Ratio Juris. Vol. 19 No. 4 December 2006 (501–17)

The Mental Path of Norms

ROSARIA CONTE AND CRISTIANO CASTELFRANCHI

Abstract. In this paper, we intend to re-examine our view of norms as two-fold objects, with a mental and social side, and a fundamental mechanism of the micro-macro link. In the light of what has been said in Volume 1 of *A Treatise of Legal Philosophy and General Jurisprudence* (Pattaro 2005), some aspects of a theory developed elsewhere by the authors will be reconsidered. With regard to the mental side of norms, the main properties of normative beliefs will be reconsidered and some issues emerging from a reading of Pattaro will be addressed, in particular, (i) normative autonomy, (ii) norm acceptance, and (iii) the interplay between norms and goals. With regard to the external side, a unitary view, bridging the gap between conventions and norms, will be put forward, and some steps will be taken towards defining a notion of external (i.e., impersonal) commands, conceived of as a dimension in which both norms and conventions are situated. Finally the applicability of this notion to the study of norm innovation will be discussed.

1. Introduction

In this paper, we intend to re-examine our view of norms as two-sided, external (social) and internal (mental), objects, representing a fundamental mechanism of the micro-macro link. We re-examine some aspects of a theory developed over the past 10 years or so (Conte and Castelfranchi 1995; Conte 1998; Conte and Castelfranchi 1999), in the light of what is said in Pattaro (2005).

In the first part of the paper, we reconsider the components of the mental processing of norms, especially addressing the main properties of normative beliefs. A number of more specific issues emerging from a reading of the *Treatise* will be addressed:

- Normative autonomy. What is needed for a norm to be internalised? Special attention will be paid to the question of acceptance. Autonomous normative actions are based on agents' decisions about norms. However, this by no means implies that agents accept and share the norms to which they are subject.

© 2006 The Authors. Journal compilation © 2006 Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden 02148, USA.

- Norm acceptance. With autonomous agents, goals are internal reasons for action; to say that norms provide reasons for behaviour does not answer the question as to how these are cognitively processed—read, represented, and reasoned upon—by agents. What does it mean to accept a norm?
- The interplay between goals and norms is essential for linking the micro to the macro. Pattaro quite rightly objects to the social contract philosophers' oversimplifying view of the origin of normativeness. There is no clear continuity between a promise and the issuing of a new norm. How can we maintain that norms are brought into existence by a mere act of volition? It will be argued that the mechanism of goal generation can shed light on this issue.

With regard to the social side of norms, we argue for a unitary view bridging the gap between conventions and norms. This means going back to the nature of obligations, which are not necessarily issued by a sovereign body, but are also emerging prescriptions. The question is what different social prescriptions have in common, and in what way they differ. We propose to find such a common property in the notion of *imperium* as an impersonal, non-subjective and therefore disinterested, easily accessible, hardly corruptible command. We propose to call this type of command *external*. An external command is one that intends to be accepted *per se*, and which is recognisable because of its claim to be accepted *per se*. No other type of command, e.g., coercion, is external in the sense defined here. External commands are transmitted as social rules, which, in addition to norms, include conventions as well as customs and rituals.

As we briefly illustrate in the second part of this paper, the view of norms based on external commands has several advantages, as it can:

- bridge the gap between conventions and norms, and more generally pave the way for a "unitary vision" of the existing plethora of social rules;
- dispense with a view of conventions based merely on conformity; this view is based upon an unwarranted incursion into the motives for acceptance (see Hart's "Postscript" to his 1961 work). The difference between conventions and norms does not seem to reside in the reasons why people accept them *de facto*, but in the reasons why they ideally ought to do so. Although norms are intended to be accepted *per se*, they are frequently complied with in order to escape sanctions, without ceasing to be and act as norms. The difference must lie in something else, possibly in the source of the obligation. Norms are issued by authorities, whether they are exercised by persons playing an institutional role or by the moral law itself. On the other hand,

© 2006 The Authors. Journal compilation © 2006 Blackwell Publishing Ltd.

conventions emerge spontaneously and gradually. Nevertheless, they might spread within a given society for the same reasons at the individual level;

- shed light on the question of norm-innovation. The conventionalist view can easily account for the proliferation of social rules, but it finds it harder to explain how they change (other than by mutation). We will endeavour to show how a view of social rules as external commands might help understand how and why innovation of conventions follows changes in social values. Indeed, the social dynamics of norms needs to be characterised as a spiral, from matters of fact to the minds of agents and then back to objective matters of fact in a recursive manner. If the progress of behaviour *b* in society *s* is accompanied and facilitated by the widespread belief that members of *s* are expected to display *b*, what spreads is not only *b* but also an external command, in the sense previously defined. We are then facing something new, which can be described neither as an observable matter of fact nor as a mental construct. Norms are neither *objektive Sollen* nor *motives for behaviour*, but a coupling of the two.

