constitutive rules can be understood also by considering that the assignment of function is made through an (actual) act – the speech act of *declaration* – which requires *collective recognition or acceptance* to be effective. As it will be better clarified in section 4 and 5 by taking into account Searle's application of the notion of collective intentionality, constitutive rule can be considered productive of institutional reality exactly because its statement depends on the realization of the social (institutional) conditions, which it forms, while it represents them.

On the other hand, “regulative rules” are typically of the form “Do X” or “if Y, do X”. The difference between the two is derived from the circumstance that “regulative rules regulate activities which can exist independently of the rule”, while constitutive rules “not only regulate, but rather constitute the very behavior they regulate, because acting in accordance with a sufficient number of the rules is constitutive of the behavior in question.”⁸ Searle's point is ontological in essence. It is aimed at underlaying that only constitutive rules create the reality to which they refer through the attribution of status functions. Hindriks and Guala overlook this point in their attempt of reduction of constitutive rules to regulative rules. In their interpretation, the collective assignment of status functions is interpreted not as the concrete creation of a new deontic power (and the relative power relations), but merely as an operation of re-naming: the assignment of a name to a (pre-existing) social institution, the vocabulary definition of a new term. This causes a crucial incoherence in their argument, which can be shown by following their “property” example.

The institution of property is thought to be explained on the ground of the agents' adoption of the pre-emption system to coordinate their action. This can be summarized as the regulative rule:

⁸Searle (2005) p. 9. Searle's examples for illustrating the difference between regulative and constitutive rules are, respectively, the convention of driving on the right-hand side, and the rules of chess. While driving exists independently on the rule which imposes to drive on the left or on the right, you cannot play chess (and in this sense chess does not exist) without following the rules which describe all the relevant behaviors of playing chess.

[R] *If one is the first to occupy a piece of land, one has the right to its exclusive use.*

The term standing for the institution of “property” is hence introduced by splitting the regulative rule in two halves with the purpose of defining separately the *base rule* which identifies the objective bearer of the name “property”

[B] *If a person first occupies a piece of land, then it is her property*

and the *status rule* which instead defines the deontological meaning (in terms of right) of the relevant behavior concerning that object

[S] *If a piece of land is someone’s property, she has the right to use it*⁹

This is hence considered equivalent to the attribution of the status function expressed by the predicate “to have the right to its exclusive use” to the state of affair “the piece of land which is first occupied” as expressed by the constitutive rule

[C] *If a person occupies first a piece of land then it is her property, and if a piece of land is someone’s property then she has the right to use it*

which in fact has the standard X counts as Y in C form, even if in this case it is transformed, via the definitional procedure of introducing the institutional term, in the double implication “If C, then X is Y, and if Y then Z” The “transformation view” applied in this way is thought to confirm authors’ claim that institutions are (an equilibrium of) regulative rules to which constitutive rules simply add general names (Hindriks, 2009).

This conclusion is incoherent for one simple reason. It actually assumes what it is supposed to eliminate, namely, constitutive rules. The definition procedure couldn’t work unless the deontic meanings of crucial notions in both the base rule and the status rule haven’t already been defined through constitutive attribution of status functions. If we consider the rule [B], the possibility of treating it as a definition of what “property” is depends on the

⁹Hindriks and Guala define status rules as rules that express “what it means to have a certain status”; they are “meaning rules and concern the definition of the status term. They concern the behavior that the status regulates, or the rights and obligations that the status entails.” On the other hand, base rules specify “what it takes to have a certain status”; they “concern the ontological basis or constitution base of statuses.” (Hindriks and Guala, 2014, pp. 12-13)

possibility of identifying what a “first occupation” is. This cannot simply be taken as given, as it were a matter of fact. It is actually already a complex institutional fact, involving, for example, the deontological definition of what objective signal (a flag, a stone border) *counts as* an occupation (and as “being first”), together with all the system of rights and duties concerning the “occupier” (has it to be a working occupier? has the occupier’s relatives the same rights over the land?). Without these definitions the term “property” to which the base rule is supposed to provide a concrete, real (univocal) meaning remains without a reference.

On the other hand, even if the nominal definition of the first part of the formula [S] was already established (as given, in extensional terms, by the relative base rule and by its deontological presuppositions), neither that status rule would serve its purpose. It would end up adding no further relevant information, unless the meaning of the deontic notion of “the right to use it” was already specified in terms of the power relations it makes available. There is no alternative but to consider the status rule itself as a constitutive rule, if we want to avoid to treat it as a tautology (*If a piece of land is someone’s property, she has the right to use it as her property*), which would be useless from the point of view of the specification of the functional meaning of the institution.

Actually, what Hindriks and Guala treat in the beginning as a regulative rule [R] has to be considered as a constitutive rule, unless what “occupy” “and the right to its exclusive use” have already been defined from a deontological point of view. In other word, the pre-emption system cannot serve its function of regulative rule, if it is not already connected to a whole network of constitutive rules which make the objects, behaviors, and states of affair exist as deontological functions and power relations. With this consideration we come to a point which will be further investigated in the second part of the paper. It becomes apparent that, at least to a certain extent, regulative rules presupposes constitutive rules within institutional contexts. If we take the example provided by Searle for explaining regulative rules – the convention of driving on the right-hand side of the road – it is plausible to admit that the applicability of that regulative rule presupposes several constitutive rules which provide with ontological consistency deontic functions and objects such as the separation line between the two carriageways, the types of vehicle which are supposed to be under the control of the rule, the legitimate driver (licenses), a recognized public traffic authority for the verification of what actually happens (traffic corps), and so on and so forth.

What I want to sustain is that there is not a straight line of separation between regulative and constitutive rules, and that, reciprocally, the ones cannot work without the others. This can be seen also from the opposite

perspective by considering that constitutive rules depends on regulative rules, in so far as the deontic power which realizes the status function (duty and rights, for example) needs to be specified in terms of the explicit reference to what is and isn't allowed in that context. The constitutive speech act which creates the status function by declaring that X counts as Y in C, needs always to be accompanied with explicit or implicit (as internalized behavioral rules) regulative rules which establish what, in relevant circumstances, can be done or cannot be done with the new institutional object or function. For example, together with the status function attribution for which a particular kind of piece of paper is attributed the status function of serving as money, regulative rules concerning who and how it can or cannot be produced, or how it can be used (for example, by forbidding the destruction of notes and coins) necessarily specify practices connected to the existence institutions.

To sum up and advance a first provisional answer to the question in the heading of this section. There cannot be a regulation of an institutionalized activity or of any institutional reality, unless this activity is constituted as the activity it is by the relevant attribution of functions. Regulative rules presupposes the existence of deontological structure of power relationships, put in existence by the collective recognition of the status functions which express them. On the other hand, deontologies, such as duties and rights as *power* relationships, require (also) to be defined by rules which specifies in terms of legitimate behaviors *actual* practices. This insight, for which rules appears as intrinsically connected with practices, will be further elaborated in the next two sections and will represent a stand point for our epistemological rethinking of the ontology of normativity in the second part of the paper. For the moment, the point I want to underline is that rules are not followed in isolation, and any occurrence of rule-following presupposes the reference to a network of rules, both constitutive and regulative, necessary for the identification of the relevant institutional object and for the performance of the institutional activity.

The incoherence of the reduction strategy adopted by Hindriks and Guala actually derives from the overlooking of the collective intentionality conditions which Searle presupposes to status function declarations¹⁰. For the purpose of game-theoretical tractability, Searlian *collective acceptance and recognition* is ironed out on the standard notions of preference, expectations and common knowledge, without enough attention on the theoretical loss entailed by this translation. The difference between the notion of common

¹⁰This is evident in the equivalence that the authors consider holding between the verb "to be" in "X is (collectively accepted as) Y", which express the status function recognition and hence an ontological attribute, and in "X is Y" in the simple extensive, vocabulary definition of the name "Y" (Hindriks and Guala, 2014, p. 13).

knowledge and collective intentionality, with respect to the conception of rules, can be appreciated by comparing Searle’s approach with Masahiko Aoki’s idea of rules as *cognitive-media*, with the function of coordinating beliefs. Searle’s adoption of collective intentionality perspective will be analyzed in section 4. In the next one (section 3), I want to explain how Hindriks and Guala reductionist rules-in-equilibrium theory is hardly compatible with Aoki’s dualist conception of institutions, on which nevertheless they rely on in order to treat rules as “representations in symbolic forms of the strategies that ought to be followed in a given game”.

3 What does it rule the rules?

Hindriks and Guala’s rules-in-equilibrium account is centered on the promising idea that “players must be able to represent the equilibrium in symbolic form”, in order to get to coordination in the context of games with multiple, asymmetric equilibria. This idea is derived from the notion of *correlated equilibrium*, which is considered the relevant solution concept in those situations¹¹. The achievement of correlated equilibrium depends on the adoption of a *correlation device*, a public signal to which players make reference in order to chose the best response action. Equilibrium is hence obtained when agents, on the assumption that the others follow the signal, have no incentives in deviating unilaterally from the strategy indicated by it.

I want to concentrate on the epistemic status of the signal. This plays a crucial role in rules-in-equilibrium account, in so far as rules are thought to play the role of such a coordination device within institutional contexts. They can do it because they provides representations of “strategies that ought to be followed” (they hence are taken as regulative rules), which players use to condition their interaction. Guala and Hindriks makes reference to Aoki (2007, 2011) in order to characterize the object of the representation provided by rules. Rules can serve the purpose of correlation devices by representing equilibria (or parts of equilibria, if the possibility of bounded rationality is admitted).

From this perspective, institutions consist in equilibria concerning behavioral strategies, but, since the possibility of the coordination in behaviors depends on the availability of rules for correlation (which dictates the

¹¹I am making reference to Hindriks and Guala (2014) pp. 6-9 in which private property is modeled as correlated equilibrium of the hawk-dove game. On the ground of the common adoption of the pre-emption sytem rule, the occupation by one player of the piece of land makes the other refrain from use. The notion of correlated equilibrium has been theorized by Aumann (1974, 1987) and it is applied for an analogous purpose by Gintis (2007, 2009).

strategies to be adopted in the relevant situations), institutions can be directly stated as equilibria in the rules themselves. To put it in Hindriks and Guala’s terms, *agent-rules* and *observer-rules*¹² can be considered as coincident in institutions as rules-in-equilibrium – with constitutive rules left with the only function of summarizing symbolically the equilibrium as given, by attributing it a name. However, this equivalence between equilibrium in the behaviors and equilibrium in the rules cannot be taken for granted. By assuming it, Hindriks and Guala are making their second step in their process of unification through reduction.

The second step of reduction consists in taking as given and as *per se* univocal the meaning of the signal and the capacity of the rule to affect behaviors in a stable and effective way. This capacity actually depends on conditions of *interpretation* which have to be specified, if we do not want to run the risk of asserting tautologically that rules (which represent equilibria) are followed because they prescribe equilibria, and equilibria are obtained because they are prescribed by rules¹³. Rule’s capacity to correlate behavioral strategies depends in fact on its capacity to condition *beliefs* by “specifying patterns of expected behaviors” (Greif and Kingston, 2011, pp. 25-28). The second step of reduction has the effect of eliminating the epistemic level of the structure of reciprocal beliefs, which accounts, not only for the equilibrium in the behaviors, but for the communal interpretation of the rule as a correlation device.

The point can be made more clear with a simple example. Traffic conventions are often supported by the introduction of signals dictating what is ought to be done in the relevant situations. Among these, traffic lights

¹²It is made clear that the notion of rule is ambiguous. Sometimes, it designates a prescription from the point of view of the agent subject to it (agent-rule); sometimes it is used for describing patterns of behavior, and in this case it expresses the point of view of an hypothetical observer (observer-rule). Equilibrium conceptions in general are expressed in terms of observer-rules (the standard game theoretic strategies are nothing but observers rules). On the other hand, it is the notion of agent-rule which provides the possibility to put the rules themselves as the object of the equilibrium in the rules-in-equilibrium view. It in fact specifies the rule as what the agents take as what ought to be followed (Hindriks and Guala, 2014, pp. 9-10).

¹³I am referring to the following passage in which the authors derive the definition by dealing with the institution of property as as equilibrium in the adoption of the pre-emption rule: “From the point of view of an external observer the convention that regulates property* takes the form of a regularity that correspond to a correlated equilibrium in the hawk-dove game. But each strategy in this profile also takes the form of a rule that dictates each player what to do in the given circumstances. Each player, therefore, will perceive the institution as a prescription to use the land if the circumstance are “right”. Since the two strategies are formulated as rules, clearly the equilibrium is a set of rules – one for each player [...]” (Hindriks and Guala, 2014, pp. 10-11)

serve the scope of regulating the transit of cars or pedestrian on crossings, by giving the signals “go” or “don’t go”. They do it by means of lights, to which the prescriptive meaning is attributed. From the point of view of correlated equilibrium, the conventional behavior for which a person or a car stops at the red light and goes at the green, can be accounted in terms of the expected payoff connected to rule-following. In particular, provided that the other do the same, there is no incentives for the agent to deviate from what is prescribed by the signal. At this point, Hindriks and Guala would make the step toward considering the equilibrium in the behavioral strategy equivalent to the equilibrium in the application of the traffic rule, which is hence considered *per se* constitutive of the institution, and sufficient for describing the condition of its stable application.

