

# A Cognitive Framing for Norm Change

Cristiano Castelfranchi<sup>(✉)</sup>

ISTC-CNR GOAL Lab, Rome, Italy  
cristiano.castelfranchi@istc.cnr.it

“Just remove a brick and the wall will sink”  
(Arab saying)

**Abstract.** Norms are within minds and out of minds; they work thanks to their mental implementation but also thanks to their externalized supports, processing, diffusion, and behavioral messages. This is the normal and normative working of Ns. Ns is not simply a behavioral and collective fact, ‘normality’ or an institution; but they necessarily are mental artifacts. Ns *change* follows the same circuit. In principle there are two (interconnected) *loci* of change with their forces: mental transformations vs. external, interactive ones. Ns change is a circular process based on a loop between ‘emergence’ and ‘immergence’; that is, changes in behaviors presuppose some change in the mind, while behaviors causal efficacy is due to their aggregated macro-result: acts that organize in stable choreographies and regularities build (new) Ns in the minds of the actors. More precisely the problem is: which are the crucial mental representations supporting a N conform (or deviating) behavior? And *which kinds of ‘mutations’ in those mental representations produce a change in behavior?* I will focus my analysis on Social Norms, in a broad sense.

**Keywords:** Norm change · Normative mind · Normative agents

## 1 Premise: Situated Normative Cognition<sup>1</sup>

I will discuss the internalized/externalized nature and working of Norms (Ns) and its impact on N change. What I have in mind is a *hybrid society* (humans and AI-Agents interacting together) with “norm sensible Agents”. On the one side the Agent mediating and supporting human interaction, exchange, organization should be able to understand human conduct in terms of Ns and to monitor and support that; on the other side Agents should be themselves regulated by true Ns (not just pre-implemented binds, executive procedures, but real deontic representations with the mission to regulate their decisions and conducts) and be able to violated them in the right situation.

The analysis and typology that I will propose (that will not be complete and fully systematized, but just *in fieri*) is focused on Social Norms (SocNs), in a broad sense,

---

<sup>1</sup> I’m in debt with my colleagues and friends (in particular Rosaria Conte, Luca Tummolini, Giulia Andrighetto) for my work on norms theory.

covering various kinds of.<sup>2</sup> Of course here I will put aside legal Ns (where there are institutional and legal ways for Ns change) although I think that several of the mechanism that I try to enlighten for SocNs also hold for legal ones.

Norms are in minds and out minds; they work thanks to their mental implementation but also thanks to their externalized supports, processing, circulation, and dynamics. This is the normal and normative working of Ns. Also because usually a N is a strange relation between a practical, effective, externalized object (the conduct of X; however mentally/internally regulated) and a cognitive artifact: a written “table of law”, a symbolic representation, a (verbal or non-verbal) message that has to pass into minds. This double face of N (cognitive and behavioral, both internal and external) is intrinsic. Ns are not simply a behavioral and collective fact, a “normality” or an institution; but they necessarily are mental artifacts [13, 22]. A N impinges on us and works *thanks to* its mental representation, (partial) understanding, and specific motivations. However, as we just said, they are not just a mental fact: this serves to determine and control the actors’ conducts and to build shared practices, scripts, messages and collective effects.

Our claim is that also Ns *change* follows the same circuit. In principle *there are two (interconnected) “loci” of change with their forces: mental transformations vs. external, interactive ones*. Of course, they are interrelated since the mental changes determine behavioral changes, which determine collective new dynamics. vice versa, behavioral changes that we observe will change our mind and our norm conception or repertoire. In other terms it is both a process of ‘emergence’ [44] and ‘self-organization’; and a process of ‘immergence’ [14, 21] and mentalization: a feedback from behavior and collective structure/phenomenon back to the individual minds layer. Not just a bottom-up and top-down, and an inside-outside and outside-inside process, but a real ‘loop’: virtuous or vicious *circles* of Ns change or confirmation or instauration. We need the same dynamics in normative Agents, able to learn and evolve SocNs, and to read the behaviors of the others in these terms for monitoring it or adjusting to it.

It would also be relevant to consider that there is no just one and unique normative role for actors with its specific mental attitudes (beliefs, goals, expectations, ...). We are not only ‘subjects’ to the N (prescribing us certain behaviors and mental states), we also have to play the role of ‘watchman’ and ‘punishers’ of the others [11, 30]; a fundamental role in N script and for the maintenance of the social order. We have to play the role of ‘issuers’ too: (either explicitly or implicitly) proclaiming Ns, prescriptively informing about them, explaining and reminding us them (for example parents towards children). I will put aside here these different normative minds and roles<sup>3</sup>, although I believe that the role of a normative ‘watchman’ will be very relevant for Agents.

What we will try to do in this work is to examine: (a) some of the *main mutation ‘events’* in particular *internal* to the *subject’s* normative minds; but also (b) as individual

<sup>2</sup> From politeness to customs, from moral norms to Ns and rules in organizations, associations, communities of practice with their “rules”. For a systematic analysis of social norms and discussion about the general theory see [5, 6, 12, 31, 35].

<sup>3</sup> I will also do not examine the other crucial phenomenon in Ns evolution: the introduction of a completely new N, and its issuing or negotiation. I will mainly focus on adherence or violation (and their reasons) in N changing, adaptation, or extinction.

conducts become signs (cues) and/or messages (signaling), and change the others and the collective emergent conducts, so becoming public phenomena and institutions. Also the other way around; I will give some hints about that: (c) how acts that organize in stable collective conducts build Ns in the minds of the actors [6] but not just as a regularity to conform to, but as expectations and “prescriptions” from the others [19, 23].

## 2 Roots of Ns into Minds

Real “norms” are based on the possibility to be violated, not obeyed. Otherwise they are not “norms” but physical barriers or ties and chains. Ns are devices for the control of “autonomous” agents that decide what to do on the basis of their beliefs, reasoning, and goals. Ns not only presuppose (accept) but also postulate a freedom in the addressees.

