

Toleration and the Design of Norms

Luciano Floridi

Abstract One of the pressing challenges we face today—in a post-Westphalian order (emergence of the state as the modern, political information agent) and post-Bretton Woods world (emergence of non-state multiagent systems or MASs as “hyperhistorical” players in the global economy and politics)—is how to design the right kind of MAS that can take full advantage of the socio-economic and political progress made so far, while dealing successfully with the new global challenges that are undermining the best legacy of that very progress. This is the topic of the article. In it, I argue that (i) in order to design the right kind of MAS, we need to design the right kind of norms that constitute them; (ii) in order to design the right kind of constitutive norms, we need to identify and adopt the right kind of principles of normative design; (iii) toleration is one of those principles; (iv) unfortunately, its role as a foundation for the design of norms has been undermined by the “paradox of toleration”; (v) however, the paradox can be solved; (vi) so toleration can be reinstated as the right kind of foundational principle for the design of the right kind of norms that can constitute the right kind of MAS that can operate across cultures, societies and states, to help us to tackle the new global challenges facing us.

Keywords Justice · Kant · Multiagent systems · Toleration · Rawls · Westphalian order

Introduction: Prehistory, History, and Hyperhistory

History has lasted 6,000 years, since it began with the invention of writing in the fourth millennium BC. During this relatively short time, Information and Communication Technologies (ICTs) have provided the *recording* and *transmitting*

L. Floridi (✉)

Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3JS, UK
e-mail: luciano.floridi@oii.ox.ac.uk

infrastructure that made the escalation of other technologies possible, with the direct consequence of furthering our dependence on more and more layers of technologies. Today, we are experiencing a radical transformation in our ICTs that could prove equally significant, for we have started drawing a new threshold between history and a new age, which may be aptly called *hyperhistory* (Floridi 2014). Let me explain.

Prehistory, history, and hyperhistory work like adverbs: they tell us *how* people live, not *when* or *where*. From this perspective, human societies currently stretch across the three ages, as ways of living. According to reports about an unspecified number of uncontacted tribes in the Amazonian region, there are still some societies that live prehistorically, without ICTs and recorded documents. If one day such tribes disappear, the end of the first, historical chapter in human life will have been written. Most people today live historically, in societies that rely on ICTs to *record* and *transmit* data of all kinds. In such historical societies, ICTs have not yet overtaken other technologies, especially energy-related ones, in terms of their vital importance. Then there are some people around the world who are already living hyperhistorically, in societies or environments where ICTs and their data *processing* capabilities are the necessary condition for the maintenance and any further development of societal welfare, individual well-being, as well as personal and collective flourishing. The nature of conflicts provides a sad test for the reliability of this tripartite interpretation of human evolution. Only a society that lives hyperhistorically can be vitally threatened informationally, by a cyber attack. Only those who live by the digit may die by the digit (Floridi and Taddeo 2014).

Such a long-term perspective should help to explain the slow and gradual process of political *apoptosis* that we are undergoing, to borrow a concept from cell biology. Apoptosis (also known as programmed cell death) is a natural and normal form of self-destruction in which a programmed sequence of events leads to the self-elimination of cells. Apoptosis plays a crucial role in developing and maintaining the health of the body. One may see this as a dialectical process of renovation, and use it to describe the development of states into information societies in terms of political apoptosis, in the following way.

Simplifying, a quick sketch of the last four hundred years of political history may look like this. The Peace of Westphalia meant the end of World War Zero, namely the Thirty Years' War, the Eighty Years' War, and a long period of other conflicts during which European powers, and the parts of the world they dominated, massacred each other for economic, political and religious reasons. The new system that emerged in those years, the so-called *Westphalian order*, saw the coming of maturity of sovereign states and then national states as we still know them today. The state became not a monolithic, single-minded, well-coordinated agent, the sort of beast (Hobbes' Leviathan) or robot that a later, mechanical age would incline us to imagine (we still speak of Heads of States). It never was. But it did raise to the role of the binding power, the network able to keep together, influence, and coordinate all the different agents and behaviours falling within the scope of some geographical borders. In modern times, the physical and legal spaces overlap and both are governed by sovereign powers, which exercise control through physical force to impose laws and ensure their respect within the national borders. Sovereign states act as multiagent systems (or MASs, more on them later) that can, for

example, raise taxes within their borders and contract debts as legal entities (sovereign debt), and of course dispute borders. Part of the political struggle becomes not just a silent tension between different components of the state-MAS, say between the king and the nobility, or between the clergy and the aristocracy, but an explicitly codified balance between the different agents constituting it. In particular, Montesquieu suggests the classic division of the state's political powers that we take for granted today. The State-MAS organises itself as a network of three main "small worlds"—a legislature, an executive, and a judiciary—among which only some specific kinds of information channel are allowed. Today, we may call this Westphalia 2.0.

With Westphalia 2.0, modern history becomes the age of the state, and the state becomes *the* information agent, which legislates on and controls (or at least tries to control), insofar as it is possible, the technological means involved in the information life-cycle, including education, census, taxes, police records, written laws and rules, press, mass media, and intelligence. The state thus ends by fostering the development of ICTs as a means to exercise and maintain political power, social control, and legal force, but in so doing it also undermines its own future as the only, or even the main, information agent. ICTs, as one of the most influential forces that made the state possible and then predominant as a historical driving force in human politics, also contributed to make it less central, in the social, political and economic life across the world, putting pressure on centralised government in favour of distributed governance and international, global coordination (Floridi 2013a). Today, we know that global problems—from the environment to the financial crisis, from social justice to intolerant religious fundamentalisms, from peace to health conditions, from terrorisms to dictatorships—cannot rely on sovereign states as the only sources of a solution because they involve and require global agents and their coordination. Politics exclusively based on sovereign states (Westphalian system) and Universal Human Rights (Enlightenment) is insufficient. However, in a post-Westphalian world (Linklater 1998) there is much uncertainty about the new MASs involved in shaping humanity's present and future.

One of the pressing challenges we face, in a post-Westphalian (emergence of the state as the modern, political information agent) and post-Bretton Woods (emergence of non-state MASs as hyperhistorical players in the global economy and politics) world, is how to design the right sort of MASs that could take full advantage of the socio-economic and political progress made in modern history, while dealing successfully with the new global challenges that are undermining the best legacy of that very progress in hyperhistory. This is the topic of the article. In the following pages, I shall argue that

- in order to design the right MAS, we need to reconsider the nature of norms not merely as rules that qualify relations among agents but as constitutive dynamic agents in themselves, which give rise to MAS;
- norms as dynamic agents are artefacts subject to the same analysis we usually adopt when we design systems such as software applications;
- the design of norms as artefacts is oriented by fundamental principles provided by the purposes which norms are envisaged to fulfil;

- toleration and justice are the two fundamental principles of normative design that have been used to pursue the Westphalian political project (toleration) and the political project of the Enlightenment (justice);
- the Westphalian political project is undermined by the paradox of toleration, whereas the political project of the Enlightenment is undermined by the circularity between justice constraining toleration which grounds justice;
- the circularity seems inescapable but the paradox of toleration can be solved, thus re-launching the viability of a neo-Westphalian, or, better, hyperhistorical political project, which is to be based on tolerant norms giving rise to tolerant MAS that can operate globally, across cultures, societies and States.

This is quite a lot of work, so the time has come to roll up our sleeves.

The Design Problem: Constructing the Right Kind of MAS

In mechanical contexts, we are used to seeing elements of a system as components, which come together to give rise to a more complex artefact capable of more sophisticated behaviour. Things come first; their relations and the processes that link them come next. A classic example is a clock. However, in network contexts, the reasoning is reversed. We take for granted that relations and processes come first, and that their static intersections or dynamic interactions give rise to nodes, which we then understand as components of the system. A classic example is a roundabout, which exists only insofar as roads intersect. The level of abstraction (Floridi 2008b, 2009) that interprets systems as mechanisms still pervades much of our socio-legal and political thinking, based as it is on the intuitive primacy of entities, being these persons, physical objects, or institutions. The schema is simple: first you have entities, such as Alice and a house, then you have relations between entities, such as that of ownership, and finally you have norms qualifying such relations, for example establishing that Alice legally owns the house. It is a bottom-up, manufacturing perspective, in which norms are seen as constraining forces or costs (second order qualifications, in logical terms, or adverbs in linguistic terms) of pre-existing behaviours or features of the world. The appropriate game analogy here is with football and its agency-constraining rules.

I do not intend to criticise such a historical and mechanical level of abstraction, but in this article I wish to explore a hyperhistorical, network-oriented approach instead. I shall adopt a broad characterization of norms as instructional information oriented to shaping the behaviour of agents, which therefore cannot be true or false but more or less effective.¹ And I shall analyse norms as agents themselves, which, together, may give rise to multiagent systems (MASs) as players in the space of socio-political and economic interactions. All this requires a clarification and the

¹ There is no agreement but (Hollander and Wu 2011) identify some convergence on a cluster of meanings attached to the concept of norm in various fields, especially AI, deontic logic, jurisprudence, philosophy, and social sciences.

elimination of some potential misunderstandings due to some powerful “semantic attractors”.