But one could further ask: what does make the green light signify “go” and the red light “stop”? To ask this is not trivial. The crucial assumption grounding the notion of correlated equilibrium, for which the agent has to take for granted that the other player is following the rule, depends on the circumstance that both agents interpret the light in the same way. How this further level of convention can be accounted for? To repeat the heading of the section: what does it rule the rule?

If we went on rule-in-equilibrium’s reasoning, we should postulate another rule or signal, to be used as a correlation device in order to explain the equilibrium in the application of the first order rule. This second order rule could be stated as a regulative rule of the type “if “green”, then interpret it as “go” and “ if “red”, then interpret it as “don’t go”, but this would in turn require further rules for defining normatively the meaning of “green” and “red” (what do the terms exactly stand for? what type of red or green light counts as a a traffic light signal, and in which circumstance?), and also of “go” and “don’t go” (to go where? Who must go?). It becomes apparent as this line of reasoning would lead us to a potential infinite regress of interpretation of interpretations, with other rules and signal necessary for establishing the meaning of adjacent rules, up to potentially involving all language rules together with the contextual definition of their practices.

We start to get into the problem for which treating the rule as representation involves the necessity of clarifying its semantics (together with syntax and pragmatics). If a rule has to be able to represent a behavioral strategy to be followed, and to be interpreted in the same way for the stable correlation in inter-subjective interactions, than we are presupposing the functioning of a language capable of pointing at meanings and practices in a stable and shared way. We will be back on this very soon, discussing the semantics problem of Aoki’s assumption of rule as a *linguistic* representation (Aoki, 2001, 2011). The example, for the moment, was only meant to suggest that

the strategy of identifying regularities in behaviors (the equilibrium in the behavioral strategies) with the regularities in the adoption of the rule, used as a correlation device in the contexts in which that behavioral regularity emerges, is not sufficient *per se* to account for normativity in institutions.

Another way to see the semantic problem implied by Hindriks and Guala's second step of reduction is to say that it leaves under-determined the *psychological conditions* for which the normativity content (namely, the interpretation of what ought to be done) of the rules is actually taken as given, shared inter-subjectively, and *acquired* to common knowledge. Aoki is aware of this problem in the moment in which he sets up his analysis of institution by treating as separate the behavioral dimension and the cognitive dimension, looking for the conditions of their correspondence and theoretical unification in the possibilities of *shared beliefs* (Aoki, 2007, 2011). This possibility can be caught by rethinking the role of institutions as *cognitive media* between strategic interactions and individual beliefs. Within this framework rules acquire the status of *social cognitive artifacts* (public representations) which mediates between salient patterns of social interactions (identified within the *recursive states of play of societal game*) and individual mental states. In this way, Aoki's analysis turns out to be shaped in four areas of inter-connection between the two couples of dimensions behavioral/cognitive and individual/society, which Aoki has also described in terms of "dual-dualities of the institutional process" (Aoki, 2010).

This approach actually maintains the dualism rule-equilibrium as irreducible and reinterprets its internal complementarity in dynamic terms. On one hand, the rule-based approach is reinterpreted as describing how rules can affect beliefs and generate (constrain) equilibria in the behavior. On the other hand, the equilibrium-based approach is reinterpreted in order to account for the origin of rules as public representation in the occurrence of the equilibria themselves. This generates the circular diagram with which Aoki illustrates both a) the dependence of common knowledge and of equilibria in the behavior on the shared beliefs as induced by rules and b) the dependence of rules on the occurrence of equilibria which they summarizes in public representation (Aoki, 2011, p. 25). This idea is obtained through the application of epistemic game theory to Lewis conception of conventions (Lewis, 1969; Aumann, 1976; Cubitt and Sugden, 2003). I restrict myself to an informal presentation aimed at showing some epistemological limitations of Aoki's choice of dealing with rule-equilibrium complementarity in a circular way.

Starting from North's definition of institutions as "the rules of the game in a society" (North, 1990), Aoki is looking for the specification of the conditions under which societal rules can become common knowledge. Once rejected

the identification of institutions with game-theoretic game form (Aoki, 2011, p. 22), Aoki comes to advance his conceptualization of the “societal rule” through the following definition:

Societal rules are commonly cognized, salient patterns of the ways in which societal games are recursively played and expected to be played.
(Aoki, 2011, p. 23)

The epistemic notions of “commonly cognized” and of “salience” referred to recursive patterns of behaviors (and expectations) make it explicit, within the definition of rule itself, that the equilibrium problem behind the normative subsistence of institutions involves the epistemic dimension of beliefs. This alone is sufficient to mark the difference with the reduction from behavior to rule, operated by Hindriks and Guala, and shifts the focus on the internal relationship between strategic interaction and interaction in the beliefs, with the rule performing the function of mediation. From this perspective, *shared beliefs* represents the condition for the establishment of common knowledge, and for the identification of the *deep structure* of institutions in societal rules, as endogenous outcomes of play of the societal game¹⁴.

Aoki points out and distinguishes several meanings of the notion of beliefs. For the purpose of this paper, it is worth recalling at least two of them:

behavioral beliefs players endogenous expectation about others’ action and expectations, and, specifically, expectation about the evolving states of the societal game recursively played;

cultural beliefs (or belief system) common knowledge about the likelihood and meanings of actions accumulated over time and inherited from the past among specific group of agents (Aoki, 2011, p.23).

The crucial problem for the emergence and maintenance of institutions concerns the formation of shared behavioral beliefs. They are the basis of common knowledge and hence of the possibility of equilibria in behavioral strategies. The formation of shared beliefs is thought to be sustained by institutional rules which have the main function of specifying agents’ cognition of the *ways in which societal game has to be played*, subject to the condition for which they can *infer in common* that others would follow the rule. This

¹⁴Societal rules can in fact constitute Nash equilibrium, provided that a) agents are rational in the sense that “they maximize their payoff subject to the cognized rule”; and b) the rules are recursively observed, so that it is not required that agents knows in common the objective structure of the game (the complete game form can not be common knowledge). In these conditions, agents do not have incentives in deviating from the rules individually (Aoki, 2011, p. 23).

latter condition translates what has been signaled above as the interpretation problem behind rule-following. Given the intermediation role played by the rule itself (which establish the condition of mutual beliefs), on what grounds agents can trust such informations?

Aoki is posing the same interpretation problem in the moment in which he asks himself to what extent the sharing of behavioral beliefs is necessary to sustain, at the same time, the common interpretation in rule/representation's object (the beliefs that others will follow the rule as I do), and common knowledge about other players' actions and beliefs, which produces equilibria in behavioral strategy (which are the object of rule's representation). More precisely, his research question is: "how and by what means salient features of recursive states of play can be informationally summarized and become common knowledge of the agents to induce some shared features of behavioral beliefs among them". The problem hence becomes defining the epistemological meaning of "commonly cognized, salient features of the way in which the game is being played" (Aoki, 2011, p.24). I concentrate on this notion to highlight two characteristic problems of Aoki's attempt to face the interpretation problem.

On one hand, we have a problem of circularity. Societal rules are thought to provide the common background of informations and the common inductive standard that agents apply in reaching shared beliefs, which are then applied in common knowledge and strategic behaviors. In this sense, rules transmit to the societal players informations and strategies for obtaining equilibria by providing a common basis of mutual beliefs. But what is, in turn, the epistemic basis of societal rules? What and how do they carry out their representation function, in order to transmit the common understanding about how the game is to be played? Without this answer we would be thrown in a infinite regress similar to that encountered above, concerning the interpretation of Hindriks and Guala regulative rules. For Aoki, within the context of societal games recurrently played in society, the rule summarily - through language and symbolization - represents the salient features of recursive states of play (i.e. equilibria or part of them). The epistemic foundation of rules as representations is thought to lie in the common experience of the game itself. On the ground of the common experience of the salient features of the game the rule is considered to hold as a statement about how the game is to be played, and at the same time the beliefs that other agents will conform to it, according to a communal interpretation of that features, can be treated as given.

In this way, Aoki's demonstration of the foundation of common knowledge in shared beliefs takes the form of a circular argument. On one hand, it is the normative proposition which conditions the possibility of shared be-

liefs and provides with a foundation the application of common knowledge in strategic choices, so recursively generating equilibria. On the other hand, it is the common experience of these recursive states of play, at providing the possibility of the rule's representation itself as a summarized public (normative) representation of its salient modalities, hence constituting the rule itself as common knowledge. In this second sense, rules, as a summarized representation of salient patterns of the way in which the game has been played (and is being played), presupposes shared beliefs concerning the salient patterns of the equilibria in the past. It becomes common knowledge because since it represents salient aspect of equilibria, its normative content appears as commonly-cognized and self-enforcing.

But this second part of Aoki's circular argument sounds problematic. The second problem I want to raise regards the notion of salience itself. Without enter into the merits of the game-theoretical notion of salience (Cubitt and Sugden, 2003; Sugden, 2011), I only suggest the epistemological problem concerning its capacity of designating the exact features of past or ongoing equilibria, necessary for rule's capacity to indicate the behavior to be followed (in order to get equilibria). This is again a semantic issue. The capacity of the rule to refer to salient features of the game are not given by any feature of the game itself, and it turns out to depend on the sharing of the beliefs about what has to be considered salient in that circumstance. In other words, it appears a necessary condition for being identifiable as a salient features to be commonly cognized. But, on the other hand, the capacity of the rule to designate salient features is what makes the common cognition of the recurrent equilibrium-state of the play the content of shared beliefs (which in turn make recurrent the equilibrium itself.) Either the criteria of salience are defined independently on the beliefs concerning the past regularities of the game itself, or we fall again in the circularity due to the necessity of assuming the sharing of the beliefs – which already depends on the commonly cognized salient features as represented by the rule – in order to explain salience.

These two aspects of circularity leaves semantics of Aoki's rules as a linguistic representation undetermined. The attempt to treat beliefs referring to past recurrent and salient aspects of the societal game as the foundation of rule-following and of the equilibrium in the societal game itself, overlooks the semantic problem consistent in the fact that experience of the past does not suffice alone for establishing a stable and univocal knowledge with regard to those experiences themselves. Other categories, notions and rules, made available by (institutional) contexts and (linguistic) practices are necessary for making the knowledge concerning past factuality, a relevant information for the constitution of institutional reality and for admitting the possibility of their acquisition to shared beliefs. Without these other contextual conditions,

beliefs and factual knowledge are not able alone to specify how the rule designate an unequivocal normative content and hence they cannot explain how rules are followed in the same way, in same circumstances, by different subjects.

The explicit choice made by Aoki of putting in circle the factors (individual and social, cognitive and behavioral) constitutive at the same time of the rule (as the foundation of shared beliefs) and of the possibility of the rule-following (as the foundation of the states of the play on which the rule's representation is applied), reflects his intention of endogenizing rule-formation and institutions itself. Aoki is aware of the epistemological risks of this circularity when he observes that:

[...] the common knowledge in a substantive form of an institution induces shared beliefs, while it summarizes and represents the salient features of recursive states of play induced by the shared beliefs. Then, the common knowledge and the shared beliefs appear to be almost a tautology.

As a consequence, Aoki finds himself in the necessity of assuming a further epistemic level of beliefs, the level of *cultural beliefs*¹⁵ in order to provide an external perspective on the dynamic processes which describes the coevolution of rules, beliefs and behaviors as embedded in wider contexts, formed by linked games of the moral domain (social norms) and of politics, among others (Aoki, 2011, pp. 29-33).

The consideration of these further level of analysis go beyond the epistemological scope of this paper. I limit myself to sum up the status of the question posed in this section and to point out the take home message which I hold for the purpose of the discussion of negative rule-following. The problem of interpretation entails a problem of epistemic foundation. What does make the rule to be followed in the determined way it states a normative content? As we have seen, this question is worth translating in the question concerning what founds the mutual knowledge that other will follow the rule in the same way as I do. This question leads Aoki to the necessity of understanding the conditions of the sharing of beliefs, which recursively (and circularly) is to be founded in the experience of the salient features of the behavioral past equilibria.