2. At the Heart of Volume 1 of the *Treatise*

It is fundamental to recognise that a norm cannot really operate as a norm and be effective without being held, if only transitorily, in the minds of the agents, by this means regulating their behaviour, and an integral part of a theory of norms should be based on an analysis of this cognitive path. This point in Pattaro's work is of great value. In particular, such a theory cannot ignore how a norm is perceived and represented in the minds of the agents, and how this internal representation affects their behaviour, usually consisting of deliberate, reason-based, and intentional actions.

Even more than this, at least in our view, a norm is aimed at prescribing specific mental attitudes, and even specific reasons for accepting it. However, we should bear in mind that compliance with an external request or command (a special case of the more general cognitive process that we have called *adoption*, i.e., taking somebody's goal as one's own), need not be based on a sharing of the reasons for such a request. Agents may accept norms without sharing their aims in whole or in part.

3. Normative Autonomy

In this section, we shall endeavour to clarify certain components of the mental processing of norms. In particular, we shall insist on those aspects that are mainly responsible for reconciling the autonomy of subjects, on the one hand, and their liability to normative influence, on the other. It is essential to render these properties compatible when seeking to explain the micro-macro link.

3.1. Normative Beliefs

In this perspective, the first fundamental point is precisely the one on which Pattaro bases rather strong and explicit claims. First of all, a norm becomes a belief in the mind of a subject, the belief that a given type of behaviour, in a particular context, for a given set of agents, is forbidden, obligatory, permitted, and so on. More precisely, the belief should be that "there is a norm prohibiting, prescribing, permitting." Indeed, norms are issued with a view to becoming such beliefs. In other words, norms must be acknowledged as such in order to work properly; this is their function.

However, what does it mean to be acknowledged as a norm? This is not a trivial question. In order to realise the existence of a norm and its influence on a particular subject, the belief that something is prescribed or forbidden is not sufficient. A norm is something more than mere volition, or the order of a private agent—say a parent or a perhaps and assailant (with their power to influence behaviour)—that obliges us to do or not to do something. Binding force is insufficient, since it also characterises non-normative commands. The point is the source of the binding force of norms, and the status of the giver of commands. For a prescription to be perceived as a norm, we need to ask who or what the source should be, and how it should be characterised.

Furthermore, the acknowledgement of the source is still insufficient. Not all the imperatives of a sovereign are norms: Some might be private interests or requests. To act as a norm, an imperative must be issued by a proper (legitimate) source in a proper (legitimate) manner. When the source is a lawgiver, such a body should play its institutional (lawgiving) role publicly and patently for the purposes and functions of that role (that are supposed to be different from private or personal interests and desires). Clearly, individual motives might be hidden behind an institutional body, according to what Pattaro calls the "hypocritical" behaviour of institutions. However, the norm will still be acknowledged as such, providing it is believed to be issued by the proper source in a proper manner. For a norm to be acknowledged, it is necessary for subjects to believe that such conditions hold.

However, the effective assumption of such beliefs is by no means necessary for an issued norm to be in force: *Ignorantia non excusat* (see Pattaro 2005, 216). Providing the sovereign's will is clear, unambiguous, public—in a word, positive—then the norm is in force, and uninformed subjects will find no excuse for failure to comply with it. This is why in any normative system, symbols and rituals are so important. It is fundamental to *signify* what should be done by whom, and to express the command in an impersonal manner. Indeed, a norm is the expression of an impersonal and either collective or super-personal will. In either case, it is an abstraction. How this fundamental feature is mentally represented is not yet fully clear, and we make no claim to fill the gap here.

To believe that a norm exists and concerns us requires at least a second set of beliefs: the beliefs of *concern*. The norm says what ought to be done by whom: (i) the obligation/permission/prohibition and (ii) the set of agents on which the imperative impinges. For example, if I am subject to a given norm (for example, the requirement to be a member of a professional body), and the norm has to take effect, I must recognise this. The prescription relates to a set or class of agents (say "Italian psychotherapists"), and since I am a member of this class, the norm applies to me. The belief that one is concerned by a norm is quite important; in fact, it is frequently a matter of dispute and legal controversy, due to subjects claiming not to be concerned by a norm, or arguing that it does not apply to them, which is the same thing.

3.2. Goals and Reasons for Acceptance

Throughout Pattaro, the role of normative beliefs is emphasised. However, a believer is not yet a decider: Beliefs are necessary but insufficient conditions for norms to be complied with. What is it that leads agents to accept a norm, which by definition prescribes behaviour that has a cost?