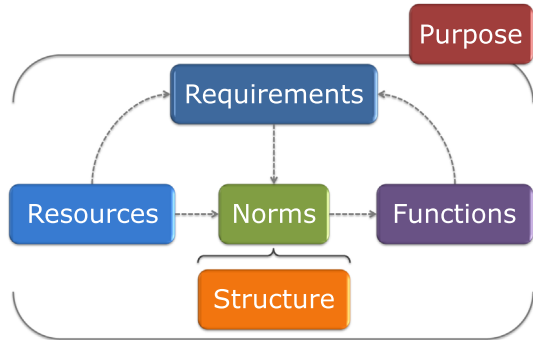
The perspective according to which norms are agency-constitutive (or, better, agent-constitutive) is more top-down. The game analogy here is with chess: MASs are like chess pieces, which are entirely and only constituted by the set of rules (norms) that establish their behaviour. At this point, a powerful “semantic attractor” may occur in the reader’s mind. The distinction between regulative and constitutive norms is well-known and popular at least since (Searle 1995).² What needs to be stressed here is that by “constitutive” I mean something a bit more radical than what Searle seems to have in mind. For I intend to refer to a relation that establishes an *entity*—as in “the pawn is constituted by...”—not just one that establishes a *practice* or a *social fact* among pre-existing entities—as in “chess is constituted by...”. The legal theorist may compare this to the difference between a constitutive, normative analysis of Alice and her house, e.g. as owner and owned good, and a constitutive analysis of the ownership relation. The computer scientist may draw on her knowledge of object oriented programming (Bohl and Rynn 2008), in which an object is constituted by state (data) and behaviour (code), in order to understand this perspective. Indeed, I shall borrow quite freely from the terminology developed in software engineering, and compare norms to software applications (apps), and entities or agents, including MAS, to applications suites. Finally, the philosopher may refer to versions of structural realism in philosophy of science, especially informational structural realism, to understand how entities may be constituted by structural relations (Floridi 2003, 2008a).

Given the change in the level of abstraction adopted, in a network-based ontology the question about the design of MASs becomes the question about the design of their constitutive norms. How can we design the right kind of norms constituting the right kind of MASs that are required today in order to deal successfully with the global problems facing contemporary information societies? This is one of the most pressing questions we are facing today about our hyperhistorical predicament. The question is daunting, and its answer, which we are constantly giving de facto by simply going about our businesses as usual, is immensely consequential. Fortunately, a more reflective approach to deal with it can leverage a reliable and piecemeal strategy, consisting in treating norms themselves as agents. In this case too, it is important to distinguish such a strategy from two other “semantic attractors”, namely two neighbouring approaches³ in the theory of normative MASs, which investigate the normative nature of MASs (how one can design and build MASs whose behaviours are regulated normatively, e.g. in order to make them more efficient and effective) or analyse social norms through models based on MASs (e.g. to understand a social organisation). Both lines of research are relevant, but what I am interested in here is rather treating agent-constitutive norms as themselves agents, by interpreting them as dynamic artefacts.

² Here I am following the Searle-inspired analysis provided by (Boella and van der Torre 2004) and (Boella and der Torre 2006) in connection with normative multiagent systems.

³ See for example (Boella et al. 2006) and (Andrighetto and van der Torre 2013). For a valuable review see (Hollander and Wu 2011).

Fig. 1 Norms as artefacts (construed agents)



The previous clarifications should explain why the question about the design of norms breaks down into a series of questions about the usual components in the design of any artefact. In the terminology common in software engineering, one can analyse norms as having a *structure*, designed according to a set of *requirements*, which are specified on the basis of available *resources*, in order to implement a set of desired *functions*, and all this in view of achieving some ultimate *purpose* (see Fig. 1). Clearly, it is the ultimate purpose that drives the design of norms as a specific kind of artefact, i.e. construed agents that constitute a MAS. Think of the design of a dishwasher: it is the ultimate purpose of cleaning dishes and kitchen utensils that drives its design. Norms are equally purpose-oriented or teleological. So the focus must be on the role played by the purpose that their design is supposed to fulfil. Only some minimal details need to be provided here about the structure, requirements, and functions of norms. This is the task of the next two sections, where I shall quickly outline these aspects of norms as a stepping-stone towards the analysis of their purpose and the design principles that shape them.

The Logical Structure of Norms

The literature on the logical structure of norms is vast and divergent (Ingram 2006), yet this is not a hurdle for the task at hand. What we need here is not a satisfactory and uncontroversial model that can capture the single logical structure underneath all sorts of existing norms. This is probably unachievable or too generic to be useful. What we need is rather a blueprint to design some kind of norms. In other words, we need a minimal articulation of the formal components that may be found commonly enough in at least some norms that we may wish to produce. Such a blueprint will suffice to help us in understanding how norms work functionally and teleologically. So the reasoning is not: “this is the logical structure of all norms, and therefore...”. The reasoning is rather: “at least some norms can be designed as having this logical structure, and therefore...”. The former reasoning is highly questionable, the latter unassailable, hence the preference. The reader who, following (Hart 2012) for example, disagrees with the structure proposed below can still concede that we may use it as a benchmark. With an analogy, we need a road (blueprint) that can take us

to the place we want to reach (the ultimate purpose and hence the principle that should drive the design of norms). The fact that it may not be the only road (on which I agree) or even the shortest (on which I disagree) is irrelevant, as long as it gets us there.

According to a classic interpretation, famously supported by Hobbes in the *Leviathan* (1651, see now (Hobbes 2012), XV. 41), and logically refined by John Austin (Austin 1832, 1995), norms are like *commands*. They convey the information that, given some circumstances C, you ought (O) to do A (A). Or, to use John Langshaw Austin’s classic analysis of locutionary acts (Austin 1962, 1975), norms are like *imperatives*. For example, if the traffic red light is showing, then you ought to stop the car behind the stop line. Slightly more formally, the Hobbes-Austin-Austin interpretation of a norm states that:

$$C \rightarrow O\underline{A} \tag{1}$$

Given (1), we can apply Hare’s Law in Deontic Logic, which states that necessarily (\Box) an ought judgement entails the corresponding imperative, namely “you ought to do A” ($O\underline{A}$) entails “do A” (A), that is:

$$(\Box O\underline{A} \rightarrow \underline{A}) \tag{2}$$

and obtain

$$C \rightarrow \underline{A} \tag{3}$$

We can then further unpack the structure of the imperative “do A” in (3) in terms of a sanction: do A (A) *otherwise* a sanction⁴ applies (S), and obtain:

$$C \rightarrow (\underline{A} \text{ otherwise } S) \tag{4}$$

There are then two ways of reading the threat implicit in “otherwise S”. One is mistaken. It interprets “otherwise” as being logically equivalent to “or else” and “or else” as being logically equivalent to “or” understood as an inclusive disjunction (\vee)⁵ thus:

$$C \rightarrow (\underline{A} \vee S) \tag{5}$$

Informally, (5) means: if C occurs, then A or S or both occur. This interpretation causes all sorts of counterintuitive problems, since it follows from (5) that you may do A and still be sanctioned, or that it is true that a sanction may occur anytime it is true that you follow the order and do A. The first paradoxical conclusion follows from the semantics of \vee , the second from the disjunction introduction rule. Both mimic Ross’ paradox, a real difficulty (Ross 1941). And because of such counterintuitive consequences, (5) has been used in order to argue against the “threat view” of imperative logic, e.g. in (Gensler 2002). Yet the argument has no real force,

⁴ A sanction is an action that punishes non-compliant (deviant) agents and/or rewards compliant (conforming) agents. In this article, I simplify using sanction as synonymous with *negative* sanction, that is, punishment.

⁵ This is suggested, for example, in (Gensler 2002), but see (Gensler 2010) for a revision.

because neither of the two consequences is a real paradox since both are based on a mistaken interpretation of “otherwise”. The correct alternative is to interpret “otherwise” still as being logically equivalent to “or else”, but then interpret “or else” as being logically equivalent to an exclusive disjunction (\oplus)⁶ thus:

$$C \rightarrow (\underline{A} \oplus S) \quad (6)$$

Informally, (6) means: if C occurs, then \underline{A} or S occur but not both. This interpretation is preferable as it captures the nature of the threat implicit in the imperative, and avoids the aforementioned difficulties. It is therefore the one I shall adopt in the rest of this article.⁷

According to (6), a norm has the logical structure of an inferential imperative (so-called epictactic inference) connected to a sanction by an exclusive disjunction. If such and such circumstances occur (C), then you, the agent(s) involved in such circumstances, ought to do A , otherwise you will incur some sanctions (S). In our example, if the traffic red light is showing, then you ought to stop the car behind the stop line otherwise you will get a fine. Note that in (6):

1. C is a set of applicability conditions that are either real or to be realised, but may also be empty, e.g. when “do A ” applies always or in any case, independently of any circumstances;
2. A , the action to be performed, is often expressed in negative terms, that is $\neg A$. I shall return to this point in the next section; and
3. S is a set of consequences of non-compliance, which may also be empty, as when there are no threats.

The final result is:

$$C \rightarrow (\neg \underline{A} \oplus S) \quad (7)$$

We can now analyse the two components in (7). The epictactic inference in (7), representable as $C \rightarrow \neg \underline{A}$, answers the “what question”: “what should I do when the light is red?” “do not keep driving!”. It is therefore the place where the agent may find the *informational component* of a norm. It is also the locus in the structure of a norm where the designer may inject some *distributive justice*, understood as the socially just allocation of goods in a society (e.g. “do not take unfair advantage of common resources”). I shall return to this point when discussing Rawls’ theory of justice. For the moment, two features are worth highlighting. First, the aforementioned “injection” is not to be taken for granted at all because, at this stage, (7) could be the source of the most unfair norms, imposed by violence. That this is not so is due to the requirements and purposes that guide its design, as I shall explain in the next two sections. And second, according to Socrates, the information about the

⁶ See for example (Holyoak and Morrison 2012), p. 137; (Johnson-Laird 2006), p. 108.