In order to avoid the circularity of the notion of salience and its semantic problem, and to provide an exit strategy from a tautological definition of

¹⁵On the ground of the availability within groups of common cultural beliefs, shared behavioral beliefs can be considered given, so that the common access to rule is sufficient to determine common knowledge without entailing the interpretation problem (Aoki, 2011, pp.26-27).

rule-following as exclusively grounded on the past occurrences and common cognition of the equilibria in rule-following itself, we can start pondering the idea that rule-formation and hence also rule-following derives its semantics from the fact of being part of wider circuits of inter-subjective and linguistic practices. Only within the context of these inter-subjective interactions, rules as institutional representations acquire the epistemological basis of language, knowledge, and practical meanings constantly defined and (re)defined in human public activity. In this way, we end up with two “context-dependency” conditions of normativity, which let us make more precise the insight advanced at the end of section 2 concerning the internal relationship between rules and practices.

As we have seen, rule-following in the regulative sense depends on the contextual availability of constitutive rules; and vice versa, constitutive rules cannot state in complete autonomy the existences of the institutional entities, but they require the regulation of the practices in which they are involved. So, the first condition of context-dependency to be maintained is represented by the network of relationships between rules (*rule-rules condition*). The second condition of context-dependency concerns the epistemic level of consideration for which rule-formation and rule-following necessarily entail the network of interactions, which as an ongoing process, while realizing emergent regularities, it provides the epistemic and linguistic conditions for their representation in rules (*subject-subjects condition*). This second condition is object of further analysis in the next section, which investigates the capacity of Searle’s application of collective intentionality, to catch commitment as the deontological essence of language and normativity. This is also the occasion to show as the problem of interpretation can be solved, not only by referring to beliefs based on the experience of the past, but by dealing with intentionality forward connected to future possible achievements. Incidentally, the possibility of (at least in part) disregarding the experience of the past, is what makes intentionality the epistemological notion apt to deal with novelty and change in rule-following.

4 The paradox of commitment

The problem of interpretation raised in the last section is ultimately due to the status of representation attributed to rule as a correlation device. As we have seen, Aoki, for the sake of epistemic foundation, interprets this merely instrumental role of rule by dividing analytically behaviors and beliefs and by attributing the rule the function of mediating between the two. In this way, rule at the same time provides the conditions for beliefs’ sharing but

it depends on belief sharing in order to provide with a foundation common knowledge and hence the equilibria that it represents. The instrumental status of rule/representation itself is responsible for this circularity and also for the potential infinite regress caused by the fact that, without an autonomous specification of its semantic capacity, a mere regulative rule alone does not designate univocally and stably its prescriptive content. It cannot do this, if its function is only conceived to be that of prescribing behavioral equilibria, and at the same time its semantic functions depends only on those equilibria to happen (and on the experience of it, via the assumption of some notion of salience).

Searle's focus on *intentionality*, escapes this problem of circularity and of interpretation because it provides the possibility of treating normativity as intrinsic to (constitutive) rules, and not instrumentally aimed at an objective separated from it (equilibrium, correlation, cooperation etc.). Through the notion of intentionality – defined as the “feature of mind by which it is directed at, or about, or of, or concerns, objects and states of affairs in the world” – behaviors and states of mind (such as beliefs, desires, intentions, etc.) can be unified in a more comprehensive (and “monistic”) conception of action – and in particular of *speech acts* – and agency can be directly described in terms of the (linguistic, grammatical) distinctive modalities of relating to (and in this sense, posing and creating) its object and of the consequent ontological implications (Searle, 1969, 2002, 2010). Only by re-placing representations and rules within the context of the individuals intentional activities to which they are interconnected, the linguistic modality through which they specify a normative content (and the relative semantics) can be referred to actual practices of rule-commitment, not (only) dependent on already-given occurrences of equilibria in rule-following.

From this perspective, Searle can state that it is not the intentionality of (mere) representation but the intentionality of *declaration* which is connected to institutional rules. While mere representation is only engaged in depicting how things are in the world (function which makes it be attributed truth values), and other speech acts, such as orders or promises, comply with the only function of making the world how they intend it, declarations assumes both the “direction of fit”: they produce a change in the reality by stating the reality of the object they are at the same time representing; in Searle's words, “you make something the case by explicitly saying that it is the case.” (Searle, 2010, pp. 11-15). This simultaneous double function makes declarations, and in particular “status function declarations”, the linguistic, logical tool for the creation of institutional reality through constitutive rules: they create the

institutional reality of status functions by representing them as existing¹⁶.

Status-function declarations and constitutive rules entails *collective intentionality*, in so far as the attribution of the status-function that they put in existence, while representing it, requires *collective acceptance* or *recognition*. I will clarify in section 5 how this collective modality grounds the semantics of constitutive rules of Searle's social ontology. For the moment, I concentrate on the notion of *collective intentionality* which is the general epistemological notion necessary for Searlian conception of institution as deontological commitment (Searle, 1990, 1995, 2010). Searle is very straight on the point that collective intentionality, as a type of intentionality, pertains to individuals (and in particular to individuals' minds) and can only determine individual's action. Nevertheless it results in collective, shared behaviors, such those described by the grammar of "first person, plural form of intentionality", for example expressed by linguistic construct such as "We are are doing ...", "We believe ...", "We intend ...", (Searle, 2010, p.43). The point to be explained then concerns how individual intentions can contain collective ends – a we-intention in each of the heads – such that coordination in the behavior can result because represented by each individual as a collective enterprise in which each of them do his/her part.

In putting the problem in this way, Searle is willing to demonstrate that collective intentionality has to be taken as a primitive notion, and its occurrence cannot be reduced to the contents of I-intentionality, plus mutual beliefs¹⁷. For Searle, this attempt of reduction counts as a misunderstanding of methodological individualism, for which from the "basic fact" that intentionality pertains to individuals (and in particular to individual brains), it is inferred that "We-intend statements" can be reduced to "I-intend statements". There are irreducible cases of genuine collective behaviors. These cases cannot be reduced to I-intentions plus mutual beliefs and common knowledge in so far as they presupposes the sharing of the representation of the communal goal to which each individual action is addressed. In Searle's examples, from the fact that two persons are driving towards the same destination, and they mutually know that the other is doing the same thing, it cannot be inferred that they are doing something together, or even pursuing the same purpose.

On other hand, the joint action of "taking a walk together" (and the

¹⁶We will return in section 5 on the problems connected to the assumption of declarations as the foundation of institutional ontology and normativity. For the moment, I only signal that constitutive rules are considered as *standing Declarations*.

¹⁷Searle's main critical objective is Tuomela and Miller (1988). Cf. Tuomela, 2005. For an alternative approach to collective behaviors in terms of team reasoning see Sugden, 1993, 2000; Bacharach 2006; Gold and Sugden, 2007.

mutual benefit deriving from it) presupposes the fact that the agents involved have in their mind an intention of the form “We are walking ...”, and this is what makes it a joint action, since each derives what he/she has to do in order to get the common purpose from that intention, at the same time, knowing that the other (who share the same intention) is doing his/her part. In the case of cooperation the logic of collective intentional behavior is even more stringent. In order to cooperation as a collective shared behavior to happen, common knowledge or mutual beliefs are not a sufficient condition¹⁸. Searle clarifies this point in favor of irreducibility of We-intentions to I-intentions, discussing the two cases of the Business school example (Searle, 1990, 2010, p.47-48).

They depict two different occurrences of the same action undertaken by two groups of people, for the purpose of underlining what distinguishes cooperation in the sense of collective intentionality. In the first case, there is a prior commonly-known end, but each of the member of the group pursues it individually, only mutually knowing that the others are doing the same. This group of Business School graduates has been convinced by the “invisible hand” story and each of them goes in the world trying to maximize individual richness, in the mutual believe that, if the others do the same, this will benefit humanity. In the second case, the same communal end (welfare optimization, through individual maximization) is adhered collectively by the graduate students, through an explicit pact. They, again, go in the world trying to be as selfish as they can, but they do it because they share an *obligation*.

Notice that also in this second case (as, of course, in the first case) there is no first level of cooperation: as prescribed by invisible hand story, they do not cooperate to achieve the goal of benefiting humanity. Collective intentionality is given by the second order cooperation concerning the deontological sense of commitment toward the fact that the end, even if carried out individually (and if embodying the individualistic ideology of the “invisible hand”), is shared collectively through the (linguistic) act of promising. In the first case, for example, defection does not result in the violation of the deontic duty, but

¹⁸Searle analyzes two main cases of cooperation in the collective intentional sense. When cooperation is obtained by the *causation* of some further effect through the combination of individual’s action (body motions), then we have a structure of action which is *causal by means of relation*. The example for this type of collective intentional action is two persons trying to start a car engine by means of one pushing and the other setting the clutch out. On the other hand, we have a collective action which is *constitutive by-way-of relation*, when the cooperation in a group endeavor *constitutes* the desire effect. Searle’s example for this second case is the performance of a music duet, where one musician play the piano and the other the violin part. In this case, the collective action does not cause the duet to be performed, but it constitutes it (Searle, 2010, pp. 48-55).

merely in the individual change in one's own end. Collective intentionality accounts for the situation in which, together with the collective achievement, individuals are normatively engaged and expect others are doing their part in the shared commitment. Without this kind of individual commitment toward collective purpose there is no cooperation in the collective intentional sense (Gilbert, 1989, 2006, 2013).

Once clarified this, the problem becomes how to justify, without making reference to super-individual (mind) intentional capacities, how-intentionality can affect (determine, coordinate) individuals' action, or to put it in Searle's terms: "How can it be the case that we intentionally can move individual bodies if the content of the "We" is not the same as the content of the "I", which constitutes doing one's part of the collective effort?" This question is crucial for the comprehension of the intentional structure of collective intentionality, and it leads Searle to specify in which precise sense collective intentionality presupposes that individuals have shared intentional content about the fact that each of them is performing his/her own part, while achieving the communal end. The notion of "intention-in-action" is crucial for understanding this point¹⁹.

On one hand, collective intentionality is given by the fact that while the joint action is performed "each has to assume that the others also have an intention-in-action which has the same goal, the same "collective B" where the singular A can be different because each person can only perform his own action A". On the other hand, this belief concerning the communal goal, does not refer to the content of the intention-in-action performed by the others individuals involved in the collective action. If the collective goal has to be part of the subject intention-in-action, the content of subjective intention-in-action can only make reference to things the subject can cause (or at least believe he can cause):

I have to believe (or assume or presuppose) that others are cooperating with me. And their cooperation will consist in their having intentions-in-action that specify the same goal as I have but need not specify the same means to the goal. I have to believe they are cooperating, but [...] it is not part of the content of my own intention-in action to cause that cooperation. [Moreover] I need not know what their contribution

¹⁹The distinction between "collective intentions in planning" (named, *collective prior intentions*) and "collective intentions in acting" (*collective intentions-in-action*) is necessary to highlight the structure of mutual intentionalities entailed by the sharing of a common, cooperative end. They respectively refers to collective intentions prior to the performance of an action (designated by propositions of the form: "We intend that we perform act A") and collective intentions during the performance of the action ("We are now intentionally performing A")(Searle, 2010, 43-45).

is. I need not know the value of the “A” in the form of their intention-in action. [Especially]for complex acts of large groups no one knows what everybody else is doing. ”(Searle, 2010, pp. 51-55)

The commonality in collective intentionality is then a commonality in the purpose (which grounds the belief that the others are doing their part) and not in the means to acquire it, i.e. the individual behavior as established by singular individual intention/action. In other words, each individual action of others is beyond the scope of subject’s intentionality. There are not intentional content which trespass the scope of individual intentions, but nonetheless the individual choice, within cooperation in the collective intentional sense, presupposes the capacity of envisioning the shared end as an irreducibly collective purpose. But then, on what grounds, the *commitment* to collective intentional action can function in establishing this complex structure of belief and intentions? The possibility of (deontological) shared commitment towards a common purpose in collective intentionality is provided by language (and in particular, but this is not a necessary condition, by communication).

Only in language, there is the possibility – vital to the constitution of institutional ontology and of its functions – of designating ends forward directed towards purposes for collective adhesion, and at least in part independent on pre-existed equilibria and on the beliefs based on them. Within the theoretical framework of collective intentionality, it becomes hence possible to treat the internal relation between language (speech acts) and the practices of institutional activities. This overcomes the problem of interpretation in so far the linguistic conditions of commitment, already imply a shared interpretation of the intentional content to which institutional practices and rules are directed, without requiring the fulfillment of further conditions of enforcement. Interpretation is intrinsic to the linguistic practices through which collective ends and institutional functions are posed and adhered in collective intentionality.