Pattaro answers this question in a rather elegant way by stating that norms are both beliefs and reasons for acceptance. However, in our view this answer, to which we turn later on in the paper at some length, is incomplete: It does not tell us what happens in the mind of a subject once she knows that a norm concerning her exists. It says nothing about her goals.

From a cognitive point of view, goals are subset of the motives or reasons for action. While motives include a large variety of (i) beliefs (for example, plans or other pragmatic beliefs) and (ii) (believed) external factors that objectively cause action (for example, the injection of a drug), goals are internal representations triggering and guiding action at the same time: They represent the state of the world that agents want to reach by means of action, and that they monitor while executing the action. Notably, unless we consider norm acceptance as an automatic reactive process, a normative belief must give rise to a normative goal for the subject to act in accordance with the norm itself. This process is more or less elaborated, and governed by different mechanisms. From the least sophisticated, where norm obedience becomes automatic, leaving little room for autonomy, to the most complex, such as instrumental adoption, i.e., the calculation of advantages and disadvantages of norm compliance, a mechanism enabling an external command to become a goal is needed.

In such a context, special attention should be paid to the non-instrumental acceptance of a norm. This is different from automation: Far

© 2006 The Authors. Journal compilation © 2006 Blackwell Publishing Ltd.

from deriving from custom or conditioned learning, this phenomenon requires the subject to accept the norm as a means for, or an instance of, a higher-level norm already internalised. Rather complex from a cognitive point of view, this type of acceptance is both belief- and decision-based; the decision is of a deliberative type. Unlike automatic responses, it may be subject to voluntary revision and falls under the system's capacity for self-reporting: Agents can say that they accept a norm because it is a norm. Indeed, they may select which norms to accept non-instrumentally, and which ones to reject, or to observe only in order to avoid punishment. Certainly less than desirable, non-instrumental compliance with norms is far more frequent among humans than would be expected by rationality theories. Again, Pattaro's view that compliance is based on an internal drive to accept norms is probably correct.

3.3. Multiple Normative Minds

There is an important aspect of the mental life of norms, which is not included in Pattaro's analysis, though not in principle it is incompatible with it: the variety of normative types.

In our view the mental representation of a norm need not be equal in all the subjects involved. Norms are tools for regulating multi-agent systems, be they artificial or natural, and presuppose cooperative plans at work. In other words, there are different normative roles enacted by one or more agents at the same time, or in different moments and circumstances. To put it differently, norms are distributed multi-faceted entities. As different roles are needed for a norm to take effect, it follows that the mental representation of the norm, the normative belief, cannot be really identical in the various role-players. For example, let up keep distinct the roles of *sovereign*, *judge*, *policeman*, and *subject*. When one plays the role of sovereign (for example, by making a promise) in order to believe (and claim) that what she is issuing is a norm, she must believe that she is entitled to do so, and that the norm is not in contradiction with other norms already issued; whereas, when playing the role of *policewoman,* monitoring whether people are conforming or not to a given norm, she must believe she is entitled (or required) to do so, and to have the faculty or the obligation to report or sanction any violations. Moreover, these different beliefs are needed for and induce different intentions and type of behaviour associated with different normative roles.

3.4. Is Norm-Obedience Just a Subjectively Unjustified Reaction?

Non-instrumental acceptance of the norm and norm obedience might also have an *affective* property. Agents not only believe that the norm must be complied with and have the aim of doing so: They also *feel* it. However, this does not necessarily transform non-instrumental norm compliance into an impulsive automatic execution. Feelings give a warmer quality to the goal of complying with the norms, but this does not prevent the goal from being subject to evaluation and deliberation, and possibly from being deliberately violated.

However, we do not intend to neglect a fundamental problem, posed by the mechanisms of acceptance. In a cognitive approach, while giving a motivated and reason-based foundation for autonomous norm-obedience, one cannot ignore the fact that even intentional actions become automatic, when habitual. The related behaviour is no longer really decided nor deliberate, but is executed in response to the recognition of a given stimulus in a particular context, and the corresponding action is performed with reduced controls and a higher threshold of attention. This is also the case with many normatively prescribed or prohibited behaviours, provided that their execution is related to specific circumstances. For example, when stopping at a red traffic light, in many cases we do not really take a conscious decision. After a while, such behaviour becomes automatic, a *reflex*, something that is accomplished without paying much attention: The recognition of the stimulus is enough for the action to be taken. However, during the learning process (necessary to establish conditioning, or simple reflex behaviour) subjects need first to explicitly formulate a goal that coincides with the norm. For example, while learning to drive and to recognise and appropriately react to signals (e.g., "give way"), subjects will formulate the goal of slowing down and letting others go past. With practice, a direct connection develops between stimulus and behavioural response.

