(Social) Norm Dynamics

Giulia Andrighetto[1], Cristiano Castelfranchi[2], Eunate Mayor[3], John McBreen[4], Maite Lopez-Sanchez[5], and Simon Parsons[6]

1 **Institute of Cognitive Science and Technologies, ISTC-CNR**
   Rome, Italy and European University Institute, EUI, Florence, Italy
   `giulia.andrighetto@istc.cnr.it`
2 **Institute of Cognitive Science and Technologies, ISTC-CNR**
   Rome, Italy
   `cristiano.castelfranchi@istc.cnr.it`
3 **LMTG/GET UMR5563, IRD-CNRS-Universite P. Sabatier Toulouse III**
   Toulouse, F-31400, France
   `eunate.mayor@gmail.com`
4 **Wageningen University**
   Wageningen, The Netherlands
   `johnmcbreen@gmail.com`
5 **University of Barcelona**
   Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain
   `maite_lopez@ub.edu`
6 **Brooklyn College, City University of New York**
   2900 Bedford Avenue, Brooklyn, NY 11210, USA
   `parsons@sci.brooklyn.cuny.edu`

Abstract

This chapter is concerned with the *dynamics* of social norms. In particular the chapter concentrates on the lifecycle that social norms go through, focusing on the generation of norms, the way that norms spread and stabilize, and finally evolve. We also discuss the cognitive mechanisms behind norm compliance, the role of culture in norm dynamics, and the way that trust affects norm dynamics.

1 Introduction

This chapter aims to identify the major aspects of norm dynamics, by which we mean the way that norms come into being and change through their life, as well as some of the relevant factors or determinants of the process that underlies this change. The need for a deep understanding of these dynamics is becoming a compelling task for the Normative Multi-Agent Systems (NorMAS) community because of the systems that the community wants to develop. Now, the members of the NorMAS community have a wide range of interests with respect to multi-agent systems so it behooves us to explain, before we get much further, what perspective the authors take on the dynamics of norms. We focus here on two view of multi-agent systems, views that we distinguish by referring to them as the 'engineering' perspective and the 'sociological' perspective.

From the *engineering* perspective, we are interested in building multi-agent systems that are flexible and open. In other words, rather than rigid systems in which all possible behaviors are determined at design-time, we are interested in developing systems that are open—in the sense that agents can enter or leave the system at any time—and systems that are able to evolve over time. This evolution will include the evolution of the mechanisms that govern the system. This desire for flexibility points to the use of norms to regulate the behavior of these systems, and so we are interested in the properties of norms in human societies, seeing this as a rich source of evidence for the development of norms in multi-agent systems. From the *sociological* perspective, we are interested in understanding human societies, in particular the mechanisms that allow these systems to self-regulate, developing and modifying rules of behavior. From this perspective, multi-agent systems are a useful research tool, one that allows models of societies to be constructed and experimented with in a way that is not possible with human subjects.

Of course, these two views are neither exhaustive (there are other perspectives within the NorMAS community) nor are they exclusive (research can contribute to both perspectives), but the reader should bear both these perspectives in mind when reading this chapter.

Now, we also need to stress that *norm* is a polysemic term with possible multiple meaning, and is used in a number of different ways. Let us distinguish three basic and fundamental meanings of the term 'norm' that we use in this chapter.

1. Norm meant as *normal* behavior or feature. In this sense 'norm' is understood as a reference to the normal (or Gaussian) distribution and we use the term in a statistical sense. What is normal? What corresponds to the norm in this sense? That which is around (standard deviation) the median value (mean). More extreme, and thus rare, values (behaviors, features) that are not *regular* are considered as *deviant*. This is just a statistical and descriptive notion of norm. It does not necessarily have a prescriptive nature nor it is assumed as ideal and model. Of course, usually very deviant features are not assumed as ideal, good. For example very small feet or very big ones. However, for example for intelligence an abnormal measure can be assumed as an ideal model, the best.

2. Norm meant as *model* or *standard, ideal-typical*. An example is provided by the normative theory of rationality in economics that indicates what would be the perfect rational (economically efficient) reasoning and decision. 'Normative' in this sense does not mean *prescriptive*; it is not an imperative or obligation to reason in such a way. It just is a useful parameter, a model, for comparing different decision processes and evaluate them. It is well known that human beings do not reason and decide in such a way [1] (consider for example, Herbert Simon's *bounded rationality*; cognitive heuristics and biases; affective intelligence, etc.). Also in other domains, it is clear that the ideal-typical characterization is not statistically dominant; for example only 15% of population has eyesight as ideally described in the textbooks of ophthalmology; but it is a precious comparative standard for evaluating the various deviations from it.

3. Finally, norm meant as an (explicit or implicit) imperative regarding the behavior of social actors, i.e. specifying how they should behave. Norms that are explicit are often those formal norms introduced by organizations or societies through some kind of legislative or regulatory process. Such norms are often handed down to societies of agents from above. Other norms, and this is what we mean here by the term *social norm*, are those that are

---

[1] Though, of course, mainstream economics continues to pretend that they do.
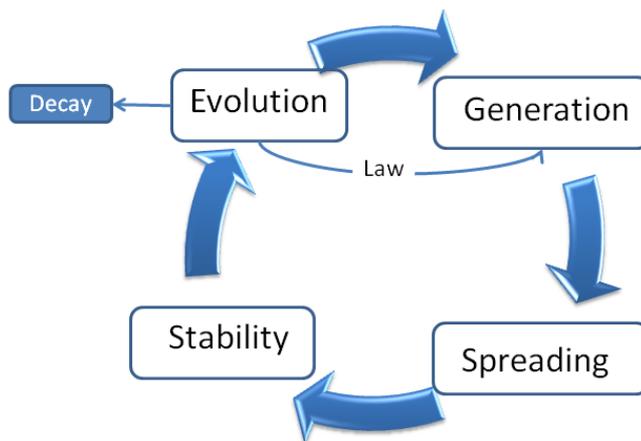
■ **Figure 1** The lifecycle for norms.

implicit and emerge from some form of evolution and self-organization by the agents that are regulated by them.

We view the lifecycle for norms as following the cycle in Figure 1, and this cycle forms the basis for the discussion in this chapter. The cycle starts at the top right. Norms are generated, and after generation may spread through a population and become widely adopted. If the norm spreads, it will eventually reach stability, having been adopted by some portion of the population and failing to grow any further within the population (though, of course, it may stabilize at a point where it has been adopted by the whole population). With stability norms will persist. Some will then decay, being superceded by new norms, some will evolve and some may be codified into law. The norms that evolve or become law can be considered as generating new candidate norms that will then spread or not.

This cycle is not identical for every type of norms. Some norms, such as social norms, emerge spontaneously from self-organizing groups of agents, and are usually implicitly (but sometimes explicitly) *negotiated* by the agents. Other norms, for example the legal and the organizational ones, are *decided* by some institution (like the parliament or the boss) and are officially *proclaimed* by an authority[2]. So the process of norm generation and the process of spreading are not identical for all kinds of norms. The spreading of a custom or social norm is based on behavioral messages (signals) and on conformity, imitation, etc.; while the spreading of a given official instruction or rule is realized through other specific channels, e.g., publication, issuing, etc. We should also note that norm lifecycle as we have sketched it mainly applies to the second and third class of norms we distinguished, but can also be seen to apply to statistical norms.

The consumption of tobacco illustrates a number of aspects of the lifecycle and it well applies to norms of different kinds. Tobacco consumption in Europe (and because of the availability of data we will consider consumption by women) through the last hundred years or so can be considered as a statistical norm, a social norm, and has ended up regulated as a legal norm (in the form of a prohibition). In Europe [43], it seems that smoking was not widespread among women until the twentieth century. It subsequently became a norm that spread, with close to 50% of the population in the UK, Denmark and the Netherlands

---

[2] This is the original etymology of *lex*, that is, officially declared.

smoking by 1970. In the UK this number then stabilized, with somewhere between 40% and 50% of women smoking by 1990 (when the data in [43] ends). At the same time, the smoking norm for men was being discarded — in 1950 70–90% of the male population in Northern Europe was smoking, and this fell to 30-40% by 1990. Looking more closely (again using data from [43]), it seems that among smokers, what was consumed changed in the early years of the twentieth century. Driven by the tobacco industry, consumption among (predominately male) smokers shifted from dark tobacco—cured in traditional ways and consumed in pipes, cigars and hand-rolled cigarettes—to manufactured cigarettes that used a lighter tobacco. In the lifecycle, this can be seen as the evolution of a norm, with the consumption of dark tobacco evolving into the consumption of blond tobacco, and the means of consumption evolving also. The new norm stabilized, and the old norm decayed.

Given the data in [43], what we are primarily talking about here is a statistical norm—the number of people smoking from which we can infer what 'typical' behavior would be. However, given that smoking is an activity under the control of the agents that indulge in it (unlike, for example, their height), it is not much of a stretch to imagine that the statistical norm is accompanied by a social norm, that is the figures we quote above are a reflection of how socially acceptable it was for European women to smoke.

Tobacco consumption also illustrates how laws can lead to new norms. Various countries have brought in smoking bans (i.e., prescriptive norms of the third kind) in the last ten years or so. To some extent these bans have had the desired effect of helping the smoking norm to be discarded by some members of the population. However, they have also led to a new norm—that of people standing outside a bar or restaurant in order to smoke while still occupying space inside the building (leaving food and drinks at the table).

Having briefly sketched the norm lifecycle in this section, we will examine it in more detail in the rest of this chapter. The principal steps of the dynamics of norms—generation, spreading, stabilization and evolution—are discussed in some detail in Sections 3.1, 3.2, 3.3. We suggest that for a well-founded and innovative study of norms, it is necessary to look at the cognitive mechanisms underlying the dynamics of norms, and we describe these in Section 2. In addition, we believe that it is also important to consider the role played by trust (see Section 4) and cultural variation (see Section 5).

2   The Minds of the Agents

If one aims to understand how social norms are able to regulate the conduct of intelligent and autonomous agents, it is necessary to analyse the cognitive mediators that make it possible to transform normative requests into actions [29, 30, 31]. This section aims to present a model of the cognitive mechanisms underlying norm compliance.

With *social norms*, we refer here to informal rules which prescribe what we ought or ought not to do. Social norms are behaviours spreading over a society on condition that the corresponding prescriptions and mental representations (namely, sets of beliefs and goals concerning the norm) spread as well. We will then refer to the third definition of norms provided in the introduction, the one that focuses on their *prescriptive* nature.

Norms are influencing mechanisms, aimed to direct the future conduct of the individuals subject to the norm. To influence the behaviour of autonomous systems, a complex mental dynamics must be activated. In order for autonomous agents to undertake (or refrain from undertaking) a certain course of action, it is not sufficient that they know (i.e., have the belief) that such a course of action is desired by someone [30]. It is necessary for them also to have the goal of performing such an action. A goal is here meant in the very general sense derived

from cybernetics, i.e., a desired state of the world triggering and driving actions [30, 68]; for an analysis of the differences between goals and other motivational states, see [23].

Since social norms are intended to guide the behaviour of agents, they are a powerful means for the generation of new goals in the minds of the individuals subject to the norm. The repeated exposure to the normative conduct of others and to their normative requests and expectations (be they implicit or explicit) generates in the minds of individuals the belief that there is a norm and there is a widespread will which prescribes that he or she should (is asked to) observe that impersonal and widespread will (normative expectation). This normative knowledge has then the power to activate or to generate the goal to observe the norm in question.

But what is the mental path or process through which the normative requests and expectations generate these goals, goals which are, in themselves, normative as well?

In order to understand the motivational power of normative requests and expectations, it is useful to sketch a cognitive anatomy of the latter type of representations. As suggested by [67], an expectation is a hybrid mental representation which consists of an epistemic component and a motivational one. Expectations are different from simple predictions (epistemic representations about the future), in that the individual not only has the belief (a prediction with a certain degree of certainty) that a certain event or state will occur, but also has the goal (with a specific value of importance) that it will (or will not) occur, and, to this end, actively monitors the process.

Expectations are not just beliefs, they imply a personal concern about the anticipated event. Having an expectation means that the agent is waiting for something (not) to happen. There is a goal involved, that is why there are positive and negative expectations. Specifically, a normative expectation simultaneously consists of (a) the belief about the existence of the norm and (b) the goal that the norm is (not) complied with (by those who are subject to it).

To reinforce the probability that a normative expectation will be fulfilled, the holder must act in such a way as to influence others' behaviour and possibly even their minds. They should provide the addressee with the relevant information, thus he must communicate the existence of the norm and that there is a (widespread and impersonal) will that it should be observed, and, in the opposite case, that there is possibility of punishment or sanction. This prescriptive nature is a crucial aspect of the normative mechanism. The normative expectation contains, in itself, a request, which individuals must decide whether to adopt or not, and in order to do this, they must recognize the normative goal which it transmits.