Without the possibility of entering all the complexities of Searle’s analysis of language’s functioning, I summarize it through the following quotation, with the purpose of highlighting a peculiar paradox which involves the internal relation between language (communication) and commitment:

We will not understand an essential feature of language if we do not see that it necessarily involves social commitments, and that the necessity of these social commitments derives from the *social character of the communication situation*, the *conventional character of the devices used*, and the *intentionality of speaker meaning*. It is this feature that enables language to form the foundation of human society in general.

If a speaker intentionally conveys information to a hearer using socially accepted conventions for the purpose of producing a belief in the hearer about a state of affairs in the world, then the speaker is committed to the truth of his utterance (Searle, 2010, p. 92, emphasis added)

The commitment is thought to be internal to language and in particular to the speech act through which the intentionality of the speaker (for example, the content of a belief or of a representation) is communicated. As underlined in the quotation, communication is a public situation in which the two components of conventional linguistic devices and of the speaker meaning interacts. For Searle, what is communicated in speech acts are *intentional states*: roughly speaking, the informational content (the representation of an object or a state of affair, called propositional content) together with the intentional attitude toward it (belief, desire, intention, etc.). To say it with Searle, the aim of communication is to try to “reproduce [one’s] intentional content in the head of the other person.” But the intentional content is first and foremost stated “privately” as a *speaker meaning*, and then it needs to be translated in the forms of conventional meanings (words and sentences), which provide language with stable inter-subjective basis for communication. If communication has to work on a regular basis “there has to be some *socially recognized conventional device*, some repeatable device, the production of which can be regularly and conventionally taken by [...] interlocutors to convey [speaker’s] message.” (Searle, 2010, pp. 73-76)

With the passage at this conventional level, language internalizes the public dimension of commitment. If the utterance of a belief, for example, takes the form of a statement, the subject gets publicly involved with the commitment toward the stated truth. The point can well be illustrated by referring to the speech act of promising (already involved in the business school example, above). Promising is not only the expression of a private intention (possibly, to be believed or not), but it involves the subject’s commitment to carrying it out. A proper constitutive rule, with its deontological implications, is introduced by the very speech act of promising (but the same hold also in the case of statements): “making such and such an utterance X in the context C count as making a promise Y”. Promising is itself an institution, which transforms the speech act of pronouncing a promise in a concrete (real) engagement in the action to be done. This deontological implication is manifested by the fact that violating a promise is not only the negation of the truth content of the promising proposition, but it damages (affects, change) the institution of promising itself (cf. Rawls, 1955; Gilbert, 2013).

Promising reflects clearly what Searle points out as the two components of commitment in general: the notion of an “undertaking that is hard to

reverse” and “the notion of an obligation” (Miller, 2007). In fact, promising, while creating an obligation, is itself an action the consequences of which are not easily reversible (for example, the consequences on the beliefs and expectation of who receives the promise). Language is intrinsically bearer of commitment in so far as every speech acts performed in accordance to rules combines these two aspects of irreversibility and obligation (Searle, 2010, pp. 81-84). In this way Searle can conclude that “once we have an explicit language in which explicit speech acts can be performed according to the conventions of the language, we already have a deontology. We already have commitment, in the full public sense that combines irreversibility and obligation. [...] there is no such deontology without language.”

This consideration makes language the fundamental institution, in Searle social ontology. Language as the source of commitment is constitutive of all institutional reality, which (only) in this sense has to be considered as linguistic²⁰. But a point remains not clear. On what grounds the speakers and hearers adhere to linguistic conventions and are committed to the rules of communication? As we have underlined, the possibility of language’s capacity of creating a deontology depends on the availability of “socially recognized device, some repeatable device, the production of which can regularly be intended by speakers to convey the message”. This regularity, the regularity in the adoption of the linguistic convention, is hence in need for an explanation, capable of accounting for the conditions of stability (equilibria?) under which social actors stick to linguistic rules. This is a type of commitment which cannot be explained by making reference to the deontological modalities made available by language, because it is prior to the possibility of language itself, and hence language and communication presuppose it.

With the phrase “paradox of commitment” I want to signal this circular situation for which language creates the public conditions of institutional commitment but at the same time depends on the individuals’ mutual commitment to conventions and the rules of language itself. This circular situation is analogous, from the opposite specular perspective, to the circular argument advanced by Aoki for dealing with the dichotomy internal to rule functioning, between equilibria in the behaviors and shared beliefs. And analogous is also the consequent need for a third principle (cultural beliefs in Aoki) in order to provide dichotomy and the circularity with an external reference capable of founding the explanation of the dual phenomenon at

²⁰From the fact that institutions imply language, does not follow that language implies institutions. For Searle, if it is true that there are no institutional facts without some form of deontology, as established by language intrinsic commitment, institutions are mainly not linguistic, and institutional facts go beyond facts about meaning (Searle, 2010, pp. 90-93).

stake on an independent factor i.e. by taking the two elements of the dualism not as elements of the relation between an explanans and an explanandum (which is then problematically reversed in circularity), but as a unique explanandum all together. This kind of third principle is ultimately provided, in Searle, by pre-linguistic intentional states, which as a-priori categories and intrinsic intentional modality of brain functioning, provide with a “naturalistic” foundation meaning’s conventions and language’s generativity (Searle, 2010, pp. 61-71).

In this way, the dualism and complementarity between rule and equilibrium re-presents itself in the form of the reciprocal epistemological dependency of the conventional conception of rule (such as that of Aoki) on the linguistic capacity to indicate the direction of collective commitments in a semantic non-ambiguous way, and of the deontological conception of rule (such as that of Searle) on the coexistence of equilibria in the practices of language itself. I will try to reinterpret this complementary as the opportunity for an attempt of integration of the discussed epistemological presuppositions behind the conventional approach and the deontological approach. This purpose leads me to address directly, starting from the next section, the problem of the origin of commitment and of its maintenance in practice. This problem exactly derives from Searle’s aprioristic and foundational conception of the relation between language and commitment, which is reflected in Searle’s treatment of the creation of “new” institutions through status-function declarations. Without the necessity of disputing Searle’s assumption of intrinsic intentional state, I directly discuss the consequences of this move at the level of his social ontology by referring to what I consider the *incompleteness of deontology*. This lets me start advancing some elements of the *relational methodology* through which I will then try to rethink the problem of rule-following and of rule-formation.

Part II

No normativity without alternatives

5 Incompleteness of deontology

The previous sections left us with the restored complementarity of equilibrium-based and rule-based approaches, in the refined versions of Aoki’s rule-as-

cognitive-media and Searle's constitutive rules. The complementarity moves one step back, with respect to the original formulation of the dualism rule-equilibrium, and it involves the epistemic basis of enforcement and commitment respectively assumed by the two approaches to rules. If, on one hand, Aoki's requirement of shared beliefs ends up depending recursively on the availability of regularity in the social interpretation of the rule's meaning, Searle's collective intentionality can provide the basis for the constitution of social ontology, if and only if it can rely on conventions at the level of language's uses in communication. If, on one hand, within Searle's framework the interpretation problem has been vanished by the circumstance that the sharing of purpose in collective intentionality directly, and intrinsically connect individuals involved in one precise interpretation of the rule in the specific circumstance of application; on the other hand, only in terms of equilibrium in inter-subjective practices, language can function on a stable and regular basis as an instrument for the creation of commitment, consequent to the public use of language.

This paradox of commitment is reflected by Searle's assumed asymmetry between cooperation and *collective recognition* within institutions (Searle, 2010, pp. 56-58). By specifying the collective intentional structure of collective recognition, Searle is able to clarify the relationship between cooperation (in the sense of collective intentionality, analyzed above) and institutional reality. Collective recognition is a form of collective intentionality (weaker than that involved in cooperation) presupposed by institutional ontology, in so far as it is the way in which institutional entities are identified with reference to their function. As signaled in section 2, the creation and maintenance of institutional reality depends on the imposition of status functions, which consists in the attribution of a structure of purpose on reality. As Searle observes, "once we have described something in terms of function we can introduce a normative vocabulary" (Searle, 2010, p. 59). This is because functions can be identified only through reference to purposes (function is only identifiable in terms of what it is function of) which are posed through intentional projection of ends (mind dependent, value relative). The creation of deontological status-functions hence consists in the identifications of powers: possibility of actions toward a purpose, and the consequent activation of commitments.

In this sense, institutional structures provide *desire-independent reasons for action*: agency's modality determined not by the logic of fulfillment of desires (on the ground of previously established tastes, preferences, or beliefs grounded in experiences) but forward-directed in terms of the intention of achievement toward an end enabled by the institution itself. Collective intentionality is involved in the attribution of the status function in so far

as it is only through the collective adhesion to the structure of purpose embodied in the institutional entity that its deontic status and function are defined. The status function are possessed by persons or other objects not (solely) in virtue of their physical structure, but in virtue of collective imposition and recognition of that status as the potentiality of acting (toward specific purposes) in the way established by the relevant deontology (power). The institutional entity is characterized by having a certain status, and the collective imposition of that status enables it to perform the institutional function.

The point Searle is willing to make is that, institutions necessarily requires collective recognition by the participants in the institution, but cooperation in the collective intentional sense is activated only in particular transactions within institutions, and, for this reason, it depends on prior collective recognition (Searle, 2010, pp. 56-58). For example, prior to a money transaction we have to acknowledge the paper notes as valuing as money, and in general the institution of money itself, together with the institution of commerce. Once the institution is established, cooperation within it depends on the previous collective recognition of the institution itself, and this recognition does not involve active cooperation²¹. It is my opinion that the assumed logical priority of collective recognition is responsible for the impossibility within Searle's social ontology of treating in a complete way the problem of the *origin* of deontology and of its actual *subsistence*. This can be shown by considering that deontic powers are not self-contained and self-sustaining.

Actually, taking deontic power relations and the constitutive rules that put them into existence as foundation of all the institutional facts is problematic for at least two reasons. The first reason is that, it overlooks the fact that the subsistence of deontic duties and rights depends on the equilibria (or at least, regularities) in the actual practices of their applications. This point can be emphasized recalling Searle's recurrent example of the twenty dollar bill. On one hand, it is true that the use of the bill presupposes the collective recognition of its status-function, via the recognition of the power relations, defined only in terms of the potential actions (duties and rights)

²¹The asymmetry between collective recognition of an institution and cooperation within it is also manifested by the fact that, the former can be reduced to individual recognition plus mutual beliefs, while the latter cannot, as demonstrated by the Harvard Business School example (above in previous section 4). In the case of collective recognition, a counterexample such that cannot be provided, because even if individuals were to recognize the institution only individually, in case of mutual knowledge that they so recognize it, there would be collective recognition. Searle can hence conclude that there are forms of collective intentionality which are reducible to I-intentionality plus mutual beliefs (Searle, 2010, p. 58).

enabled by it (as for example intrinsic to the notion of “value”). But, on the other hand, this definition is not univocal and viable, unless the conditions of the actual uses of the institutional object or activity, are at the same time specified and acknowledged in practice.

The functionality of a bill is not only defined by the a-priori attributes which identifies formally its abstract, general, institutional status of counting as “(a representation of) value” or as “a means of exchange”. A twenty dollar bill is not only a twenty dollar bill (its strictly deontic content in terms of duties and rights). It is also what you actually do (or do not do) with it. These, in general, are practices which are defined in terms of the multiple purposes which subjects attribute to institutional objects in the moment in which they use them (being involved in a wider network of institutional practices). For example, the actual purchases that an individual can make, has to be considered part of the meaning of money, and it is at least in part independent from duties and rights, as defined by the deontological specification of what ought to be done with money. The institutional essence of money, cannot in fact be separated from the actual actions of purchase (of lending, of saving, of exchange, etc.) in which money is involved in economic history, together with the deontological definition of its status in terms of general duties and rights. In other words, powers enabled by institutions are not only deontic potentialities, but they are actual choices among alternative possible purposes. We will be back on this point later in this section, in order to specify the conditions of purposiveness (the choice between alternative ends) as fundamental for normativity and for rule-following (overlooked by a purely deontological approach)²².

The second reason for which Searle’s deontological foundation is problematic is that it cannot alone explain its own birth. If, on one hand, collective recognition of status functions is a necessary condition for an institution to exist, the recognition, as such, presupposes the *pre*-existence of the institutional function to be recognized (a recognition is always a recognition of

²²In the *Construction of social reality* Searle presents one aspect of the relation between deontological meanings and practices in this way: “The word “money” marks one node in a whole network of practices, the practices of owning, buying, selling, earning, paying for the services, paying of debts, etc. [...] The word “money” functions as a placeholder for the linguistic articulation of all these practices.” (Searle, 1995, p. 52). In Searle (2010) the relation between practices and deontological meanings seems weaker, as a result of the foundational strategy embraced by the author in the moment in which he presupposes the speech act of declaration as the necessary condition of constitutive rules and deontological ontology. The problems connected to the assumption of this foundation provide me with the (critical) theoretical reference for elaborating my notion of practice as individual-and-collective purposive choice. And this represents the main reason for which I keep on referring to Searle (2010) in what follows.

something that already exists). This circumstance reproduces the paradox which, as we have seen above, involves commitment in its circular relationship with language. Circularity prevents the possibility to account for the origin of deontology because, if its origin depends on collective recognition, deontological relations of power has to be already available (already originated) in order to be recognized. The fact that Searle is aware of this problem is manifested by his assumption of the speech act of Status Function Declaration as the foundational moment (the origin) of institutional status functions:

[...] Institutional reality is created and maintained in existence by (representations that have the same logical form as) Status Function Declarations, including the cases that are not speech acts in the explicit form of Declarations. [...]