Nevertheless, as pointed out above, we can bring this action back under conscious control and real deliberation to evaluate the circumstances and decide to comply with the norm after considering the possibility of not doing so. Norm obedience cannot be reduced to an instinct or an uncontrolled automatic response, though sometimes it takes on these characteristics.

In sum, all the classical and basic mechanisms controlling purposive behaviour in cognitive agents apply to norm-related behaviour: From automatic reflexes, to impulsive and conscious urges, and from these to reasoned decision and real intentions. This is possible because norms elicit (and are made for eliciting) goals and goal-directed behaviours, and these goals obey all the normal constraints relating to the processing of goals in a cognitive framework.

3.5. The Interplay between Goals and Norms: Generation and Activation

In this section, we describe at some length a crucial aspect of the micro-macro link, that is the mutual influence between goals and norms. In particular, we go back to the cognitive process leading from a normative belief to a new goal.

However, the opposite process also needs to be considered. In fact, where do norms come from? Clearly, it is not the intention here to enter the dispute between imperativists and conventionalists, which Pattaro describes at length. We believe that a cognitive account of how agents generate goals may shed light on the opposite process, putting obligations on oneself, and more generally creating obligations.

3.5.1. From Norm to Goal

As noted above, the mental representation and implementation of norms is not only a matter of beliefs, but also a matter of goals. On the basis of certain beliefs, norms are aimed at activating or creating goals, and even intentions (post-decision goals). This is crucial for understanding the functioning both of the human mind and of norms.

One of the reasons why the current so called BDI (Beliefs-Desires-Intentions) approach to the study and implementation of the human mind is inadequate is that it places only *desires* at the origin of intentions and actions. On the contrary, a lot of our actions are not elicited by desire but by duty, as well as external pressures and prescriptions. Norms are precisely one of those external sources of our goals. How is this possible? How can norms generate goals? Is it possible to have an effective theory of norms without modelling this part of the process?

Norms work through social *goal-adoption*, i.e., the fact that x believes that y wants p, is a reason for x to adopt p as a goal, since it is also the goal of y.

Goal-adoption can be unilateral and spontaneous, as in benevolent actions. In such a case, x ascribes a given need or desire to y and without any (implicit or explicit) request from y, x decides to realise y's goal. But, of course, y may expect or ask x to do something for her (and x understands that y is expecting this from him). We use the term *adhesion* for a goal-adoption determined by an implicit or explicit request. Norms are aimed at obtaining adhesion, and this is why they can be communicated, informing subjects of the will of the authority that something be done by someone. If y's goal is that x does something (an expectation about x's behaviour, a request, an order), x's adhesion to y's goal will lead to x performing the expected action.

The problem is that autonomous cognitive agents need internal reasons, i.e., goals, for doing something; their actions are intentional and motivated, and follow some decision. Hence, the question is, on what basis do the norms aim to be adhered to by competing (within the decision-making process) with other active goals, stemming from personal desires, needs, and so on? Why should agents decide to conform to a recognised norm?

Autonomous agents are not driven from the outside; and the belief that they are addressed by an imperative is not sufficient for them to carry out the action immediately. An autonomous cognitive agent acts always for her own final motives and purposes, and has to have reasons for choosing to act as she does.

An agent can decide to do her duty, to adhere to and comply with a norm for several higher motives:

- instrumental reasoning: incentives or sanctions enforcing the norm, including the approval of others and reputation;
- cooperation: The subject shares, is convinced by and wants to cooperate with the purpose, the function of the norm; she would do this in any case (for example, helping a person injured in a road traffic accident),
- terminal goal: She has the goal (motive) or value that "norms should be respected" (Kantian morality).

The agent reasons about a norm, and assesses the consequences of violating it especially in the first two cases. For example, she can decide whether or not to stop at the red light at a deserted crossroads, where no vehicles or policemen are passing by, in which the chances of a collision or sanctions for failing to respect the norm are negligible.

However, in our view this is not how norms are ideally intended to work. Rather, they are sub-ideal cases based on norm enforcement.

The normative imperative is intended to be complied with as such, because it is a norm and it is the will of an acknowledged authority. Requests are aimed not only at having the requestor's goal adopted, but also at making the addressees build up in turn a plan in their minds. When I ask you something as a personal request or as a matter of courtesy, my aim is that you adhere to my request with a specific mental set, for specific motives, for example, out of pity, generosity, or courtesy. Conversely, if I give you an order, I intend you to recognise it *as an order*, based on your perception of my power or authority, and to do as requested out of fear or obedience (nor for reasons of pity, courtesy, sympathy, cooperation, or so on). Generally, we expect a specific response for certain goals and reasons.