Our main claims are the following ones:

- A N is not just aimed at regulating our conduct, at inducing us to do or not to do a given action; it is aimed at inducing us to do that action *for specific motives*, with a given mental attitude (belief, goal, expectation). The ideal-typical *Adhesion* (see Sect. 3.2) to a N is for an intrinsic motivation, for a “sense of duty”, recognition of the authority, because it is right/correct to respect Ns, etc.; and only sub-ideally one should respect Ns for avoiding external or internal sanctions (see below). Also normative education goes in this direction [18].
- We agree with Bicchieri’s theory that an “empirical expectation” and the perception of the existence of a “normal” diffused behavior is not enough for creating a real N in “normative” sense (to use Kahneman’ terminology [39]). A merely “descriptive” N is not “injunctive” [42]; a N implies for us a prescriptive character: it is for inducing us to (not) do something. There is a social pressure: expectation and *prescription*.
- As we said, our object is “norms” in the “normative” (prescriptive) meaning/sense, not in the “normality” (descriptive or statistic or standard sense). However there is an important and bidirectional relation between N in *normative* sense and N in *normality* sense:
  - (a) Normality-N creates and becomes a Goal for the actors and even a normative-N (a prescription, something “due”), in order to conform, to be like the others. This conformity is either a need of the individual or a need (and request/pressure) of the group, or both.
  - (b) Normative-N creates a statistical normality-N, a normal conduct in the community, if it is respected: N conformity is “normal”. Moreover:
    - Normative-N has the *goal* and the *function* to be respected and thus to create a normality-N, a normal behavior (at the individual, internal level this helps N also to become an automatic response, just an habit);
    - If normative-N doesn’t become/create a normality-N it is weakened and perceived as less credible and less binding [6, 22].
- In order to *perceive* a social practice as a N we have to guess, presume, or understand some “end” in it: the protection of the interest or rights of somebody, of the community; from that a deontic “should”, an obligation. Not conforming is an harm, is noxious, not just something irregular, strange. I’m at least frustrating your

prescription to maintain regular practices; you count on that and plan to regulate your behavior on that; so I'm upsetting and betraying you, not just amazing you. I'm harming social order, and the natural 'suspension' of uncertainty, the assumption of normality: a fundamental good [32], a "common".

- Ns have to be "impersonal" and depersonalized (and perceived as such) on both sides: the issuer's and the addressee's side. It is not a conflict between you and me; it is not "my" personal request (for me, for my desires, etc. for my personal will that you have to adopt); and it is not a request to "you". The message is:

*"I do not talk, monitor, sanction, in my name"; "I'm not addressing to you "ad personam", but as an instance of a class, a member, a citizen, ... like any other in the same conditions". Also for that "You have no reasons for rebelling".*

This really is a crucial point in the perception of Ns as Ns; thus it is something that must be signaled in some way (for official Ns: uniform, role symbols, specific documents, etc.; for Social Ns by collective practice or attitude or explicit messages) or at least contextually presupposed and assumed in the script.<sup>4</sup>

- As we said, Ns are social devices controlling behaviors through minds [14] but in a specific way; through a partial understanding. They require (for their existence and effectiveness) their *explicit mental representation*, their (partial) understanding and recognition "as Norms"; specific cognitive representations and motivational processes ("*Cognitive Mediators*": [22, 24]); differently from other social phenomena like *social functions*, that can be played by social actors even without understanding - and even less intending - them [16]. Not necessarily the agent supporting the N in some role has as his/her mental goals ("intention", "motive") the aims and utility of the N; these are the goals (and functions) of the N not of the individuals.
- Ns have to build in us an "ought", a "duty", "you have to"; with a rather constrictive feeling, a negative "frame", an avoidance orientation (even when it elicits "you have to do this action"). And this "ought" is a non-technical "ought", not instrumental to and planned for a given outcome/goal. This entails a process of Adhering *without sharing* the 'instrumental' nature of the N, and without (necessarily) understanding/adopting its 'function' or end. My 'plan' is different from the authority's 'plan'. Citizens are not real "cooperators" but "subjects". They have to "alienate" their own powers and products [18].

### 3 N Internalization

Anyway, all this requires a specific "translation" of Ns into the minds of the addressees such that they recognize a N as such, and – on the basis of various motives – decide

---

<sup>4</sup> The fact that Ns are always relative to a "class" of subjects, not just to one specific person and it holds "for all the values of X" is one reason why the violation has not an individual meaning. X the violator is just "one of all/many", is a representative, an "example"; that's why his (bad) behavior can be a (bad) "example"; and the impact of the behavior is more that "individual": It is not longer true that "for any value of X, X has to, will do, and does action A".

whether to conform or not to it. Let's sketch the basic constituents of Ns internalization in our theory [18, 24]. Ns are based on a *specific* process of *Goal-Adoption* or better *Adhesion*; since they have the nature of an “imperative”.

### 3.1 Goal Adoption and Adhesion

Ns induce new goals through “adoption”. *Goal-Adoption* is how an autonomous agent is not an isle but becomes social, or better pro-social<sup>5</sup>; that its s/he does something *for* the others; *puts her/his autonomous goal-pursuing (intentional action), her/his cognitive machinery for that, and her/his powers and resources into the service of the others and of their interests*. What is needed is the architecture of a social Agent able to import goals from outside (and to influence other agents by giving them goals and relying on them) but remaining ‘autonomous’. S/he is able to arrive to set up an intention not only from her own endogenous ‘desires’, but also from imported goals.

**Goal-Adoption** means that:

*X believes that Y has the goal that p and comes to have (and possibly pursue) the Goal that p just because he believes this.*

“I do something ‘for’ you” (which doesn’t mean ‘benevolence’!); I want to realize this since and until you wants/ needs this; because it is your goal.

Of course there are different kinds of Goal-adoption, *motivated* by different reasons: merely selfish and instrumental, like in exchange; altruistic; or strictly cooperative, for a common goal. Ns prescribe a specific motive for accepting the injunction: in Bicchieri’s view’s a “normative expectation”, for us also the recognition of the *prescription* by the others and their authority (see below).

A stronger form of G-Adoption is **Adhesion**: *when I adhere to your (implicit or explicit) ‘request’ (of any kind: prey, favor, order, law, etc.). In other words, you (Y) have the goal that I adopt your goal p, that I do something (action a of X) realizing that goal, and I adopt your goal p or of doing a, (also) because I know that you expects and wants so.*

In Adhesion one of the *reasons* for Adopting the goal of the other is that the other wants so:

- She also has the (meta-)goal that we adopt her goal;
- We adopt her goal by adopting the meta-goal.

In a sense, there is a double level of adoption (a meta-adoption): *I know and adopt your goal that I adopt*. Moreover, in case of Adhesion there is a (presupposed) agreement between X and Y about X’s adoption, X doing something as desired by Y. Other forms of adoption (like help) can be unilateral, spontaneous, and even against Y’s desire. Ns require from us not just adoption but adhesion.

---

<sup>5</sup> Not to be used as synonym of “altruistic”, “benevolence”, etc.

### 3.2 Normative Adhesion

Adhesion obviously presupposes specific beliefs into the mind of the agents (and this is the first aim of the N: to be conceived/perceived as such). In particular the recognition of the N as a N, in force on me, and valid in that context.

It is implied a ‘generalized’ G-Adoption where:

- X believes that there is a goal impinging not directly on a single individual but on a class or group of agents:
  - if X believes to belong to that class,
  - she believes to be concerned by the norm, and
  - she instantiates a Goal impinging on her; adopts it.