How is it possible that the goal (need, desire, objective, request, order, ...) of somebody else succeeds in regulating behavior of an autonomous agent? Autonomous agents have the ability to choose whether or not to have a given goal and this choice is conditioned to the belief that such a goal is a means by which they can achieve other goals that they already possess. An autonomous agent acts to achieve her own goals and must have reasons for choosing whether to act as she does. In particular, if she accepts another's request, she must have good reasons for doing so. The general mechanism by which an autonomous agent adopts external requests, called goal adoption, has been described at some length in [30, 68] . Here, suffice it to say that an agent (the adopter) will adopt another agent's (i.e., the adoptee's) goal as hers, on condition that she, the adopter, comes to believe that the achievement of the adoptee's goal will increase the chances that she will in turn achieve one of her previous goals. For example, I will accept your request to lend you my laptop, if in return you allow me to borrow your car tonight. When the external request is a prescription, a special application of this process occurs, i.e., norm adoption. I will adopt the norm if, say, I think that by doing so I avoid getting a fine, obtain others' approval, build a good

reputation, etc. General adoption leads to the formation of social goals (achieve somebody else's goals). Norm adoption leads to the formation of normative goals (comply with the norms), i.e., a goal that an agent happens to have because and to the extent she has the correspondent normative belief.

An agent can decide to adopt a norm, and form a normative goal for several reasons (for a detailed analysis, [30, 68]):

- Instrumental reason: the subject adopts the normative goal if she believes she can get something in return (avoid punishment, obtain approval, praise, reward, etc.).
- Cooperative reason: the subject adopts the normative goal to achieve a common goal. Norm-adoption is cooperative when it is value-driven, that is, when the subject shares both the end of the norm and the belief that the norm achieves it. For example, an agent may decide to conform to the refuse-recycling norm because she believes that, by doing so, she helps to reduce the negative impact of not doing so on the environment.
- Terminal reason: The subject wants to observe the whole set of norms she is subject to, as ends in themselves. She has the terminal goal or value that "norms be respected" (Kantian morality).

It is interesting to notice, that normative goals can be formed for different reasons, also for self-regarding reasons, as in instrumental norm adoption. However, this does not mean that the generated goal is not normative in its nature, it is. A normative goal is a goal relativised to a normative belief, i.e., held because and to the extent that it is believed to be exacted by a norm. For a goal to be normative, all that is necessary is that it is based upon internal representations of a normative nature [27].

In this section, we have described the standard path of norms in agents' minds. However, this path is rarely followed in its totality and it is more plausible to consider that shortcuts take place during the computation of the normative input. Norm conformity in humans is usually due to the automatization and simplification of a such a rich cognitive decision process. After a repeated normative learning and especially for simple behavioral rules (like stopping at the red light) the subject just reacts automatically to the normative stimulus. For example, when you are driving and you see a red traffic light you will automatically stop. In this situation, it is not necessary that input follows the complete mental path.

The normative beliefs and reasoning remain inactive, but they remain present in the subject's mind and they will be re-activated if needed. For example, a given routine must be blocked when given conditions activate inconsistent prescriptions. A car driver stopping at a red traffic light might see a policeman asking her to move on. In such a case, the car driver needs to be able to retrieve control of her action, block the automatism, and decide which normative input should be given priority. Here the norm is explicitly taken into account and we reason about violating it or not [5, 29].

3 The Lifecycle of Social Norms

This section goes more deeply into the *Norm Lifecycle* introduced in Section 1. Lifecycle models of norms have been also proposed by [85] and [92] (see also the chapters "Normas in MAS: Definitions and Related Concepts" and "Simulation and NorMAS" in this volume). [85] takes a simulation perspective, while [92] examines norm dynamics from a lower-level operational point of view, i.e., in terms of its key states and transitions. The lifecycle model we present in this chapter is not identical for all kinds of norms, we believe that it is widely applicable. As already described, the lifecycle starts with a *generation* phase. Generated norms may then *spread* through the population and eventually reach *stability*

(i.e., persistence) if they are adopted by enough individuals in the population. Afterwards, different evolution stages may occur: norms may *decay*, being superceded by new norms; norms may evolve; or alternatively, they may be *codified into law*. We discuss how the last two cases close the lifecycle.

3.1   Norm Generation

The life cycle in Figure 1 starts with the norm generation process. This process may be triggered when a brand new organisation lacks norms to structure individual interactions or when an existing organisation requires the addition of new norms. The intrinsically dynamic nature of Internet and its commonalities with organisational MAS approaches offer many opportunities for the emergence of organizations with specific objectives. Some examples can be found in economic coalitions, working teams, social communities, etc. Defining the rules of the game for brand new organizations may not be a straightforward process, so we claim that the automatic generation of organisational rules will become an important topic in the Normative MAS area if we understand the term "norm" in its broad sense of social conventions.

Norm generation can be tackled from different approaches. In this section we provide a brief overview of the main approaches that have been taken so far by the research community, their assumptions and the open issues that still require further consideration.

Formal approaches to norm generation have been generally referred to as *norm synthesis*. These approaches exhaustively enumerate the full state space in order to define norms that ensure access to goal states whilst disallowing access to undesirable states in the state space. Therefore, exhaustive formal approaches explicitly define at design time which actions are allowed and which actions are forbidden within a specific scenario. By definition, their results are proven to be complete. Nevertheless, the intrinsic complexity of these formal methods causes them to be intractable for real world systems. In fact, Shoham and Tennenholtz [90] have shown this complexity to be NP-Complete by performing a reduction to 3-SAT. Furthermore, a typical assumption of these approaches is that the system will be static (or at least that the system dynamics will be known and included at design time in the state space being explored).

An attempt to reduce complexity is the work by Christelis and Rovatsos [25], which proposes an automated method for synthesising prohibitive norms in planning-based domains at design time. This method includes norm generalisation and performs a local search around the state space, disallowing access to undesirable states but ensuring accessibility to goal states. Despite its improvements, this approach remains intractable for real world systems. Regarding system dynamics, it is also important to consider the agents that participate in the organisation, since it may be the case that agents decide not to comply with the generated norms. Ågotnes et al. [1] have formally (logically) studied the extent to which a normative system is robust, i.e., its effectiveness when populated by non-compliant agents. Still at design time, they investigate different robustness notions such as identifying those key agents that are necessary and/or sufficient to guarantee the effectiveness of a normative system or the proportion of an agent population that must comply in order to ensure success.

On the other hand, some research approaches define systems that require participant agents to be involved in *agreement* processes on the norms to follow. These systems constitute a democratic mechanism that reaches agreements (i.e., converge) eventually. Nevertheless, they may require participating agents to be enriched with abilities that go beyond its social basic capabilities (that is, those required for performing specific tasks or for participating in the community to obtain an individual or common goal). An example is the work of

Artikis [6], where agents are able to update norms by using an argumentation protocol to support discussion. Furthermore, they can even discuss about the argumentation protocol itself in order to update it. Design-time requirements for these systems are thus related to the argumentative capabilities of the agents, the specification of argumentation protocols to follow, and the knowledge agents need in order to be able to successfully complete the generation process. Nevertheless actual norm generation occurs at run-time, and therefore, the time and effort (in terms of reasoning resources) required for proposing and agreeing upon a new norm needs to be taken into consideration. Generation time may also be influenced by specific protocol parameters such as the degree of consensus required for a norm to be generated.

Nowadays, *norm* (or convention) *emergence* provides the approach to norm generation that is attracting most attention from the research community. This mechanism does not require any oversight or centralized control and thus it is defined in settings where agents lack of an explicit organisation. Norm emergence is based on the autonomy of agents for choosing individually a solution (i.e., a specific behaviour convention) from a space of alternative solutions. Afterwards, they consider a convention has emerged when a majority of agents actually choose the same actions (that is, the society converges towards a convention when agents perform the same actions). For example, in a traffic scenario, agents may decide whether to drive on the left or on the right, and we will say that a norm has emerged when most agents drive on the same side of the street.

Agents are generally considered to be self-interested, like in [44], where Griffiths et al. study how to establish a suitable set of norms in a decentralized population and the problem of group recognition by using observable tags as markings, traits or social cues attached to individuals. These approaches usually consider a number of repeated interactions involving pairs of agents (repeated two-player games). Differences lie in the assumptions about agent interaction: some consider general interaction patterns such as uniform random one-on-one interaction probabilities [89, 91] whereas others study the impact of societal topology that constrains the interactions of individuals [33, 60, 99]. This norm generation process is therefore intermingled with the subsequent steps of spreading and stability in the norm life cycle. Since these steps are further detailed in the remaining of this section, rather than covering all approaches exhaustively, we will just mention that it is also possible to consider some kind of observation as a requirement or instrument to adopt norms [33, 81, 99] and that internalisation is also proposed as a mechanism to guarantee norm acquisition [28].

Overall, it is worth mentioning that norm emergence approaches include some parameters that may be decided at design time. For example, one that is common to all of them is the threshold of compliance in the society that will be used to consider a norm to have emerged. Additionally, although most works on norm emergence are empirical, there are exceptions such as the one by Brooks et al. [18] that presents a mathematical model of the emergence of norms based on utilities. Specifically, they model the emergence of norms in societies of agents who adapt their likelihood of choosing one from a finite set of options based on their experience from repeated one-on-one interactions with other members in the society. Their goal is to study both the process of emergence of norms as well as to predict the likely final convention that will emerge if agents had preconceived biases for certain options. All these approaches to norm emergence pose very interesting research questions, nevertheless, it is important to notice: i) how much they rely on the fact that the initial set of alternative norms have to be known (by the agents) at design time; as well as that ii) convergence often depends on initial conditions. From our point of view, the research community can transform these limitations into research opportunities when trying to overcome them. As for any other

research area, the proposed approaches should be as general (and domain-independent) as possible. And this should apply also to the empirical approaches by taking advantage of those general MAS characteristics that are common to most systems.

An alternative approach to norm generation is that from Morales et al. [73] which learns norms based on run-time experience. This solution follows a 'division of concerns' paradigm where the majority of agents participating in the organization can act by simply conducting their domain specific activities without enduring the process of establishing new norms, which is left for specialized regulatory agents. These regulatory agents are staff (i.e., organisational) agents devoted to: first, proposing (a set of) rules for the system; and second, to monitoring to what extent these rules are effective in regulating the organization. This empowerment distribution is not meant to generate a centralized totalitarian system though. Instead, their aim is that these regulatory agents propose norms that smooth organizational activity and agent interactions. This is done by detecting when conflicts arise and by including new norms with the aim of preventing those conflicts from happening in the future. Once these regulations have been established to the system, these regulatory agents will be also in charge of evaluating their adequacy based on the organizational run-time experience. This is measured in terms of agent compliance and the resulting conflicts: agents may decide whether to observe norms or not and these decisions may have different consequences in the system. Therefore, the system does not focus on an enforcement mechanism that invalidates agent actions whenever they do not conform with the established rules. On the contrary, regulatory agents observe the agents and "listen" to them, since agents may not comply with norms if they consider they are not necessary (see 3.2.2 for a more detailed discussion on violation). The underlying assumption here is that it is possible to identify an unnecessary norm whenever agents violate it and no bad consequences (i.e., conflicts) arise. In fact, if this is the case in a number of experiences along the system execution, regulatory agents can then consider they have gathered enough evidence against the norm so that they can discard the norm from the set of currently established norms.

Overall, it is worth noticing that this empirical approach does not explore the complete state space but that it is flexible enough to deal with system dynamics. It can be seen as a non-intrusive, autonomy preserving, norm generation mechanism. Moreover, the amount of knowledge it requires at design time is rather limited. The basic requisites for the regulatory agents are conflict identification, and monitoring of agents' norm compliances and violations. Nevertheless, some parameters still need to be defined at design time. Some examples are the quality threshold that is required for a norm not to be discarded or the amount of evidence that each norm violation/compliance provides. Convergence time in this case depends on the conflict frequency.

As in Savarimuthu et al. [84], previously introduced approaches for including norms in a multi-agent system can be classified into two different categories. The first one is the top-down approach, where an institutional mechanism specifies (prescribes) norms that regulate agents' behaviour. Formal approaches in which norms are specified at design time could also be included in this category. The second one is the bottom-up approach, where agents locally interact in order to spread the adoption of suitable norms within the society. Both emergence and agreement approaches fit in this category. Nevertheless, a balanced mixture between top-down and bottom-up approaches may lead to the best results. This is the reason Morales et al. [73] advocate for methods where norms are proposed by regulatory (top) agents and validated against the experience of domain (bottom) agents.

3.2   Norm Spreading and stabilization

Once norms have been generated, transmission is the process by which norms spread from one agent to another. As shown by a great number of works on the classification of norms [101, 79], there is no single modality through which a norm can be transmitted. A norm can, for example, be communicated through explicit commands, orders, or requests, both written and oral, "do this", "don't do that", or by means of declarations which mention or imply deontic predicates, such as "x must/must not do y", or "you must/must not do y". Evaluations are also powerful means through which norms are transmitted, in phrases such as "paying your taxes is right/tax evasion is wrong", an evaluation relative to the state of things which is derived from the execution (or not) of normative action is expressed (for a more detailed analysis of the communication of norms, see [27]). [54] refers to this type of transmission techniques as *active transmission.* "Active transmission occurs when one agent purposefully broadcasts a set of norms to neighboring agents" [54]. The authors suggest that this type of transmission is usually accompanied by social enforcement, such as social sanctions aimed to deter others from violation. In multi-agent systems, examples of the use of active transmission is the use of norm entrepreneurs [51] and that of sanction [3, 20, 61, 63, 98]. Over time this process favours the diffusion of norms throughout the entire group.