The same holds for norms; they are aimed at being adopted out of obedience, for compliance to a normative authority, and because they are norms and "norms must be obeyed." Of course this motive can be absent from or weak in the minds of agents (it is a matter of socialisation and education and of the current *credit* of institutions); hence, agents might not always follow a norm, and sometime may even reject it. This is why norm-enforcing mechanisms are often useful corollaries of normative imperatives.

© 2006 The Authors. Journal compilation © 2006 Blackwell Publishing Ltd.

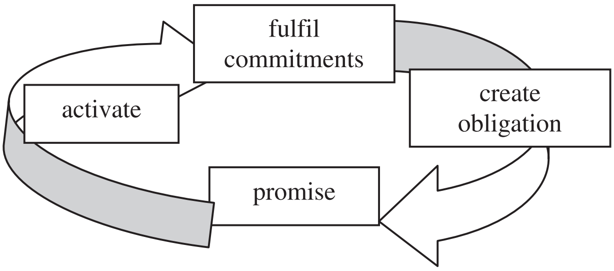


Fig. 1: The promise circuit

3.5.2. From Goal to Norm

In discussing the general act of a promise, Pattaro shows how and why it has been taken as a basic form of norm issuing.

> The promise confers on the promisee the normative power to order the promisor what to do. [. . .] the promisee issues binding commands on the basis of a *particula libertatis* transferred to him or her through a promise made by the promisor. This is the reason why the promisee's commands create obligations binding upon the promisor. (Pattaro 2005, 50)

Later on in the volume (Pattaro 2005, 244ff.), the author carefully argues against such a view (and consequently against the account of the social contract "offered by various seventeenth- and eighteenth-century natural-law theories.").

For our part, we believe that promises in their general form exemplify how interpersonal obligations are created. A promise creates an obligation by activating a higher-level obligation, i.e., to fulfil one's commitment or to keep one's word. This in turn creates an expectation on the part of the promisor, namely, to keep to the specific promise and fulfil the commitment. This lower-level obligation is therefore an instance of, and a means for, the observance of the meta-obligation activated by the promise (see Figure 1):

There are two interesting aspects of this model. First, it may be said to apply to social and non-social contexts: One may also make a commitment to oneself to do something (see Castelfranchi 1996). In this case, the 1 promisee coincides with the promisor, and it makes no sense to state that the promisor transfers part of his freedom to the promisee, who in turn creates an obligation on the promisor. Rather, the obligation comes into existence as a means for a meta-obligation, which is activated by the promise and in turn creates the sub-obligation to keep to the current promise as a means.

© 2006 The Authors. Journal compilation © 2006 Blackwell Publishing Ltd.

The second interesting aspect of the model is that a promise is a voluntary act, indeed an intention. Hence, agents want to create obligations on themselves. Why and how is this possible?

The motive for such behaviour requires specific analysis. One possible answer has to do with the promisor's inability or reluctance to do something (adopt the promisee's goal), which is either obligatory or instrumental for obtaining something else in return.

The second question, concerning the means, is more interesting for the purpose of the present analysis. In particular, we would like to stress that what is true for goals is also true for norms: One cannot generate goals from scratch, so to speak. Not, at least, by cognitive influence. As noted above, for a new goal to be generated, a specific mechanism ought to operate (means-end reasoning): A new belief (q is a means for p) interacts with an old goal (p), thereby generating a new goal q as a means for p.

Analogously, one cannot generate obligations anew: for this to occur, a specific mechanism ought to operate (normative reasoning). By this means, a new belief (o_j is a means or activates o_i) interacts with an old normative goal $G(o_i)$, generating a new normative goal $G(o_j)$ as a means for $G(o_i)$.

4. External Commands

As shown systematically and analytically in Pattaro (2005), the study of norms has been constantly characterised by different forms of "dualism." From time to time, different aspects of norms have given rise to different dichotomies, such as their origins (social contract *v.* the state of nature), binding force (conventions *v.* obligations), ontological status (objective *v.* subjective), formation (emergence *v.* deliberate issuing), and so on.

If norms express a will, it is essential to account for the relationship between the will expressed in the norm and that generated in the subject. We also agree with the assumption that an accurate and satisfactory theory of norms cannot fail to account for the profound difference between legal norms and other social rules, for example, social conventions.

Norms are instruments for multi-agent coordination and regulation. However, if not nature, at least society *facit saltus*: A major discontinuity occurs between the issuing of norms and the gradual emergence of social conventions. Of the various dichotomies mentioned above, the one between deliberately issued and spontaneously emerging norms is perhaps the most challenging.

