

# Distributed Morality in an Information Society

Luciano Floridi

**Abstract** The phenomenon of distributed knowledge is well-known in epistemic logic. In this paper, a similar phenomenon in ethics, somewhat neglected so far, is investigated, namely *distributed morality*. The article explains the nature of distributed morality, as a feature of moral agency, and explores the implications of its occurrence in advanced information societies. In the course of the analysis, the concept of *infraethics* is introduced, in order to refer to the ensemble of moral enablers, which, although morally neutral per se, can significantly facilitate or hinder both positive and negative moral behaviours.

**Keywords** Distributed morality · Information and communication technologies · Information ethics · Infraethics · Moral enablers · Multiagent systems

## Introduction

In recent years, the scope of the concept of “moral agent” has been expanded to include both natural and legal persons (Allgrove 2004; Barfield 2005; Koops et al. 2010). The debate is not entirely new (Donaldson 1982; May 1983), nor devoid of controversial points (Ewin 1991). Its revival is due to the increasing pervasiveness and autonomy of artificial agents and of hybrid multiagent systems (Wooldridge 2009), both in every day contexts and in business environments (Andrade et al. 2004, 2007; Hildebrandt 2008, 2011, #93; Wallach and Allen 2009; Verbeek 2011; Kroes and Verbeek forthcoming). In (Floridi and Sanders 2004), I argued that standard perspectives on “mindless morality”—ethical issues involving artificial,

---

L. Floridi (✉)

School of Humanities, University of Hertfordshire, de Havilland Campus, Hatfield,  
Hertfordshire AL10 9AB, UK  
e-mail: l.floridi@herts.ac.uk

synthetic or hybrid agents, from companies to webbots—run the risk of remaining unduly constrained by an anthropocentric conception of agency. More recently (Floridi 2010a), I have shown how Business Ethics might be approached fruitfully from an Information Ethics’ perspective (Floridi 1999) that seeks to overcome such anthropocentrism. In this paper, I shall develop and defend the view that limiting the ethical discourse to individual agents hinders the development of a satisfactory investigation of *distributed morality*. This research expands some of the conclusions reached in my work on the foundations of information ethics (Floridi 2008a, 2010b, forthcoming).

I introduced the concept of distributed morality (DM) in (Floridi and Sanders 2004), where I used it in order to refer to the macroscopic and growing phenomenon of global moral actions and non-individual responsibilities, resulting from the “invisible hand” of systemic interactions among multiagent systems (comprising several agents, not all necessarily human) at a local level. Insisting on the necessarily human-based nature of the individual agents involved in any moral analysis means undermining the possibility of understanding not only DM but also another major transformation in contemporary ethics, the appearance of artificial agents (AAs). These are sufficiently informed, “smart”, autonomous artefacts, able to perform morally relevant actions, independently of the humans who engineered them, causing “artificial good” and “artificial evil” (Gips 1995; Floridi and Sanders 2001, 2004; Simon 2012). AAs are most relevant here because they play an important role in the dynamics of DM. They can be legitimate sources of im/moral actions, so the ethical discourse should include the analysis of their design, deployment, control, and behaviour, as part of a larger strategy to understand a range of new ethical issues not only in Information and Computer Ethics (ICE) but also in ethics in general, especially in the case of DM. As I anticipated, this is the specific topic investigated in this paper, which builds on (Floridi forthcoming). The following is a brief outline.

In “[The Basic Idea of Distributed Morality](#)” section, I introduce the basic idea of DM, by relying on a comparison with the well-known phenomenon of distributed knowledge in epistemic logic. I then explain the difference made by the occurrence of DM by discussing the moral scenario before and after its introduction (“[The Old Scenario without Distributed Morality](#)” and “[The New Scenario with Distributed Morality](#)” sections respectively). Next (“[Some Examples of Distributed Morality](#)” section), I provide some elementary examples of DM that should help to illustrate the phenomena in question more vividly and intuitively. In “[The Big Challenge: Harnessing the Power of DM](#)” section, I argue that the biggest challenge posed by DM concerns the possibility of harnessing its power in the right way. In “[Distributed Morality and the enabling Infraethics](#)” section, I outline a theory of moral enablers that can facilitate the occurrence and dynamics of DM. This calls for much more work on what I shall label infraethics. In the concluding section, I stress that the scope and magnitude of the ethical issues that we are, and will be, facing is such that it requires equally powerful multiagent systems (MAS)—capable of dealing with them through the impact of their proper DM-based actions—as well as infraethical environments that are friendly towards the sort of moral enablers that can facilitate MAS’ distributed morality.