3.2.1   Implicit normative communication

In this work, we aim to stress how the spreading of norms does not occur exclusively through active transmission and how a privileged role in this process of norm spreading is played by communication realised through actions or behavioral implicit communication ([54] refers to this process as *passive transmission*). With Behavioral Implicit Communication (BIC), we refer to a specific type of communication in which there is no specialized signal (neither arbitrary acoustic signal nor codified gesture), but the message is conveyed through a practical action. In BIC, the subject performs a usual practical action (e.g., walking, drinking, etc.), knowing that somebody is observing her and is able to understand the behavior (or the result of the behavior) she performed, and this is one of the goal for performing the practical action (for a more complete taxonomy of the many forms of implicit behavioral communication, [24]). In other words, "X performs the behavior b in order Y perceives it and on such a basis believes that p". Consider, for example, a friend who, during a dinner party at her own house, places the ashtrays only on the balcony. By this action, she is communicating to the guests that they can only smoke outside and that smoking is not permitted inside her apartment. This form of communication is *behavioral* because it exploits usual practical behavior that is not conventionalized (e.g., an arbitrary acoustic signal or a codified gesture); and it is *implicit* because, not being codified and specialized, its communicative character is unmarked, undisclosed, non-manifest and thus deniable. Behavioral implicit communication plays a key role in the spreading of norms. In particular, in the next subsection, we will show how the acts of obeying, violating and defending norms in terms of BIC provide interesting insights for understanding the dynamics of norm spreading.

3.2.2   Obedience and Violation

The action of obeying a norm is an act which can convey important information. It can convey, for example, to whoever observes (and monitors), the information that the conduct prescribed by the norm has been performed. From this information, the observer can infer new details regarding the normative actor, such as that the actor is an obedient individual

and is therefore trust-worthy. When individuals realise the expressive (i.e., demostrative) power of their own actions, they can intentionally decide to perform those actions in order to influence the minds of others in a normative way. They can, for example, decide to observe a norm (also) in order to communicate the fact that they have obeyed it, thereby avoiding being punished and in the hope that they will be considered as obedient and trust-worthy citizens. In the same way, that they can decide to observe the norm simply to *set an example* and thereby influence the others to do the same. As the action of observing a social norm involves a cost, whoever obeys does not want to be the only one to uphold it, and wants the normative costs to be distributed equally [27, 30]. Thus, behavioral implicit (normative) communication simultaneously has both an informative and a prescriptive nature. On the one hand, it transmits information about (1) who performed the normative act; (2) the existence of the norm; and finally (3) the fact that the norm should be respected. On the other hand, the behavioral implicit communication goes beyond mere informing and also prescribes the correct conduct to follow, asking the addressees themselves to comply with the same norm and indicating the consequences if not. Through behavioral implicit communication normative expectations and requests are transmitted and spread within the population.

The violation of a norm can also be a communicative act. For example, the desire to communicate to others (especially to those in authority, be their parents or the state) that one has violated a norm is characteristic of the provocative behavior of adolescents, and of revolutionary movements. One recalls Gandhi publically burning his South African pass[3], a document which all Indians had to have in their possession at all times, but which no white South African did. On this occasion, Gandhi communicated by means of a gesture, not with words, his indignation at the injustice, racism and exploitation to which Indians were subjected.

A norm can be perceived by an agent as more or less salient. With salience, we refer to the perceived degree of importance and strength of a norm [5, 13, 26, 61, 98, 102]. The actions of others, e.g., their compliance or violation of a norm, are important cues from which an individual can infer how salient a norm is. This is particularly true when no punishment follows the violation. Even if it is not the actor's intention, the act of violation signals that a norm is poorly salient, and the lack of punishment reinforces even more this perception. Conversely, by observing a large number of acts of compliance with a norm, agents can infer that the norm is highly salient and deserves respect. Psychological evidence suggests that the more a norm is perceived as salient, the more likely it will be complied with [26]. The perception of a lack of salience or of the weakening of salience can cause a reduction of the motivation (be it instrumental or terminal) to conform to the norm. If an individual, for example, observes a norm for reasons which are purely instrumental (or to avoid punishment), the perception that both the norm and the normative will have lost their strength allows him to infer that the motivation to defend the norm has also diminished, and, as a consequence, the probability of being punished has decreased too. If, however, an individual obeys a norm for terminal reasons, because a norm, as such, must be respected, then the belief that the social norm is slowly loosing its salience reduces the motivation to comply with it. When norm transmission is not explicit it is possible that the observer misunderstands the norm thus activating a process of norm change or norm innovation [4, 48]. A simplified approach

---

[3] Under the 1906 Transvaal Asiatic Registration Law all Asians who were eight years old or more and resident in the Transvaal were required to register with the Transvaal government and carry a registration document. This was an extension of the exiting 'pass laws' which already restricted the movement of the black population.

to the implicit transmission of norms has been implemented in simulations in which agents copy the norms of the more successful agents in their group [17, 37].

3.2.3 Punishment

As with norm obedience and violation, the reactions to these actions can be communicative acts through which normative requests are transmitted within a group. Punishment, if properly modelled, communicates to the offender (and also to observers) that through his conduct, he has violated a norm and that such violation is not approved of. In the large part of existing work, punishment is looked at from the classical economic perspective as a way of changing wrongdoers' conduct through the infliction of material costs [9]. As suggested by [61, 98], this way of looking at punishment is incomplete and a more insightful understanding of this practice is available once its norm-signaling nature is identified and properly exploited. The norm communicative power of punishment has been supported largely by legal theorists, who claim that a well designed punishment mechanism should explicitly express disapproval for norm violations and should provide cues for appropriate conduct [76, 93].

As claimed in [61, 98], if properly designed, punishment not only imposes a cost for the wrongdoing, but also informs violators (and the public) that the targeted behaviour is not approved of because it violates a social norm. [42] have referred to this mechanism as *sanction*, thus distinguishing it from material punishment. Since sanction communicates the presence of norms and asks people not to violate them, it allows agents to learn of the existence of norms and their relative salience and that their violation is not condoned. In particular, it will generate the belief that the violation of the norm in question will result in a sanction (this way making explicit the causal link between violation and sanction "you are being sanctioned because you violated that norm") that can be more or less severe depending on the salience of the violated norm. Thus, the information conveyed by sanction is twofold: on the one hand it communicates to the offender (and possibly to the observers) the *existence* of a norm, the consequences resulting from this violation and the *legitimacy* of this reaction; on the other hand, it indicates the specific salience of the norm and the *seriousness* of the violation. This normative information has the effect of creating in the mind of the sanctioned agent a set of normative beliefs and possibly of generating the normative goal to comply with (and possibly enforce) the norm in the future. The severity, legitimacy and frequency of sanctions are important cues from which to infer the salience of a specific norm and the seriousness of the violation, information that directly affects the cogency of the normative goal. An agent endowed with normative goals will compare them with other goals of hers and to some extent autonomously choose which one will be executed. The more cogent the normative goal is, the more likely it will outcompete other goals of the agent.

The norm-signaling component of sanction allows social norms to be activated, to increase their salience and to spread more quickly in the population than if they were enforced only by mere punishment. [61, 98] show by a simulation experiment that the use of sanction, the enforcing mechanism that supplements material punishment with normative information, promotes a higher level of norm compliance in a group, makes it more stable and reduces promotion and maintenance costs compared to the use of material punishment alone. [86] shows by simulation experiments the signalling power of punishment and its effect in favoring norm learning, while several mathematical investigations and agent based models have explicitly studied the role of punishment in the transmission and evolution of norms.

Often sanctions are accompanied by explicit messages, often oral, such as "you don't behave in this way" or "you shouldn't have done it". These messages do not necessarily have to be transmitted through explicit communication (oral or written). The fact that a form of

behavior or conduct has violated a norm and that such a transgression is not approved of can also be communicated by practical actions, which can be more or less violent. Consider, for example, a pedestrian who decides to communicate his rage and indignation at the owner of a car which is illegally parked in a space reserved for disabled motorists by deliberately damaging the car's windscreen wipers. In addition, there are contextual factors that facilitate and, in some cases, amplify the behavioral implicit communication of the normative request thereby increasing its motivational power [102]. For example, when punishment is not meted out by a single individual, but by a group of people (or part of it) who co-ordinate themselves to do so, it is easier for the person sanctioned to interpret distributed punishment as aimed to defend a norm rather than driven by a personal interest [16, 97]. A cross-methodological study by [97] shows that a constant punishment level has a stronger effect when it comes from more punishers than when it comes from fewer punishers.

3.3 Norm Evolution

In his introduction to *Law and Revolution*, H. J. Berman refers to law as 'law in action', a "living process of allocating rights and duties and thereby resolving conflicts and creating channels of cooperation." [11, page 5]. This 'ongoing character' of law and institutions, its self-conscious continuity in time, appears to be built upon a conscious process of continuous development, conceived as a process not merely of change but of organic growth.

This concept of 'conscious organic development' of law has been largely applied to eleventh- and twelfth-century institutions, which "were expected gradually to adapt to new situations, to reform themselves, and to grow over long periods of time." [11, page 6]. However, neither can every change be seen as growth, nor does growth necessarily mean the expression of a deliberate will to achieve particular goals.

In Western legal tradition, law is conceived as an integrated system, a 'body' (of law), a *corpus iuris*, which continuously develops over time and generations. Unfortunately, this dynamic character is not self-sustaining. The body of law, as a coherent whole, only survives because it contains 'a built-in mechanism for organic change', that is, an internal logic. In that sense, changes are not only adaptations of the old to the new, but they are also part of a pattern of changes that is coherent with the system over time. Thus, those changes do not occur at random, but are the subject of a development process that relies on the existence of 'certain regularities and, at least in hindsight, reflects an inner necessity." [11, page 9].

We are, however, far from substantially understanding the 'mechanical' machinery lying behind the emergence and stability of social and customary norms. Although the essence of law lacks, by nature, the verifiable characteristics that are inherent to the subjects of study of exact, or 'natural' sciences, in the following sections, we examine the essential elements to configure this 'bootstrapping formula' about the stability and evolution of codified norms.

3.3.1 Stability and codification of norms

Some legal rules are not enacted by a legislator but grow instead from informal social practices[4]. Legal scholarship is divided about the relevance and consistency of such type

---

[4] When referring to customary law, it is often said that courts do not 'create' law, but they rather 'find' it. An example of this can be found in Cooter [32, page 216], where he relates how English merchants in the medieval trade fairs developed their own rules. In some cases, they even had their own courts. However, when those courts were unable to deal with the increasing amount of disputes generated as commerce grew, English judges where responsible for assuming jurisdiction. Since judges where outsiders, with limited knowledge of the special issues concerning trade, "instead of imposing rules,

of norms. Thus, unlike positive or natural law, customary law has received little scholarly analysis and its dynamics remain highly unknown. Part of the doctrine seems to be reluctant to accept its significance, considering customary law a secondary object of study, a 'minor' source of law that grows only when it is necessary to fill legislative gaps left by the legislator. On the contrary, however, some jurists and philosophers consider that all manifestation of law is based on a pre-existing custom; that is, that custom provides the indispensable frame of shared moral and legal reasoning in which law is embedded.

The question is how to fill this gap? The lack of a proper analysis of the dynamics that determine the normativity of customary norms constitutes an important *lacuna* in our understanding of the normative conformity of individuals, both from a social and a cultural perspective.

3.3.2 The normativity of 'informal' sources of law: open questions

When dealing with norm acquisition, the role of cognition is generally poorly explored and mostly refers to the individual motives that lead to compliance with norms. That is, the common individual traits that lead the addressees of a norm to recognize and internalize the normativity embedded in the latter. This may lead us to conclude that, from an individual point of view, human nature presents essential characteristics that predispose us for normative conformity; however, such elements are, albeit crucial pieces of the great puzzle of norm compliance, not the only ones.

Constant behavior and belief in its obligatoriness

The characterization of human beings as 'legalis homo' is not only a product of the individual spirit of men, but it is also in great measure a part of their sense of belonging to a group. Strangely enough, however, until the mid-1990s[5] the attention paid to the 'social' source[6] of law's normativity has been scarce.

Undeniably, group-awareness and group-membership play a key role in normative conformity. Many aspects of our daily lives witness behaviors which are motivated not just by internal psychological drives, but that may also be influenced by the environment in which the individual resides[7], in the form of social or peer pressure.