Having adopted the ‘generalized’ goal X doesn’t limit her mind and her behavior to this (self-regulation); she will also worry about the others’ behavior:

- X is also able to have Goals about the others’ behavior: she adopts the Goal not to do but that for any z (DOES z A).
- Given such an Adoption she has *expectations* (predictions + prescriptions) about the others behavior, and is not only surprised, but also ‘disappointed’ by their non-conformity.

Also because she is paying some cost for respecting the norm and the authority, for maintaining the prescribed social “order”, which is supposed to be a “common”. She wants the other be fair, reciprocates, contributes.

### 3.3 Equity and Spreading

Conte and Castelfranchi [23] claim that the decision to conform to what is perceived to be an obligation plays a relevant role in N spreading over a population of cognitive agents. While the conventionalist view derives social norms from the spreading of conformity, in our view *conformity is derived*, so to speak, *from the spreading of obligation-recognition and -adoption*.

“The very act of accepting an obligation implies and turns into enforcing it. The agent respecting the obligation turns into a supporter. Conforming leads to prescribing. The agent undergoing an obligation becomes a legislator. The more an obligatory behavior is believed to be prescribed, the more it will be complied with, and the more, in turn, its prescription will be enforced. Rather than acting only through a behavioral contagion or a passive social impact, the spreading of norms is affected by *cognition* in a variety of ways and attitudes”:

- (i) *It leads to implementing effective conformity.* When an autonomous agent recognizes a norm as a norm and decides to conform to it, the number of conformers will be increased, and the norm is more effective.
- (ii) *Effective conformity contributes to the spreading of normative beliefs.* The larger the number of conforming agents and the more likely the observers will form normative beliefs and the strength/certainty of the belief will increase.
- (iii) *The spread of normative beliefs contributes to the spreading of normative actions.*

- (iv) *The spread of normative actions contributes to the spreading of normative influence.* The larger the number of agents conforming to one given norm, and the more distributed will be the want that other agents will conform to the same norm. “This is due to:
- An *equity* rule. People do not want others in the same conditions as their own to sustain lower costs - benefits being equal (this is, indeed, one the most probable explanations of the Heckathorn’s [36] group sanction control: the more agents respect the norms, and the more likely they will be to urge others to do the same).
  - “*Norm-sharing*”. Agents are likely to “share” the respected norms, that is, to believe that those norms are sensible, useful, necessary, etc. This is also a powerful self-defensive mechanism (agents share the norms they happened to respect). Agents will defend the norms they share, implementing the number of agents who want those norms to be respected.” [17].
- (v) *The spread of normative influence contributes to the spreading of normative beliefs,* and the whole process is started again in a circular way.

The same cognitive mediation holds for an *observed* violation, deviance, and their crucial *interpretations* and meanings by the observer (see also Bicchieri and Mercier [7]).

Also for Agents this might be relevant: do we want/need just agents doing as expected/ordered or agents able to violate but also able to conform to the norm as a decision and for specific deontic motives/reasons (N-Adhesion)? Don’t want we to “share” norms (social, moral, legal) with our Agents? To really have a hybrid society regulated by values and norms?

## 4 *Internal Locus: Kinds of N Mutation Within Subjects’ Mind*

Let’s identify the various though and ‘reasons’ of the ‘subject’ (S) for *abandoning* or *violating* a given N. We will distinguish between:

- (i) ‘Unintentional’ effects; where changing or weakening that N (or Ns) is not the end or an end of S, and
- (ii) ‘Intentional’ act; where S understands, expects, and intends to jerk the N.

### 4.1 *Norm Decay, a Close Approach*

It is useful to cite a recent work on N decay, also in order to underlining some differences with our proposal.

In Hammoud et al. [35] we find a good perception of the role of N decay (not studied enough), and an important formal and simulation study, also with a nice ontology of different forms and reasons for Norm decay.

In their perspective: “*Norms decay* refers to the case in which *a norm is not practiced or adopted by any of society’s members, and eventually deleted and forgotten.*” They introduce a framework that contains three cases of norms decay which are: *Norms*

*Removal, Norms Disappearance, and Norms Collapse.* The first case needs an intervention from a powerful authority, while the latter two cases happen when society members stop adopting or violate a norm. That is, there is a change starting from the decision of the agents.

- “*Norms disappearance* is the result of abandoning a norm from the majority in a society. Abandoning a norm means not practicing it without being sanctioned from the authority. Norm abandoning happens when it loses its benefit.”
- “*Norm collapse* is the case of norm vanishing from a society due to agents’ violation of this norm and violation sanction decay. An agent checks the benefit of violating a norm, and the sanction of violation ... if the benefit of violation is more than the sanction of violation, the agent violates the norm and gets the benefits he wants”.

The main difference with our view is of course our more systematic analysis of the specific changes in the mental aspects of Ns. However, there are also other differences. On the one side, we have a broader view of the ‘reasons’ why agent respect Ns; not mainly “for” avoiding sanctions, or “for” the social ‘utility’ of the N.

On the other side, the authors are a bit optimistic on the collective/community ‘benefit’ of the N. In our approach Ns are not necessarily *well conceived* by the authority or *fair*; and also social norms are *not necessarily good* for the community. They are simply self-maintaining just because they are “social order”, reduction of uncertainty, identity, even if on the practical side they can be not so good. In our vocabulary they can be badly ‘functional’ [16].

So not necessarily a N “is abandoned when it loses its benefit”. It can remain there. vice versa, N can be abandoned although it was and would be useful.

We also admit that N can be there even if not respect in practice by any agent, but they know the N’ and are aware of the systematic violation. In a sense a social meta-N is emerging, a shared practice/habits of violating N’. And we admit that not necessarily when “An agent checks the benefit of violating a norm, and the sanction of violation ... if the benefit of violation is more than the sanction of violation, the agent violates the norm and gets the benefits he wants” this induces to a “N collapse”. This self-interest violation can be there for one or few agents, since agents are in different conditions and with different preferences; what might be convenient for an agent can be not convenient for another one. The mental processing is the crucial device and cannot be so simplified and ‘rationalized’.

## 4.2 Unaware Violations

S does not realize that her behavior is an N violation. Mental conditions for such a conduct:

- Ignorance of the N (*beliefs*); or
- A mistaken interpretation or instantiation (*beliefs*): S does not realize to be a member of the set of the addressees of that N or that it does apply in those circumstances and context; or
- No memory retrieval of the N in those circumstances, lack of attention, absent-mindedness (*beliefs*).



The violation is unintended since it is fully unaware, but - given the observable behavior (“bad example”) - it equally injures the N.