⁷ For a more sophisticated approach to the difficulties caused by the disjunction, in terms of reparation and “ordered” disjunction, in which S is activated when A is negated, see (Governatori and Rotolo 2006). In this context, the exclusive disjunction adopted suffices to convey the essential structure of a norm, but for some negative effects see (Governatori 2014).

right thing to do, if indeed it is conveyed by (7), is sufficient for doing the right thing, so Socratic agents require only the epitactic inference in a norm as *intrinsic motivation* for coordination, collaboration, and fair competition. This is Socrates' positive anthropology. The same effect is achieved when threats are internalised, e.g. through education or cultural influence ((Durkheim 1976) originally 1912, and (Scott 1971)).

The exclusive disjunction in (7), representable as $\neg A \oplus S$, answers the “why question”: “why should I stop when the light is red?” “because otherwise you will get a fine”. It is therefore the place where the agent may find the *motivational* and *reparational component* of a norm. It is also the locus in the structure of a norm where the designer may inject some *retributive justice*, understood as the socially just allocation of punishment proportionate to the failure to perform a specified action (non compliance). The same consideration about the potentially unfair and violent nature of (7) applies, but now, according to Hobbes, the information about the right thing to do is insufficient for doing the right thing, so Hobbesian agents require both the epitactic inference and the disjunctive sanction for co-existence. This is Hobbes' negative anthropology.

(7) is not a norm, but a schema that generates an infinite set of actual norms. As such, (7) raises at least three fundamental questions. The first is genetic, and it concerns the old tension between positivism versus naturalism: what has *logical priority*, the design of (7) and then its top-down, almost deductive application that populates the initially empty set of norms, or a set of primordial norms from which to extract bottom-up, almost inductively, the design of (7) as a general schema? The second question is semantic: who has the *legitimacy* to interpret (7), that is, to interpret (7) and thus populate the set of norms (further or originally, depending on the answer to the previous question)? The last question is pragmatic: who has the *authority* and the *power* to apply (7), in its various and specific interpretations as norms? Each question points in the direction of classic debates in the philosophy of law. Fortunately, in this context we do not need to opt in favour of one answer rather than another. The reader should feel free to choose the answers that she finds most convincing because our task is to consider the *requirements* that must be met by the design of norms, independently of foundational issues about logical priority, legitimacy, and authority. In other words, we should not get distracted in our journey.

The Requirements of Norms

In system engineering, a specification for a system to be designed, e.g. an app, contains an explicit description of the requirements that must be satisfied by the system once it is built. Such requirements are usually distinguished into non-functional and functional ones.

Non-functional requirements—also known as *requirements tout court*—define a system's *architecture*, that is, what the system is supposed to *be*. For example, a refrigerator's non-functional requirement is that its compartment is thermally insulated and easily accessible in order to store food. In our case, when the artefact to be designed is a norm, we have already encountered Austin's non-functional

requirement of the *universality* of norms: unless otherwise specified, *any* agent in circumstances C ought to do A. Kant's formulations of the categorical imperative have often been seen as (formal) motivational features of an agent's behaviour, that is, as interpreting the exclusive disjunction in (7). However, from a system design perspective and a more consistent interpretation of Kant's autonomous dictates of reason, they seem to be better understood as non-functional requirements of (7), namely *universalizability* and *non-instrumentalisation* ((Kant 1786), see now (Kant 2011)). Other non-functional requirements include Fuller's classic conditions: norms must be *public*, *not contradictory*, *comprehensible*, and *not ephemeral* (Fuller 1969). The list is not exhaustive and further non-functional requirements may easily be added, such as the preservation of an internal *balance* between distributive and retributive justice. What matters here is that we have a clear idea of some of the necessary non-functional requirements that a good specification for the design of norms should include.

Functional requirements—also simply known as *functions*—define a system's *behaviour*, that is, what the system is supposed to *do*. In our previous example, the function of a refrigerator is to maintain a constant temperature a few degrees above the freezing point of water, inside its easily accessible, thermally insulated compartment. In the case of norms, they implement at least three kinds of functions: a *signalling* (Banks 2001), a *hermeneutical*, and of course a *pragmatic* function. Whether intentionally or not, each of these functions is about conveying some kind of information. Since the pragmatic function is the most important for our current task, let me first comment very briefly on the other two functions.

A norm has a *signalling* function insofar as it works as a message, sent by the agent issuing it, to convey some information about the agent itself. For example, in 2013, the British Government proposed new regulations according to which EU migrants would have to wait for 3 months before applying for Jobseeker's Allowance and other out-of-work benefits. The signalling function of the norm was obvious: the government wanted to reassure the British electorate that it took very seriously the public unease regarding the alleged risk of immigrants coming to the UK just to claim benefits, and that it was ready to take tough actions to prevent such alleged risk.

A norm also has a *hermeneutical* function insofar as it is a means to interpret itself and the social environment that it is shaping. Using the previous example, Sarah Teather, a former Coalition minister in the British Government, criticised the latter for making Britain "hostile" to immigrants, stating that the new proposed regulations were "nakedly political" (Hall July 13 2013). This shows that sometimes the signalling and the hermeneutical function of norms may generate tensions, in our example between home and foreign politics.

Finally, a norm has a *pragmatic* or action-shaping function. Here we find the illocutionary force and perlocutionary effects of (7) (Austin 1962, 1975). These correspond to what the source of a norm wants and what it actually obtains by issuing it. In the running example, the British Government wanted to force EU migrants to wait for 3 months before applying for Jobseeker's Allowance and other out-of-work benefits. As a result, EU citizens may be more reluctant to migrate to the UK. This pragmatic function is usually taken to be the main or even the only

function of norms. Such a view is justified in terms of *logical priority*: a norm can have a signalling and a hermeneutical function only if it has a *prima facie* pragmatic function to begin with, even if this is actually redundant (e.g. another norm already in place achieves the same function) or ineffective (e.g. it cannot work and it may even be known to be unworkable). Yet the same view is mistaken in terms of *explanatory primacy*: it is not always the case that the pragmatic function is actually the reason why a norm is designed. The example above illustrates that the main function of the new regulations proposed by the British Government was not that of regulating the behaviour of EU migrants, especially if it was known that there was no real risk, or that the new regulations in question were going to be ineffective anyway (they would make no difference), or redundant (the difference they would make would overlap with the difference already made by previous legislation). Their main function was signalling, with the pragmatic function a necessary means, and the hermeneutical function a perhaps unwanted but inevitable consequence.

As in the case of the signalling and the hermeneutical function, the pragmatic function is also a matter of information. Recall that we have modelled norms as commands. This may be controversial but it is not implausible: norms are there to communicate to agents what to do. The fact that norms then seek to entice or force agents to adopt a specified behaviour is secondary and, in the case of Socratic agents, may even be unnecessary. Indeed, agents may often demand and welcome norms, in order to know what to do (e.g. “on which side of the road should I drive?”).

It is commonly assumed that the pragmatic function (action-shaping communication) of norms is to regulate social and not merely individual behaviour. We have seen above that this leads to a negative formulation of the action to be undertaken: an agent is not told what to do in isolation, but how to adapt or restrain his or her behaviour when interacting with other agents. However, since the individual and the social sphere cannot always be clearly distinguished, heated disputes regularly ensue on the interpretation of the social aspects, implications, or ramifications of norms about personal choices concerning abortion, assisted suicide, divorce, euthanasia, polygamy, recreational drugs, same-sex marriage, and more generally all those cases in which the rivers of individual behaviours flow into the sea of social behaviour, forming an estuary hard to navigate. If, for the sake of our current investigation, we simplify and assume a clear distinction between individual and social behaviour, then it becomes easier to see that, through their pragmatic function, norms are designed

- (a) to avoid or resolve conflicts among agents, and
- (b) to help agents to coordinate their actions, collaborate, and compete fairly.

Note that only (a) has Hobbesian agents as its main referent, yet norms are as much about (b) as about (a). It is simply false that “If men were angels, no government would be needed” (Minogue 2000, p. 94). Even in a society of angels or Socratic agents, norms and their governance would be necessary to coordinate all the well-intended interactions. Indeed, one may conceive a perfectly peaceful and tolerant world in which (a) is unnecessary, but not one in which (b) is avoidable.

Given the structure and functions of norms, and what they can achieve in their application, what is their ultimate purpose? Answering this question means identifying the most general principles that guide the design of norms as informational agents, a task for the next section.

The Purposes of Norms

In the best circumstances, norms are designed—or at least are presented as being designed—to improve the co-existence of agents. At different times in political history, this has meant delivering a *peaceful society* (PS), a *tolerant society* (TS), a *liberal society* (LS), or (an inclusive or) a *just society* (JS). In the terminology adopted in this article, PS, TS, LS, and JS are the principles that guide the design of “normative artefacts” as their ultimate purposes.

As design principles, PS, TS, LS, and JS are not logically independent. Famously, Locke provides a convincing and influential argument in favour of the link between a peaceful and a tolerant society in his “A Letter concerning Toleration”. As he writes:

No peace and security, no, not so much as common friendship, can ever be established or preserved amongst men so long as this opinion prevails, that dominion is founded in grace and that religion is to be propagated by force of arms (Locke 1689), p. 18, see now (Locke 1991).