However, and this is where perhaps our analysis diverges to some extent from Pattaro's, we would not see this discontinuity as an irreducible gap. Rather, we would like to put social rules on a continuum of binding force, or obligation, generated by external commands. In this section, we endeavour to define external commands in a more systematic, and yet pre-formal, manner. We turn next to the question of norm innovation.

© 2006 The Authors. Journal compilation © 2006 Blackwell Publishing Ltd.

4.1. A Step toward Unification

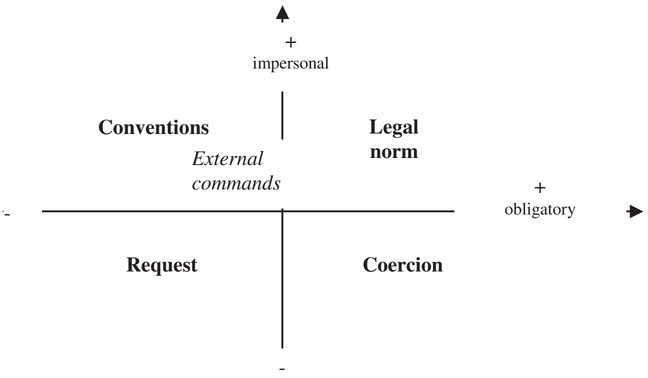
In line with the analysis presented in previous work (Conte and Castelfranchi 1995; Conte 1998), we maintain that despite important differences, legal norms and conventions are extremes on a continuum. We propose that this continuum be seen as a gradient of obligation. In a more accurate model, the more anonymous, impersonal, abstract the command, the more institutional, entitled, formal the appearance of the norm. In other words, we claim that social rules, including rituals and social conventions, are mandatory, albeit to a varying degree. The two-fold nature of such a binding role corresponds and is incorporated into a linguistic ambiguity: (i) social rules "bind" people together, reinforcing their social ties, and (ii) they "oblige" people to comply with them. It is significant that in everyday life, customs, rituals and conventions are described and conveyed in no uncertain term: "Don't open your mouth while you're eating!"; "Remember: To answer a greeting is a must!"; "But my dear, you should never wear red for a wedding!"; "Serving fruit before the dessert is allowed, but not recommended." Although no specific sanction is mentioned, these sentences are expressed as imperatives.

The question is, what kind of *imperium* do they express? More precisely, in what sense do they differ from ordinary commands? In the examples examined here, we cannot simply solve the problem by making reference, as done before, to the will of an institutional authority. *Imperium* (in the Latin sense referred to by Pattaro 2005, 53) is an external command perceived as legitimate, as it is (i) impersonally addressed (anyone in certain circumstances is required to comply with it) (ii) impersonally issued (it is expected that . . .), or issued by a *third super-agency*, that we may characterise as an intended abstraction from the original ones.

One tricky theoretical problem concerns the interplay between interests and will. A personal request may be disinterested and even benevolent: "Why don't you go out and take a walk? I'll do the dishes for you." Analogously, a command may be parental in tone: "I want you to do your homework before going out!". In contrast, unless hypocritically pursuing private goals, impersonal will is by definition disinterested. One of the main properties of institutional *persons* is to represent or act in the interests of the whole, whether they are identified or not with the interests of all of its components or with the supreme interest of the collectivity itself (see also Conte and Turrini 2006).

Although a self-interested command may be an extortion or abuse, a command that is not in the interest of the person making the request need not be a social rule. It might consist of a parental command, inspired by a more or less personalised form of benevolence towards the addressee. We claim that the less this is the case, the more the command is normative. Such a conclusion matches one of the best-known, but most controversial, views of norms as tools of coordination. At least as far as conventions are

Table 1: Internal and External Commands 2



concerned, Lewis's classical interpretation requires them to be Pareto-optimal equilibria (Lewis 1969), in which all players, to put it in game-theory terms, turns out to be better off in the end. To this view, imperativists obviously object that norms may be unfair and still maintain their binding (obligatory) force. We do not claim to say anything conclusive on this dispute, since both views find room within the normative domain. Whereas conventions often emerge as coordination tools for the benefit of all the parties involved, though not necessarily perceived as such, norms are not necessarily inspired by distributive justice, but by the well-being of the collectivity as a whole, which might result in conflict with the interests of parts of society. They are officially presumed to be in the interests of the collectivity they represent.

In this perspective, social rules do not differ from legal norms. They are mandatory, though to a lesser extent than legal norms, and their binding force is based on the same grounds: Both are rooted in on *imperium*, or impersonal will. Clearly, institutions are also endowed with organised and possibly coercive power. But without *imperium*, coercive power does not provide sufficient ground for building institutions.