There are also extra-mental conditions facilitating or inducing such a “mistake”. For example, the N and its local pertinence should have been appropriately and explicitly signaled, not given for obvious: “Please, do not park more than one car in our courtyard; this is our polite convention”.<sup>6</sup>

### 4.3 Aware Violations

#### A. *Without the goal of injuring/weakening the N*

As we do not intend the supportive ‘function’ of our conforming to the N, equally we do not necessarily intend the destructive ‘function’ of out violating it.

There are several *reasons for dropping* a N-goal, do not adhere to it and formulate a conform intention:

- (a) *Goal-conflict*: the N-goal contrasts with another goal of the agent; Apart from the *belief* that the N is in conflict, what matters are the following parameters:
  - value of the goal based on the value of the meta-goal of respecting Ns;
  - value of the contender goal;
  - value of the negative expected consequences of violation, including feelings associated to N-violation; and in particular the perceived threat: estimated probability and weight of ‘punishment’ and blame (*beliefs*).<sup>7</sup>
 A sub-case of (a) is a N-conflict: N contrasts with other Ns accepted by the agent (see below).  
 The decision to violate if I can a N that is not convenient for me now and here (not necessarily “in general”) can just be for my private interests. However, not necessarily the goal in contrast with the N is a private/personal one; it might be a goal formulate for efficiently performing S’s role or mission [17]: violating for functional reasons, for an intelligent problem-solving in our work.
- (b) *N Application & Instantiation disagreement*: S is aware of N but he contests to be a member of the set of the addressees or that it *does apply* to that circumstances and context.
- (c) *Material impossibility*: S forms a N-goal but cannot comply with it (*beliefs*); the intention would be impossible (*beliefs*).
  - i. As we said, a remarkable case of (a) – but in a sense close to (c) (in terms of not “material” but of “deontic” impossibility) – is:
- (d) *Norm conflict*: the N I should apply and respect is in contrast (*beliefs*) with another N:

<sup>6</sup> An interpersonal example may be: X: “You can not go around in underwear!” Y: “But you had to say me that there were guests in our house!”.

<sup>7</sup> This expectation should be part of what Bicchieri calls “empirical expectation” (“what we expect the other do”). However, we should distinguish between “to expect that the other conform” and “to expect that the others monitor and sanction”. Two different predictions based on different experiences that might also don’t be fully correlated.

- Either another social N (social Ns are not so coherent and non contradictory, especially in their application). For ex. the social N about our male group meeting for drinking beer implies the possibility or prescription to burp in public (just for funny and be deviant), while I would desire – due to my “education” – do not burp;
- Or a conflict with legal or organizational N.

In all these cases S will not conform to the N but she is *not* motivated by the aim of weakening it. For sure that violation (given the message to myself and to the stakeholders) weakens the N, however the agent’s intention is not necessarily this.

- (e) *Expectation of not sanctions*: Either due to some reason in the others of not sanctioning; or just because I expect to not be detected, to hidden: “I will get away with it; they will not see me; nobody will know that”; or “They do not catch any violator, they never punish”<sup>8</sup>. Of course, these beliefs are relevant in particular for agent motivated to respect Ns just by the fear of sanctions.
- (f) *Indifference to sanction*: There are cases and individuals where the fact that other people respect N and that there will be a negative judgment by the others (sometime even publically expressed), is not a sufficient reason for not violating: an important sub-kind of conflict. Consider for example a young guy sited in a waiting room where there are quite old waiting people standing up, and not giving up his seat to them, although he knows that he “should do” that, and that he is disapproved. Either there is in this guy (and context) indifference to the judgment and sanction from the others (*goals*), since “I do not care of these guys”, “who knows them?” “I will never meet them again ...” (*beliefs*). Or there might even be a provocation attitude (*goals*): “Yes! I’m not like you, I do not care of you”, “I’m underbred, so what!”. Or the attitude is “motivated” by an opposition specifically to the N, as a meaningless N: a value opposition (like in people violating the rule of giving priority to women).

All these are (more or less sincere and not self-deceptive) beliefs and motives of the violator.

Sometime we (unconsciously) find a new interpretation of framing of our action and circumstance, and of the N, in order to facilitate our violation. Consider the very famous and beautiful case of people “interpreting” the monetary sanction for the violation of the N as a fair, a price, and thus deciding to systematically violating it, and just pay what they have to pay [34]. Let’s rewrite in our mind as a tax what in fact would be a fine! But this morally facilitates our decision to violate.

- (g) *Violation as epistemic act*: I know and intend (in case) to violate, but my motive is to “see”: to see if that N is there or if I correctly understood it; or to see if the violation will be noticed/punished; to see your reaction. Even to see if you know that N, not in order that I know the N, but in order to know if you know it.<sup>9</sup>

Of course, there are other kinds of assumptions and reasoning that induce or facilitate (intentional) N violation; in particular interpretations of observed deviant behaviors,

<sup>8</sup> This is a change in our “empirical expectations” in Bicchieri and Xiao terminology [8].

<sup>9</sup> My behavior is like an exam question, where I in fact already know the answer but I want to know if you know it.

changing our mind. We will see some of them below: the effect of external changes (observed deviant behaviors) on our mind and conduct.

B. *Aware violations with the goal of harming, breaking down the N*

Violation is not just intentional but motivating: I violate in order to violate (Ns or that N).

(h) *Violating for changing*: Intentional and public violation of N for rebellion and opposition to that N, for rejecting and breaking it; to send a *message* to the others, to the “authority”. Like Gandhi that rips in a central place of Johannesburg in front of the police the special document obligatory for Indian people. The message (and belief) is “This N is discriminatory, unacceptable, unfair; it has to be abolished: rebel to it!”<sup>10</sup> Notice that I can violate an N as unacceptable, not fair even if it does not directly damage me.

(i) *Violation against stigma, for changing values, building our identity*: I violate for provocation and rebellion towards stakeholders’ values and attitudes. There are two different cases.

A possible aim is to build our collective identity, to remark that “we” are different, not like you, and we do not want be part of you (like Punk’s provocation; or adolescent deviant attitudes). We are not In-group, but Out-group; it is an “exit” or *secession move* from your value and community.

Another possible aim is to change your values, to obtain respect: like in the provocation of the “Gay pride” and exhibition: “Our aim is not splitting from you; on the contrary we want to be accepted, integrated, and respected; you have to change your conservative values and thus your social Ns on that”.