Locke is careful in formulating the connection between PS and TS negatively. He does not argue that TS is sufficient to deliver PS, only that it is necessary. Realistically, there is no set of individually necessary and jointly sufficient conditions that can ensure the delivery of PS. If nothing else, human insanity may just be unavoidable. But there are plenty of conditions without which PS is unachievable or at least less likely to occur and, for Locke, TS is one of the most important. This seems to be correct. TS is a fundamental principle that should orient the design of norms because toleration is required to build the kind of peaceful society that all agents, by default, wish to enjoy and hence have an interest in establishing. Without toleration there is no peace, in short:

$$\neg TS \rightarrow \neg PS \quad (8)$$

Mill argues in a similarly negative way when he connects the lack of tolerance to the lack of liberty (LS):

it is not the feeling sure of a doctrine (be it what it may) which I call an assumption of infallibility. It is the undertaking to decide that question for others, without allowing them to hear what can be said on the contrary side. And I denounce and reprobate this pretension not the less if it is put forth on the side of my most solemn convictions. (Mill 1859) p. 45, see now (Mill 1963), vol. 18, pp. 213–310.

It turns out that toleration is also a necessary condition for liberty, so:

$$\neg TS \rightarrow \neg LS \quad (9)$$

Adding together the consequents of the same antecedent, (8) and (9) suggest that an intolerant society is neither peaceful nor liberal, that is:

$$\neg TS \rightarrow (\neg PS \wedge \neg LS) \quad (10)$$

At this point, it would be a logical fallacy to assert that, unless norms are designed to deliver a tolerant society, they will fail to deliver a peaceful and liberal society. Yet one only needs to refine such inference in terms of conditions of simple possibility, or qualified probability (that is, transcendentally⁸) to reach a logically sound and convincing principle of normative design. A tolerant society makes it possible, or indeed probable (or more probable, or highly probable... views diverge on the qualifications) to deliver a peaceful and liberal society. This has seemed highly plausible to most, yet a difficult question still remains: what is then the relation between toleration and justice? Does one design principle deliver the other, or are both necessary in order to deliver peace and liberty? In a world so torn apart by religious and cultural clashes, this is not an idle question.

The Enlightenment tradition, from Kant to Rawls,⁹ seems to argue in favour of a grounding of toleration on the basis of justice. “Seems”, because the line of reasoning is more complex. The following passage by Rawls illustrates very well the difficulty:

The characteristic feature of these arguments for liberty of conscience [our TS] is that they are based solely on a conception of justice [our JS]. Toleration [our TS] is not derived from practical necessities or reasons of state. **[a]** Moral and religious freedom [our TS] follows from the principle of equal liberty [our LS], and assuming the priority of this principle, **[b]** the only ground for denying equal liberties [our LS] is to avoid an even greater injustice [our $\neg JS$], an even greater loss of liberty [our $\neg LS$] (Rawls 1999b), p. 188.

Looking at the bare logical structure of the argument, the first half, which I marked as **[a]**, simply rephrases (9): TS follows from LS, which means $\neg TS \rightarrow \neg LS$. Rawls agrees with Mill, as expected: toleration is likely to deliver liberty. But there is more. The second half of Rawls’ reasoning, which I marked as **[b]**, may seem to argue for the priority of liberty (LS) first and then of justice (JS). Not so. It is actually arguing for the primacy of toleration (TS). Let see why.

The passage can first be simplified into:

⁸ Much philosophy goes astray because it loses sight of the transcendental, the thin line between the immanent, which is knowable but not theoretical, and the transcendent, which is theoretical but not knowable.

⁹ “The theory that results [i.e. *A Theory of Justice*] is highly Kantian in nature”, first Preface to (Rawls 1999b), p. xviii.

(i) the only ground for denying LS is to avoid $\neg JS$ and/or $\neg LS$

which can then be further simplified into:

(ii) not LS only if (if LS then $\neg JS$ and/or $\neg LS$)

which is formalised into:

(iii) $\neg LS \rightarrow (LS \rightarrow (\neg JS \vee \neg LS))$

which is equivalent to:

(iv) $\neg(LS \rightarrow (\neg JS \vee \neg LS)) \rightarrow LS$

which is equivalent to:

(v) $(LS \wedge \neg(\neg JS \vee \neg LS)) \rightarrow LS$

which is equivalent to:

(vi) $LS \wedge JS \wedge LS \rightarrow LS$

which can then be further simplified into:

(vii) $JS \rightarrow LS$

which is equivalent to:

(viii) $\neg LS \rightarrow \neg JS$

which, combined with [a]/(9): $\neg TS \rightarrow \neg LS$, leads to

$$\neg TS \rightarrow \neg JS \tag{11}$$

What Rawls is actually arguing is that toleration is a necessary condition also for justice. In the transcendental language adopted above, a tolerant society is a condition of possibility of a just society, and can make the latter more likely. Toleration should be the ultimate principle in the design of norms.

(11) is not a mere logical accident, due to (my) peculiar choice of an unusual text in Rawls' vast production. It is actually evidence of a deep and significant tension, in Rawls' philosophy and more generally in Modernity, between justice and toleration as the two principles shaping the construction of (norms constituting MASs giving rise to) a peaceful and liberal society. A better understanding of the logic of such a tension will enable us to progress with our analysis. Three points will suffice.

First, why is the tension not more evident? It did take quite a bit of manoeuvring above to make it explicit. The fact is that, in *A Theory of Justice*, Rawls manages to keep the tension under control by nuancing the meaning of "justice as fairness" in such a way as to make it largely encompass that of toleration. The very opening of his book is quite indicative:

Toleration is the first virtue of social institutions, as truth is of systems of thought. A theory however elegant and economical must be rejected or revised if it is untrue; likewise laws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are intolerant. Each person

possesses an inviolability founded on toleration that even the welfare of society as a whole cannot override. For this reason toleration denies that the loss of freedom for some is made right by a greater good shared by others. It does not allow that the sacrifices imposed on a few are outweighed by the larger sum of advantages enjoyed by many. Therefore in a tolerant society the liberties of equal citizenship are taken as settled; the rights secured by toleration are not subject to political bargaining or to the calculus of social interests. The only thing that permits us to acquiescence in an erroneous theory is the lack of a better one; analogously, an act of intolerance is just only when it is necessary to avoid an even greater act of intolerance. Being first virtues of human activities, truth and toleration are uncompromising.

If the passage sounds convincing and familiar, yet not entirely right, that is because it is actually a fake: I replaced in it every instance of “justice” and its cognate terms with “toleration” and its cognate terms.¹⁰ The fact that it seems to work perfectly fine indicates how close Rawls’ concept of justice is to his concept of toleration. This is not surprising. Truth and toleration have been strictly related at least since Mill (Stetson 2005).

The second point is that the tension is there nonetheless. Toleration and justice are almost interchangeable, but Rawls did not write *A Theory of Toleration*. And he did not because in the end he considered toleration too problematic a foundation for a liberal and peaceful society. In his view, it is too blunt an instrument for the careful and fine design of norms. To see why, and why this is crucial, let us take a step back.

Consider (10) and (11). Adding together the consequents of the same antecedent, they suggest that:

$$\neg TS \rightarrow (\neg JS \wedge \neg PS \wedge \neg LS) \quad (12)$$

(12) means that toleration is the ultimate principle of design of a fair (in the sense of just), peaceful, and liberal society. Let us call (12) the Westphalian political project.

The Westphalian political project takes full advantage of two features of toleration as a principle of normative design. First, toleration appears to be able to ground justice, peace and freedom and, second, it does not seem to require any further grounding, because it is interpretable as synonymous with socialisation, at least normatively (“socialisation ought to mean toleration”) if not factually. One may argue that an intolerant society is, strictly speaking, a contradiction in terms: it is not a society at all, in the same sense in which a violent peace is not a peace. Accordingly, TS may be presented as a self-booting design principle, the social

¹⁰ Here is the original text: “Justice is the first virtue of social institutions, as truth is of systems of thought. A theory however elegant and economical must be rejected or revised if it is untrue; likewise laws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust. Each person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override. For this reason justice denies that the loss of freedom for some is made right by a greater good shared by others. It does not allow that the sacrifices imposed on a few are outweighed by the larger sum of advantages enjoyed by many. Therefore in a just society the liberties of equal citizenship are taken as settled; the rights secured by justice are not subject to political bargaining or to the calculus of social interests. The only thing that permits us to acquiescence in an erroneous theory is the lack of a better one; analogously, an injustice is tolerable only when it is necessary to avoid an even greater injustice. Being first virtues of human activities, truth and justice are uncompromising” (Rawls 1999b) pp. 3–4.

equivalent of the Cartesian *cogito*. Justice, peace and freedom may require toleration, but toleration requires only itself to ground itself because it is logically equivalent to the exit from a potential condition of conflicting interactions. Socialising means tolerating and vice versa. I am not the first to argue this:

Tolerance is an end in itself. The elimination of violence, and the reduction of suppression to the extent required for protecting man and animals from cruelty and aggression are preconditions for the creation of a humane society (Marcuse 1965), p. 101.