This 'social' feature that emerges from culture is even clearer in the case of customary law. The idea of the necessity of some sort of inherent in-group homogeneity in order to make it easy to overcome coordination issues —or any other kind of problems resulting from the coexistence of different individuals— seems intuitively appealing. However, a crucial distinction should be made before we continue, regarding the conceptual difference between regularity of behavior and rule-following. Certainly, the existence of a social rule

---

[English judges] tried to find out what practices already existed among the merchants and enforce them. Thus the judges dictated conformity to merchant practices, not the practices to which merchants should conform." We can find the same argument in W. Mitchell [69].

[5] It is in the 1990s when, especially due to scholars influenced by the law-and-economics approach, there was a boomlet of interest in the topic.

[6] Even if we talk about the 'social' aspect of law, in the present text we focus on the relationship between norms and single individuals, when the latter belong to a social group. No attention is given to norms that bear on the conduct of organizations of individuals, such as customary international law. The dynamics of organizational behavior may differ, of course, from the dynamics of individual behavior and those of individual behavior influenced by the feeling of membership to a social group.

[7] See, for example, the definition provided by M. Hechter and K. D. Opp in their volume entirely dedicated to the study of social norms: "[n]orms are cultural phenomena that prescribe and proscribe behavior in specific circumstances." [47, page xi].

does presuppose a regularity of conduct in the relevant group, but the latter element is not sufficient to constitute the former. Just as certain is the fact that custom can be a source of law.

Not all our habits and customs are based on rule-guided deliberations; not all habits become customary rules, and not all customary rules come from habits[8]. The justification of the normativity of customs is the outcome of a process in which individuals are aware of having an obligation to act in that particular way and act accordingly, either out of a belief in the importance of the rule or as the result of social pressure [56, page 157]. In addition to this, a rule is a social norm of a group only if non-conformity to the rule is met by a degree of adverse reaction from other members of the group. Therefore, it requires conscious following of the rule and, furthermore, critical reaction to departure from it. This is what gives it normative force in that society.

Enforceability and internal point of view

Legal norms are defined as normative rules of conduct prescribed by an authority invested with legal legitimacy. Their connection with social norms comes into sight when we study the emergence, adoption and compliance of norms not from the external point of view of the authority, but from the internal point of view of the agent[9], whose choice constitutes the last word in the implementation of normative standards.

In both cases, the adoption or rejection of those standards depends on the individual choice of the agent, based on his own expectations, beliefs, priors and preferences[10]. However, social norms do not necessarily imply enforceability in the sense in which legal norms do. They "allow for a variety of *social* mechanisms which induce norm-following behaviour and promote autonomous acceptance" [83, page 11], such as spread of reputation, social monitoring, normative influencing and rights and entitlements.

Furthermore, in order to be able to adopt a flexible approach towards normative standards, individuals must be endowed with "mechanisms for recognizing, representing, accepting norms and for solving possible conflicts among them." [83]. However, the observance by the subject of a norm as a mental object with clear, determined features is only possible if the norm remains in use for a minimum amount of time. This prevalence does not necessarily imply, though, an anchylosis of the norm, but rather something akin to -at least temporal-stability.

Stability and evolution of norms

Traditionally, at least two reasons have been considered essential for the stability of a norm. *Firstly,* a large part of the socialization process is the definition, for the individual, of prevailing social norms; since these habits or customs generally lead to a relatively seamless integration into a society or social group, few refuse to adopt them. *Secondly,* even if a particular instance of normative behavior goes against an individual's instincts or previous

---

[8] As Ibbetson explains, '[i]t is easy [...] to say that habit represented a factual regularity while custom was normative. This is true, but it does not explain how habit became custom, nor does it explain why custom is normative." [56, page 156].

[9] For a broader discussion on this issue, we refer the reader once more to Sartor et al. [83, page 11]. We must also note that the emphasis on this 'internal' point of view in the legal scholarship gained its force with Hart's *The Concept of Law* [45].

[10] [96] made one of the first attempts at explaining norms and norm compliance in a rational choice framework.

habits, the weight of social pressure may induce that person, nevertheless, to follow the norm. Many issues of law are about whether the addresses actions "have conformed to expectations which other parties had reasonably formed because they corresponded to the practices on which the everyday conduct of the members of the group was based." [46, page 96].

This leads us to an open question: the study of social norms and customs has led to deeper insights into the ultimate intrinsic forces behind human normative behavior and yet, "[t]he existence of social norms is one of the big unsolved problems in social cognitive science." [35, page 185]. We still know very little about how social norms are formed, the forces determining their content, and the cognitive and emotional requirements that enable a species to establish and enforce those social norms.

Issues regarding the emergence of social norms, their stability over time and their evolution are indeed questions that remain open in legal literature. In addition, the outcomes of evolutionary models of social norms are extremely sensitive to the postulated rules of transmission[11]. How is culture (and with it the social norms that evolve around culture) 'transmitted'? There are contrasting opinions in the scholarship about this topic. Here, we will not dig deeper into the subject, but rather just remember the reader what Goethe once said, "a tradition cannot be inherited —it has to be earned."

A positive 'double-feedback loop'

Unquestionably, the intrinsic nature of social norms involves change. Its link to the life of the community makes it the product of a continuously changing process, consisting in the aggregation of in-group decisions over time. However, in this accumulation of responses to everyday situations and conflicts, are we more likely to accumulate right or wrong answers? Is there a way to change the path, in cases where the wrong one is taken, or are previous choices inevitably binding?

The mechanisms by which some customs are selected and others disappear are rather obscure. The choice might not necessarily or consciously be outcome-oriented and, indeed, the 'socially optimal' solution is not always the one to survive over time[12]. Many factors affect the selection process, and understanding the causes and the course of action of that complicated selection procedure might indeed be more important than the actual outcome.

Customary law as such may be a function of institutional influence on individual and group preferences and behavioral patterns; it may as well be the end product of background facts and social norms, or the solution to coordination problems as discussed by [87, page 22]. Some scholars treat the development of customary rules as the fruit of some sort of 'social evolution', dealing with "populations of [replicating] entities, including customs and social institutions that compete for scarce resources." — see the arguments exposed by Hodgson and Kundsen [50, pages 477–486]. Other scholars consider that customs emerge as solutions to adaptive problems.

Furthermore, customary law, as opposed to civil law, is not equipped with mechanisms of systematization or codification[13], or any order which may assist in eliminating internal sources of inconsistency or conflict[14]. Despite this unsystematization, social and customary

---

[11] "[S]ince there is no firm basis on which to choose the rules, almost anything is possible." [38, page 73].

[12] For further discussion on path-dependance and persistence of bad choices recall, for example, [40, page 217].

[13] See Berman [11, page 328], who points out that manorial law "lacked the high degree of logical coherence and the consciously principled character of canon law."

[14] An argument in the same direction is proposed by F. [87, page 30]: "the existing array of entrenched

norms must be incorporated into the bigger picture of an integrated, coherent 'body' of law; with the additional difficulty that, although the *corpus iuris* is also flexible and relatively receptive to changes in the 'environment'[15], customary and social norms are driven by a different dynamic. This leads to an on-going tension between the different dynamics and sources of legitimacy of positive and customary law and the necessity for a sustainable and stable equilibrium.

However, this apparent 'chaos' that characterizes social and customary norms has a positive side: a higher grade of flexibility and adaptability[16]. Furthermore, insofar social and customary norms are 'closer' to what their addressees consider legitimate, to their 'internal point of view', norm-abidance is usually easier to achieve[17]. The existence of this 'double-feedback loop' between custom and culture reinforces the positive dynamics between them: "[f]ar from being a minor adjunct to the law properly so called, custom is [...] one essential component of any legal system, sufficient to sustain a rule of law under some circumstances, and one essential component of the rule of law under any and every circumstance." [75, page 100].

Notably, when the deviant behavior is 'naturally' agreed upon[18], according to the prevailing distribution of values in the community, the social forces will find it necessary to severely discourage any undesired conduct. In the case of social norms, we can consider that these attitudes of endorsement or rejection towards a particular action form part of the cognitive map through which the members of the group interpret their environment. An interpretation that is consensually constructed and commonly shared; consequently, the implementation of the enforcement mechanisms' is smoother than it would otherwise be.

An example: smoking ban in pubs

As we have explained in previous sections, when dealing with norm compliance, we can talk about two different layers or perspectives. In the first place, those addressed by the law shall consider their potential legal liability in case they don't abide the norm. Monetary sanctions, incentives and other deterrence measures are essential. However, there is also a second aspect, the morally persuasive role that is inherent to law. People obey the law because 'it's the right thing to do'. Sometimes the 'threat' of punishment is not the main reason for conformity, and non-pecuniary considerations must be taken into account.

Individuals belong to communities, which share a general respect for a system of values and norms. Norms are thus seen as a reflection of the set of behaviors that are seen as desirable of legitimate in the shared view of societal member, and whose violation elicits at least informal disapproval. This socio-cultural aspect of law has deep implications in terms of policy: instilling a norm in countries where it currently does not prevail and is culturally accepted may be a daunting task. People in such countries may find its content incompatible with the social environment in which they live, and therefore unfeasible to act accordingly.

---

normative customs is so complex, so large and so unsystematized that mutually exclusive customs abound."

[15] Here 'environment' is understood as the aggregate of social and cultural conditions that influence the life of individual in the community

[16] For the opposite argument, cf. Hart [45, page 89], who considers that custom tends to be static and inefficient and that, given that customary law is not under anyone's rational control, it cannot serve policy making ends.

[17] "What is more, written law will have no purchase on a community, unless it reflects the practices of that community in some way; even a law that sets out to correct custom will necessarily reflect other aspects of the customary practices of a community, or it will lack purchase in the community for which it is intended." [75, page 100].

[18] 'Naturally' as opposed to 'artificially' imposed by a legal authority.

In the case of smoking ban in pubs, the latter seems of extraordinary impact. The enforcement of such a ban is highly ineffective, and its compliance depends therefore on the underlying systems of norms and values that is embedded in that group's culture. Indeed, smoking habits are deeply rooted in greater cultural traits, such as social structures, gender roles, autonomy, authority considerations or distribution of power, to mention some of them.

Concern for risks related to smoking has raised considerably during the last decade. Not only for smokers, but also for the ones surrounding them, due to 'environmental tobacco smoke' (ETS, also called 'second-hand' smoke or 'passive smoke'). Hence, it does not seem implausible to believe that a majority of EU citizens support smoke-free public places, such as offices, restaurants and bars. This fact would be consistent with broader findings in sociological literature, that individuals exhibit similar value preference rankings across cultures. But even if this is generally true, there are also cultural differences that lead to cross-country specificities in citizen's reactions to the smoking ban in pubs. Let us take for example, three sample countries: Italy, Portugal and The Netherlands.

(a) *Italy*: is considered *par excellence* the European country in which informal ties and social relations are fundamental. This would lead to a higher compliance with the ban, due to informal enforcement mechanisms and social control. Stigmatization of smokers is higher and has further consequences. Furthermore, we believe, though restaurants and bars are an important part of cultural life, they are easily replaced. The culture of 'open-air' events is also large, due obviously to the warm Mediterranean weather, and many social meetings take place around close family and friends at private locations, making the displacement of smoking at home easier, and other adverse social consequences such as the increase in the noise in the streets not so relevant.

(b) *The Netherlands*: should present radically different reactions to the ban. Dutch citizens are considered more independent from their social ties. The informal enforcement mechanisms of the ban are hence deemed to be effective in a lesser extent. A second aspect that could be considered relevant, is that a great part of the socialization process takes place in pubs and bars. Citizens should be highly opposed to the ban, which would clearly diminish social and cultural exchange, as well as harming values such as individual freedom or autonomy.

(c) *Portugal*: represents somehow an intermediate example. In addition, the ban in Portugal was implemented in different, less radical terms. Bars were allowed to create separate spaces for smoking and non smoking in case they were bigger than 70 square meters. For bars smaller than that, it was up to the owner to decide whether to constitute the place as either smoking or non-smoking environment. This should have facilitated the transition, but it could have lead, however, to a majority of the bars becoming 'smoking allowed'.

As yet there is little published evidence that we can use to verify these predictions, though what there is suggests that they agree with what has happened. [14], for example, suggests that in Italy the ban has led to a big reduction in smoking in bars and restaurants without the need for much enforcement. There is less direct evidence for the effect of the ban in Portugal, but [64] describes a small-scale study on Portuguese restaurant workers that shows a reduction in exposure to ETS following the ban, suggesting an appreciable degree of compliance. In the case of the Netherlands, the only data we are aware of is from [71], which suggests that there is a smaller increase in support for smoking bans than in other countries.

4   Trust and the Dynamics of Norms

In this section we discuss the relationship between trust among a group of agents and the dynamics of norms within that group. We will argue that trust is essential for the adoption of a norm, and describe how we believe that trust fits into each stage of the norm lifecycle. We start by offering an overview of definitions of trust, before advancing our own position.