A crucial construct in human mind is the “sense of justice” and the related sufferance for iniquitous situations (not only harming us personally but even favoring us, or harming others: we can play the role of the victim, of the privileged guy, or of the stakeholder, but always with some discomfort) (“equity theory”), the *need* for equity (a “value” and a “motivation”<sup>11</sup>). We can consider a given N with this perspective, by evaluating its “equity and justice”. This changes very much our disposition in obeying to it, or in supporting/defending it as punisher (Sect. 3.3). I feel “justified” in my violation; not a bad guy but a good guy; I do not feel guilty but proud of me.<sup>12</sup> If I consider a given N unfair I can have a serious conflict between two internal values, intrinsic motivations: the sense of duty/obedience vs. the sense of justice. The conflict is within my own values.<sup>13</sup> Sometimes this mental justification and motivation in terms of “sense

<sup>10</sup> This nice example is about a legal N, however similar examples exist also for social ones; like the “provocation” acts of courageous women in Arabic countries.

<sup>11</sup> For a rigorous cognitive notion of “value” and its strict link with evaluations, prescriptions and Ns see [40].

<sup>12</sup> Agents too should have some moral value and should be able at least to interpret our behavior and reasons in these terms, and possibly mediate our interaction caring of moral norms.

<sup>13</sup> This is Antigone tragedy. This also is Socrates’ message to us while taking the poison: respecting Ns and authorities (even when their decision is incorrect and harming us) may/should be a prevalent value.

of injustice” is just a convenient alibi (in front of the others, or in front of myself) for allowing my violation for personal advantages and desires (like the “sense of injustice” sometimes used for covering/hiding our envy).

- (j) *Violation to be noticed, to innovate:* Sometime we violate a social Ns or consuetude’s just to emerge, to be noticed, and to be original; like women first wearing a bikini or a mini. These provocative guys (actually innovators that may create a new “fashion”, but not necessarily with this intention) are aware of and ready to cope with criticism and even insults.

Two examples about previous cases: I violate the N that on the beach one cannot be nude, and (with other people) I use “topless”; so I create or converge a new use, imposing tolerance to the others (they can no longer blame and reproach me). Or I’m completely nude; but this is too disturbing, intolerable for that group, so this creates a scission of groups and places: you nudists must have your own beach (and we will not come there!), but you cannot stay in “our” beach and be nude. If you become part of the new group and go to the nudist beach it become not just tolerate to be nude (the old N doesn’t constrains you any longer) but there even is a new N of “being nude”. Similar path for vegans: they want not just be permitted to refuse current food without objection, ridiculous, blame, but they are trying to build new Ns - based on new values - (“Do not eat animals!” etc.) on such a basis to criticize, blame the violator (although they are the majority) and make propaganda. Their aim is not just to build a separate culture and community, but also to change the practices and the Ns of the big community.

Notice that this kind of N change requires (and is grounded on and aimed at) a change of “value” which is first of all a specific mental object.

- (k) *Against the authority as such:* It is also possible to violate in order to rebel, but not against a given set of N that we want to reject or change, but against the normative authority A. To impair A, independently from the specific N. What matters is to violate; to show to myself or to my peer or to A that I do not respect A, do not submit: this is the message and motive. Like a “rebel” child that rejects any parents’ prescription or restriction to his desires; like some political movement or demonstration where what matters is to broken something, to do something prohibited, not what to broken and why.

The crisis of the authority (see Sect. 6.2) can be due to various assumptions and motives; like the fact that A is no longer credible, trustworthy, correctly and competently playing its role; so I do not want longer depend on and delegate to it. Or a crisis of identity and membership: I do not any longer feel one of “you”. Or for a crisis of values grounding that A: I do not any longer feel morally “obliged”. And so on.

Again; it is not necessarily a matter of sanctions, power, and fear.

## 5 *External Locus: The Others' Observed Behaviors*

Which and how many observed changes in normative behavior are necessary for changing our conform conduct? Not necessarily we need diffused and spreading practices. Even a single violation act or meta-violation (for example do not monitoring or punishing) can call into question a given N in my mind (for example, a single resounding act of euthanasia); a single provocation can be enough for discredit authority (see Gandhi's example).

To know that somebody has violated N is an important factor in the crisis of that N. However, this works through our mind and what matters is the *interpretation* we give of that behavior: Accidental? Intentional? And why? And which are the consequences?

Let's first see some examples/kinds of assumptions and reasoning that induce or facilitate (intentional) N violation; in particular interpretations of observed deviant behaviors, changing our mind:

### (1) *Interpretations of observed deviant behaviors:*

- “If he (they) is doing that, me too I can do so! It is not *fair* that he does that and I cannot!”
- “If he (they) is doing that it *means* (it is a *sign*) that it is permitted/possible: there is not a N or is no longer in force here”
- “If he (they) is doing that it *means* (it is a *sign*) that this is the right way; what we have to do (he expects that I do so)”.<sup>14</sup> Actually this is an intentional action entailing a violation, but not intentional *as* violation.
- “In fact he is right! He is courageous. It is correct to violate this N!” (Thanks to his violation behavior I change my value-attitude towards N; this goes in the direction of N criticism).

### 5.1 A Single Bad Example

The impact of an external, observable deviating behavior does not depend only from the *number* of violators: the many the violators the more impaired the N.

A single guy's deviant behavior can be sufficient for a large impact. It depends on the network, on the number of stakeholders and – of course – on his/her role and influence.

It also is important the fact that (a) *not all violations are equivalent, although behaviorally identical*; and (b) that sometimes a single deviating example (not a multitude) be enough for; but of course it depends on its visibility and significance and interpretation. The single violation of a leader is not the same of the one of a follower; the violation a well-known person is not like the violation of an anonymous person, and so on.

The number of violator is of course a relevant factor because one principle for the strength of our persuasion is *the number of converging sources or examples*. But also the single's reliability - as model or authority – and prestige has a precise impact on the degree of our persuasion.

---

<sup>14</sup> This case and the previous one change our “normative expectation” in Bicchieri and Xiao [9] terminology.

## 5.2 The Others (Deviant) Behaviors as Messages

Since minds are typically read off behavior “it is impossible not to communicate” about our minds even those prescribed by a specific role. Our behaviors or their traces inevitably “signify” our mental attitudes. And we use our everyday behavior or its traces (practical actions not “expressive” ones or conventionalized gestures) on purpose to send this information to others; for *signaling*. This is a special form of communication crucial for human social coordination, and conventions and institutions establishment via “tacit” negotiation and agreement, not to be mixed up with gestural or other forms of non-verbal communication [43].<sup>15</sup>

Also N maintenance or innovation “circles” (observation-interpretation-change-action-observation- and so on) (Sect. 6) works thanks to the fact that a cognitive agent “reads” the others’ conducts, and they signify/inform about the existence, respect, or violation of Ns [3]. Thus a violation conduct may acquire either the communicative function or the communicative intention of impairing the N or of explaining my reasons. Demolition or establishment of SocNs is mainly based on such a kind of not explicit communication, negotiation, and tacit agreements.