The most general form of the creation of an institutional fact is that we (or I) make it the case by Declaration that the status function Y exists. Constitutive rules of the form “X counts as Y in C” are what we might think of as *standing Declarations*. [...]

Thus, the existence of a Declaration is itself an institutional fact and thus a status function. But does it itself require a further Declaration to exist? It does not. Indeed, if it did, we would have an infinite regress. (Searle, 2010, pp. 13-14)

In the quoted passage, it is made explicit that Declaration works as a foundation: the linguistic foundation of the nonlinguistic institutional reality²³. But this foundation, as such, is problematic (at least, it is problematic as long as we want to look for an alternative to Searle’s naturalization of intentionality). This further level of foundation has to be independent and autonomous in order to escape the circularity of commitment and of status functions recognition. But the assumption of the speech act of status function declaration can do nothing but reproducing the circularity at the deeper ontological level of the relationship between the institutional function and its ontological (concrete) bearer. I illustrate this point by making reference to Searle’s example (thought experiment) of the constitution of (the ruins of) a wall as a boundary²⁴.

²³Searle needs to make a distinction between the circumstance that all institutional facts are in a sense linguistic (in so far they are constituted and maintained by the linguistic functions of constitutive rules), and institutional facts which are not linguistic, because their objects and activities becomes real component of ontology, at least in part independently from their linguistic statement (Searle, 2010, pp. 90-93).

²⁴The following discussion can be applied to the more complex cases, presented by Searle as further step of his analysis, of a) the establishment of the status function through an explicit constitutive rules and b) of the creation of a corporation which necessarily entails explicit rules, a complex legal structure, and language (Searle, 2010, pp. 96-100). Cf. Smith, 2003.

Searle imagines a tribe which builds a wall for defensive purpose. In the beginning of the story, the defensive function of the wall (the function of restricting the access) is carried out by the physical structure of the wall alone (the wall is too high to climb over easily). Then, the wall decays and only a line of stones remains. Searle goes on supposing that nonetheless inhabitants, as well as outsiders, continue to recognize the line of stones as having the status of a boundary: the deontological delimitation of a territory expressed in terms of rights and duties. In fact, peoples continue to recognize collectively that it is forbidden to cross the boundary, unless authorized. In this sense, declaration functions in establishing the fact that, independently on its physical properties (which are lost in the decay of the actual wall), the line of stones represents “symbolically” the status function of the boundary with all its deontological implications (Searle, 2010, pp. 93-96).

Two correlated observations can be made. Firstly, what does it make the original wall an impassable limit. Searle would answer, its physical characteristics (for example, its height or depth). This is obviously true, but I would suggest that this is not enough. In order the wall to function as an effective defense, it has already to *count as* the deontological demarcation of tribe’s territory. No matter how high or depth the wall could be, this would not prevent (at least) the attempts by other intentional subjects to pass it, unless its presence is already assigned the deontological function of a boundary. What does make the wall to perform its physical function (a part from the physical structure itself), is, for example, the fact that a possible attempt to destroy it by outsiders, would already be recognized (both by the insiders, and by the intentionality of the outsiders) as the violation of the tribe’s right to preserve its own resources, with the possible consequences in terms of defense and punishment of the violation.

Secondly, it is not completely true that the line of stones in the second part of the story is attributed of the deontological function of being a boundary only in virtue of the status function declaration. The status function declaration is not a sufficient condition, because it presupposes the past collective recognition of the original wall as a boundary. If it is true that, in any case collective recognition is a necessary requirement for the attribution of the deontological status function, this latter does not come out from nothing. If the line of stones wasn’t already recognized as what in the past was, both physically and deontologically, an impassable limit, the collective recognition wouldn’t have nothing to be attached at²⁵. Not only this. Together with the

²⁵In fact, not every line of stones which can be found in nature, neither the highest mountains, can count as an object of collective recognition, unless there is an history of previous attributions of status functions, and related practices, which make them a possible object of deontological identification.

collective recognition of that piece of reality (including its history), all the deontological functions of objects, persons, activities, and events related to it have to be given collective recognition²⁶. For example, both the original wall and the line of stones cannot be recognized as representation of deontological relations, unless a relevant collective cognition of what can be considered a “territory” and the legitimacy of its possession are already available²⁷.

This last point reminds us one of the provisional conclusions obtained in section 2 and put us in the condition of starting rethinking Searlian notions of rule: rules, both constitutive and regulative, are not followed in isolation. This insight can be reformulated by considering that the relation between deontology and its objective bearer is not only top-down: from the collective recognition of the status function via declaration, to the attributed passive object. It is, at the same time bottom-up, to the extent that that object, person, state of affair, together with all the regulative rules which *already* disciplines their functioning represent a condition of applicability of (new) constitutive rules. This can suggest that there is not a strict line of separation between regulative and constitutive, but the distinction can be traced only case by case, with the functions defined on one side, always calling for the functions defined on the other side, along with the dynamics of possible changes in time of these relations.

If we consider also the problem of subsistence – i.e the dependence of deontology on practices, as the necessary completion of the status function potentialities by the actual uses which constitutes stable and concrete network of practices – we have reconstituted the framework of context dependencies conditions (rule-rules condition and subject-subjects condition) derived in the first part of the paper, which I consider necessary conditions for dealing with rule-following. The lack of consideration for the dependence of status functions on interactive practices and of institutional determinations on the coexistence of networks of rules, is responsible for what in the title of this section I summarized as the “incompleteness of deontology”. Incompleteness

²⁶Incidentally, a defensive wall cannot perform its physical and deontological function, without having a door, which in turn has to be attributed of its own physical and deontological functions, for example, by defining who can be admitted in or not. Obviously, it is vital for the tribe to be able to get in, and hence, it is a fundamental condition for the functioning of the wall itself and for its deontological status that at least the members of the tribe are admitted to get in.

²⁷The decision concerning where to build up a wall is secondary, from a logical point of view, to the deontological connotation of the notion of territory. If for example, the territory is collectively recognized as including all the spaces from which the resources necessary for the tribe sustenance can be extracted, including for examples, rivers and forests, there could not be no physical nor symbolical wall which could correspond to this more fundamental deontological attribution of status function.

is the result of the attempt of founding deontology (circularly) on itself, and it is reflected on the impossibility within the framework of Searle's social ontology of accounting for the modalities of origination, the maintenance in actual practices, and, as we will see in the following, the change and variation of deontology within institutional evolution. These shortcomings call for a rethinking of the internal relation between normativity and institutions capable of reintegrating Searlian notion of commitment (as intrinsic to deontological collective intentionality) within the interactive practices of cooperation which actually provide deontology with an historical and concrete content.

In order to think at an alternative to Searle's problematic foundational strategy, it is worth re-formulating the problem of the origin of deontology, and of its subsistence in actual practices, assuming the point of view of rule-formation. Recalling that at every function corresponds the assumption of a purpose, the research question then becomes: what is the origin of the purpose embodied in status functions and collectively recognized through status function declarations? The focus on purpose provides, as we will see, the possibility to interpret the intentional modalities of rule change as representative in general of the conditions of rule formation and of rule maintenance. This approach makes available an alternative strategy for solving the paradox of commitment, without appealing to a more fundamental cause behind language and deontological status itself (intrinsic intentional state in Searle), but interpreting deontology as part of actual interactive (inter-subjective) practices of posing (and choosing) an end among alternatives. In this way, Searle's deontological foundation of institutional rules and its incompleteness, will be complemented with a conception of rule-following as the result of dynamics and properties of networks of purposive relations.

I prepare this discussion by highlighting the epistemic conditions behind the possibility of identifying purposes as the determinant of rule-following. In this way, I start considering the principle which defines the field of discussion of this second part of the paper for which there cannot be normativity without alternatives. The decision to follow a rule depends on the availability of alternative ends. Only with reference to these alternative ends, complying to the end entailed by the rule is meaningful. Without these alternatives, the rule (and its statements) would be simply unnecessary and the behavior dictated by it would simply result as the consequence of non-deontological (for example, bio-physical, deterministic) causes. In other words, there is no logical necessity of stating duties and rights, if the potentialities opened by them are already practiced as choices (given as chosen).

This consideration leads us to rethink rule-following behaviors in terms of the relation between the subject and alternative ends. This is not equivalent

to sustain that rule-following is a subjective, arbitrary deliberation made by individual in isolation and independent on anything else. I am interested in logic-epistemic conditions of purposiveness and in this respect the description of rule-following in terms of the relation subject-ends, directly entails the description of the conditions which make this relation possible. These further conditions of rule-following can be schematically summarized as the relations between the subjects and other subjects and the relation between the subjective end and the other alternative ends.

In this way, we get a reformulation in relational terms of the two context dependency conditions discussed so far. On one hand, the purposive relation subject-ends depends on the interaction between the subject and other subjects. Only within this context, in fact, the rule-following decision and the correspondence (or not correspondence) of the consequent action to its deontological content is realized in practice. Practice is the dimension in which deontological rules are specified as the actual interactive trade-off between subjective purposive behaviors, and this is the condition in which duties and rights are stabilized (or changed) in/by inter-subjective actions. This, on the other hand, requires that ends are in relations between themselves. As just observed, the relation subject-ends presupposes the availability of a multiplicity of alternative ends. A deontological declaration cannot pose the purpose enabled by the function it establishes as an isolated autonomous end. As we have considered, it presupposes a whole network of other regulative and constitutive rules for being effective, and these expresses just as many power relationships.

The two contextual conditions subject-subjects and end-ends are connected by a mutual implication, which makes it possible to overcome the dualism between methodological individualism and methodological collectivism²⁸. By focusing on relations instead of on objects – and in this way advancing what in the introduction I have defined as *relational methodology* – it is possible to avoid the necessity to choose between the logical-ontological priority of (the properties of) individuals over collectivities, or vice versa, of

²⁸Cf. Davis, 2003, and Hodgson, 2007. I consider both individualism and collectivism as methodologies centered on objects. In both cases, agency is seen as realized by an object and its attributes. It can be practiced by individuals and then it has consequences on the totality of the whole (assumed as a distinct object) of which individuals are parts; or it can be due to the fact that the totality has its own capacities independent from individuals' properties and hence able to influence them. What I am proposing is viewing agency as an attribute of relations, in this way trying to overcome the alternative between methodological individualism and collectivism, and between agency and structure. However this methodological option is only suggested here and left to a further discussion. In any case, it has to be cleared that by assuming relations as methodological reference I am not sustaining any ontological relativism, but only a different specification of realism.

(the reification of) the whole over its parts. In this way, we can obtain a unified ontology (and epistemology) in which speaking of the relations between alternative ends is only another name and another aspect of the interactions among individuals in so far as both are identified and described as modes of relations, attributes of the same network of connections.

From this relational perspective the choice which connects the individual subject to one determinate end among alternatives is not anymore to be considered as subjective (in the sense of the opposite of objective and of collective) because the possibility of the specific practical and epistemic relation between one subject and one end is only possible in the context of the whole network of relations between the other subjects and the other ends. This approach provides my rethinking of Searle's deontological ontology with the theoretical grounds for adopting purposiveness as the reference for describing the intentional modality of rule-following. Purposiveness, elaborated through the notion of *interactive intentionality*, is the element which makes the adhesion to a given deontological structure, not only a matter of individual commitment and collective recognition (necessary but not sufficient conditions of rule-following), but also an interactive practice, with the potential capacity of transforming institutions and its cognitive basis.

6 Interactive intentionality

With the purpose of complementing Searle's social ontology perspective with the epistemological conditions of rule change as representative of rule formation, I elaborate the notion of *interactive intentionality* in order to incorporate cases in which agency is the result of the *interactive formation and choice of an end among multiple alternatives*, and not merely the joint sharing of a given purpose, as in collective intentionality (at least in the Searlian version). This provides some suggestions for the interpretation of negative rule-following, and for representing cases of generation of new normativity, through interaction and possibly conflicts between constitutive and regulative rules, deontology and practices. The comprehension of these cases will in turn offer insights concerning the essence of rule-following and the subsistence of normativity in human institutions in general: in any occasion in which a rule is followed (or not followed) it is re-created as the rule it is by the practice of a decision which, while fixing the relation between a subject and an end, it reconstitutes the whole context of relations and interactions in which all the possible ends are defined and organized in institutions. Interactive intentionality represents the epistemological prerequisite for the rethinking of rule-following, rule-formation, and change, which will be put through in

the next section and in the conclusion.