Social rules may be placed on a gradient of obligatoriness, from the least binding, social conventions, to the most binding, legally sanctioned norms, but all of them share at least one property: They are *external commands,* insofar as they express an impersonal will and are addressed to an impersonal recipient.

4.2. Back to Acceptance

We have so far argued that social rules are external commands, impersonally addressed and impersonally (or super-personally) issued. Unlike private commands, following from particular interests, external commands

© 2006 The Authors. Journal compilation © 2006 Blackwell Publishing Ltd.

ground their binding force in impersonal, disinterested will. But what does this mean, more explicitly?

In his "Postscript" to his 1961 work (see Pattaro's careful and clear reconstruction in Pattaro 2005, 174ff.), Hart states that social conventions differ from norms in the reasons why they are followed. "Rules are conventional social practices if the general conformity of a group to them is part of the reasons which its individual members have for acceptance" (Hart 1997, 255–6 as reported in Pattaro 2005, 176). In other words, a rule is a convention if it is followed out of conformity.

Indeed, people frequently conform to conventions to the extent that others do the same. But the same is true for many norms: People see an incentive to comply in the general respect of norms. Notably, however, conformity is not always a gregarious behaviour: As argued elsewhere (see Conte and Dignum 2001), agents often infer the existence of norms by monitoring others. They replicate the behaviour observed only if they have reason to believe that this is expected by the norms.

In both cases, it is true that observing the behaviour of others plays a strong part in people's acceptance of social rules. The question, however, is not only why people accept social rules, but why they ought to do so, i.e., the ideal reasons for norm acceptance. We fully agree with Pattaro when he says that norms are, or claim to be, motives for behaviour: not goals, we should like to add, but reasons for generating new (normative) goals. As stated in the previous section, norms are external commands with ideal reasons for their acceptance incorporated into them. In a deontic enunciation, one finds both what is to be done and why it is to be done. Both the content of a command (the action to be taken, the conduct to be carried out), and the reason for adopting it are expressed in one and the same sentence. By way of illustration, think of the impotence of a parent or tutor answering the question: "Why should I do this?" with the inconclusive statement: "Because you *must*!". No other social artefact has a comparable impressive power. The obligatory nature of the norm is ideally self-contained, grounded in itself.

However, the ideal reasons for acceptance are not effective enough to make people observe the norms. Norms are not self-enforcing *per se*, and often need the support of incentives and deterrents. The point is not so much whether the real reasons for acceptance match the ideal ones. Rather, it is that norms are expressed and recognised as self-justified commands. External commands are impersonally issued and addressed, and this seems to provide a reason for their acceptance.

4.3. Proliferation v. Innovation

As often stated by Pattaro, norms are not only motives for acceptance but also beliefs. Earlier in this paper, we reproposed our view of normative beliefs, which can easily be extended to social rules at large. We do not follow Hart, when he states that a command is believed to be a social rule, if it is generally conformed to (although conformity might provide an indicator that it is a social rule). On the contrary, we maintain that something is a social norm when it is believed to be expressed as an external command, that is, when it is believed to be impersonally issued and addressed, thereby providing ideal reasons for its acceptance. This normative belief is essential for the acknowledgement of any sort of social rule. In the case of norms in the full sense, the command is traced back to a specific normative source, the sovereign, whose properties were examined earlier in the paper.

However essential, a normative belief is not sufficient for an external observer—even if she is aware that such a belief is held by a given entity—to say that a norm exists anywhere else than in the mind of such a believer. In this respect, our view is perhaps less constructivist than Pattaro's. What else is needed, then?

We believe that a norm is a complex construct that includes both mental and social constructs and processes, and that this view has great potential for the study of norm innovation. Notably, in fact, the classic view of conventions as based upon general conformity, though well suited to explaining the proliferation of norms, seems to encounter more difficulties in accounting for their innovation. How conventions spread is comprehensible, but how and where do they originate? What is the factor at the origin of a new social rule?

A possible answer to this difficult question might be found in the interplay of observed behaviour, and the way it is mirrored in the beliefs of the observers. If any new behaviour *b* is interpreted as obeying an external command, a process of normative influence will be activated (cf. earlier in the paper; see also again Conte and Dignum 2001). The new behaviour *b* is more likely to be replicated than would be the case if no normative belief were formed. Hence, the replication of *b* will increase as a function of the probability that *b* is believed to be dictated by an external command. Indeed, such a probability will be increased by an active normative influence exercised by the normative believer. A good example of this is to be seen in the practice of flashing headlights on the highway for cars coming in the opposite direction, a practice that started to appear with the increase in speed limit enforcement measures.