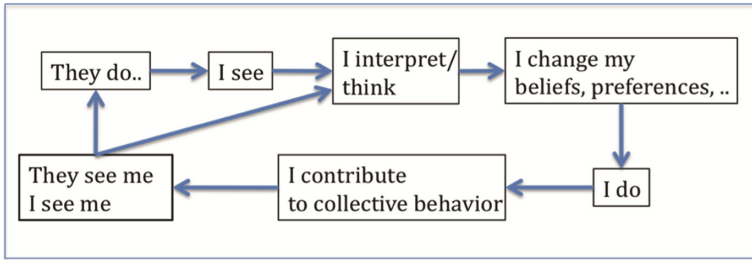
This factor contributes to the explanation of a crucial issue. As remarked by Christine Cuskley<sup>16</sup> “frequency and stability exhibit an interesting relationship in language: *the more frequent a linguistic construction is, the less it tends to change over time.*” In my view this might be generalized to behaviors, and in particular to normatively regulated behaviors. Also linguistic constructions are “norms” and “rules” for people aimed at using that language; just a sub-case (with its specific additional dynamics). “Despite the evident relationship between frequency and stability, it is still unclear what specific social and cognitive factors underlie this relationship.” As for social Ns, I would say that part of these factors is rather clear: the more diffused a (normative) behavior, the greater the probability to be observed and imitated/learned (a very strong and repeated “message”!), and thus *not just to spread around but to be “reinforced” in its prescriptive character*. Moreover, the more it is diffused the greater the absolute number of necessary “exceptions” and “violations” for its change or elimination. Thus the more widespread the more stable. And vice versa: the more stable in time and people, the greater the probability to be diffused and repeated (frequency). And so on.

## 6 Collective Destruction/Construction: Emergence-Immergence Cycles

On the basis of this analysis of internal mutations and their behavioral consequences, let’s focus on the description of the internal-external, mental-behavioral, individual-collective loops, and on the description of the phases of Ns change (vicious) ‘circles’ (Fig. 1).

<sup>15</sup> On the relevance of Norm-signaling, and of explicit communication, not just of punishment, see also [2, 3].

<sup>16</sup> Christine Cuskley “Frequency and stability in linguistic rule dynamics”, Invited seminar at ISTC October 2014.



**Fig. 1.** Internal-external cycle

### 6.1 External $\Leftrightarrow$ Internal Circles

Obviously – as for the “external” observed events (single or regular) – what matters is the Intentional Stance *interpretation*, the ascribed mind and reasons. I observed an individual violation by S or by W (not blame, no sanction); is it by accident, ignorance, or lack of attention? Or was it intentional? And “why”? Was S just egoist and self-maximizing, or is he violating because disagrees about the N or for invalidating the A? As we saw in Sect. 5 there are various possible interpretations and effects. And about norm ‘watchman’ role: was he indulgent because lazy or corrupted or familiar with S? Or was he thinking that N doesn’t apply in that circumstance or is bad and unfair?

The effect on my mind and on my view of the N in the various cases is very different. The external event impact depends on our subjective interpretation of it.

That’s why also a very clear collective behavioral regularity is not always and automatically interpreted (and complied) as a N. There are “vicious” and “virtuous” circles, from the point of view of normative behavior. Both, the vicious one (that is, violation, behavioral messages, N impairment, and collapse) and the virtuous one (N emergence, implicit negotiation, establishment, and maintenance) are due to the same internal-external cycle (Fig. 1).

There is also a very interesting self-referential feedback: the violating or conforming subject is observing his/her own behavior, and interpreting it, and confirming or changing his/her beliefs and preferences and feelings (as we saw in Sect. 4), and so on. Our behavior signifies a lot to us, and we send (intentional or unintentional) messages to ourselves. Also because, if I act on the basis of some implicit, presupposed, assumptions or choices, and the action is successful (good results), this automatically reinforces the presupposed mental conditions for that act, and increases the probability to take the same path next time.

## 6.2 The Crisis of N Authority

A nice example of a multilayer vicious circle between normative behavior and norm-related mental attitudes is the crisis and discredit of the “authority”. To work well authority requires not only respect/submission for authoritarian strength, threats, coercive power (credible sanctions), but “prestige” or more precisely “authoritativeness”. That is, A’s “credibility”. An A requires trust for its role; without trust it cannot work. Information authority, source of knowledge must be “credible” in strict sense: it has to be perceived (evaluated and felt) as “competent” in that domain and honest, not cheating for some private interest. Analogously the norm-A must be “credible” and trustworthy, its Ns should be perceived/given as the right one (from a technical and a justice point of view) and not due to private interests. If the A is authoritative, I accept its information or prescription, without need for prices or threats, without conflict, rebellion: I have a *generalized adoption disposition*; in a sense I obey for intrinsic motivations.

However this authoritativeness can collapse, and A can have a crisis of credibility, be discredited and no longer “automatically” respected. Which are changes in individual mind that might start (or reinforce) this process?

- (a) I no longer *believe* that A or its behavior is respectable, that A is authoritative, credible; thus
- (b) I do not adopt its prescription/N, I start do not conform to (*decision*);
- (c) this feedbacks, and reinforce my belief about violability of N and my right to violate, and - since my deviating behavior can be observed
- (d) it discredits the A in the others’ eyes; diffuses the same evaluation about A (and probably also its perceived capacity or right of sanctioning); it builds a “collective belief”<sup>17</sup>
- (e) it infects, diffuses deviating behaviors; but
- (f) this spreading of the evaluations and of the deviating behaviors confirms and reinforces my perception of A, of that N, and my behavior; and so on.

The collapse of A’s authoritativeness is a mental and behavioral, and internal and external, and individual and collective, fact.<sup>18</sup>

## 7 Concluding Remarks

Three issues.

- As we said, Ns are *based on* the possibility to be violated, not obeyed. They are devices for the control of “autonomous” agents that decide what to do on the basis of their beliefs, reasoning, and goals. Ns not only presuppose (accept) but also postulate a freedom in the addressees. Is this just a not so good but unavoidable feature? Or violability in this regulating device of social conduct has some advantages? N

<sup>17</sup> Not in the sense of a “collective mind” but in the more basic sense of a collective of minds; many minds sharing certain assumptions and infecting each other.

<sup>18</sup> It is clear that such an internal/external dynamics of Ns change might be fully simulated only with cognitive Agents in MAS.