All this appears cogent. There is, however, a “but”. Toleration may be self-booting, but it is not, contrary to the *cogito*, self-reinforcing. The more I doubt, the more I think, the more I can be certain that I am thinking, and hence that I exists as a thinking I. But the more I tolerate, the more I approach a threshold of diversity beyond which toleration becomes unacceptable. This is a well-known “paradox of toleration”.¹¹ There are many formulations of it,¹² and I shall argue that it is really just one horn of a dilemma. For the moment, let us follow Popper, who presents it in a way that is particularly helpful in this context:

Unlimited tolerance must lead to the disappearance of tolerance. If we extend unlimited tolerance even to those who are intolerant, if we are not prepared to defend a tolerant society against the onslaught of the intolerant, then the tolerant will be destroyed, and tolerance with them. In this formulation, I do not imply, for instance, that we should always suppress the utterance of intolerant philosophies; as long as we can counter them by rational argument and keep them in check by public opinion, suppression would certainly be most unwise. But we should claim the *right* [italics in the original] to suppress them if necessary even by force; for it may easily turn out that they are not prepared to meet us on the level of rational argument, but begin by denouncing all argument; they may forbid their followers to listen to rational argument, because it is deceptive, and teach them to answer arguments by the use of their fists or pistols. We should therefore claim, in the name of tolerance, the *right* [my italics] not to tolerate the intolerant. We should claim that any movement preaching intolerance places itself *outside the law* [my italics], and we should consider incitement to intolerance and persecution as *criminal* [my italics], in the same way as we should consider incitement to murder, or to kidnapping, or to the revival of the slave trade, as criminal. Originally (Popper 1945), see now (Popper 2013) Ch. 7, Note 4, p. 581.

Note how too much toleration (Popper’s “unlimited tolerance”) leads to the need of some justice (Popper’s “right”) that must counterbalance its misapplication. This is problematic because TS was supposed to ground JS, not vice versa. In *A Theory of Justice*, Rawls acknowledges such a difficulty, calling it a “practical dilemma”, and makes it quite explicit:

¹¹ Strictly speaking, this is only one of the three paradoxes of toleration discussed in the literature, for an excellent overview see (Forst 2012). See also (Horton 1994).

¹² See (Oberdiek 2001), chapter two for an overview. For a philosophical study of the justifications of toleration see (Mendus 1988). Ricoeur discusses some paradoxes of toleration in (Ricoeur 1996).

Let us now consider whether justice requires the toleration of the intolerant, and if so under what conditions. There are a variety of situations in which this question arises. Some political parties in democratic states hold doctrines that commit them to suppress the constitutional liberties whenever they have the power. Again, there are those who reject intellectual freedom but who nevertheless hold positions in the university. *It may appear that toleration in these cases is inconsistent with the principles of justice, or at any rate not required by them* (Rawls 1999b), p. 190, my italics.

In the end, Rawls is forced to go much further than just accepting some inconsistency or mere independence between justice and excessive toleration. Like Popper, he ends up arguing that justice provides some grounding constraints for the proper exercise of toleration. The ultimate principles of design of a peaceful and liberal society are toleration and justice, but toleration is undermined by its own paradox (there is such a thing as too much toleration) whereas justice is not (there is no such thing as too much justice) and so justice can actually help us to manage toleration satisfactorily. Ultimately, it is really justice that should ground toleration. The result of this shift can be summarised thus:

$$\neg JS \rightarrow (\neg TS \wedge \neg PS \wedge \neg LS) \quad (13)$$

Let us call (13) the political project of the Enlightenment. (13) means that justice is the ultimate principle of design of a tolerant (but not over tolerant), peaceful, and liberal society. The relation of dependence of tolerance on justice may be broadly construed in the same transcendental terms we adopted previously: a just society makes it possible, or indeed probable (or more probable, etc.) to deliver (the right kind of) a tolerant, peaceful and liberal society.

The political project of the Enlightenment, with its reliance on a rationalist understanding of justice, takes full advantage of two features of justice as a principle of normative design. First, justice appears to be able to solve the paradox of toleration by offering the right constraints to avoid its distorted application, while grounding peace and freedom. And, second, justice can act as a self-reinforcing design principle, the juridical equivalent of the Cartesian *cogito*: the more justice is exercised the easier it is to exercise further justice. Justice begets more justice, in a virtuous circle that offers only a positive strengthening and faces no threshold of “the other”.

In this case too, however, there is a “but”. And the “but” is that, like (12), (13) too is not entirely satisfactory. Toleration is self-booting but not self-reinforcing; but justice, which is indeed self-reinforcing, is not self-booting, because it requires toleration to be grounded. Recall that we already reached the conclusion that there is no justice without toleration. An unjust society is not a contradiction in terms, and to avoid it, justice must be framed within a toleration approach towards diversity. Thus, we are caught in a circular reasoning in which justice presupposes toleration, which presupposes justice. This leads to the third and final point I wish to make on the tension between toleration and justice.

The problematic interaction between the two design principles emerges in its full strength in Rawls’ account of international toleration, which has been largely and convincingly criticised for its inability to defend a robust principle of toleration (that is, one that does not collapse into over toleration) when confronted by “decent” but

non-liberal societies. As (Kuper 2000) has pointed out in his criticism of (Rawls 1999a):

the question is not whether to tolerate [other, nonliberal] cultures but rather how to do so. [...] Only an original position that includes all the persons of the world as free and equal persons can be tolerant in the right way (p. 648).

Rawls' theory of justice needs to be adapted to make room for a robust principle of toleration. Otherwise, Rawls' work, squarely placed in the political project of the Enlightenment—justice is the most fundamental principle of normative design—ends up supporting “toleration of the wrong kind” (Kuper 2000, p. 648), which undermines justice itself in its attempt to solve “the paradox of tolerance”. Too much or the wrong kind of toleration corrodes justice.

The difficult relation between toleration and justice is not escaped when we extend the reasoning to include Habermas' “deliberative account”,¹³ Popper's exchange of “rational arguments” or, more generally, some version of the rational discourse. Within the Enlightenment tradition, the dependence of toleration on justice in (13) may be grounded by an appeal to the public use of reason (PUR), that is, to a specific interpretation of a global information society based on open-minded, rational and evidence-based communication.¹⁴ In short:

$$\neg\text{PUR} \rightarrow \neg\text{JS} \rightarrow (\neg\text{TS} \wedge \neg\text{PS} \wedge \neg\text{LS}) \quad (14)$$

This is indeed how O'Neil interprets Kant's position about the relationship between justice and toleration, from a “communication-based” (or, as I prefer, “informational”) perspective (O'Neill 1986). As she writes:

For Kant, the importance of (some sort of) toleration is connected with the very grounding of reason, and so in particular with the grounding of practical reason. His arguments suggest that liberal political thinking can vindicate practices of toleration without commitment either to a strong form of individualism or to the view that we can distinguish “self-regarding” acts, and without claiming that reasoning either has a “transcendent” vindication or is groundless. (p. 523)

Indeed, Kant makes the public exchange and discussion of information a necessary condition for the freedom of thinking:

We do admittedly say that, whereas a higher authority may deprive us of freedom of *speech* or of *writing*, it cannot deprive us of freedom of *thought*. But how much and how accurately would we *think* if we did not think, so to speak, in community (Gemeinschaft) with others to whom we *communicate* our thoughts and who communicate their thoughts to us! We may therefore conclude that the same external constraint which deprives people of the freedom to *communicate* their thoughts in public also removes their freedom

¹³ For an accurate reconstruction and a convincing criticism of Habermas' position on tolerance see (Thomassen 2006).

¹⁴ For another example of Kantian “space of reasonable communication”, see Herman's concept of “a community of moral judgement” in (Herman 1996).

of *thought*, the one treasure which remains to us amidst all the burdens of civil life, and which alone offers us a means of overcoming all the evils of this condition (Kant 1991 (originally 1786)), p. 247.

But the trouble is that communication, and the free, open-minded exchange of information and arguments must ultimately be grounded on toleration of different views that should be expressible, within limits to be specified in a non-circular way. As O'Neill herself acknowledges:

Toleration of public use of reason is on this account necessary for the emergence and maintenance of the increasingly shared standards of reasoning which fully public communication requires. Practices of intolerance may damage the partial standards of reason on which restricted communication also depends. If we undermine the public use of reason by intolerance, all uses of reason are ultimately in jeopardy, [...] (O'Neill 1986), p. 534.

Yet, this merely shifts the problem one step back: what guarantees the tolerant but not over tolerant exercise of the public use of reason, if not some justice? We are back to the difficulty already encountered: the inclusion of the public use of reason in (14) only makes the circular reasoning “wider”, it does not break it.

The time has come to summarise the previous analysis. There seem to be two main principles for the design of norms (as constituting MAS) that can deliver a peaceful and liberal society: toleration and justice. Toleration can ground itself, but runs into the paradoxical problem of its excessive scope. Justice can reinforce itself through its own application (this is what O'Neill calls “recursive use”), and can constrain excessive toleration, but then it requires toleration itself to be grounded. The Westphalian political project is the construction of a society on the basis of toleration first, and hence the delivery of justice, peace, and liberty on its basis. This project has resurfaced in practice whenever fresh starts have been advocated, from the American Revolution to various student movements. It is not mainstream because it has always been blocked by its own paradox: unjust toleration is not a contradiction and seems inevitable. The political project of the Enlightenment may be interpreted as a way of by-passing that obstacle, by proposing the construction of a society on the basis of justice first, and hence the delivery of toleration, peace, and liberty. This project has been the blueprint of mainstream Western liberal democracies for some time, but it is challenged by another difficulty: intolerant justice is also not a contradiction. By opting for the Enlightenment project, Modernity has been struggling with the implementation of tolerant justice since the French Revolution.

What I am going to argue is that the Westphalian project can and should be revived. We should and could reconsider toleration as the primary principle for the design of norms as agents constituting the right kind of MAS needed in a globalised world. Of course, to be viable, such a “neo-Westphalian” or rather hyperhistorical proposal requires a solution of the paradox of tolerance that does not rely circularly on justice. This is what I am going to show in the next section.