## The Basic Idea of Distributed Morality

There is a sense in which cases of distributed morality have always been with us. *Collective responsibility*, for example—according to which a whole group of people is held responsible for some of its members' actions, even when the rest of the group has had no involvement at all (not even passively) in such actions—is a rather familiar concept in the *Old Testament*. The same applies to *social* or *group actions* and to (the theory of) *unintended consequences*. However, if these and similar phenomena are understood as being entirely reducible to the sum of (some) human, individual, and already morally-loaded actions—and I agree with Narveson (2002) that they might—then this is not what I will be concerned with in this article. As explained in the introduction, in the following pages I intend to use “distributed morality” (DM) to refer only to cases of moral actions that are the result of otherwise morally-neutral or at least morally-negligible (more on this distinction below) interactions among agents constituting a multiagent system, which might be human, artificial, or hybrid. A comparison to a very elementary, classic case of distributed knowledge in epistemic logic (Halpern and Moses 1990; Fagin et al. 1995) may help to clarify the basic idea.

Consider the case in which A knows only that  $[P \vee Q]$ , e.g. that “the car is in the garage or Jill got it”, whereas B only knows that  $\neg P$ , i.e. that “the car is not in the garage”. Neither A nor B knows that Q, only the supra-agent (with “supra” as in “supranational”)  $C = AB$  knows that Q.<sup>1</sup> It is the aggregation of A's and B's epistemic states that leads to C knowing that Q. Now, suppose A causes a set of actions  $\{a_1, \dots, a_n\}$ , and B causes another set of actions  $\{b_1, \dots, b_n\}$  to the effect that the supra-agent C causes a set of actions  $\{c_1, \dots, c_n\}$ . The question about “distributed morality” is this: can “big” morally-loaded actions (in our example, C's actions) be the result of many, “small” morally-neutral or morally-negligible interactions (in our example, A's and B's actions)? I hold the answer to be yes, and the rest of the paper is dedicated to supporting and explaining it. A good step forward is to start from a scenario in which there is no DM and then see what difference its introduction makes.

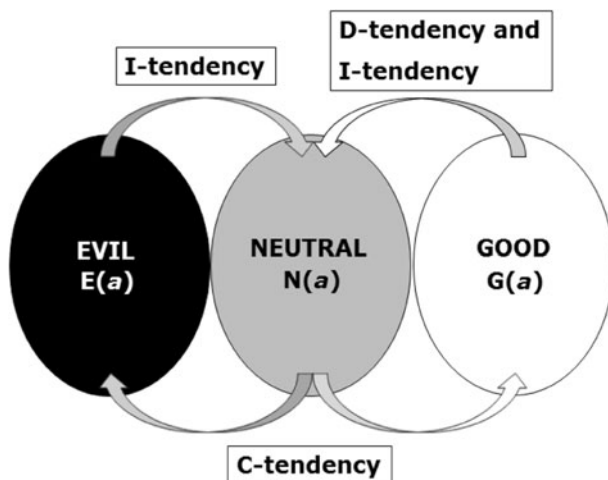
## The Old Scenario Without Distributed Morality

Let us follow common practice and assume that, for every action  $a$ ,  $a$  can be either morally Evil ( $E(a)$ ), Good ( $G(a)$ ) or Neutral ( $N(a)$ ). A moment of reflection shows that, for the deontologist, it is quite easy to fill up the grey oval (see Fig. 1), representing the set of all actions that are morally neutral. This is because, as is well known, morally good actions done out of a sense of convenience, or interest, or inclination or any other heteronomous reason, to use Kant's terminology, are stripped of their positive moral value. Slightly more formally,<sup>2</sup> let us represent the

---

<sup>1</sup> More precisely, C is the agent that is perceived to know that Q at the level of abstraction at which we do not have A and B as observables. On the method of levels of abstraction see Floridi (2008b, 2011).

<sup>2</sup> For the logically-minded reader, these are not formulae but mere abbreviations. They could be transformed into formulae by adopting a quantification ranging over the domain of all actions occurring in the system under observation, but this would be cumbersome and provide no further insights. The same holds true for an analysis in terms of deontic logic.



**Fig. 1** The old scenario without DM

deontologist's evaluative tendency to demote actions from  $G(a)$  to  $N(a)$  with the symbol  $\rightarrow$ , thus:

- i.  $G(a) \rightarrow N(a)$

Graphically, (i) is represented by the D-tendency in Fig. 1.