“violation” usually has a negative connotation, since to “violate” is an evil in itself (as harm at a general and meta-level, of order, authority, trust; as we explained). However – actually – not only it can be morally justified and even noble and courageous, but also it plays a key function. It is one of the mechanisms and pressure for N change, adaptation, and evolution<sup>19</sup> [16].<sup>20</sup>

- I’m not sure that the current theory and definitions of social norms (see for example [6, 37]) fully captures some of the aspects we have discussed<sup>21</sup>. For example, there are social norms (not only legal ones) that are still there even if systematically violated by a large part of people. The norm *is still in force since it is perceived as such by that people*, although they violate it. They actually know/decide to “violate” it, thus, in a sense, that N still “regulates” their conduct. For example, in several part of Italy it is very frequent that people throw papers on the street or do not collect the excrements of his dog; however, they know (and even agree) that this is bad, not “correct” (N violation), but since it is tiring do not do so, and since a lot of people does the same ... Is that N “in force” in this group? Yes: everybody knows what one “should” do. In our view a social norm to be there doesn’t require to be a behavioral norm, a stable practice. It is *sufficient* that the large part of the group knows it, reminds and considers it, although regularly or frequently violating it. It is perceived *as a N, taken into account* in the individual cognitive process and mentally shared in the group, although ineffective on the conduct. It is a strange N state: an still in force but ineffective N. We shouldn’t forget that first of all a N is into the (shared) mind of the agents; this is its presupposition.

Of course it is fully true – coherently with Bicchieri’s theory – that:

- (i) On the one side the norm not only is ineffective but is probably in “decadence”, close to disappearing also from the mind of people, for example for the learning process of the new guy or for the mental automatization of the bad practice without no longer considering/perceiving that you are violating.

This is reasonably a possible and rather typical *intermediate step in the path of N extinction*: N respect and sanctioning; bad practices but the N is still considered as such; non longer taken into account as a N, no longer impinging on us.

<sup>19</sup> This obviously shouldn’t be an excuse for the selfish violator just for his own private interests (although – as Adam Smith has explained – even this guy plays his social function, beyond his personal motives).

<sup>20</sup> I worry about the rigorous computational (intelligent) coordination and surveillance on human work and organization. At least in “critical states” we need violations, although not foreseen in the program; but just opportunistic and reactive to a given contingency.

<sup>21</sup> For example, the motto of Bicchieri for synthesizing the spirit and working of social Ns “Do the right thing: But only if others do so” could create some misunderstanding. This might be the mental rule, the prescription that the individual gives to himself in front of a N (it can explain his conformity or violating behavior) but is not the prescription of the N: the N says, prescribes, just “Do the right thing!” Ns want to be obeyed and respected in any case; this is their imperative. I may decide or be leaning to respect this absolute imperative only “if”, under certain condition, but the “normative expectation” also by the others doesn’t say “only if the others do so”.

- (ii) On the other side, it is true that the fact that several guy systematically violate that N encourages ignoring it, to consider that it is possible and not so terrible to violate it. We live in a rude world and we adapt/belong to it.<sup>22</sup>

In a sense the norm is still there in the mind of the agents; they know that there is such a norm. However, they are no longer *committed* to respect it [28]; they do not formulate the intention to respect it. One might say that knowledge about the others' conformity to a norm is not only or necessarily the origin and basis of our *believing* that a norm is there (Bicchieri's theory), but is more the basis of our "*commitment*" (and its strength) to that duty.

- Agents are relevant in two ways: for modeling the complexity of such a dynamic and immergent/emergent process, by Agent-based Social Simulation; but also because we need non-passive normative and moral agents in Hybrid Societies where Artificial Intelligences (Agents, robots,) will work and cohabit with humans. In particular *N change* processes (internal and external) should be present in both MAS with cognitive Agents, and in Hybrid Societies. We have even to allow and exploit violations of rules and practices in organization, coordination, and work, but only when it is the case and by understanding "why" (reading behavior and mind) [17]. Actually there is a strong and advanced tradition in AgMAS on Agent architecture for Ns, in N based MAS and organization, in MAS simulation of Ns efficacy<sup>23</sup>, however – in my view – we still need some advancements in theoretical modeling of cognitive and collective aspects of Ns dynamics. This work is a partial attempt in this direction.

## References

1. Andrighetto, G., Governatori, G., Noriega, P., van der Torre, L.W.: Normative Multi-Agent Systems. Dagstuhl Follow-Ups 4, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik (2013). <http://dblp.uni-trier.de/db/series/dfu/index.html>
2. Andrighetto, G., Brandts, J., Conte, R., Sabater-Mir, J., Solaz, H., Villatoro, D.: Punish and voice: punishment enhances cooperation when combined with norm-signalling. PLOS (2013). <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0064941>
3. Andrighetto, G., Castelfranchi, C.: Norm compliance: The Prescriptive Power of Normative Actions. *Paradigmi* 2, 152–168 (2013)
4. Artikis, A.: Specifying norm-governed computational societies. ACM (2009)
5. Bendor, J., Swistak, P.: The evolution of norms. *Am. J. Sociol.* 106(6), 1493–1545 (2001)

<sup>22</sup> It is even possible that a meta-norm emerges: the idea that to conform to this N (for example, of politeness) by antequite, ridiculous, or snob, and this elicits negative attitudes in the others, that I want to avoid. A sort of *meta N of not conforming to the traditional N* is emerged; sometime even justified by new value (for example, "do not give precedence to women" as sign of women discrimination). In this case, for those people the previous N is no longer there, is no longer considered and accepted as a social N. The emergence or formulation of a *meta-N* about the violation (and then abandon) of a previous specific N is one of the processes of N abandoning and N innovation. It requires specific mental changes and contents; including a value-based justification of the "criticism" to the previous impinging N.

<sup>23</sup> See for example: [1], [40], [10], [26] [27], [37], [44], [29] [30].