A Solution of the Paradox of Toleration

So far I have refrained from defining exactly what I mean by “toleration”, for a general notion of it was sufficient. However, it is well known that the concept may mean different things to different people. In what follows, I need to use “toleration” in two distinct senses, so let me introduce a simple convention. I shall use “toleration” and “tolerance” in the following way:

- *toleration* A is informed about B’s ϕ -ing (believing, thinking, acting, behaving...), disapproves of it, could prevent it, but refrains from doing so. e.g. A exercises toleration towards B’s marrying someone of the same sex.
- *tolerance* C is informed about the imposition of B’s ϕ -ing (believing, thinking, acting, behaving...) on C, C disapproves of it, could resist it, but refrains from doing so. e. g. C exercises tolerance towards B who does not allow C to wear a hijab in public spaces.

According to the previous definitions, toleration is more akin to *respect*,¹⁵ whereas tolerance is more akin to *forbearance*.¹⁶ The distinction facilitates the clarification of four important features sometimes missed by discussions of the paradox. All are needed in order to present a solution of it.

First, each definition has four components: each of them is necessary and together they are jointly sufficient. If there is no information, then there is no toleration/tolerance. The “don’t ask, don’t tell” official United States policy on service by gays and lesbians in the military between 1994 and 2011 was not an act of toleration/tolerance. If there is no disapproval, then there is either indifference or indeed approval, so no toleration/tolerance can be in question. If there is no possibility of preventing or resisting what one disapproves, then there is only unwilling acceptance. The question of toleration/tolerance arises only if one could actually do something about what one is informed about and disapproves. And finally, it is the restraint that shows that toleration/tolerance is being exercised. The intolerant satisfies the first three conditions but not the last.

Second, the *primary* referent of the exercise of toleration/tolerance is some ϕ -ing (believing, thinking, acting, behaving...) not the *source* of that ϕ -ing. In our example, Alice tolerates same-sex marriage primarily and directly, as a social practice or institution, and hence secondarily and derivatively those who support it or engage in it, because of their behaviours or beliefs. The difference in focus is crucial in order to understand how the paradox can be solved. And to explain why, I need to clarify in what sense Popper’s “paradox” is really only one horn of a dilemma. Consider the fourth clause in the previous two definitions: the question there is, when is A’s or C’s refraining morally good? The question may concern the *scope* or *limits* of toleration, and hence lead to Popper’s paradox: sometimes it is wrong to be tolerant. But the question may also concern the *justification* or

¹⁵ For an interpretation of Kant’s conception of toleration as respect, see (Abellán 2012).

¹⁶ The distinction resembles the one endorsed by Mary Warnock between “strong” and weak” toleration (Warnock 1987), although she develops it only from an agent-oriented perspective: in her analysis it is A that can exercise both weak and strong toleration.

groundings of toleration¹⁷: if someone disapproves of something as morally wrong and could prevent it, how can that someone be morally right in tolerating it? This, which is the other horn of the dilemma, is unfortunately also known in the literature as “the paradox of toleration”. It is the only issue on which Bernard Williams concentrates, for example, when discussing the “impossibility” of toleration:

the difficulty with toleration is that it seems to be at once necessary and impossible (Williams 1996), p. 17.

And it is with this second horn of the dilemma (or second paradox) that the fact that the primary referent of the exercise of toleration/tolerance is some ϕ -ing and not someone ϕ -ing helps significantly. For what Williams interprets as a contradiction is a direct consequence of the inability to solve the second horn of the dilemma (Williams’ other paradox of toleration) when the latter is misunderstood as a mere conflict between two parties that need toleration to get along, but cannot implement toleration because they do not get along. Yet this is too coarse an analysis. As already convincingly argued by Susan Mendus,¹⁸ the resolution of such an apparent impasse lies in distinguishing very clearly the primacy of respect for B and C as autonomous agents when compared to the toleration of B’s ϕ -ing. The real difficulty that remains to be solved is therefore the first horn of the dilemma, which I have introduced above as *the paradox*. We shall see that one advantage of the proposed solution is that it relies on the same conceptual resources, namely the primacy of respect for autonomous agents. We can solve two paradoxes with one stone, so to speak.

Third, toleration/tolerance is best modelled as a ternary rather than a binary or even unary relation (especially when treated as a virtue Heyd 1996), between A, B, and C, where $A \neq B \neq C$ (this is justified by the asymmetry in power between A, B, and C introduced above)¹⁹ and

- A is an agent that has the power to prevent B’s ϕ -ing;
- B is the agent towards whose ϕ -ing A can exercise toleration; and
- C is the agent who is at the receiving end of the effects of A’s toleration towards B’s ϕ -ing and needs to exercise tolerance towards it (the ϕ -ing).

Finally, the paradox is best approached from a patient-oriented perspective, that is, by placing at the centre of the analysis C, the agent who is ultimately at the receiving end of B’s ϕ -ing, which is going to be subject to A’s toleration and C’s tolerance.

¹⁷ The difficulty is analysed at length in (McKinnon 2006), who suggests, as a solution, to weaken the “tight connection” between disapproving of something (“judgements of evaluative disapproval”) and preventing something from happening (“acting upon them”), see p. 27.

¹⁸ “The fundamental requirement of respect for persons thus incorporates and explains the necessity of allowing actions of which we disapprove morally” (Mendus 1988), p. 5. See also (Mendus 1989). For a similarly “layered” approach, but epistemological rather than ethical in nature, see (King 1998) who argues in terms of priority of arguments and doubts, rather than values.

¹⁹ The case in which toleration/tolerance is still analysed as a ternary relation but $A = C$ is simpler and I shall not analyse it here.

With these clarifications in place, let me introduce the solution. Recall that the paradox is one of scope (Popper) no longer one of grounding (Williams): there is a limit to how far toleration (now in the specific sense introduced above) should be exercised, but how can one cope with the intolerant, without resorting to some sort of justice, and hence entering into a vicious circle? The solution lies in constraining A's toleration of B's ϕ -ing (not primarily of B, as Williams mistakenly assumes) by means of a patient-oriented analysis of C's tolerance of B's ϕ -ing, in the following way:

A should exercise *toleration* towards B's ϕ -ing (believing, thinking, acting, behaving...) only insofar as any patient C of B's ϕ -ing is not coerced to exercise *tolerance* towards B's ϕ -ing.

This is toleration without coerced or uninformed tolerance, so let us call it the TT-principle. Figure 2 provides an illustration.

Rephrased in terms of freedoms, the TT principle may be interpreted as stating that A's tolerance of B's "freedom to" is constrained by C's "freedom from". For example: A should exercise *toleration* towards B's eavesdropping only if any C subject to B's eavesdropping does not have to exercise coerced or uninformed *tolerance* towards it. Since there is a C, C is *significantly affected*, and C did not provide uncoerced and informed consent, A should not tolerate B's eavesdropping. Toleration of eavesdropping promotes the structural inhibition of some of the most important individuals' positive liberties and their ability to participate in the information society, thus fulfilling their own potential as informational agents. It is a form of "informism", comparable to other forms of inhibition of social agency such as classism, racism, and sexism, and should not be tolerated unless freely and informatively agreed upon by those affected.

There are three main cases in which the TT principle leads to the conclusion that A should exercise toleration towards B:

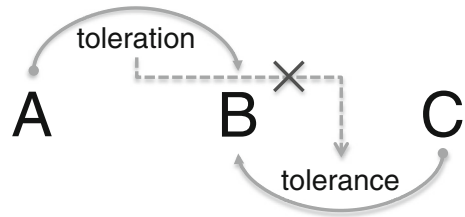
- (i) whenever *there is no C*;
- (ii) whenever there is a C, but C is not *significantly affected* by B's ϕ -ing; and
- (iii) whenever there is a C, C is significantly affected by B's ϕ -ing, but C provides *uncoerced and informed consent* about B's ϕ -ing.

Before examining some difficulties, let me briefly illustrate each case using some examples more controversial than the one just discussed.

According to (i), A should exercise toleration towards B's option in favour of assisted suicide, if there is no third party. According to (ii), A should exercise toleration towards B's option in favour of personal use of recreational substances (alcohol, tobacco, marijuana or other drugs), if any third party is not significantly affected by this (e.g. not while driving). According to (iii), A should exercise toleration towards B's option in favour of polygamy if any third party significantly affected by this provides uncoerced and informed consent.

Each example is open to debate, yet not in terms of the *logical* application of the TT principle, but because, *epistemologically* and *pragmatically* it shows that it may be difficult to establish

Fig. 2 The TT principle: no toleration with coerced or uninformed tolerance



- *whether there are third parties significantly affected.* For example, what if, in (i), the two children of the person who wants to opt in favour of assisted suicide are against it, or perhaps disagree with each other?
- *what it means for a third party to be significantly affected.* For example, what if, in (ii), one can show that society is significantly affected (e.g. in terms of a significant increase in the crime rate) by the consumption of recreational substances and corresponding addictive habits? And
- *what it means for a third party not to be coerced and/or uninformed with respect to the consent to be given.* For example, what if, in (iii), one can show that in some cases consent to enter in a polygamous relationship is socially or uncritically forced upon third parties?

Relying on concrete cases, one may argue that toleration should not be exercised, despite the apparent application of the TT principle, when

- after 24 weeks of pregnancy, a baby has a fighting chance of survival if born early, hence representing a new third party whose interests must be protected by the law, so that one of the conditions to tolerate an abortion is that the pregnancy has not exceeded its twenty-fourth week. This is the case in UK legislation;
- hygiene standards and measures to prevent the spreading of disease affect third parties. This is why the European court of human rights rightly dismissed the case brought by Shirley Chaplin, a nurse who wanted to wear a dangling cross at work, against the British National Health Service (NHS) policy, which forbids personnel to wear any jewellery because of the risk of infection. Note that the same court decided in favour of Nadia Eweida, who won her religious discrimination case against British Airways because her wearing a cross conflicted for a while with BA's uniform policy but did not really affect significantly any third party;
- in the case of the wearing of the hijab in French public schools, no uncoerced consent can really be (expected to be) given by the women involved.