Following a similar reasoning, it is easier for the intentionalist to demote good to neutral ("great, but was not meant"), as in (i), but also evil to neutral ("sad, but was not meant"), so we have:

- ii.  $E(a) \rightarrow N(a)$

Graphically, both (i) and (ii) are represented by the I-tendency in Fig. 1.

As for the consequentialist, it is quite difficult to ensure that ultimately there is any  $a$  that is neither  $E$  nor  $G$ , but  $N$ . This is so because all actions have consequences and the latter inevitably have some moral value, so we have two tendencies to promote actions:

- iii.  $N(a) \rightarrow G(a)$
- iv.  $N(a) \rightarrow E(a)$

Graphically, both (iii) and (iv) are represented by the C-tendency in Fig. 1.

Now, trend (i) is one of the traditionally counterintuitive aspects of Kantian ethics, which requires a theory of praise in order to make (i) more palatable. Trend (ii) might be welcome in many contexts of "mindful morality", where it grounds the concepts of exculpation and forgiveness. Trends (iii) and (iv), in their full strength and if left unmodified, lead to the unacceptable conclusion that there are really no neutral actions at all, but only actions that are more or less (but never zero) morally-loaded, either positively or negatively. This is too implausible to be acceptable as it is, for it would force us to consider as morally significant a boundless number of *prima facie* non-moral actions, from the way someone scratches her head to how she

opens the door of a car. In order to rescue the consequentialist, we need to ring-fence both (iii) and (iv).

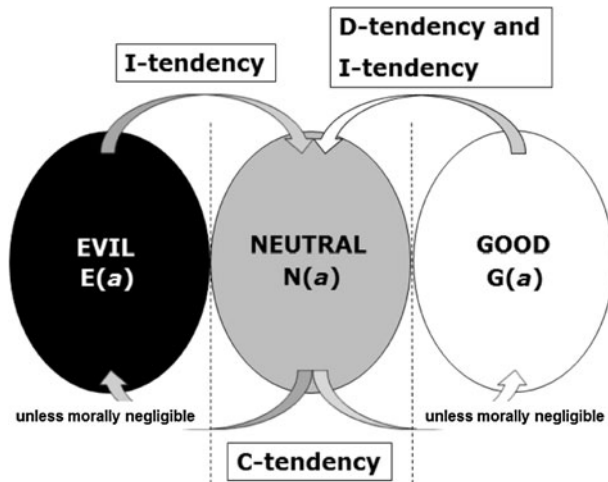
An obvious safety measure is provided by the concept of the *morally negligible* (the drop in the ocean effect, to oversimplify): many, if not most, actions are actually neither morally good nor morally evil (they are not subject to either trends described in (iii) and (iv)) because their actual effects are too small to be morally significant or because they mutually cancel each other. A spy scratching her head might be intentionally decreeing the death of an individual, but that is an extraordinary case. Likewise, you might open the door of a car in such a way, or in such circumstances, that your action might count as morally commendable (perhaps you helped someone in real difficulty) or disastrous (you wilfully triggered a bomb), yet this too seems to be the exception rather than the rule. Finally, we often do and not do things in such ways that the end result is still negligible.

In order to implement the plausible idea of morally negligible consequences, and thus ground the possibility of morally neutral actions, let us introduce two moral thresholds in our model: one makes it more difficult to apply  $N(a) \rightarrow G(a)$ , while the other has the same function with respect to  $N(a) \rightarrow E(a)$ . Now actions need to be morally significant in order to move from being neutral to being morally loaded. More formally, the two arrows that graphically describe the C-tendency become vectors: they have not only a direction but also a strength, which needs to be sufficiently high in order to overcome the relevant threshold (Fig. 2).

Specifying how actions can be, or become, morally significant—conversely, establishing the right level at which the thresholds can be overcome—is a serious difficulty, comparable to the problem of identifying individual utilities when single rational agents need to make personal choices. It is certainly a major issue for the consequentialist, who probably needs to deal with it more contextually than she might be happy to admit initially. Such a difficulty is also what lies behind the debate on rule consequentialism. Luckily, all this need not concern us here, since our goal is to gain a better understanding of DM. For this purpose, it is interesting to note that, once we model the applications of (iii) and (iv) as being constrained by some thresholds the value of which can be left unspecified here, we obtain one more new concept, that of *moral inertia*: most actions are morally neutral and tend to stay that way either because of the two thresholds, if one is a consequentialist; or because of the I-tendency, if one is an intentionalist; or because of the D-tendency, if one is a deontologist. These are all the details we need from the old scenario. We shall now use the concepts of morally negligible, moral threshold, and moral inertia in order to introduce a new variable in our model, that of distributed morality, a task for the next section.