4.1   What is trust?

Trust is a concept that is both complex and rather difficult to pin down precisely. As a result of this slipperiness, there are a number of different definitions of trust in the literature. Sztompka [94], for example, suggests that "Trust is a bet about the future contingent actions of others", while Mcknight and Chervany [65], drawing on a range of existing definitions, offer the definition "Trust is the extent to which one party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible." Gambetta [41] states that "Trust is the subjective probability by which an individual, A, expects that another individual, B, performs a given action on which its welfare depends", while Mui *et al.* [74] define trust as " a subjective expectation an agent has about another's future behavior based on the history of their encounters."

As the alert reader will have noticed, these definitions, though different, do overlap somewhat. All four definitions given here focus on trust as a mechanism for making predictions about the future actions of individuals. That is if one individual trusts another, then that first individual can make a (more or less) accurate prediction about what the other will do in the future. One might, as [41] does explicitly and [94], does implicitly[19], decide that trust can be quantified as a probability. Or one might, as Castelfranchi and Falcone [22] argue, decide that trust is more complex (and in the case of [22], decide that trust has a rational basis in reasons for beliefs about the future actions of others). This distinction need not concern us here — for our purposes it suffices to think of trust as an abstraction, a summary perhaps of some complex pattern of reasoning, for some model of how others will behave.

4.2   Trust in the norm lifecyle

The fact that trust allows agents to predict the behavior of other agents is exactly why trust, or the lack of it, plays an important part in the norm lifecycle in Figure 1. To see why this is the case, consider the spread of norms. Once a tentative norm has been generated, it will either spread through a population and eventually stabilize, or it will be discarded. Which happens depends on individuals' attitudes to the norms and their ability to predict each other's behavior, and this latter depends on the trust between individuals. As Bicchieri [12] argues, "a social norm depends for its existence on a cluster of expectations. Expectations, ..., play a crucial role in sustaining a norm". Trust is important because of the way it helps agents to handle those expectations[20].

---

[19] Subjective probability having a natural interpretation as a propensity to make bets at particular odds [57, 655].

[20] One can also argue the reverse, that because norms exist, it is possible for agents to develop expectations based on their beliefs that others will follow the norms, and hence agents will trust others. While this is clearly the case, we believe that the cycle that takes observed behavior, infers trust from it, and then develops norms as a result of the trust is more general, allowing the establishment of trust and then norms in the absence of any existing norms. In other words it plays a role in norm generation as well as

Consider the norms that govern how people wait for and then board a bus. There are at least three distinct ways in which this is done. In some populations, people who arrive at a bus-stop form a queue. When the bus arrives, people board the bus in the order of the queue. In other populations, people who arrive at the bus-stop do not form a queue, but each remembers the people who were there before them, and when the bus arrives, people do not board until everyone who was at the stop before them has boarded. In yet a third population, people arriving at the bus-stop don't form a queue, and when the bus arrives, everyone boards the bus as quickly as they can with no regard for the order in which people arrived at the bus-stop. (Of course, there are populations in which some mixture of these behaviors co-exist, but we will consider simpler cases for now).

Now, consider how someone reacts when they arrive at a bus-stop where they don't know what the norm is. A typical reaction is to try to infer the norm by observation. For example, if there is a queue, take this as an indication that people will board in the order of the queue, so the right thing to do is to join the queue and let the people in front board first. In doing so, the new arrival at the bus-stop is trusting other bus riders to do the same. There is risk in this. By waiting when the bus arrives, our rider is both allowing later arrivals who don't believe in queuing to push ahead, and denying themself the opportunity to push ahead of others. However, if the queue operates as it should, our bus rider is ensuring that while they can't exploit others by pushing ahead, they will board before later arrivals, and they will not suffer whatever sanctions the bus queue imposes on the violators of norms (which in the authors' experience ranges from sharp intakes of breath and disapproving looks to loud lectures on appropriate behavior).

What we have here is an example of how trust allows norms to operate. If everyone trusts that everyone else will follow a norm[21], and in the face of a population that observes a norm and enforces there is no net benefit in violating the norm, then the norm will be stable. Of course, if there is no bad consequence of violating the norm (for example, the rest of the queue just ignores the violation), or there is no advantage in following the norm (somehow waiting in the queue means that our rider always ends up being the last person to get on the bus), then there is no incentive to following the norm and it will be discarded.

To take this example a little further, we can also see how trust is involved in the spread of a norm. Say we have a population of bus-queuers in a country with small bus shelters and a high likelihood of rain (any similarity with the situation in the United Kingdom is entirely coincidental). In such a situation, there is a disadvantage to queuing if one is not one of the first in the queue — one has to stand in the rain on rainy days because the queue extends beyond the shelter. Imagine that on such a rainy day, at one bus stop, someone who would, under the prevailing norm, have to take their place at the back of the queue, outside the shelter, decides instead to stand somewhere dry (under the awning of a nearby building), but when the bus arrives insists on boarding in order of their arrival. If they are not sanctioned (and why should they be — they may have violated the standing-in-line norm, but they haven't violated the first-in first-out spirit of the queue), then it is possible that their action will be remembered by those damp members of the queue who get to board later, and recognized as superior. These damp queuers may then copy this behavior at other bus-stops, so spreading this new norm through the population. Of course, this happening

---

norm spreading.

[21] Or, more correctly as we shall see below, as long as a population trusts that a sufficiently large fraction of the population will follow the norm. For the perspective that trust is based on a normative expectation see also [58].

requires trust on the part of the old-norm violators that they will not be sanctioned, and that they will still be able to board the bus in the order in which they arrived at the bus-stop. If that trust is not possible, then the new norm will never become established.

4.3 How trust develops, how trust decays

Given the essential role of trust in ensuring that norms spread and stabilize, it is worth thinking a little about how trust develops between agents. As pointed out above in the quotation from [74], trust between individuals develops as a result of the history of interactions between them. When those individuals are representative members of a population, then the trust develops as a result of all the interactions between members of those populations. The development of trust is therefore dependent on the ability of the individuals to learn, and on the repeated nature of their interactions.

One well-cited example of the development of trust even between agents that might be thought to have opposing interests is the story of the unofficial truce[22] that took place around Christmas 1914 between British and German troops (and, to a lesser extent between German and French troops) stationed in the trenches along the Western front. The truce was sufficiently widespread that a number of football games took place between groups of British and German soldiers. As Axelrod [7] points out in his description of this incident, at the time that this truce developed, units were stationed opposite each other for relatively long periods. This meant that these soldiers had a chance to establish some form of relationship with their opponents and to learn that these opponents would not take advantage of friendly actions. For example, Weintraub [100] reports that both sides had been taking a 'live and let live' attitude for some period before Christmas, not firing on each other during mealtimes, and observing ceasefires when both sides could retrieve the dead and injured from the no-mans land between the trenches. Here we see the importance of repetition in the development of trust.

Axelrod also gives a nice example [7] of how individuals can be representative of populations, reporting a case in which one group of Saxon soldiers from the German army apologized to the British unit they were facing for the fact that the British were shelled by a different unit:

> . . . a brave German got up on to his parapet and shouted out "We are very sorry about that; we hope noone was hurt. It is not our fault, it is that damned Prussian artilary." (the words of an unnamed British officer quoted in [80, page 34] which in turn is cited by [7, page 84].)

Clearly the Saxons wished to distinguish themselves from the Prussians in order not to undermine the trust the Saxons had established with the British.

It is not clear how much this case represents the development of trust, and from it the adoption of a new norm, in the few months between the start of trench warfare on the Western front in September 1914 and Christmas, and how much it represents a continuation of an existing norm for opposing sides in a conflict to observe ceasefires on humanitarian grounds. Figes [36], for example, describes regular armistices at the siege of Sevastopol in the Crimean War, while Weintraub [100] points out that such incidents go back at least as far as the siege of Troy. However, it is clear that the official response was to outlaw further

---

[22] So unofficial that commanders actively tried to stop it happening.

unofficial truces and that this and mandatory raids on enemy lines[23] led to the steady decay of this norm — while there seem to have been further incidents during World War I, they were much less widespread. In this we see the importance of learning. Trust was adjusted as a result of experiences in the repeated interactions.

In many situations, individuals do not have the chance to develop trust in others, either at the individual level, or the population level. In Axelrod's [7] term, there is no 'shadow of the future' in such scenarios — if two individuals are going to interact just once, there is no way for one to repay the other for a kindness (not shooting during a mealtime to extend our military example), and so no motivation for the other to act kindly.

As Resnick *et al.* [78] argue, it is exactly in such situations that reputation systems come to the fore. By providing a mechanism for recording the outcome of one-off interactions, reputation mechanisms create a shadow of the future. Now the kindness or otherwise of two individuals who interact just once will be recorded, and when one of them interacts with a third, that third individual can respond to the outcome of the previous interaction. (In the trench warfare case, the reputation of the local enemy units was passed along as troops rotated in and out of a sector.)

Of course, reputation systems are not infallible. [77] describes a problem with the reputation system on eBay (a system that in general performs rather well), in that in the empirical analysis in [77], there is little negative feedback of sellers. This in turn leads to a rather over-optimistic estimate, and means that reputation isn't the best predictor. Furthermore, as [62] shows empirically, reputation systems can be manipulated, and [82] backs this up with an theoretical analysis of how reputation systems can be manipulated.

Finally, we should note that while the reputation of an individual, and hence trust in that individual, is most strongly affected by the behavior of that individual, it can be affected by others as famously argued by Akerloff [2]. Akerloff studies the 'market for lemons'[24], a market in which sellers offer indistinguishable goods some of which are of poor quality. In such a market, trust in all sellers is decreased by the presence of these poor quality goods — buyers translate the risk of loss through purchasing a lemon into a loss of trust in every seller, even though they have no personal experience of lemon purchase.

4.4   Implicit trust

Several of the authors (for example in [21, 95]) have developed formal models that can be used to explicitly represent the trust that one individual has in another. Indeed, the model from [21] is widely adopted in the multi-agent systems community and has been adapted [34] and extended [15] by a number of authors. However, while it is possible, and we believe useful, to build explicit models of trust, it is not necessary. Trust can be implicit.

Indeed we can consider several levels of implicit trust. When it uses an explicit model of trust, agent A considers the trustworthiness of agent B, and makes decisions based on that trustworthiness. One form of implicit trust is when A, rather than modelling B, just reasons about what action B will take, perhaps on the basis of B's past behavior. If A learns to predict B's behavior in interactions using fictitious play [19, 39], in which A takes B's past behavior to be a representative sample of its full range of behaviors, then the resulting probability distribution over actions can be used to help A determine what to do (which is

---

[23] Which, according to Axelrod [7], were not intended to reduce 'live and let live' but ended up doing so.
[24] In the US, the term 'lemon' is used to refer to a badly constructed car with many faults, and many states have a 'Lemon Law' which provides financial assistance to car buyers who have been sold a such car.

(a) Prisoner's Dilemma   **(b)** Battle of the Sexes   **(c)** Matching pennies
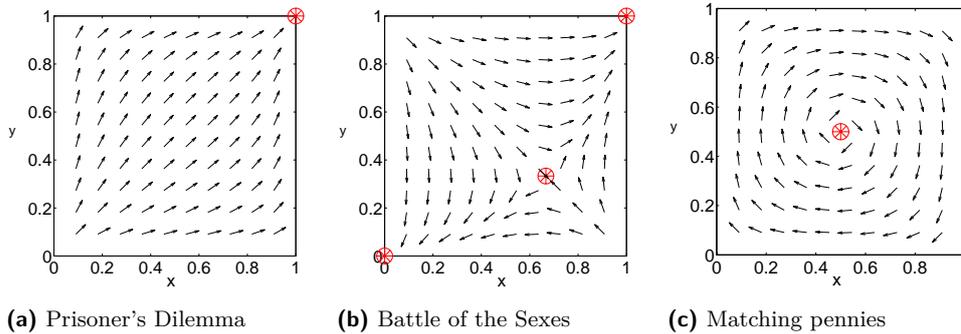
■ **Figure 2** Replicator dynamics for three two player, two strategy games under frequency-adjusted Q-learning. Figures by kind permission of Michael Kaisers.

functionally what trust does) without A having to invoke any notion of trusting B. In this kind of scenario, as long as the agents have converged to equilibrium — so that their notion of what the other will do is no longer changing — it seems reasonable to consider that they trust one another since they have accurate predictions of what the other will do[25]. Fictitious play allows this to happen for some scenarios, for example , Miyazawa [70] showed that fictitious play converges in two-player, two strategy general sum games such as the iterated Prisoner's Dilemma, the scenario studied by Axelrod in the classic work cited above [7].