Consider, as a further example, the controversial case in which the UN (agent A) should not tolerate (and hence intervene against) a sanguinary tyrant B if the population C is forced to exert tolerance toward B. These concrete cases indicate that the “what ifs” in (i)–(iii) are subject to interpretation, disagreement, and hence rational discussion. Clearly, the TT principle is not an algorithm to be applied mechanically. It is a general rule of normative design that enables the public use of

reason in the application of schemas such as (7) to generate actual norms.²⁰ At the same time, it would be a mistake to think that the serious hermeneutical, argumentative, and practical difficulties just seen affect the value of the solution of the paradox. On the contrary, we now have a way of constraining toleration by means of tolerance, without a circular recourse to the principle of justice. The need for interpretation through public debate assumes that, by default, toleration is legitimate and should be exercised whenever it is not constrained by tolerance or unless the interpretation of the conditional convincingly shows otherwise. In other words, a just, peaceful, and liberal society, in which the public use of reason can be exercised, requires a tolerant but not tolerating (in the sense indicated by the TT principle) society, or TTS, in short:

$$\neg TTS \rightarrow (\neg PUR \wedge \neg JS \wedge \neg PS \wedge \neg LS) \quad (15)$$

Let us call (15) the hyperhistorical political project. (15) means that toleration without coerced or uniformed tolerance can work as the ultimate principle of design of norms that constitute MASs that can deliver a reasonable (public use of reason), just (fair), peaceful, and liberal information society. As before, the relation of dependence of reasonableness, justice, peace and liberty on toleration (now understood in terms of the TT principle) may be broadly construed transcendently. A tolerant information society (in the not over-tolerating sense specified above) makes it possible, or indeed probable (or more probable, etc.) to deliver (the right kind of) a peaceful and liberal information society, though the public use of reason, a point consistent with the “Kantian” analyses developed by (O’Neill 1986) and (Herman 1996).

The hyperhistorical political project now takes full advantage of two features of toleration without coerced or uniformed tolerance as a principle of normative design. As before, toleration can ground justice, peace and freedom without requiring any further grounding. But now toleration that requires no coerced or uniformed tolerance is not undermined by its own paradox of scope because it embeds the right constraints that prevent its own distorted application. In short, as a principle of normative design, TTS is both self-booting and, because self-regulating, also self-reinforcing, thus enjoying the two features that we previously saw to be divided between unconstrained toleration and justice. TTS as a principle of normative design is the political equivalent of the Cartesian *cogito*.

Are there any difficulties faced by (15)? Yes, but they are not as damaging as the ones we encountered above. The first and perhaps most important difficulty is the following criticism: (15) seeks to solve the problem affecting the Westphalian political project but, in so doing, fails to escape the circularity affecting the political project of the Enlightenment. If third parties are significantly affected and yet toleration must still be exercised, this requires their tolerance to be based on their informed and uncoerced consent. However, “uncoerced” means free, and a free,

²⁰ On this point, and on the robust way in which the solution offered in this paper can deal with cases in which C is represented by a group within which there may be a degree of disagreement regarding the tolerability of B’s behaviour see (Pagallo 2014).

that is, liberal society, is what toleration was supposed to deliver in the first place, but is now presupposing. Clearly this cannot work.

The criticism would be serious, if it weren't mistaken. There is no circularity because (15) does not presuppose the existence of a liberal society, but the possibility that significantly affected third parties may give or withhold their consent autonomously (freely and informatively). This ethical issue is not trivial (Solove 2013), but it is also very different. Autonomy, in a Kantian sense, is taken to be a characteristic of the agents involved, not of the society that they form, which will be liberal only if tolerant, so there is no circularity, only consistency between a given philosophical anthropology and a liberal political philosophy based on (15).²¹ As Kuper rightly stresses in his criticism of Rawls:

It is this most basic internal feature—the respect for persons captured by the idea of ethical toleration—that must be the cornerstone of a consistent global liberal regulatory framework (Kuper 2000), p. 652.

It is here that the single “cornerstone” of respect for persons kills the two birds represented by the two horns of the paradox of toleration. The ethical respect for individuals enables us to avoid the paradox of grounding (see above the discussion of Williams' position), while making the circularity objection against the solution of the paradox of scope inapplicable.

Other criticisms concern the need for clarification of crucial aspects of the TT principle. In particular, what is the relation between asymmetric power, toleration and tolerance? What is “affectedness”, how can it be identified and—if it comes in degrees, as one may plausibly assume—how can thresholds of significance be drawn non-arbitrarily? And what is the relation between the hyperhistorical political project and a theory of pluralism about individuals' choices concerning different conceptions of the good (as in “morally good” and hence “good life”)? These are important questions, but they do not amount to objections. They are requests for further work that do not undermine the conclusion reached in the previous pages: toleration without coerced or uninformed tolerance should be the principle driving the design of norms in hyperhistorical societies. What the previous questions do stress is how difficult the task ahead is, even after having identified it clearly. Yet nobody should have had any illusions about this in the first place.

A significant consequence of the solution of the paradox of toleration is that the grounding of the information society on toleration, rather than justice, as the most fundamental principle of normative design, is no longer blocked but perfectly viable. Much still needs to be done, but the strategy is now clear and available.

²¹ I have defended a specific philosophical anthropology in which values are the counterpart of specific forms of flourishing in (Floridi 2013b). To be fair, the reader convinced by the criticism of Kant's and Rawls' alleged assumption of a specific “metaphysics of the self” may consider the possibility of extending it to my ontocentric approach as well. Personally, I believe it is unjustified, but I could not wish for better company.

Conclusion

I opened this article remarking that, in a globalised world in which we have not learnt yet to live together peacefully and tolerantly (or indeed sustainably), we need powerful MASs that can help us to tackle the immense difficulties facing us. The Westphalian state, challenged by globalisation and the development of transnational information societies, is still a necessary part of the solution, but it is insufficient. We need more powerful and better-coordinated MASs. I then suggested that the question about the design of the right kind of MASs becomes the question about the design of the right kind of norms that, as agents, may constitute the right MASs. This pushes the problem further back: how can we design the right kind of norms? In the previous pages, I offered an analysis of norms as artefacts, and sought to identify the most fundamental principles that can guide their design. I showed what it means to treat norms as designed artefacts, and indicated that Rawls is right; the modern political project has been grounded on the concept and practices of justice. However, that is not the only alternative. The modern political project could have been based on the concept and practices of toleration.²² There are historical reasons why that did not happen, but they have not been the topic of this article, in which I focused on the philosophical justifications. Justice, rather than toleration, has been the most fundamental guiding principle in the design of agents throughout Modernity because of the paradox of toleration. Once the latter is removed, the reason for not adopting toleration, as the main principle for the design of norms, is also removed. The conclusion reached is that toleration constrained by uncoerced and informed tolerance is a viable approach to the design of norms that constitute the right kind of MASs. Justice is crucial, to ensure that toleration does not become unfair, but toleration without coerced or uninformed tolerance must underpin it, if we wish to have a serious chance of succeeding in laying the foundation for cosmopolitan, peaceful, and liberal information societies. This, I argued, should be our political project in hyperhistory, in line with *The Declaration of Principles of Tolerance* adopted by acclamation on 16 November 1995 at the twenty-eighth session of the UNESCO General Conference.

Acknowledgments I discussed several versions of this article on many occasions in the past years and I am deeply indebted for their feedback to too many colleagues and members of the audiences to thank them all here. Some, however, deserve to be named explicitly because of their input and sustained effort in saving me from the most obvious mistakes: Anthony Beavers, Guido Boella, Elizabeth Buchanan, Schafer Burkhard, Juan Carlos De Martin, Massimo Durante, Guido Governatori, Patrick Grim, Patricia Mindus, Ugo Pagallo, Giovanni Sartor, and Mariarosaria Taddeo. The following occasions were especially useful for presenting and improving a draft that stayed with me for too long: the Weizenbaum Award Address, CEPE Conference—Computer Ethics: Philosophical Enquiry—11th edition (Autónoma University, Lisbon, 1–3 July, 2013); the workshop on “Protection of Information and the Right to Privacy: a New Equilibrium?” European University Institute, (Fiesole, 21 June 2013); an invited talk at the seminar at the Law School Stockholm University in collaboration with IRI Institutet för Rättsinformatik (Stockholm, 15 May, 2013); an invited talk at the Philosophy Department, Stony Brook,

²² Let me remark here that this hypothesis is close to Weber’s ideas on the “polytheism” of value-fragmentation. The point is too complex to be analysed here and deserves its own specific analysis. However, the *pars destruens* of Weber’s analysis shows why tolerance must underpin justice, and not the other way round.

SUNY (New York, 30 March 2 April 2013); an invited talk at the EU Commission “Onlife Project” workshop (Brussels, 8 February, 2013); the Ethics Lecture—Institute for Global Studies (University of Evansville, 10–13 November, 2012); an invited talk at the Beijing Forum (Beijing, 1–4 November, 2012); an invited lecture at the ISI Foundation—Institute for Scientific Interchange (Turin, 13 September, 2012); an invited talk at the IACAP Symposium, IACAP-AISB 2012 World Congress, University of Birmingham (Birmingham, UK, 2–5 July 2012); the CLS Annual Lecture, Centre for Law and Society, University of Edinburgh (Edinburgh, 8 June 2012); and an invited talk at the Camera dei Deputati (Rome, 23 March, 2012). David Sutcliffe carefully copyedited the last version and provided several helpful suggestions on some unclear or ambiguous passages. Finally, a special thank goes to the two anonymous referees who provided insightful and helpful comments that significantly improved the published version of this article.