6. Bicchieri, C.: *The Grammar of Society: Nature and Dynamics of Social Norms*. Cambridge University Press, New York (2006)
7. Bicchieri, C., Mercier, H.: Norm and beliefs: how change occurs. In: Edmonds, B. (ed.) *The Dynamic View of Norms*. Cambridge University Press, Cambridge (2009)
8. Bicchieri, C., Xiao, E.: Do the right thing: but only if others do so. *J. Behav. Decis. Making* **22**, 191–208 (2009)
9. Bicchieri, C., Xiao, E., Muldoon, R.: Trustworthiness is a social norm, but trusting is not. *Polit. Philos. Econ.* **10**, 170 (2011)
10. Boella, G., Noriega, P., Pigozzi, G.: HarkoVerhagen: introduction to the special issue on NorMAS 2009. *J. Log. Comput.* **23**(2), 307–308 (2013)
11. Boyd, R., Gintis, H., Bowles, S., Richerson, P.J.: The evolution of altruistic punishment. *PNAS* **100**(6), 3531–3535 (2003)
12. Brennan, G., Eriksson, L., Goodin, R.E., Southwood, N.: *Explaining Norms*. Oxford University Press, Oxford (2013)
13. Campennì, M., Andrighetto, G., Cecconi, F., Conte, R.: Normal = Normative? The role of intelligent agents in norm innovation. *Mind Soc.* **8**, 153–172 (2009)
14. Castelfranchi, C.: Through the minds of the agents. *J. Artif. Soc. Soc. Simul.* **1**(1) (1998). <http://www.soc.surrey.ac.uk/JASSS/1/1/contents.html>
15. Castelfranchi, C.: Prescribed mental attitudes in goal-adoption and norm-adoption. *AI Law Spec. Issue Norms MAS* **7**(1999), 37–50 (1999)
16. Castelfranchi, C.: The theory of social functions. Challenges for multi-agent-based social simulation and multi-agent learning. *J. Cogn. Syst. Res.* **2**, 5–38 (2001). Elsevier. <http://www.cogsci.rpi.edu/~rsun/si-mal/article1.pdf>
17. Castelfranchi, C.: Formalising the informal? *J. Appl. Log.* N° 1 (2004)
18. Castelfranchi, C.: Cognitivizing “Norms”. Norm internalization and processing. In: Faro, S., Lettieri, N. (eds.) *Informatica e Diritto*, Special Issue in “Law and Computational Social Science”, vol. XXII, no. 1, pp. 75–98 (2013)
19. Castelfranchi, C., Tummolini, L.: Positive and negative expectations and the deontic nature of social conventions. In: *Proceedings of the 9th International Conference of Artificial Intelligence and Law (ICAIL 2003)*, pp. 119–125. ACM Press (2003)
20. Cialdini, R.: Descriptive social norms as underappreciated sources of social control. *Psychometrika* **72**(2), 263–268 (2007)
21. Conte, R., Andrighetto, G., Campennì, M., Paolucci, M.: Emergent and immergent effect in complex social systems. In: *Proceedings of AAI Symposium, Social and Organizational Aspects of Intelligence*, Washington (2007)
22. Conte, R., Castelfranchi, C.: *Cognitive and Social Action*. UCL Press, London (1995)
23. Conte, R., Castelfranchi, C.: From conventions to prescriptions. Towards a unified view of norms. *Artif. Intell. Law* **3**, 323–340 (1999)
24. Conte, R., Castelfranchi, C.: The mental path of norms. *Ratio Juris* **19**(4), 501–517 (2006)
25. Conte, R., Castelfranchi, C., Dignum, F.: Autonomous norm acceptance. In: Mueller, J. (ed.) *Proceedings of the 5th International workshop on Agent Theories Architectures and Languages*, Paris, 4–7 July 1999
26. Criado, N., Argente, E., Noriega, P., Botti, V.J.: MaNEA: a distributed architecture for enforcing norms in open MAS. *Eng. Appl. AI* **26**(1), 76–95 (2013)
27. Criado, N., Argente, E., Noriega, P., Botti, V.J.: Corrigendum to ‘Human-inspired model for norm compliance decision making’. *Inf. Sci.* **245**, 218–239 (2013). *Inf. Sci.* **258**, 217 (2014)
28. Dastani, M., van der Torre, L., Yorke-Smith, N.: Commitments and interaction norms in organization. In: *Autonomous Agents and Multi-Agent Systems*, pp. 1–43 (2015)
29. Dignum, F.: Autonomous agents with norms. *Artif. Intell. Law* **7**(1), 69–79 (1999)

30. Dignum, F., Dignum, V.: Emergence and enforcement of social behavior. In: Anderssen, R.S., Braddock, R.D., Newham, L.T.H. (eds.) 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, July 2009, pp. 2377–2383. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation (2009)
31. Fehr, E., Fischbacher, U., Gächter, S.: Strong reciprocity, human cooperation, and the enforcement of social norms. *Hum. Nat.* **13**, 1–25 (2002)
32. Galoob, S., Hill, A.: Norms, attitudes, and compliance. *Tulsa Law Review* (2015, Forthcoming)
33. Garfinkel, H.: A conception of, and experiments with, ‘trust’ as a condition of stable concerted actions. In: Harvey, O.J. (ed.) *Motivation and Social Interaction*, pp. 187–238. The Ronald Press, New York (1963)
34. Gneezy, U., Rustichini, A.: A fine is a price. *J. Leg. Stud.* **29**(1) (2000). [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=180117#%23](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=180117#%23)
35. Hammoud, M., Tang, A.Y.C., Ahmad, A.: A norms decay framework in open normative multi-agent systems. *Br. J. Appl. Sci. Technol.* **12**(4), 1–15 (2016)
36. Heckathorn, D.: Collective sanctions and the compliance norms a formal theory of group-mediated social-control. *Am. J. Soc.* **94**, 535–562 (1988)
37. Hechter, M., Opp, K.-D. (eds.): *Social Norms*. Russell Sage Foundation, New York (2001)
38. Hübner, J.F., Matson, E., Boissier, O., Dignum, V. (eds.): *COIN@AAMAS 2008*. LNCS, vol. 5428. Springer, Heidelberg (2009)
39. Kahneman, D., Miller, D.T.: Norm theory: comparing reality to its alternatives. *Psychol. Rev.* **80**, 136–153 (1986)
40. Miceli, M., Castelfranchi, C.: A cognitive approach to values. *J. Theor. Soc. Behav.* **19**(2), 169–193 (1989)
41. Noriega, P., Chopra, A.K., Fornara, N., Cardoso, H.L., Singh, M.P.: Regulated MAS: social perspective. *Normative Multi-Agent Syst.* **2013**, 93–133 (2013)
42. Schultz, P.W., Nolan, J.M., Cialdini, R.B., Goldstein, N.J., Griskevicius, V.: The constructive, destructive, and reconstructive power of social norms. *Psychol. Sci.* **18**(5), 429–434 (2007)
43. Tummolini, L., Castelfranchi, C.: Trace signals: the meanings of stigmergy. In: Weyns, D., Van Dyke Parunak, H., Michel, F. (eds.) *E4MAS 2006*. LNCS (LNAI), vol. 4389, pp. 141–156. Springer, Heidelberg (2007)
44. Ullmann-Margalit, E.: *The Emergence of Norms*. Oxford University Press, Oxford (1977)
45. Vasconcelos, W., García-Camino, A., Gaertner, D., Rodríguez-Aguilar, J.A., Noriega, P.: Distributed norm management for multi-agent systems. *Expert Syst. Appl.* **39**(5), 5990–5999 (2012)

Coordination, Organizations, Institutions, and Norms in  
Agent Systems XI

COIN 2015 International Workshops, COIN@AAMAS,  
Istanbul, Turkey, May 4, 2015, COIN@IJCAI, Buenos  
Aires, Argentina, July 26, 2015, Revised Selected  
Papers

Dignum, V.; Noriega, P.; Sensoy, M.; Sichman, J.S. (Eds.)

2016, XII, 437 p. 119 illus., Softcover

ISBN: 978-3-319-42690-7