As shown elsewhere (Conte and Castelfranchi 1999), when a normative believer replicates *b*, she will influence others to do the same not only by ostensibly exhibiting the behaviour in question, but also by explicitly conveying an external command. An example of this type of influence is offered by "politically correct" conventions, which include a wide vocabulary of unacceptable and strongly disapproved linguistic usages. The appearance of such conventions is associated with a change both in behaviour and in social "feelings" with regard to minorities and certain social categories, which is indicative of a profound transformation in social values. But corresponding conventions do not spontaneously follow the new moral code. Rather, people have been increasingly imposing them on one another by means of external commands ("One should never use those words when speaking about others") and explicit evaluations ("I cannot stand this language!"; "It upsets me to hear people speak like that").

Clearly the difficult question here is to formulate a predictive model: What behaviour, on the grounds of what factors and processes, is likely to become the object of external commands? An interesting question, to which we have no answer as yet. However, it is our impression that it would be impossible to answer it without a two-fold view of social rules, both as types of behaviour and external commands. Moreover, it is our position that nobody can answer such a question without considering the issue of what conventions have in common with norms.

5. Concluding Remarks

We would like to conclude this discussion by stating once again that Pattaro's approach to the theory of norms is truly enlightening. Clearly, as cognitive social scientists, we would like the mental path followed by norms in regulating human behaviour to be stressed even more.

In particular, the main issues that we have endeavoured to examine in this paper are as follows:

- The complex and rich nature of the crucial normative belief, as identified by Pattaro, must be further analysed, by specifying the particular conditions and the assumptions needed for the acknowledgment of impinging norms.
- The conditions under which normative beliefs activate a corresponding type of behaviour also need to be further investigated and modelled; our claim here is two-fold: (i) behavioural activation implies a process of goal generation (or activation) based on a process of goal-adoption (from will to will); and (ii) there are different norm-related minds and types of behaviour (that is, partially different beliefs and goals).
- It is precisely on this explicit cognitive side that we see more similarities between social rules (and their emergence) and institutional norms: Although norms dramatically differ from social rules (like conventions), the properties they have in common ought to be recognised. As expressions of an impersonal will, they both seem to be located on a cultural evolutionary continuum: also conventions entail external commands.

- Such a unifying model is not only valuable *per se*, but also seems necessary to account for the process of innovation in norms and social rules, which is a crucial and still open aspect of the social life of norms.

(for Rosaria Conte and Cristiano castelfranchi)
National Research Council
Institute of Cognitive Science & Technologies
Via S. Martino della Battaglia, 44
I-00185 Roma
Italy
E-mail: rosaria.conte@istc.cnr.it; cristiano.castelfranchi@istc.cnr.it

References

Castelfranchi, C. 1995. Commitment: From Intentions to Groups and Organizations. In *Proceedings of the First International Conference on Multiagent Systems* (June 12–14, 1995). Ed. V. R. Lesser and L. Gasser. San Francisco, CA: MIT.

Castelfranchi, C., and R. Conte. 1999. From Conventions to Prescriptions. Towards a Unified Theory of Norms. *Artificial Intelligence and Law* 7: 323–40.

Conte, R. 1998. *L'obbedienza intelligente*. Bari: Laterza.

Conte, R., and C. Castelfranchi. 1995. *Cognitive and Social Action*. London: University College London Press.

Conte, R., and F. Dignum. 2001. From Social Monitoring to Normative Influence. *Journal of Artificial Societies and Social Simulation* 4: http://jasss.soc.surrey.ac.uk/4/2/7.html

Conte, R., and P. Turrini. 2006. Argyll-Feet Giants: A Cognitive Analysis of Collective Autonomy. *Cognitive Systems Research* 7: 209–19.

Hart, H. L. A. 1997. *The Concept of Law*. 2nd. ed. With a Postscript edited by P. A. Bulloch and J. Raz. Oxford: Clarendon.

Lewis, D. K. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Pattaro, E. 2005. The Law and the Right. A Reappraisal of the Reality that Ought to Be. In *A Treatise of Legal Philosophy and General Jurisprudence*. Ed. E. Pattaro, vol. 1. Berlin: Springer.

© 2006 The Authors. Journal compilation © 2006 Blackwell Publishing Ltd.

SNP Best-set Typesetter Ltd.	
Journal Code: RAJU	Proofreader: Emily
Article No: 342	Delivery date: 31 August 2006
Page Extent: 17	

AUTHOR QUERY FORM

Dear Author,

During the preparation of your manuscript for publication, the questions listed below have arisen. Please attend to these matters and return this form with your proof.

Many thanks for your assistance.

Query References	Query	Remark
q1	Au: Should the **Castelfranchi 1996** here be changed to **Castelfranchi 1995** so as to match the reference list?	
q2	Au: Table 1 has not been found in the text.	