For interactive intentionality to work a specific logical-ontological condition is required: (at least potential) multiplicity of foreseen ends. In the case of one, already given purpose – implicitly or explicitly assumed by both the notions of convention and of collective intentionality – the problem of enforcement and of commitment, and the necessity of founding circularly one into the other, emerge²⁹. In particular, the incompleteness of deontology depends exactly on the fact that the necessary condition of collective recognition can be applied only to what is given as a status function, already determined with respect to the implicit structure of purpose and of commitment which is supposed to be its univocal and autonomous logic-epistemic basis. Exactly for the necessity of assuming one given shared function/purpose as the condition of applicability of collective intentionality, Searle’s deontological ontology remains incomplete and unable to explain its own origin and subsistence, i. e. the origin of the structure of purpose recognized and practiced through individual and collective commitment.

The reason behind the hypothesis of interactive intentionality is to reinterpret Searlian status-functions declaration as an interactive practice³⁰ which, from time to time, can select and point alternative ends, and not exclusively on the ground of given institutional purposive functions. Multiplicity of ends is given in language, conceived not as the locus of pre-established conventional meanings, but as the activity through which individuals and communities cognize and practice their institutional life. If, following Searle, it is true that deontic functions are enabling powers which provides the possibility of desire-independent reasons for action, than we can suppose that this independence can have a role in institutional evolution, by exerting a feedback influence on the institutional and linguistic basis from which it springs up.

Abandoning the exclusive logic of the fulfillment of pre-given desire (tastes, preferences) means opening the possibility of at least some degrees of auto-determination of intentional purposiveness. Interactive intentionality is the

²⁹One given end is assumed by the theory of convention through the standard rationality assumption of utility maximization, which poses one homogeneous type of end indiscriminately behind every types of preferences (for example, both pro-social and self-regarding ends), and reflected in the fact that payoffs are supposed to be given and commonly known. On the other hand, collective intentionality, as already enlightened, is explicitly construed on the assumption that an unique end is the shared objective of each individual action, and then the psychological condition behind the shared action are studied starting from that assumption.

³⁰Searle has been charged of solipsism, for which his conception of intentionality and of collective intentionality couldn’t account for inter-subjective relations and hence also for joint activity in a proper sense (Zaibert, 2003; Meijers et al., 2003).

lens through which investigating the conditions of possibility and the limits of this capacity of auto-determination³¹. Moreover, it offers the possibility of rethinking both the problems of commitment and enforcement, by subverting the epistemological basis on which they have been theorized in terms of the backward orientation towards past, pre-existing, deontological structures of duties and rights, in the case of commitment, and towards preordained equilibria or recurrent states of the play (given “rationality” and common knowledge), in the case of self-enforcement *a la* Aoki. On the contrary, interactive intentionality is forward oriented and it can describe situations in which, without bounding agreements, agents commit themselves to (future) joint action, interactively planned.

This orientation toward the future is made available by the notion of *intentionality*. As we have seen, Searlian conception of collective intentionality clarifies in what sense a commitment toward a joint action depends on the sharing of a communal end, which does not require mutual knowledge concerning the other’s action. By providing the reference of the end, which is contained and shared in intentions-in-action, intentionality can depict what, in the present course of action, points at a future collective achievement, without needing any further rationality postulation, for example for inferring through backward induction mutual best responses. But, in Searle’s argument, this type of commitment toward the future presupposes a training in linguistic conventions (experience about past equilibria in the stable use of the public signs), and circularly, commitment towards them.

By making reference to *interaction* in the notion of interactive intentionality, I want to sustain that it is not necessary to assume neither the necessity of a foundation in past experience, nor the conventionality of linguistic devices and interpretations, for dealing with some interesting types of collective intentional commitments. Interaction, in interactive intentionality, involves beliefs, but it does not depend on common knowledge. The interactive procedures of posing a shared (future) purpose is a linguistic procedure which can make reference to past experience but it is engaged in another sort of forward-looking representation. The shared construction of the representation of the further end to be achieved works in shaping the mutual expectations about the others doing their part. In this way mutual self-enforcement is already included into the linguistic shared planning: the structure of the interaction to come is pre-construed through the performance of the linguistic practice itself³².

³¹The hypothesis of interactive intentionality could be developed by referring to the models of cognitive interaction and decision-making grounding the analysis of collective intentionality by Tomasello et al. (2005); Tomasello (2009, 2010).

³²The idea of interactive intentionality, conceptualized in this way, can be considered

The linguistic nature of this interactive procedure of representation is of a particular kind. It does not necessarily require explicit communication, since it can be based on implicit co-representations concerning the others participation in end's selection³³. But above all, its semantics depends only on the same interactive speech act of posing the end for collective achievement. Interpretation and meaning (the meaning which identifies the selected end) is intrinsic to the language practice through which it is construed. Of course, context dependency conditions of the relational sort envisaged in the previous section contribute in this process of creation of a new meaning. Purposiveness involved in interactive intentionality is not an arbitrary pick of an end, but it is decisively affected by the structure of relations between subjects and ends in which it takes place, and in this sense in past deontological and practical meanings. But by making the hypothesis of interactive intentionality what I want to stress is that a new meaning is created as intrinsic to the procedure of interactive representation through which the (new) end is selected as a possible alternative. In this way, interaction, unlike in Searlian collective intentionality, prevents the paradox of commitment because it avoids the necessity of a foundation on an underlying linguistic conventions or pre-existing structure of intentionality (or deontology), by construing interactively the linguistic tool and its intentional content, through which the shared agency is put in existence as a planned and self-enforcing course of communal action.

In this way, interactive intentionality provides the possibility of internalizing enforcement and of externalizing commitment. It is its inter-subjective procedure of selection of one end among alternatives at solving *ex ante* the problem of how mutual beliefs hold in the shared planned action. This is because the interactive procedure of selection itself internalizes the selected end as the object of individual intention (collectively pursue). But at the same time, it makes commitment an inter-subjective endeavor, dependent on conditions of equilibrium in the interactive creation of the shared end which enable deontological adhesion. Assuming alternative ends we obtain a more complete framework for dealing with normativity. The problems of commitment and enforcement are subsumed under the more general problem of purposiveness as the choice among alternative ends and dealt with from the perspective of relational methodology applied to the structure of subjects and ends.

My attempt to conceptualize interactive intentionality can be viewed as a

compatible with enactive conception of mind and with interactionism in social psychology (De Jaegher and Di Paolo, 2007; Thompson and Stapleton, 2009).

³³For the notions of “co-representation” and “cognition in the we-mode” cf. Gallotti and Frith (2013).

contribution to the theory of motivations, aimed at describing one particular type of cooperation neither exclusively based on intrinsic, a priori, commitment to one shared given end, as in the case of collective intentionality, nor consistent only in the a posteriori coordination in a behavioral equilibrium, on the ground of enforcement, determined by common knowledge and rationality, and by their implicit assumption of only one homogeneous type of purpose (maximization of utility, both in the self-regarding sense or in the sense of group's benefit via reciprocity). The structure of motivations behind this type of cooperation can be properly described, if we do not try to trace it back either to one given a priori internal type of subjective intrinsic motivations (commitment to right and duties, to fairness, to altruism, etc.) or to the external motivators realizing conditions of inter-subjective (self-)enforcement (systems of punishment, rewards, incentives, rules-as-cognitive-media, etc.). Actually, both intrinsic commitment and extrinsic enforcement needs in turn to be motivated with respect to purposes in order to be able to fulfill their function.

Some hypothetical situations can be imagined in order to try to describe the relational and purposive conditions presupposed by cooperation, when it is subordinated to the choices concerning an end among alternatives. The cases in which cooperative commitment and mutual enforcement requires a prior interactive specification of the purpose toward which the cooperative action is directed, reveal the logic of processes of generation of *novelty* and *change*. These processes are seen from the perspective of interactive intentionality as endogenous to the implicit or explicit (through communication) structure of relations through which individuals relates to each other in envisioning the possibility of a new (or changed) end for action. As remarked, these conditions of possibility are at the same time limiting the degree of autonomy and the freedom of variation of novelty and change. The actual practices of the selection of new ends are limited by the same linguistic procedure of end selection and by its concrete capacity of designating an end as an effective purpose for inter-subjective action.

This is, for example, the case of the collective endeavors toward scientific-technological innovations. A plurality of ends can motivate the joint action of research and development agents engaged in the realization of a new product (without considering the commercial ends ascribable to the firm itself). In this case the choice among the alternative ends (in many cases exclusive) is functional to the achievement of the collective product. The interaction for determining which end is to be pursued (or which combination of ends), at the same time satisfies both requirements of commitment and enforcement. In fact, if and only if the interactive determination of the specific purpose ends well, each individual is automatically committed to the shared task,

which can also be thought as collectively obtained as the equilibrium result of team reasoning (Sugden, 1993, 2000). The point I want to suggest is that (the search for) novelty imposes the prior selection of an end among alternatives, which in turn implies both the commitment to the end and the mutual beliefs concerning how to collectively get it.

An interactive intentional modality can also be enlightened in the specular circumstances of the annulment of an already available end, i.e. already embodied in an object, a person, a service, an activity, etc. as attributed of a deontological function or reflective some conventionalized practice. The annulment results in an individual and collective act of renouncing. It can be imposed by an external factor, such as, for example, a governmental decision to withdraw a determinate object or service from the market. Or it can be determined by the purposive choice made, for example, by consumers aimed at a public boycott. The choice (passive or active) of renouncing at the end or cluster of ends embodied in the given good or service (definable in terms of powers, functions and possibilities of action connected to that good) implies (implicitly or explicitly) the selection of an alternative, incomparable end, so that it can be seen as another case of interactive intentionality. Also in this case, negative (determined by actively or passively renouncing) procedure of end's selection among alternatives works as a factor which fosters both individual commitment and (the possibility of) the new equilibrium in conventions (in this case conventions concerning patterns of consumptions).

These modifications both in individuals' intentions and in collective behaviors can be described as conditioned by a reorientation of the whole structure of purposes linking subjects and ends in the community (of consumers, in the example). The apparently subjective individual's choice of deselecting the power relations enabled by that given good or service is already (implicitly or explicitly, in the case of public campaign) an interactive procedure of choice, in so far as, it necessarily involves a public recognition of the meaning of that act of renouncing and hence a commitment. This interactive modality represents at the same time the condition on which the possibility of an actual shift in the convention can be pursued and can work as a collective self-enforced re-orientation toward a new alternative end. Actually, it is acted as a practice which receives its meaning from the same fact of being conceptualized (individually and collectively) as a (possible) shift in the conventions concerning the uses of the substitute item or service, or in general, the priority of an alternative end over the abandoned one.

The analysis of conditions of purposiveness can also be applied to interpret the intentional structure of motivation-crowding, through a rethinking of intrinsic motivations and extrinsic motivators (incentives or disincentives,

punishment and rewards)³⁴. The interactions between the two motivational factors, can be seen not as the comparison and possibly the conflict between an internal (intrinsic) commitment (taken as given and hence unexplained in terms of the intentions composing the motivation) and an external (extrinsic) causal factor capable of affecting positively or negatively that commitment relation (and respondent only to external and independent conditions of equilibrium, which regulates the type, the practical meaning, and the value, represented by the incentives). It is rather worth interpreting them as cases of relations and conflicts between alternative ends, because this approach could provide a wide spectrum of applicability and of specificities by revealing the context-dependency conditions under which the intentional content (status functions) associated to different kinds of incentives (possibly not only respondent to the instrumental purpose typical of the material incentive) affects, positively or negatively, one of the multiple possible types of individual or collective purposive agency.

From the perspective of conditions of purposiveness, both motivations and motivators are put on the same ontological level, and motivation crowding is seen as a phenomenon endogenous to the network of purposive relations internal to individuals and communities. A mutual effect and reciprocal influence between commitments and enforcements can be hypothesized on this ground. On one hand, the possible effect of the motivators depends on the pre-existent deontological functions and purposes already collectively attached to it, and determining individual commitments in that institutional context. For example, it can be thought that material incentives can crowd out intrinsic pro-social motivations only in contexts in which the attributions of a specific value (function, purpose) to material incentives is already determined and shared within the population to which the policy is addressed. In this sense, extrinsic factors are also intrinsic, because, while corresponding to determinate collective attributions of purposes to material objects (such as money) as institutionalized objects (together with the purposive practices involving them), they are also the objects of individual intrinsic motivations and commitments.