References

- Abellán, J. (2012). Immanuel Kant—Tolerance seen as respect. In J. C. Laursen & M. J. Villaverde (Eds.), *Paradoxes of religious toleration in early modern political thought* (pp. 207–219). Lanham, MD: Lexington Books.
- Andrighetto, G., Governatori G., Noriega P., & van der Torre, L. W. N. (2013). *Normative multi-agent systems*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik GmbH.
- Austin, J. (1832). *The province of jurisprudence determined*. London: J. Murray.
- Austin, J. L. (1962). How to do things with words. In O. Urmson (Ed.), *The William James lectures delivered at Harvard University in 1955*. Oxford: Clarendon Press.
- Austin, J. L. (1975). *How to do things with words* (2nd ed.). Oxford: Clarendon Press.
- Austin, J. (1995). *The province of jurisprudence determined*. Cambridge: Cambridge University Press.
- Banks, J. S. (2001). *Signaling games in political science, Fundamentals of pure and applied economics positive political economy 2*. London: Routledge.
- Boella, G., & der Torre, L. (2006). Constitutive norms in the design of normative multiagent systems. In F. Toni & P. Torroni (Eds.), *Computational logic in multi-agent systems* (pp. 303–319). Berlin: Springer.
- Boella, G., & van der Torre, L. W. N. (2004). Regulative and constitutive norms in normative multiagent systems. *Principles of Knowledge Representation and Reasoning*, 4, 255–265.
- Boella, G., Van Der Torre, L., & Verhagen, H. (2006). Introduction to normative multiagent systems. *Computational and Mathematical Organization Theory*, 12(2–3), 71–79.
- Bohl, M., & Rynn, M. (2008). *Tools for structured and object-oriented design: An introduction to programming logic* (7th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Durkheim, E. (1976). *The elementary forms of the religious life (first published in French in 1912)* (2nd ed.). London: Allen and Unwin.
- Floridi, L. (2003). Informational realism. In J. Weckert & Y. Al-Saggaf (Eds.), *Selected papers from conference on Computers and philosophy-Volume 37* (37: Australian Computer Society), (pp. 7–12).
- Floridi, L. (2008a). A defence of informational structural realism. *Synthese*, 161(2), 219–253.
- Floridi, L. (2008b). The method of levels of abstraction. *Minds and Machines*, 18(3), 303–329.
- Floridi, L. (2009). Web 2.0 vs. the semantic web: A philosophical assessment. *Episteme*, 6(01), 25–37.
- Floridi, L. (2013a). Distributed morality in an information society. *Science and Engineering Ethics*, 19(3), 727–743.
- Floridi, L. (2013b). *The ethics of information*. Oxford: Oxford University Press.
- Floridi, L. (2014). *The fourth revolution—How the infosphere is reshaping human reality*. Oxford: Oxford University Press.
- Floridi, L., & Taddeo, M. (Eds.). (2014). *The ethics of information warfare*. New York: Springer.
- Forst, R. (2012). Toleration. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. <http://plato.stanford.edu/archives/sum2012/entries/toleration/>.
- Fuller, L. L. (1969). *The morality of law* (rev ed.). New Haven, CT: Yale University Press.
- Gensler, H. J. (2002). *Introduction to logic*. London: Routledge.
- Gensler, H. J. (2010). *Introduction to logic* (2nd ed.). London: Routledge.
- Governatori, G. (2014). Thou shalt is not you will. *arXiv preprint arXiv:1404.1685*.
- Governatori, G., & Rotolo, A. (2006). Logic of violations: A gentzen systems for reasoning with contrary-to-duty obligations. *Australasian Journal of Logic*, 4(1), 193–215.

- Hall, M. (2013). Ex-minister Sarah Teather: Government wants to make UK 'hostile' to immigrants. *The Telegraph* available online.
- Hart, H. L. A. (2012). *The concept of law* (3rd ed.). Clarendon law series Oxford: Oxford University Press.
- Herman, B. (1996). Pluralism and the community of moral judgment. In D. Heyd (Ed.), *Toleration: An elusive virtue, chapter 4*. Princeton, NJ: Princeton University Press.
- Heyd, D. (1996). *Toleration: An elusive virtue*. Princeton, NJ: Princeton University Press.
- Hobbes, T. (2012). *Leviathan*. 3 vols, *Clarendon edition of the works of Thomas Hobbes*. Oxford: Oxford University Press.
- Hollander, C. D., & Wu, A. S. (2011). The current state of normative agent-based systems. *Journal of Artificial Societies and Social Simulation*, 14 (2). <http://jasss.soc.surrey.ac.uk/14/2/6.html>.
- Holyoak, K. J., & Morrison, R. G. (2012). *The Oxford handbook of thinking and reasoning*. Oxford: Oxford University Press.
- Horton, J. (1994). Three (apparent) paradoxes of toleration. *Synthesis Philosophica*, 9(1), 7–20.
- Ingram, D. (2006). *Law: Key concepts in philosophy*. New York: Continuum.
- Johnson-Laird, P. N. (2006). *How we reason*. Oxford: Oxford University Press.
- Kant, I. (1786). *Grundlegung zur Metaphysik der Sitten*. 2e Aufl. ed. Riga.
- Kant, I. (1991) (originally 1786). What is orientation in thinking? In *Political writings*, 237–249. Cambridge: Cambridge University Press, originally Was heißt: Sich im Denken orientieren? First published in *Berlinische Monatsschrift*, VIII (October 1786), 304–330.
- Kant, I. (2011). *Groundwork of the metaphysics of morals: A German-English edition*. German text from the second original edition (1786) edited by Jens Timmermann, English translation by Mary Gregor, revised by Jens Timmermann. ed. Cambridge; New York: Cambridge University Press.
- King, P. (1998). *Toleration* (2nd ed.). London: Frank Cass.
- Kuper, A. (2000). Rawlsian global justice: Beyond the law of peoples to a cosmopolitan law of persons. *Political Theory*, 28(5), 640–674.
- Linklater, A. (1998). *The transformation of political community: Ethical foundations of the post-Westphalian era*. Oxford: Polity.
- Locke, J. (1689). *A letter concerning toleration*. Translated from the Latin of John Locke by W. Popple: London.
- Locke, J. (1991). *A letter concerning toleration*. In J. Horton, & S. Mendus (Eds.), *Routledge philosophers in focus series*. London: Routledge.
- Marcuse, H. (1965). Repressive tolerance. In R. P. Wolff, B. Moore, & H. Marcuse (Eds.), *A critique of pure tolerance* (pp. 95–137). Boston: Beacon Press.
- McKinnon, C. (2006). *Toleration: A critical introduction*. London: Routledge.
- Mendus, S. (Ed.). (1988). *Justifying toleration: Conceptual and historical perspectives*. Cambridge: Cambridge University Press.
- Mendus, S. (1989). *Toleration and the limits of liberalism*. Basingstoke: Macmillan.
- Mill, J. S. (1859). *On liberty* (2nd ed.). London: John W. Parker & Son.
- Mill, J. S. (1963). *The collected works of John Stuart Mill*. Toronto: University of Toronto Press.
- Minogue, K. (2000). *Politics: A very short introduction*. Oxford: Oxford University Press.
- O'Neill, O. (1986). The public use of reason. *Political Theory*, 14(4), 523–551.
- Oberdiek, H. (2001). *Tolerance: Between forbearance and acceptance*. Lanham, MD: Rowman & Littlefield Publishers.
- Pagallo, U. (2014). *Il diritto nell'era delle società ICT-dipendenti*. Turin: Giappichelli.
- Popper, K. R. (1945). *The open society and its enemies*. London: Routledge.
- Popper, K. R. (2013). *The open society and its enemies*. New introduction by Alan Ryan, essay by E. H. Gombrich, single volume ed. Princeton: Princeton University Press.
- Rawls, J. (1999a). *The law of peoples: With, the idea of public reason revisited*. Cambridge, MA: Harvard University Press.
- Rawls, J. (1999b). *A theory of justice*. Rev. ed. ed. Cambridge, MA: Belknap.
- Ricoeur, P. (Ed.). (1996). *Tolerance between intolerance and the intolerable*. New York: Berghahn Books.
- Ross, Alf. (1941). Imperatives and logic. *Theoria*, 7, 53–71.
- Scott, J. F. (1971). *Internalization of norms: A sociological theory of moral commitment*. Upper Saddle River, NJ: Prentice-Hall.
- Searle, J. R. (1995). *The construction of social reality*. New York: Free Press.

- Solove, D. J. (2013). Privacy self-management and the consent paradox. *Harvard Law Review*, 126, 1888–1893.
- Stetson, B. (2005). *The truth about tolerance: Pluralism, diversity, and the culture wars*. Downers Grove, IL: InterVarsity Press.
- Thomassen, L. (2006). The inclusion of the other? Habermas and the paradox of tolerance. *Political Theory*, 34(4), 439–462.
- Warnock, M. (1987). The limits of toleration. In Susan Mendus & D. Edwards (Eds.), *On toleration* (pp. 123–140). Oxford: Oxford University Press.
- Williams, B. (1996). Toleration: An impossible virtue? In David Heyd (Ed.), *Toleration: An elusive virtue* (pp. 18–27). Princeton, NJ: Princeton University Press.