## **The New Scenario with Distributed Morality**

Recall that we wish to consider actions that might be performed by human, artificial or hybrid multiagent systems, so that you and I, as well as artificial agents (e.g. some kinds of webbots online), a corporation, an individual driving a car with the help of a GPS, or a drone-network-pilot-command system, may all count as

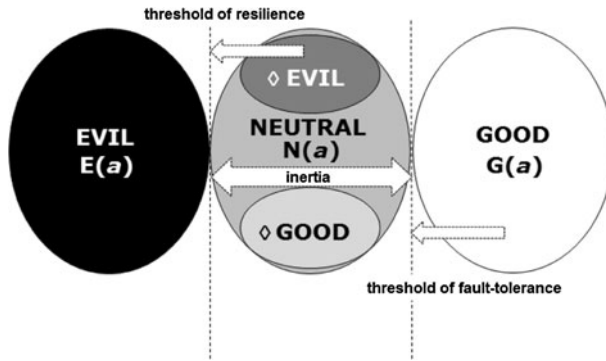


**Fig. 2** The old scenario with moral thresholds and morally negligible actions

potential sources of possibly morally-loaded actions. Because we are adopting such a MAS-oriented approach, we cannot rely on a framework of moral evaluations based on intentionality or motive-related analysis. After all, the MAS in question might be totally mindless, so that any talk of beliefs, desires, intentions and motivations would be merely metaphoric. Indeed, we are interested in adopting a uniform, minimalistic level of abstraction (Floridi 2008b) such that even human individuals might be treatable as mindless agents. The consequence is that we need to evaluate actions not from a sender but rather from a receiver perspective: actions (including MAS', artificial and supra-agents') are assessed on the basis of their impact on the well-being of the environment at large and its inhabitants specifically. With these adjustments in place, let us return to the three concepts introduced in the previous section.

Because most actions are morally negligible, that is, because they remain below a given moral threshold, it follows that possibly evil actions (the subset of neutral actions labelled  $\diamond$ Evil in Fig. 3) may be ineffective. From a receiver's perspective, this is another way of saying that environments can be morally resilient, or, to paraphrase Paul of Tarsus (1 Corinthians 13), that goodness (understood as absence of evil, hence including also neutrality) is fault-tolerant. An elementary example is provided by speeding on the motorway: a potentially evil action fails to become actually evil thanks to the resilience of the overall environment. The driver is morally irresponsible not because of the effects of his action—we assume that his reckless driving turned out to have no nasty consequences—but because of his unwarranted reliance on the fault-tolerance of the rest of the system. This is why his behaviour cannot be universalised<sup>3</sup>: the system can bear only so much pressure before collapsing.

<sup>3</sup> Universalization is an obvious factor that can help in such a strategy. By universalization I refer here to the normative coordination of the possibly good, distributed actions of a multiagent system: agents



**Fig. 3** The new scenario of distributed morality

At the same time, precisely because most actions are morally negligible and remain below a given moral threshold, possibly good actions (the subset of neutral actions labelled  $\diamond$ Good in Fig. 3) can equally fail to be effective. Environments can be morally inert: below a given threshold, possibly good actions never actually make a (significant, lasting or indeed any) difference, but remain neutral. In other words, potential goodness can be too weak to become actual goodness. In this case, some forms of charity provide a good example.

To summarise (Fig. 3), on the one hand, environments are morally inert or morally fault-tolerant. On the other hand, many MAS' actions often turn out to be morally neutral, in the sense of having insufficient strength to overcome the two thresholds introduced in the previous section. This might be because

- a. they are morally-unloaded (value-free, in a different vocabulary); or
- b. they are insufficiently morally-loaded (have some moral value, but still fail to overcome the threshold); or
- c. they mutually off-set each other.