In fictitious play, A is still conscious of the choices made by B — that is what A models. A could also just concentrate on its own actions and learn what works best for it. If A does this, then since the outcomes of its actions depend to what B does, it is implicit that A learns to predict B's actions, and so very implicitly can establish trust in B. Kaisers discusses how Q-learning and its variants [59] can be used to do this learning, and shows how learning converges (and hence how A establishes trust in B). Figure 2 shows this convergence. The axes in each plot give the probability of the agents playing one of its strategies (and since there are only two strategies, by extension the probability of playing both strategies). Figure 2a shows the Prisoner's Dilemma, where the agents converge to the Nash equilibrium at $(1, 1)$[26], Figure 2b shows the Battle of the Sexes, where the agents converge to the Nash Equilibria at $(0, 0)$, $(1, 1)$ (the fixed point at $\left(\frac{2}{3}, \frac{1}{3}\right)$ is not stable). Figure 2c shows the Matching Pennies game which has a fixed point at $\left(\frac{1}{2}, \frac{1}{2}\right)$ to which agents do not necessarily converge[27].

4.5 Trust as a normative phenomenon

Finally, given that trust can develop and decay just like the norms in the norm lifecycle, it is worth pointing out that trust itself follows parts of the norm lifecycle. Again Figure 2 shows us how this might happen.

Above we interpreted the plots in Figure 2 as giving the probability of a specific pair of

---

[25] To quote BIcchieri [12] again, "norms of cooperation . . . emerge as equilibria of learning dynamics in small-group interactions"

[26] One might, of course, argue that given the details of the scenario, this outcome does not represent one in which the agents trust each other, but in accurately predicting the action of the other agent, it meets our definition of trust.

[27] The fixed point only has Lyapunov stability in this particular case, so while points that are close to the fixed points will not diverge, they will not necessarily converge to the fixed point.

strategies being played. We can also interpret each point in one of the graphs as indicating the proportion of a population of agents that will pick a specific strategy. Under such an interpretation, the point $(\frac{1}{5}, \frac{3}{5})$ in Figure 2b captures what happens when, in a population of agents playing the Battle of the Sexes, 20% of the agents picking the row in the game matrix play B and 60% of the agents picking the column play B. The arrows on the plots show how the population changes its strategy from this mixture — in this case the arrow indicates that a population at $(\frac{1}{5}, \frac{3}{5})$ will change such that more agents picking the row will tend to play B, and fewer agents picking the column will tend to play B.

Under this interpretation, we can see the plots as showing trust, in terms of the ability of one agent to predict the behavior of another, spreading through the population. Look at Figure 2a, and consider a point near $(0, 0)$. Here, any given agent can be pretty sure that any other agent it encounters will play C, but it will also know that this will change over time — the choice of playing D will tend to appeal to any agent in the population so this ability to predict is not stable. The direction field tell us that the choice of playing D will tend to spread until the population reaches $(1, 1)$ and every agent knows for sure that any agent it interacts with will play D. Thus the Prisoner's Dilemma is an example where trust can spread through the population and become stable. In contrast, the fixed point at $(\frac{2}{3}, \frac{1}{3})$ in the Battle of the Sexes (Figure 2b) illustrates that it is possible for equilibria to be unstable. Without the equilibria at $(0, 0)$ and $(1, 1)$, the population playing this game would not be able to develop trust in one another, and the lack of an attracting fixed point in the Matching Pennies game is a further illustration of trust not becoming stable.

5   Culture for modifying norms' dynamics

Cultural differences in normative dynamics are considerable, as argued above. Group dynamics are an essential part of social interaction and are critical to the evolution of social norms, as already argued in Section 3.2. Relationships to others in a group context usually affect one's willingness to emulate, provoke, forgive, reproach, oppose, admire etc. the adoption of new social norms by other group members. [55] argues that recognition of the role of social relations will improve our understanding and our predictions, with regard to social norms. There is no universal mechanism for the transmission of social norms across all human cultures. It follows that at some stage in the development of NorMAS, we should tackle how cultural differences affect the dynamics of social norms. Indeed, the approaches of researchers to modelling NorMAS are undoubtedly influenced by their own cultures, and we should try to be aware of these biases when postulating mechanisms of norm transmission. An example of this comes from [8], who investigate beliefs about what it means to be human in various cultures. The dangers of ignoring culture in research into social processes has been emphasised by [49]. Members of different cultures are socialised to obey very different social norms and integrate very different value structures in their ways of living. They also have different propensities to intepret behaviour as normative.

We shall focus in this section on the differences in the processes of norm dynamics across cultures. Comparative quantitative studies of culture have usually been at the level of the nation state, due primarily to issues of data availability. While such analyses may miss many of the subtler nuances of culture, we believe that they are at a manageable level of abstraction for incorporation into the next generation of NorMAS. The empirically derived Hofstede Dimensions of Culture [52, 53], shall form the basis of our discussion of culture in NorMAS. Another cross national analysis of culture has been conducted by [88], who analysed self-reported values to elicit dimensions of values. This approach contrast with the

■ **Table 1** Dimensions of Culture.

IND	INDividualism
PDI	Power Distance Index
MAS	MASculinity
UAI	Uncertainty Avoidance Index
LTO	Long Term Orientation
IVR	Indulgence Versus Restraint

Hofstede theory, which was derived from questions about everyday practices.

We begin with a brief introduction to the Hofstede Dimensions of Culture. We then discuss the preceding sections of the chapter from the point of view of the culture theory outlined. In the final part of this section we discuss how some of these differences can be included in the next generation of NorMAS.

5.1 A Brief Introduction to Hofstede Culture Theory

Hofstede and co-workers conceptualize culture as a limited number of major societal issues, to each of which a society finds a shared solution. These issues are conceptualized as continua, as scales with a lower and an upper end. [52, 53] call these dimensions and they describe six of them that vary across nationalities, see Table 1. These dimensions have been shown to correlate with a wide range of empirical data in the social sciences [52], notably marketing data [72].

The Hofstede model is based on questions about everyday work practices; the dimensions of values were a serendipitous finding. They refer not to convictions or beliefs but to broad tendencies to perceive the social world in a certain way. The model has grown over time, as more sources of data were consulted. The dimensions were derived using factor analysis of survey data.The latest model [52, 53] consists of six dimensions. Each of them is modelled as a continuum running along a scale from 0 to 100. This means that almost all actual values will share characteristics of both extremes.

Here they are, together with some explanation and also an impression of the perceptual capacities that agents need to have in order to accommodate the dimension in a model:

Identity: individualism versus collectivism. (IND) Essentially this is the extent to which members of a society feel responsible for themselves, or for the larger group they belong to. In the first case, rights and obligations should be the same for all people, while in the second, the boundary of the in-group is also a moral boundary beyond which rights and obligations do not hold. Agents in collectivistic cultures will act differently depending on whether other agents are in group members or not.

Hierarchy: large power distance versus small power distance. (PDI) That is the extent to which the less powerful members of a society expect and accept that power and rights are distributed unequally. Large PDI divides a society depending on positions within that society, i.e. there is stratification of social groups based on status, those with low and high positions do not mix. Agents in cultures of large power distance will respond differently to others depending on how they perceive their status to be relative to their own.

Aggression and gender: masculinity versus femininity. (MAS) This dimension is about assertive dominance and emotional gender roles. It contrasts a strong-handed, competitive orientation in 'masculine' cultures, in which people in general cannot be assumed to be

trustworthy, men are supposed to be tough and women subservient and tender, versus a consensus-seeking and care-taking orientation for both women and men in 'feminine' cultures. In masculine cultures, it is desirable that agents be gendered and recognise gender.

Otherness and truth: uncertainty avoidance versus uncertainty tolerance. (UAV) In uncertainty avoiding societies, anxiety levels are high. In defence against it, strict rules, rituals, and taboos govern life. Distinctions between categories should be sharp and the unknown is considered dangerous. Out-group members and institutions will not be trusted by agents from such cultures. In uncertainty tolerant cultures, relaxation is the rule and actions are results-based rather than anxiety based. The level of fixity of all kinds of rituals goes up with uncertainty avoidance.

Immutability vs. pragmatism: short-term versus long-term orientation. (LTO) In short-term oriented societies immediate gratification of needs and keeping up social appearance, behaving well and respecting tradition are seen as virtues. In long-term oriented societies, reasoning is pragmatic and planning, foresight and perseverance are valued. Persons from such societies are more likely to learn from experience and to change their norms accordingly.

Gratification of drives: Indulgence vs. Restraint. (IVR) Indulgence stands for a society where people feel happy and healthy, and like to enjoy life and to spend time with friends, but could also slip into violence if they feel like it. Restraint stands for a society in which people feel the burden of duties heavily, and keep both positive and negative impulses in check. This dimension will heavily impact the exuberance of greetings.

Note that the six dimensions are not personality traits, but societal patterns! This means that, unlike personality traits, they will be shared by the persons with the same cultural background. Yet culture, as an unconscious set of basic values, should not be confused with conscious group affiliation.

5.2   A Discussion of Culture in the Topics Raised in Previous Sections

Three definitions of the polysemic term 'norm' are given in the introduction to this chapter, Section 1. The first, which is simply that which is most normal or most usual need not detain us here. The second is norm as an ideal. The ideals of how to behave as social actors vary significantly across cultures. In some, the more long term oriented, being pragmatic and reasonable is considered ideal, in others, the more short-term oriented to be strong, resolute and unyielding in one's convictions is more admired. The ideals of a culture, as exemplified by the heros of that culture, vary with each of the dimensions of culture, in ways that can be straightforwardly inferred from the brief description of each dimension above. The third definition of norm that was discussed is that of an "imperative regarding the behavior of social actors". These imperatives, or the resultant tendencies to behave in certain ways are, along with the manner of perceiving the world, strongly moulded by culture. However, culture is more than an exhaustive list of norms and practices, it is the deep structure embedded within these norms and practices, part of which is reflected in the six Hofstede dimensions of culture.

Section 2 discusses the cognitive architecture required to model normative processes in multi-agent systems. It forms part of the universal basis for social behaviour on the part of agents, which in the real world is modified by culture. How best to get a norm adopted in the mind of another is discussed in Section 2, and how culture can change this is discussed as the end of Section 3.3.2, with regard to real experiences of the smoking ban. It is postulated

quite reasonably that the adoption of a new goal (such as a goal to observe a norm) is conditioned on the new goal being as means, in some manner, of achieving a pre-existing goal. These pre-existing goals will almost always be influenced by the culturally derived disposition to value certain states of the world over others. Individuals can reject the overt values of their society, but the traces of their culture of upbringing always remain in their way of perceiving and interpreting their social world.

Among the reasons given for motivating the adoption of a norm, in Section 2, are instrumental, cooperative and terminal reasons. Cooperative reasons apply when the norm adopter holds the values that are supported by the norm. Another more social reason to adopt the norm is simply as it is the done thing in the social group one belongs to. This comes close to envisaging the desire for compliance to the group, or being well respected by certain individuals, as a terminal reason to adopt certain norms. We may have the goal that the norms of a certain group should be respected, but which group, and why? The subtle processes of group dynamics can be influential in these decisions, decision that are not always fully conscious. To refer this to the dimensions of culture, the desire to simply fit in is strongly influenced by collectivism (or low individualism) and Restraint.

Section 4 discusses the role of trust in norm dynamics, and it is clearly a critical element for any functioning human society. Indeed, trust is so basic that it cuts across cultures. It's an essential part of social cooperation and no culture could survive without it, though how cultures manage, preserve, enforce and nurture trust, vary greatly. In low power distance, feminine cultures people are expect to be good without the presence of tough rules and supervision. In high power distance, masculine cultures that are also restrained it is usually thought very important to enforce norms strictly or else they shall not be obeyed.

Section 3.1 concerns norm generation, and shows a number of processes proposed to emulate the real world emergence of new norms. This section highlights the admittedly large gap that separates the complexity of the evolution of norm in human societies and current methods for the run-time evolution of norms. This raises the question of whether the time is right to start considering real world cultural variations in NorMAS, and for some of the models discussed this is clearly not appropriate. However, as mentioned in the introduction to this section researchers beliefs about what norms are, and how they work, is in part a product of their culture, and a broader view of these processes is likely to be helpful in the development of successful NorMAS. Also, when the adoption of a new norm is considered in different cultures, as discussed above for the smoking ban in Europe, consistencies within (and differences between) cultures related to the interpretation of behaviour as normative, the manner of enforcing norms and the readiness to adopt certain types of norms can be incorporated in currently feasible NorMAS.

Section 3.2 makes the case that observed behaviour can be interpreted normatively. If an agent sees many people perform an action that agent can interpret that this is more than a simple empirical fact, but something that can carry normative message. Here are some ways in which this process is likely to vary with national culture. In societies strongly stratified by hierarchy (large power distance), the behaviour of others of differing status is rarely a reliable guide to the social norms prevailing for those of one's own status. For example, a student may think they should use another staircase to that used by a professor. The dimension of Indulgence versus Restraint is linked to how ready people are to infer social norms from behaviour. Restraint cultures have many strict norms and are primed to interpret new behaviours in that light. Indulgent cultures are much freer in their behaviour and are more likely to see new behaviours as the result of personal preferences than normative obligation. Finally the level of individualism, as opposed to collectivism, is likely to affect

the interpretation of observed behaviour. In individualist cultures personal preferences are more freely followed, while in collectivist cultures conforming the one group's behaviour and outlook is a much stronger force.