On the other hand, we can also think at situations in which intrinsic deontological commitments, in so far as they can affect actual inter-subjective practices, crowds out material incentives, subverting the given equilibrium of the system of external references and of incentives. In other words, intrinsic motivations can be viewed as dynamically connected, via interactive intentionality, to others' intentionality and purposes, such that subjective choices of ends can exert a feedback influence on the external, collective environments

³⁴Bowles (2008); Bowles and Polania-Reyes (2012); Frey (1997); Frey and Jegen (2001).

in which ends are shared and put in practices as extrinsic material references and uses. This is the case of one modality of collective action which will be analyzed in the next section as a crucial case of negative rule-following. It will provide arguments for sustaining the idea that, viewing motivation-crowding from the perspective of the context-dependencies of purposiveness fosters the study of conditions of *crowding-in* based on interactive intentionality. The interactive selection of an end among alternatives can represent one modality of activation of individual commitment and inter-subjective enforcement in the formation and maintenance of stable practices, which does not require neither the pre-adhesion to given deontologies, nor the stimulus of external material payoffs. From this perspective, the possibility of novelty and change in given normative structure is *per se* the possibility of re-generation of the motivational basis of institutional life.

7 How we don't follow rules

All the cases just discussed – cooperation toward innovation, the decision and action of renouncing to some object or service, motivation crowding – can be referred to for rethinking *collective action* in terms of interactive intentionality³⁵. In particular, interactive intentionality can represent the purposive, relational conditions of individual commitment towards rule-change and its collective, practical co-implications in the change of institutions. In this way, we enter into the merit of the crucial issue for this paper: how don't we follow rules? What I call negative rule-following represents a crucial case of interactive intentionality, which is worth studying from the perspective of the structure of the purposive choice entailed by it.

In doing this, my aim is not to identify rational/motivational causes or evolutionary dynamics accounting for collective action. It is not to provide an explanation in term of *what* commits or alternatively *what* enforces agents to deviate from a given rule, as if these commitment or enforcement factors were separated (acting as the explanans) from the action which performs them. My strategy is rather descriptive, and it focuses on purposiveness in order to show as, in the case of interactive intentional collective action, both commitment and enforcement depends on the modality of interactive intentionality,

³⁵The problem of collective action is that “unlike idiosyncratic play, participation in collective action is not only intentional (rather than accidental) but also conditional on one's beliefs about the likelihood and consequence of a substantial number of [agents] changing behaviors”; moreover, it has to be taken into account of “incentives for each to free ride when others act in pursuit of commonly shared objectives.”(Bowles, 2004, pp. 427 ff.)

in such a way that neither the problem of interpretation, nor the paradox of commitment reappear. The first thing to be noticed, for the purpose of description, is that there are several ways in which (or several motivations for which) agents do not follow rules. These depends on circumstances (context-dependency), which can be schematically gathered in four types. I derive these four types from the examples of interactive intentionality above:

1. New rule for the constitution/regulation of a new object or service (innovation)
2. New rule for compensating the fact that an object or service has been (or has to be) removed (renouncing)
3. Change of convention through deontological commitment (*resemantization*)
4. Change of deontology through uses (*practice*)

The discussion of these four cases entails the clarification of intentional modalities of rule-change with reference to *novelty* (circumstances 1 and 2) and to *conflict* (circumstances 3 and 4). These two aspects provide insights for complementing the incompleteness of deontology, by rethinking the problem of the origin as the problem of *normativity of novelty*, and the problem of the relation between deontology and practice by explicitly dramatizing the *conflict between regulative and constitutive*. Starting with the first two cases, in which sense does the necessity of a new rule (both in the positive case of innovation and in the negative case of renouncing) represent an occurrence of negative rule-following? In both cases agents do not follow rules simply because there is no rule to be followed. What are hence the consequences of this negative rule-following on the deontological structure of normativity and practices in which the object or service is introduced or removed?

In the first case, the introduction of a new object or service within the realm of institutional life (for example, because of technological innovation) depends, logically and ontologically, on the purposive conditions of its creation which set up its own normative status both in terms of regulative and constitutive rules. As we have noticed above, if it's true that no rule has been followed for the creation of the new object (otherwise it wouldn't be considered as a new object), its realization depends on the interactive choice of a purpose which entails both commitment and enforcement as the condition of its production. In this way, the creation of a new object intrinsically carries its own normativity, which start interacting with the whole structure of deontological obligations and purposive practices from the moment of its

production, ending up modifying and being modified by it, through mutual adjustments.

This simplified reconstruction of the genesis of the normative status of a genuine new object via technological innovation, has the only purpose of signaling that there are cases, not considered in Searle's analysis, in which institutions and status functions, depends on cooperation. Cooperation can represent the locus in which, through the denial of given procedural and constitutive rules (or at least, in their absence), new institutional purposiveness (statuses and functions) are materially put in existence, and provided to inter-subjective practices. Considering this circumstance, interactive intentionality can complement social ontology with a description of the conditions in which a new function and deontology is realized through cooperative planning and production, before its institutional and normative status can be established through collective recognition in constitutive rules and equilibria in practices.

In the second case, negative rule-following is the consequence of the choice (private or centralized) of renouncing (actively or passively) to the use (for example, consumption) of the object or service. This counts as the negation of a rule in so far as it (possibly) causes the removal of the object or service from the institutional realm and hence makes the constitutive and regulative rules defining the normative status of that object or service redundant. This loss in normativity is determined or compensated through the introduction of a new rule (or new rules). When the object or service is removed via governmental action this is made through the introduction of rules which terminate the existence of the object by forbidding its use or denying its status function (while reorganizing the whole systems of deontologies and practices connected to the lacking of that object or practices). When the object or service decays because of the collective action of agents (consumers) renouncing to its use, this is done on the basis of the interactive intentional pursue of an alternative end and it is intrinsic to that agency the (at least implicit) affirmation of the opportunity of a new rule, a new set of duties, right and legitimate practices, consistent with the negation of the use of that object or service (affirmation which in turn, potentially, could affect the whole system of deontologies and practices).

This last case introduces us to the second aspect of negative rule-following, for which it entails instances of conflict between a) systems of commitments, as expressed by constitutive rules which hold as collective recognized duties and rights, and b) systems of enforcements (and related incentives), as expressed by regulative rules which hold in practice. In particular, the case of the repugnance and avoidance to use a certain object or service with the (implicit or explicit) fostering of collective action, can be read also as a specific

case of conflict between commitment to a certain deontology, in contrast with the subsistence of a given convention in society (and the relative structure of enforcements in regulative rules and incentives), concerning, for example, patterns, of consumptions³⁶. In general, negative rule-following, in the circumstances of conflict between deontology and conventions, re-proposes the logic of motivation-crowding (rethought in terms of purposiveness), which makes available an analysis of conflicts, not only in the specific case of the motivations towards the suppression of the institutional object or service and of its embodied deontological and practical functions (point 2), but more in general, of all the possible occasions of contrast between mutually exclusive attributions of instrumental function (pragmatic conditions of use) or deontological status to a given (existent) institutionalized object or activity.

Conflicts of this sort manifest themselves in the two cases at point 3 and 4 of the above list, which are maintained distinct for the only purpose of analyzing separately the cases in which a given convention in practice is contrasted by the assumption of an alternative commitment; and the cases in which a constitutive attribution of a deontological status to an object is contrasted by an alternative practice concerning that institutional object or function. Apparently, this separation distinguishes cases of collective action based on intrinsic *subjective* deontological commitment (subjective and intrinsic, in the sense of independent on the consequences of the actions on others and in terms of further material benefits), and cases of collective action based on the performance of an alternative *inter-subjective* practice (the actual shift in the convention due to coordination in non-best responses). But the epistemic conditions sustaining the negative-rule-following action in 3 and 4 actually co-implicate each other and their separation must be considered fictional³⁷.

As suggested in the previous discussion, I maintain that systems of regulative rules (conventions) and systems of constitutive rules (deontologies) always coexist within an interconnected network. Being interested in epistemic conditions of rule change, conflict is actually interesting because it reveals the *unitary* form of interactive intentionality as inclusive of the nec-

³⁶Or vice versa, the subsistence of collective pattern of consumptions connected to the conditions of enforcements exerted within equilibria in the market systems can be considered as affecting commitments, for example, crowding out the motivations towards a certain type of commitment-based (for example, ethical) consumption.

³⁷Moreover, the logic of possible conflicts among individual-and-collective adhesion to alternative rules is not completed by the consideration of the cases of the opposition of constitutive to regulative, and vice versa. The conflicts between regulative rules and regulative rules end the conflicts between constitutive rules and constitutive rules are also significant from the point of view of institutional change. However, I signal only the two cases of crossed conflict, because they are revealing of the two aspects of resemantization and change in practice essential to rule-following in its negative occurrence.

essary conditions of collective action both in terms of subjective commitment to deontology and of inter-subjective enforcement in practice. The two types of conflicts identified are actually two components (or two aspects) of negative rule-following, in so far it always involves at the same time a (potential) change in the convention and a (potential) change in the deontology. From this perspective, negative rule-following entails both an attribution of a new purpose/function on object previously identified in practice (in this sense, a *resemantization*) and, on the other hand, the capability of a new mode of interaction for realizing in practice that new purpose. In this sense, negative rule-following implies both the imposition of constitutive rules over regulative rules and the expression of the new deontological function as a new equilibrium in regulative rules.

Through this double move the deviation from a given normative asset produces change. Interactive intentionality is the linguistic/epistemic modality with which new normativity, both in the constitutive and the regulative sense, springs from conflicts which can be interpreted both as internal to individual's structure of motivations and involved in interactions between individual choices and purposes. Viewed from the perspective of my relational and descriptive methodology, the (potential) conflict between convention and deontology, regulative and constitutive, is only a way of providing a representation of the context dependency conditions of institutional agency. The possibilities of collective actions toward rule-change depends on conditions of purposiveness, which have to be specified both in terms of practices (uses) and commitments (meanings) because they have to be able to represent the modality of choice of an end, alternative to that stabilized in previously given structure of conventions and deontology. Interactive intentionality can represent modalities of negative rule-following exactly because it entails the idea that when we do not follow a rule we do it at the same time committing ourselves to an alternative end, which only can be identified through the interaction at the level actual uses which realize in practice the linguistic procedure of selection and planning towards the pursuing of the alternative end.

Negative rule-following' agency can be described as a property of the network of relations which connects subjects and ends within institutions³⁸. It is motivated by intrinsic commitment, but it requires interactions based on mutual expectations which prefigure, through language and imagination, toward what alternative purpose commitment can be addressed. It is acted as

³⁸This is equivalent to say that agency toward rule change is not on an different ontological level with respect to the (normative) structure of institutions, because if we look at this structure as a network of purposive relations (both individual and collective) agency is only the dynamic representation of those relations.

a collective action, but it presupposes the internalization of the attribution of a new value/function to the external (material) references, which reinterpret (re-semanticize) the given system of incentives making the action itself self-enforcing (according to the new meaning). Negative rule-following is at the same time intrinsically and extrinsically motivated, in so far as its epistemic basis and the modality through which it is performed individually-and-collectively reflect both the deontological motives for commitment and the common re-attribution of a new value (in terms of functions) to the external goal to be put in practice. Not to follow a given rule for the purpose of changing it involves individuals, both in terms of the personal connection to one's own end and in the inter-subjective mutual expectations about the planned collective end to be achieved. Both the types of relation are determined at once by the very same linguistic interactive procedure through which the alternative end is selected and envisaged in the sharing of intentionalities and values.

In this sense, rule can be viewed as the way in which motivations crowd-in themselves through the interactive procedures of selection of an end among alternatives. Through the double interconnected process of internalization (creation of habits as intrinsic subjective commitments) and externalization (inter-subjective enforcements in re-semanticized practices), interactive intentionality specifies the conditions of rule-formation (and transmission) and negative rule-following represents the constitutive modality of rule variation. Conflicts are only the internal engine of the motivational ongoing process internal to institutions and affecting at the same time individuals-and-collectivities. In many occasions, the creation of institutions and their vital change is not the result of pre-ordained agreement or pre-constituted recognitions, but it is the (often conflicting) discursive procedure of comparisons and selection between alternative ends, according to the logic of negation. Relations, on the ground of this irreducible multiplicity of mutually excluding ends, are the modes of motivations, and if we want to catch the dynamics involving them we need to depict the networks of purposes through which they are cognized and practiced. Only in this way rules can be seen, not as a given fact – neither as regularities in behaviors self-enforced by given and already internalized rationality rules (reflected also in the systems of interpreted and known payoffs and incentives), nor as auto-consistent deontological command, based on the abstract logic of given intrinsic and aprioristic motives and meanings – but as endogenous processes which have in the internal modality of their change the reasons of their subsistence.