We have seen in “[The Basic Idea of Distributed Morality](#)” section that, unless A and B interact properly, their distributed knowledge cannot emerge, for it is held neither by A nor by B alone. Likewise, unless A and B interact properly, their distributed action remains below the threshold of the morally negligible. The overall result is that, in this new scenario, neutrality works as a powerful attractor and many actions are simply unable to escape their neutral status. In many cases, it is only by aggregating and merging individual courses of action that a moral difference is made. Note that, at this stage, such difference could be for the best (moral goodness) or for the worst (moral evil). Note<sup>4</sup> also that the aggregation in question is not one-way. Some evils emerging from DM might be further aggregated to such actions as

---

Footnote 3 continued

constituting a MAS ought to implement, optimise and coordinate their actions in such a way as to make them converge on the achievement of a morally good output. There are of course several other ways of understanding ethical universalization, see my reply to Stahl in Floridi (2008a).

<sup>4</sup> I am grateful to Massimo Durante for having called my attention to this important point.

to generate morally good outcomes. Likewise, some morally good actions reached through DM might be further aggregated in such a way as to cause evil. Clearly, in all these cases, the correct management of DM is both a challenge and an opportunity. Before discussing it, let me complete the description of DM by briefly presenting a few concrete illustrations.

### **Some Examples of Distributed Morality**

A classic and well-known example of negative DM is represented by the tragedy of the commons (Hardin 1968; Hardin 1998). However, since I have already analysed its digital version insofar as it applies to the infosphere in (Greco and Floridi 2004), I shall not discuss it here, where I wish to focus instead on some examples of positive DM. Just for the sake of illustrative simplicity, they are all based on quantitative analyses, in terms of moral benefits that can easily be quantified economically. In each of the following cases, MAS' actions, which are morally negligible in themselves, give rise to aggregated morally good actions:

1. the shopping Samaritan: (RED);
2. plastic fidelity: the Co-operative Bank;
3. the power of giving: JustGiving;
4. socially oriented capitalism: P2P lending.

Let's have a look.

#### **The Shopping Samaritan: (RED)**

Perhaps the best way to present (RED) is by quoting the website of the project:

(RED) is a simple idea that transforms our incredible collective power as consumers into a financial force to help others in need.

Each time you buy a (RED) product or service, at no extra cost to you, the company who makes that product will give up to fifty-per cent of its profit to buy and distribute antiretroviral medicine to our brothers and sisters dying of AIDS in Africa. Every dollar goes straight to Africa. Straight to the people who need it. Straight to keeping them alive so that they can go on taking care of their families and contribute socially and economically to their communities. [...]

Since (RED)'s launch in 2006, over 5 million people have been impacted by HIV and AIDS programs supported by your (RED) purchases. (<http://www.joinred.com/aboutred>).

Partners in the program include American Express, Apple, Armani, Converse, Dell, GAP, Motorola, Nike, Starbucks and many others.

#### **Plastic Fidelity: The Co-operative Bank**

Our next example of positive DM is represented by a customer loyalty scheme, promoted by the Co-operative Bank in the UK. The bank offers a number of credit



cards, linked to specific charities, including Amnesty International, Oxfam, and Greenpeace. By using the card, the customer ensures that

- the chosen charity receives £15 for every account opened;
- a further £2.50 is received if the account is used within six months;
- plus 25p for every £100 spent using the card and 25p for every £100 transferred to the card.

These might seem drops in the ocean, but, for example, between 1994 (the year the scheme was launched) and 2007 the Co-operative Bank's Oxfam-affiliated credit cards contributed £ 3 million towards Oxfam's work around the world.<sup>5</sup>

### The Power of Giving: JustGiving

It can be expensive to run charities. In the UK, their management and administration typically represents between 5 and 13 % of their total expenditure.<sup>6</sup> So a company that provides online fundraising tools to enable the electronic management of charitable donations, like JustGiving in the UK and its twin organisation FirstGiving in the US, can make a huge difference. Not only can it facilitate the process of fund raising and lower its costs, it can also provide visibility and support, as well as suggestions and solutions for extra funding opportunities. Here is some evidence.

JustGiving provides its service for more than 5,000 UK registered charities and 300,000 fundraising pages for users, collecting over £450 million. The administrative function includes the automatic reclaiming of Gift Aid on all donations from UK taxpayers. JustGiving's stated goal is to 'allow ordinary people to raise extra ordinary amounts of money'. More than £450 million for over 8000 member charities has been raised through JustGiving since its launch. Charity Times claimed the company had "transformed the face of donating in the UK". (<http://en.wikipedia.org/wiki/Justgiving>)

In the "business of beneficence" (Rockefeller) agents need to be frugal with their wasteful consumption but generous with their fruitful interactions.