Section 3.2 also discusses obedience and violation. Notably, masculine societies feature more punishment for transgressions, while feminine societies have higher levels of forgiveness for transgressions. Restrained and collectivist cultures are likely to see far high base levels of obedience to a larger number of norms while in indulgent and individualist cultures people are much less constrained in appropriate social behaviours they may adopt. Indeed, in such cultures 'following your own star is admired', while in others obedience to social demands is admired [10].

The two-way relationship between formal laws and informal social norms is examined in Section 3.3, where the role of culture and the importance of modeling humans as actors within groups are referred to on a number of occasions. One word of caution, which is appropriate here and also applies to the examples and statements in this Section, is the possibility to read too much into the determining role of culture in any given example. The confirmation bias is an inherent part of human reasoning, and seeing a plausible explanation does not mean that it is *the* explanation[28]. This also applies to different interpretations of the role of culture in a particular phenomena, such as the varied success of the smoking ban in bars in different European countries, and well as the many institutional influences of the evolution of this particular norm. One of the major problems in getting a handle on culture in a model is its sheer pervasiveness. It's even present in how pervasive we consider culture to be. Culture is present in the smoking bar example in the institutions that created the laws and regulated their implementation. It is also present in the subtle interactions by which individuals test the the prevailing norms, and accept, reject or attempt to influence them.

5.3   Culture in the Next Generation of NorMAS

We outline some essential elements of group dynamics before discussing how these may affect norm transmission, and then how well-understood cultural variations could influence this process.

Intra-group Behaviour is critical to the transmission of social norms, and it varies considerably across cultures. The idea that people learn differently from others depending on who they are is called selective attention by [48]. When explaining why norms are enforced in some situations and not others, social context matters [55]. Some elements of human social behaviour can reasonably be assumed to be universal, such as the role of trust, see Section 4. Most social behaviour is anchored in a group context, where the relationships between individuals, and their relationships to the wider group, matter. Section 3.1 already mentions the importance of societal typology. Social norms are first and foremost a mechanism for regulating behaviour in groups. We now outline how culture can begin to be implemented in NorMAS to modify these behaviours.

5.3.1   Suggested Group Level Primitives

Each agent in the NorMAS we envisage would have a perception of each group it belongs to, including the membership of that group, the rules of behaviour appropriate to that group, and both its own and others position in the group. What do we mean by position? In order

---

[28] [66] argue that the fruitful clash of arguments which results from this universal bias is the main evolutionary reason for what would otherwise be a straightforward flaw in human reasoning.

to create implementable computational models this needs to be broken down into a limited set of primitives that agents can understand and reason over when making decisions. In the following we introduce one possible set of such primitives.

One of the first elements agents should look to as a guide to social behaviour is hierarchical status. This usually means age, often role (e.g. employer/employee), and sometimes gender. Hierarchical status is important in determining who can take initiative, who should look to whom for leadership, whose social cues should be followed etc..

Another important element is reputation. Who is a good member of the group, and whose behaviour is a reliable guide to the right thing to do in social situations. This is a critical element for imitation behaviour, and hence the spread of social norms. Agents will be more likely to emulate those they admire, i.e. those with a higher reputation, and they may change their criteria for judging behaviour (their social norms) if those with a currently high reputation change their behaviour. Reputation within a group is hence a critical element for altering the social norms of that group.

In addition to hierarchical status and reputation, another element that can be central to social behaviour is the relative importance of other group members. We shall refer to this as centrality, though a related idea has been named interdependence in the literature [55]. It echoes the idea of a central node in network theory, but here mean how much another individual matters in a more general sense. The importance of close family members is not necessarily derived from their position in a network. This primitive of centrality can influence the weight given to sanctions or messages from other group members. A sanction from a very central person is likely to be more harder to ignore than one from a peripheral group member.

These three elements of internal group structure can all be important in the spread of social norms, and can all operate differently across cultures.

5.3.2 Culture and Intra-group Behaviour

We hold that the empirically derived Hofstede Dimensions of Culture are at the correct level of abstraction for incorporation into the next generation of NorMAS. The Hofstede theory was derived from questions about everyday practices rather than self-reported values as in the Schwarz system of values [88].

We shall here discuss how three of the Hofstede dimensions may affect the operation of the three group primitives introduced above in the transmission of social norms. The three dimensions will shall focus on are Individualism, Masculinity and Power Distance. The full list of dimensions can be seen in Table 1, and are described in detail in [52, 53].

Power distance determines the degree of importance of hierarchical status in social behaviour. It influences the extent to which the less powerful members of a society expect and accept that power and rights are distributed unequally. In large power distance societies, high status members can disobey social norms with little resistance from those of lower status. Also, the behaviour of others of different status is not a good guide to appropriate behaviour for oneself. In low power distance societies social norms apply much more equally across the population.

Individualism is the degree to which members of a society feel responsible for themselves, or for the larger group they belong to. This dimension relates to the centrality primitive introduced above. Individualistic agents would grant more equal levels of centrality to other agents, while at the other end of the dimension, collectivist agents would be more discriminating in attributing higher centrality to others. It is harder to change one's centrality in collectivist societies where group boundaries are both more relevant and harder to pass.

The dimension of Masculinity represents, in part, the distinction between a strong-handed, competitive orientation, and a consensus-seeking and care-taking orientation in 'feminine' cultures. This would relate most directly to the reputation primitive, with masculine cultures having more unequal distributions of reputation, and social norm infringing acts having significant affects on reputation. In 'feminine' cultures reputation is more evenly assigned and not so easily lost, as such societies are more forgiving than their 'masculine' counterparts. Masculine societies feature more punishment for transgressions, while feminine societies have higher levels of forgiveness for transgressions.

We have argued that one important challenge for NorMAS is to incorporate social context into models of norm transmission. As a further extension, and as a warning against excessive generalisation of culture specific social contexts, we argue that cultural differences in social behaviour are important and that a first representation of some of these differences is possible in the next generation of NorMAS. As stated by Hollander and Wu [54] in their review of normative multi-agent systems, the degree to which relationships influence norm adoption is still under investigation with the social sciences. However, implementations of NorMAS that address this issue can help clarify the questions to be addressed by both social scientists and the NorMAS community.

6   Conclusions

A clear definition of a norm, or definitions of the types of norms that exist, is a necessary first step for a fruitful exploration of the question of norm dynamics. We discussed the definitions of three types of norms and applied these in the later sections of the chapter. The essential cognitive constructs required for agents to reason about norms have been introduced, along with references to more detailed discussions of this issue. The major point to note is that norms can only spread in a population if the relevant mental constructs (i.e., normative beliefs, goals and intentions) propagate as well through the minds of the individuals in that population. The reasons agents may adopt new norms are also surveyed.

This chapter presents a *Norm Lifecycle* that, although it is not identical for the different kinds of norms, starts with a *generation* phase. Briefly, norm generation corresponds to a process where new norms are proposed (either prescribed or emerged). Generated norms may then *spread* through a population and eventually reach *stability* (i.e., persistence) if they are adopted by enough individuals in the population. Afterwards, different evolution stages may occur: norms may *decay* being superceded by new norms; they may evolve; or alternatively, they may be *codified into law*. We interpret last two cases close the lifecycle, since resulting norms can be considered as norm candidates that are being generated. Thus, norm generation is a phase that can be instantiated through different methods. In this chapter we have reviewed a number of alternative approaches such as norm synthesis, agreement, emergence, or empirical learning. Strengths and limitations are discussed in terms of their complexity, required domain knowledge, or time to convergence.

The individual interactions by which norms can spread in a population have been discussed at some length. These include the norms implicitly communicated by practical actions, and more direct signals such as reproaches for the neglect of norms. The complex relationship of social norms with the law has also been elucidated. This is an unduly neglected issue in the legal domain, and one where the multi-timescale, simultaneously micro and macro approach possible with NorMAS is very promising.

We have analyzed the relationship between trust among a group of agents and the dynamics of norms within that group. An overview of definitions of trust was presented,

before we advanced our own position. We hold that for the purpose of analyzing norm dynamics, it suffices to think of trust as an abstraction, a summary of some (perhaps complex) pattern of reasoning, for some model of how others will behave. Trust enables individuals to form expectations regarding the behaviour of others. We have argued that trust is essential for the adoption of a norm, and described how we believe that trust fits into each stage of the norm lifecycle.

We have finally examined how culture might be expected to effect the normative processes presented in this chapter and finally we sketched a method by which variations in human cultures can be incorporated in the next generation of NorMAS.

Acknowledgements

References

1 T. Ågotnes, W. van der Hoek, and M. Wooldridge. Robust normative systems. In Padgham, Parkes, Muller, and Parsons, editors, *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems*, pages 747–754, Estoril, Portugal, May 2008. IFAMAAS/ACM DL.

2 G. A. Akerlof. The market for 'lemons': Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, August 1970.

3 G. Andrighetto, M. Campennì, F. Cecconi, and R. Conte. The complex loop of norm emergence: A simulation model. In Hiroshi Deguchi and et al., editors, *Simulating Interacting Agents and Social Phenomena*, volume 7 of *Agent-Based Social Systems*, pages 19–35. Springer Japan, 2010.

4 G. Andrighetto, M. Campennì, R. Conte, and M. Paolucci. On the immergence of norms: a normative agent architecture. In *Proceedings of AAAI Symposium, Social and Organizational Aspects of Intelligence Washington DC*, 2007.

5 G. Andrighetto, D. Villatoro, and R. Conte. Norm internalization in artificial societies. *AI Communications*, 23(4):325–339, 2010.

6 A. Artikis, D. Kaponis, and J. Pitt. *Multi-Agent Systems: Semantics and Dynamics of Organisational Models*, chapter Dynamic Specifications of Norm-Governed Systems. 2009.

7 R. Axelrod. *The Evolution of Cooperation.* Basic Books, 1984.

8 P. Bain, J. Vaes, Y. Kashima, N. Haslam, and Y. Guan. Folk Conceptions of Humanness: Beliefs About Distinctive and Core Human Characteristics in Australia, Italy, and China. *Journal of Cross-Cultural Psychology*, 43(1):53–58, August 2011.

9 G. S. Becker. Crime and punishment: An economic approach. *The Journal of Political Economy*, 76(2):169–217, 1968.

10 R. Benedict. *The Chrysanthemum and the Sword : Patterns of Japanese Culture.* Houghton Mifflin, Boston, 1989.

11 H. J. Berman. *Law and Revolution. The Formation of the Western Legal Tradition.* Harvard University Press, Cambridge, Massachusetts, 1983.

12 C. Bicchieri. Learning to cooperate. In C. Bicchieri, R. Jeffrey, and B. Skyms, editors, *The Dynamics of Norms*, pages 17–46. Cambridge University Press, 1997.

13 C Bicchieri. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press, New York, 2006.

14 N. Binkin, A. Perra, V. Aprile, A. D'Argenzio, S. Lpresti, O. Mingozzi, and S. Scondotto. Effects of a generalised ban on smoking in bars and restaurants, italy. *The International Journal of Tuberculosis and Lung Disease*, 11(5):522–527, May 2007.

15 J. Bourdon, G. Feuillade, A. Herzig, and E. Lorini. Trust in complex actions. In D. M. Gabbay and L. van der Torre, editors, *Logics in Security*, Copenhagen, Denmark, 2010.

16 R. Boyd, H. Gintis, and S. Bowles. Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare. *Science*, 328(5978):617–620, April 2010.

17 R. Boyd and P. J. Richerson. Group beneficial norms can spread rapidly in a structured population. *Journal of theoretical biology*, 215(3):287–296, April 2002.

18 L. Brooks, W. Iba, and S. Sen. Modeling the emergence and convergence of norms. In Toby Walsh, editor, *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 97–102. IJCAI/AAAI, 2011.

19 G. W. Brown. Iterative solution of games by fictitious play. In T. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 347–376. Wiley, New York, 1951.

20 M. Campennì, G. Andrighetto, F. Cecconi, and R. Conte. Normal = normative? The role of intelligent agents in norm innovation. *Mind & Society*, 8:153–172, 2009.

21 C. Castelfranchi and R. Falcone. The dynamics of trust: from beliefs to action. In *Proceedings of the Workshop on Deception, Fraud and Trust in Agent Societies*, Seattle, WA, 1999.

22 C. Castelfranchi and R. Falcone. Trust is much more than subjective probability: Mental components and sources of trust. In *Proceedings of the 33rd Hawaii International Conference on System Science*, Maui, Hawai'i, January 2000. IEEE Computer Society.

23 C. Castelfranchi and F. Paglieri. The role of beliefs in goal dynamics: Prolegomena to a constructive theory of intention. *Synthese*, pages 237–63, 2007.

24 C. Castelfranchi, G. Pezzulo, and L. Tummolini. Behavioral implicit communication (BIC): Communicating with smart environments. *International Journal of Ambient Computing and Intelligence (IJACI)*, 2(1):1–12, 2010.