8 Conclusions: on rule-formation and change

Negative rule-following depends on the availability of alternative ends and on the possibility for agents to choose them. This circumstance accounts for normativity in general, in so far as rules are not necessary, if there is not the possibility of deviating from them. The paper is engaged in explaining this possibility. If we want to comprehend theoretically it, we have to understand the conditions, both ontological and epistemological, through which alternative ends are connected in the (dynamic) network of individual-and-collective purposive actions. Speaking about alternative ends is speaking about purposive actions and this consideration moved the focus of my analysis on conditions of purposiveness. These conditions can be schematically summarized in two interwoven context-dependency principles. The agency which connects a subject and an end depends logically and ontologically on a) cognitive relations among ends and b) on pragmatic relations among subjects. Hence, purposiveness has to be treated as a property of this network, and normativity, which depends on the possibility of identifying purposive actions, is at the same time the result and the condition of these interconnected relations.

Negative rule-following is revealing of this relational structure of normativity. It in fact cannot be interpreted if the two conditions are not taken into account at once by the analysis. This is the reason why the purely conventional approach and the purely deontological approach to rules cannot interpret alone negative rule-following, and stick to the assumption of one given end. In dealing with only mutual beliefs (and common knowledge) as the epistemic condition of the emergence of regularities in behaviors, the conventional approach cannot deal with the intentional dimension of deontological commitment and cannot solve what has been identified as the problem of interpretation (on which ground it can be established that others will apply the rule as I do?). On the other hand, an approach exclusively based on deontological intentionality meets with the paradox of commitment because within its framework it is not possible to understand how the ends, as the condition of possibility of commitment, are posed as the result of interactive practices.

Interactive intentionality solves at once both the problem of interpretation and the paradox of commitment. The linguistic procedure of selection of a purpose among alternatives commits individuals toward a shared end (as in collective intentionality), and it does not require a further mutual enforcement structure. It is based on co-representation (co-imagination and planning), which derives its capacity to designate its objects univocally from the very same fact of being construed through an interaction capable of pro-

viding with a new semantics the system of material incentives connected to the possible payoff/result of the new practice. The problem of interpretation is solved by the fact that the obtainment of the new possible equilibrium in behaviors does not necessarily depend on pre-existing behavioral equilibria, but it plays the role of a precondition, since it is prefigured *ex-ante*, with respect to cooperation in practice, through the interactive designation of the purpose and the intentional, shared attribution of a value/meaning to it.

On the other hand, the notion of interactive intentionality binds deontological designation of functions and powers to practices, as the realm in which purposive actions are constantly rearranged and realized. This avoids the paradox of commitment, since practices are (at least in part) independent on pre-established linguistic conventions and can produce change and novelty in institutional deontology, while elaborating inter-subjectively the same linguistic resources through which it points at alternative ends. Practice is the (pragmatic) space in which the network of alternative ends are created, maintained and re-created through inter-subjective actions. In this way, deontological commitment does not require anymore a pre-recognition of language and of already established institutional powers/functions, but it has its logical-ontological basis in interactive purposive choice, and hence in the very same possibility to deviate from commitment. Normativity is not only based on the commitment toward a given end, but it must admit the possibility of several alternative ends, and hence it is intrinsically constituted by negative rule-following. This complements the intrinsic incompleteness of deontology, by clarifying the origin of deontological normativity and its dependence on change in practices.

The main theoretical advantage of admitting the possibility of negative rule-following in the sense of the interactive, linguistic procedure of construction of (and commitment toward) an end among alternatives is that, in this way, it becomes possible to deal with genuine cases of rule-change. I referred to this possibility by speaking in the introduction of the logical and ontological asymmetry of rule-following. Merely conventional and merely deontological conceptions of rules are not able to deal with negative rule-following exactly because they assume – either in the conception of rationality and common knowledge, or as the object of commitment – one only homogeneous type of end. In those conceptions, the occurrence of deviation (individual or collective) cannot but reconfirm the logic of that structure of enforcement or of commitment and does not realize change in any case. Novelty and conflicts, as essential factors of rule-change, can be conceptualized only by considering the incommensurability and incomparability of a plurality of possible purposes, both among individual intrinsic motivations and within the inter-subjective structures of external incentives and enforcements. If we

want to study rule-change, we need a notion of interactive intentionality capable of representing the interactive procedure of purposive choice, through the network of relations between ends and subjects, and both in deontological constitutive definitions and in regulative practices of regularities.

The purposiveness of negative rule-following and the descriptive, relational methodology implied by interactive intentionality fosters a change in perspective as regards rule-following in general. If normativity essentially depends on the availability of alternative multiple ends and on the possibility of choosing them in individual-and-collective practices and deontologies, we have to consider that the same epistemic conditions of negative rule-following are necessary conditions also for positive case of rule-following. When we follow a rule, either regulative or constitutive, we do it as a purposive choice which is based on the logical-ontological possibility of not following it. This means that when we comply to a given normativity we are (implicitly or explicitly) adopting in the interactive intentional way an end in contrast to alternative possible others. This also means that the same context-dependency conditions of negative rule-following jointly activate in making rule-following both an inter-subjective procedure (with its self-enforcement) and a subjective intention (with its intrinsic commitment). We follow rules as always engaged in a collective action, which connects our personal purpose (to comply with the end dictated by the rule), with all the structure of ends embraced by other subjects in interactions. Rule-following always intrinsically involves (the possibility of) novelty and conflicts. Change is endogenous to rule's regularity and permanence.

From this perspective, normativity appears as essentially dynamic. By asking the question about the origin of deontology and its relation with practice, I didn't want to provide an explanation of the emergence of regularities in behaviors or of any given normative deontology. My aim has been the description of the linguistic-cognitive modalities behind the logical-ontological possibility of (negative) rule-following, and this purpose led me to highlight the structure of purposive relations on which normativity depends, and, in particular, changes. But, if normativity essentially depends on the same ontological and epistemic conditions which determine the possibility of its negation through the interactively intentional modalities of deviation from rule, then we could start thinking at rule-formation as intrinsically dependent on rule-change. And this can in turn provide some suggestion for further research, oriented by the consideration that normativity persists and evolves, through a type of transmission which is not only replication, conformism, or imitation, but depends (also) on the same interactive intentional modality which determines its variation in time and circumstances.

References

- Masahiko Aoki. *Toward a comparative institutional analysis*. MIT press, 2001.
- Masahiko Aoki. Endogenizing institutions and institutional changes. *Journal of Institutional Economics*, 3(01):1–31, 2007.
- Masahiko Aoki. Between game theory and institutional studies: The dual-dualities of the institutional process. *Available at SSRN 1624003*, 2010.
- Masahiko Aoki. Institutions as cognitive media between strategic interactions and individual beliefs. *Journal of Economic Behavior & Organization*, 79(1):20–34, 2011.
- Robert J Aumann. Subjectivity and correlation in randomized strategies. *Journal of mathematical Economics*, 1(1):67–96, 1974.
- Robert J Aumann. Agreeing to disagree. *The annals of statistics*, pages 1236–1239, 1976.
- Robert J Aumann. Correlated equilibrium as an expression of bayesian rationality. *Econometrica: Journal of the Econometric Society*, pages 1–18, 1987.
- Samuel Bowles. *Microeconomics: behavior, institutions, and evolution*. Princeton: Princeton University Press, 2004.
- Samuel Bowles. Policies designed for self-interested citizens may undermine” the moral sentiments”: Evidence from economic experiments. *science*, 320(5883):1605–1609, 2008.
- Samuel Bowles and Sandra Polania-Reyes. Economic incentives and social preferences: substitutes or complements? *Journal of Economic Literature*, pages 368–425, 2012.
- Robin P Cubitt and Robert Sugden. Common knowledge, salience and convention: a reconstruction of david lewis’ game theory. *Economics and Philosophy*, 19(02):175–210, 2003.
- John B Davis. *The theory of the individual in economics: Identity and value*. Routledge, 2003.
- Hanne De Jaegher and Ezequiel Di Paolo. Participatory sense-making. *Phenomenology and the cognitive sciences*, 6(4):485–507, 2007.

- Bruno S Frey. *Not just for the money: An economic theory of personal motivation*. Edward Elgar Pub., 1997.
- Bruno S Frey and Reto Jegen. Motivation crowding theory. *Journal of economic surveys*, 15(5):589–611, 2001.
- Mattia Gallotti and Chris D Frith. Social cognition in the we-mode. *Trends in cognitive sciences*, 17(4):160–165, 2013.
- Margaret Gilbert. *On social facts*. Princeton University Press, 1989.
- Margaret Gilbert. *A theory of political obligation*, volume 189. Oxford University Press Oxford, 2006.
- Margaret Gilbert. *Joint commitment: How we make the social world*. Oxford University Press, 2013.
- Herbert Gintis. The evolution of private property. *Journal of Economic Behavior & Organization*, 64(1):1–16, 2007.
- Herbert Gintis. *The bounds of reason: game theory and the unification of the behavioral sciences*. Princeton University Press, 2009.
- Avner Greif. *Institutions and the path to the modern economy: Lessons from medieval trade*. Cambridge University Press, 2006.
- Avner Greif and Christopher Kingston. Institutions: Rules or equilibria? In *Political economy of institutions, democracy and voting*, pages 13–43. Springer, 2011.
- Frank Hindriks. Constitutive rules, language, and ontology. *Erkenntnis*, 71(2):253–275, 2009.
- Frank Hindriks and Francesco Guala. Institutions, rules, and equilibria: a unified theory. *Journal of Institutional Economics*, pages 1–22, 2014.
- Geoffrey M Hodgson. What are institutions? *Journal of Economic Issues*, 40(1):1–23, 2006.
- Geoffrey M Hodgson. Institutions and individuals: interaction and evolution. *Organization studies*, 28(1):95–116, 2007.
- Pierre Jacob. Intentionality. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter 2014 edition, 2014.
- David Lewis. *Conventions*. Harvard University Press, 1969.

- Anthoñew Meijers et al. Can collective intentionality be individualized? *American Journal of Economics and Sociology*, 62(1):167–183, 2003.
- Seumas Miller. Joint action: the individual strikes back. In Savas Tsohatzidis, editor, *Intentional Acts and Institutional Facts*, pages 73–92. Springer, 2007.
- Douglass C North. *Institutions, institutional change and economic performance*. Cambridge university press, 1990.
- Douglass C North. *Understanding the process of economic change*. Princeton University Press, 2005.
- Elinor Ostrom. *Understanding institutional diversity*. Princeton university press, 2005.
- Andrew Schotter. *The economic theory of social institutions*. Cambridge University Press, 1981.
- John R Searle. *Speech acts: An essay in the philosophy of language*. Cambridge university press, 1969.
- John R Searle. Collective intentions and actions. In *Consciousness and language*, pages 90–105. Cambridge University Press, 1990.
- John R Searle. *The construction of social reality*. Simon and Schuster, 1995.
- John R Searle. *Consciousness and language*. Cambridge University Press, 2002.
- John R Searle. What is an institution. *Journal of institutional economics*, 1(1):1–22, 2005.
- John R Searle. *Making the social world: The structure of human civilization*. Oxford University Press, 2010.
- John R Searle. Status functions and institutional facts: reply to hindriks and guala. *Journal of Institutional Economics*, pages 1–8, 2015.
- Robert Sugden. *The economics of rights, co-operation and welfare*. 2004 2nd edition, 1986.
- Robert Sugden. Thinking as a team: Towards an explanation of nonselfish behavior. *Social philosophy and policy*, 10(01):69–89, 1993.

- Robert Sugden. Team preferences. *Economics and Philosophy*, 16(02):175–204, 2000.
- Robert Sugden. Saliency, inductive reasoning and the emergence of conventions. *Journal of Economic Behavior & Organization*, 79(1):35–47, 2011.
- Evan Thompson and Mog Stapleton. Making sense of sense-making: Reflections on enactive and extended mind theories. *Topoi*, 28(1):23–30, 2009.
- Michael Tomasello. *Why we cooperate*. MIT press, 2009.
- Michael Tomasello. *Origins of human communication*. MIT press, 2010.
- Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, 28(05):675–691, 2005.
- Raimo Tuomela and Kaarlo Miller. We-intentions. *Philosophical Studies*, 53(3):367–389, 1988.
- H Peyton Young. *Individual strategy and social structure: An evolutionary theory of institutions*. Princeton University Press, 1998.
- Leo A Zaibert. Collective intentions and collective intentionality. *American Journal of Economics and Sociology*, 62(1):209–232, 2003.