25 G. Christelis, M. Rovatsos, and R. Petrick. Exploiting Domain Knowledge to Improve Norm Synthesis. In *Proceedings of he Ninth International Conference on Autonomous Agents and Multiagent Systems*, pages 831–838, 2010.

26 R. B. Cialdini, R. R. Reno, and C. A. Kallgren. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6):1015–1026, June 1990.

27 R. Conte. *L'obbedienza Intelligente*. Laterza, 1998.

28 R. Conte, G. Andrighetto, and M. Campenní. Internalizing norms: A cognitive model of (social) norms' internalization. *International Journal of Agent Technologies and Systems*, 2(1):63–73, 2010.

29 R. Conte, G. Andrighetto, and M. Campennì, editors. *Minding Norms. Mechanisms and dynamics of social order in agent societies*. Oxford University Press, Forthcoming.

30 R. Conte and C. Castelfranchi. *Cognitive and Social Action*. UCL Press, 1995.

31 R. Conte and C. Castelfranchi. The mental path of norms. *Ratio Juris*, 19(4):501–517, 2006.

32 R. Cooter. Structural adjudication and the new law merchant: A model of decentralized law. *International Review of Law and Economics*, 14:215–231, 1994.

33 J. Delgado, J. M. Pujol, and R. Sangüesa. Emergence of coordination in scale-free networks. *Web Intelligence and Agent Systems*, 1(2):131–138, 2003.

34 R. Demolombe and E. Lorini. A logical account of trust in information sources. In *Proceedings of the 11th International Workshop on Trust in Agent Societies*, Estoril, Portugal, may 2008.

35 E. Fehr and U. Fischbacher. Social norms and human cooperation. *Trends in Cognitive Sciences*, 8(4):185–190, 2004.

36 O. Figes. *Crimea.* Penguin, London, 2010.

37 F. Flentge, D. Polani, and T. Uthmann. Modelling the emergence of possession norms using memes. *Journal of Artificial Societies and Social Simulation*, 2001.

38 J. Fracchia and R. C. Lewontin. Does culture evolve? *History and Theory*, 38:52–78, 1999.

39 D. Fudenberg and D. K. Levine. *The Theory of Learning in Games.* MIT Press, Cambridge, MA, 1998.

40 F. Fukuyama. *The Great Disruption. Human nature and the reconstitution of social order.* The Free Press, New York, 1999.

41 D. Gambetta. Can we trust them? In D. Gambetta, editor, *Trust: Making and breaking cooperative relations*, pages 213–238. Blackwell, Oxford, UK, 1990.

42 F. Giardini, G. Andrighetto, and R. Conte. A cognitive model of punishment. In *COGSCI 2010, Annual Meeting of the Cognitive Science Society 11-14 August 2010,*. Portland, Oregon, 2010.

43 H. Graham. Smoking prevalance among women in the European Community 1950–1990. *Social Science & Medicine*, 43(2):243–254, 1996.

44 N. Griffiths and M. Luck. Norm Emergence in Tag-Based Cooperation. In *9th International Workshop on Coordination, Organization, Institutions and Norms in Multi-Agent Systems. 79-86*, 2010.

45 H. L. A. Hart. *The Concept of Law.* Clarendon Press, Oxford, 1961.

46 F. A. Hayek. *Law, Legislation and Liberty. Volume I. Rules and Order.* University of Chicago Press, Chicago, 1973.

47 M. Hechter and K. D. Opp. *Social Norms.* Russell Sage Foundation, New York, New York, 2001.

48 J. Henrich, R. Boyd, and P. Richerson. Five Misunderstandings About Cultural Evolution. *Human Nature*, 19(2):119–137, June 2008.

49 J. Henrich, S. J. Heine, and A. Norenzayan. The weirdest people in the world? *The Behavioral and brain sciences*, 33(2-3):61–83; discussion 83–135, June 2010.

50 G. M. Hodgson and T. Knudsen. The nature and units of social selection. *Journal of Evolutionary Economics*, 16:477–489, 2006.

51 M. Hoffmann. Self-organized criticality and norm avalanches. In *In Proceedings of the Symposium on Normative Multi-Agent Systems. Hatfield, UK: AISB.*, 2005.

52 G. Hofstede. *Culture's Consequences.* Sage Publication, 2nd edition, 2001.

53 G. Hofstede, G. J. Hofstede, and M. Minkov. *Cultures and Organizations: Software for the Mind, Third Edition.* McGraw-Hill, 2010.

54 C. D. Hollander and A. S. Wu. The current state of normative agent-based systems. *Journal of Artificial Societies and Social Simulation*, 14(2):6, 2011.

55 C. Horne. Explaining Norm Enforcement. *Rationality And Society*, 19:2, 2007.

56 D. Ibbetson. Custom in medieval law. In A. Perreau-Saussine and J. B. Murphy, editors, *The Nature of Customary Law. Legal, Historical and Philosophical Perspectives*, pages 151–176. Cambridge University Press, 2007.

57 E. T. Jaynes. *Probability Theory: The Logic of Science.* Cambridge University Press, Cambridge, UK, 2003. (Edited by G. L. Bretthorst).

58 A. J. I. Jones. On the concept of trust. *Decision Support Systems*, 33(3):225–232, 2002.

59 M. Kaisers and K. Tuyls. FAQ-learning in matrix games: Demonstrating convergence near nash equilibria, and bifurcation of attractors in the battle of sexes. In *Workshop on Interactive Decision Theory and Game Theory*, 2011.

60 J. E. Kittock. The impact of locality and authority on emergent conventions: Initial observations. In B. Hayes-Roth and R. E. Korf, editors, *Proceedings of the 12th National Conference on Artificial Intelligence*, pages 420–425. AAAI Press / The MIT Press, 1994.

61 B. Kokinov, A. Karmiloff-Smith, and N.J. Nersessian, editors. *Beyond the Carrot and Stick Approach to Enforcement: An Agent-Based Model.* European Conference on Cognitive Science, New Bulgarian University Press, 2010.

62 J. Lang, M. Spear, and S. F. Wu. Social manipulation of online recommender systems. In *Proceedings of the 2nd International Conference on Social Informatics*, Laxenburg, Austria, 2010.

63 U. Lotzmann, M. Mohring, and K. Troitzsch. Simulating the emergence of norms in different scenarios. *Artificial Intelligence and Law*, forthcoming.

64 J. Madureira, A. Mendes, S. Almeida, and J. P. Teixeira. Positive impact of the portuguese smoking law on respiratory health of restaurant workers. *Journal of Toxicology and Environmental Health, Part A: Current Issues*, 75(13–15):776–787, 2012.

65 D. H. McKnight and N. L. Chervany. The meanings of trust. Working Paper 96-04, Carlson School of Management, University of Minnesota, 1996.

66 H. Mercier and D. Sperber. Why do humans reason? Arguments for an argumentative theory. *The Behavioral and brain sciences*, 34(2):57–74; discussion 74–111, April 2011.

67 M. Miceli and C. Castelfranchi. The mind and the future the (negative) power of expectations the mind and the future the (negative) power of expectations the mind and the future. the (negative) power of expectations. *Theory and Psychology*, 12:335–366, 2002.

68 G. Miller, E. Galanter, and K.H. Pribram. *Plans and the structure of behavior.* New York: Holt, Rinehart and Winston., 1960.

69 W. Mitchell. *An Essay on the Early History of the Law Merchant.* Cambridge University Press, Cambridge, 1904.

70 K. Miyazawa. On the convergence of the learning process in a 2×2 non-zero-sum two-person game. Econometric Research Program, Research Memorandum 33, Princeton University, Princeton, 1961.

71 U. Mons, G. E. Nagelhout, R. Guignard, A. McNeill, B. van den utte, M. C. Willemsen, H. Brenner, M. Pötschke-Langer, and L. P. Breitling. Comprehensive smoke-free policies attract more support from smokers in europe than partial policies. *European Journal of Public Health*, 22, 2012.

72 M.d. Mooij. *Consumer Behavior and Culture: Consequences for Global Marketing and Advertising.* Sage, Thousand Oaks, California, 2004.

73 J. Morales, M. Lopez-Sanchez, and M. Esteva. Using experience to generate new regulations. In Toby Walsh, editor, *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pages 307–312. IJCAI/AAAI, 2011.

74 L. Mui, M. Moteashemi, and A. Halberstadt. A computational model of trust and reputation. In *Proceedings of the 35th Hawai'i International Conference on System Sciences*, 2002.

75 J. Porter. Custom, ordinance and natural right in grantian's decretum. In A & J B Murphy Perreau-Saussine, editor, *The Nature of Customary Law. Legal, Historical and Philosophical Perspectives*, pages 79–101. Cambridge University Press, 2007.

76 E. A. Posner. *Law and Social Norms.* Cambridge MA: Harvard University Press, 2000.

77 P. Resnick and R. Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system. In M. R. Baye, editor, *The Economics of the Internet and E-Commerce*, pages 127–157. Elsevier Science, Amsterdam, 2002.

78 P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation systems: Facilitating trust in internet interactions. *Communications of the ACM*, 43:45–48, 2000.

79 A Ross. *Directives and Norms*. Routledge and Kegan Paul, London, 1969.

80 O. Rutter. *The History of the Seventh (Service) Battalion, The Royal Sussex Regiment, 1914-1919*. Times Publishing Company, London, 1934.

81 N. Salazar, J. A. Rodríguez-Aguilar, and J. L. Arcos. Convention emergence through spreading mechanisms. In W. van der Hoek, G. A. Kaminka, Y. Lespérance, M. Luck, and S. Sen, editors, *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, pages 1431–1432. IFAAMAS, 2010.

82 A. Salehi-Abari and T. White. Trust models and con-man agents: From mathematical to empirical analysis. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, Atlanta, Georgia, 2010.

83 G. Sartor, R. Conte, and R. Falcone. Introduction. agents and norms: How to fill the gap? *Artificial Intelligence and Law*, 7:1–15, 1999.

84 B. T. R. Savarimuthu and S. Cranefield. A categorization of simulation works on norms. In G. Boella, P. Noriega, G. Pigozzi, and H. Verhagen, editors, *Normative Multi-Agent Systems*, number 09121 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2009. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany.

85 B. T. R. Savarimuthu and S. Cranefield. Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. *Multiagent and Grid Systems*, 7(1):21–54, 2011.

86 B. T. R. Savarimuthu, S. Cranefield, M. A. Purvis, and M. K. Purvis. Obligation norm identification in agent societies. *Journal of Artificial Societies and Social Simulation*, 13(4):3, 2010.

87 F. Schauer. Pitfalls in the interpretation of customary law. In A & J B Murphy Perreau-Saussine, editor, *The Nature of Customary Law. Legal, Historical and Philosophical Perspectives*, pages 13–35. Cambridge University Press, 2007.

88 S. H. Schwartz. A Theory of Cultural Value Orientations: Explication and Applications. *Comparative Sociology*, 5(2-3):137–182, 2006.

89 S. Sen and S. Airiau. Emergence of norms through social learning. In M. M. Veloso, editor, *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 1507–1512, 2007.

90 Y. Shoham and M. Tennenholtz. On social laws for artificial agent societies: off-line design. *Journal of Artificial Intelligence*, 73(1-2):231–252, February 1995.

91 Y. Shoham and M. Tennenholtz. On the emergence of social conventions: Modeling, analysis, and simulations. *Artificial Intelligence*, 94(1-2):139–166, 1997.

92 M. P. Singh. Norms as a basis for governing sociotechnical systems. *ACM Transactions on Intelligent Systems and Technology*, (to appear), 2013.

93 C. R. Sunstein. Social norms and social roles. *Columbia Law Review*, 96(4):903–968, 1996.

94 P. Sztompka. *Trust: A Sociological Theory*. Cambridge University Press, Cambridge, UK, 1999.

95 Y. Tang, K. Cai, P. McBurney, E. Sklar, and S. Parsons. Using argumentation to reason about trust and belief. *Journal of Logic and Computation*, (to appear), 2012.

96 E. Ullmann-Margalit. *The Emergence of Norms*. Oxford University Press, Oxford, 1977.

97 D. Villatoro, G. Andrighetto, J. Brandts, J. Sabater-Mir, and R. Conte. Distributed punishment as a norm-signalling tool. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, Valencia, Spain, 2012.

98 D. Villatoro, G. Andrighetto, R. Conte, and J. Sabater-Mir. Dynamic sanctioning for robust and cost-efficient norm compliance. In Toby Walsh, editor, *Proceedings of the 22nd*

International Joint Conference on Artificial Intelligence, pages 414–419, Barcelona, 2011. IJCAI/AAAI.

99 D. Villatoro, J. Sabater-Mir, and S. Sen. Social instruments for robust convention emergence. In Toby Walsh, editor, *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pages 420–425. IJCAI/AAAI, 2011.

100 S. Weintraub. *Silent Night: The Story of the World War I Christmas Truce.* Simon and Schuster, New York, 2001.

101 Von G.H. Wright. *Norms and action.* Routledge and Kegan Paul, London, 1963.

102 E. Xiao and D. Houser. Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 102(20):7398–7401, May 2005.