### Socially-Oriented Capitalism: Peer-to-Peer Lending

Our last example concerns P2P lending, also known as social lending, person-to-person lending or community lending. This is lending online occurring between individuals directly, without the intermediation of an institute (usually a bank). P2P lending as a macroscopic phenomenon is really made possible only by the Internet, the enabling technology. There are two models, each illustrating a case of DM:

- a. in the marketplace model, online intermediaries, such as Prosper Loans Marketplace in the US or Zopa in the UK, put in touch lenders and borrowers, who go through an auction-like process to negotiate a loan;

---

<sup>5</sup> Source: Oxfam, [http://www.oxfam.org.uk/get\\_involved/companies/downloads/oxfamatwork\\_web08.pdf](http://www.oxfam.org.uk/get_involved/companies/downloads/oxfamatwork_web08.pdf).

<sup>6</sup> Source: CharityFacts, [http://www.charityfacts.org/charity\\_facts/charity\\_costs/index.html](http://www.charityfacts.org/charity_facts/charity_costs/index.html).

- b. in the community model, lenders and borrowers are already acquainted with each other, and online intermediaries such as Virgin Money US (formerly CircleLending) only help them to formalise a personal loan.

In both models, we see distributed actions being aggregated to make a difference in the life of the beneficiaries.

### **The Big Challenge: Harnessing the Power of DM**

The previous examples show how actions that are morally negligible in themselves may become morally significant, if properly aggregated. At the end of “[The New Scenario with Distributed Morality](#)” section, I mentioned that harnessing the power of DM is a challenge but also an important opportunity. This is represented by the possibility of strengthening environmental resilience and fault-tolerance, while weakening inertia, so that possibly evil, but still neutral, actions are blocked below the moral threshold, while possibly good, but still neutral, actions are enhanced above the moral threshold. Such a twofold manoeuvre requires ethical policies of

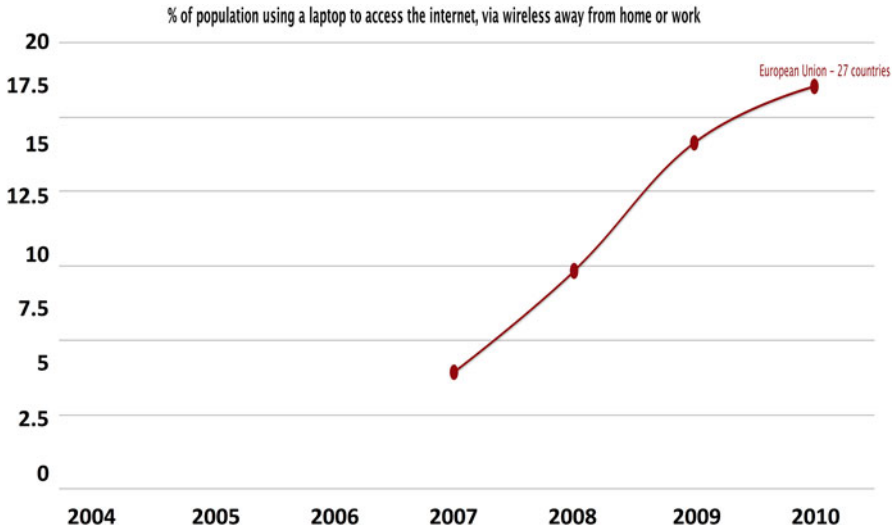
- a. *aggregation* of possibly good actions, so that the latter might reach the critical mass necessary to make a positive difference to the targeted environment and its inhabitants; as well as
- b. *fragmentation*, so that possibly evil actions might be isolated, parcelled and neutralised.

Such policies are socially furthered by

- c. incentives and disincentives, which represent the political and legislative side of the ethical discourse, and
- d. technological mechanisms that work as “moral enablers”.

Regarding (c), since the moral behaviour of large number of agents has always been a concern, there is a long tradition of trial and error, social and political thinking (under the label social or public choice theory), legislation, ethical norms and mass behaviour (think of the phenomenon of “social pressure” or “peer pressure”) that can help significantly in shaping and orienting DM in the right direction. I shall not expand on this well-known point in this article, but it is obviously of crucial importance.

Regarding (d), however, much work still needs to be done, for the following reason. DM is made *increasingly* possible by multiagent systems, which in turn are made *increasingly* possible by extended, pervasive and intensive interactions. These interactions are *increasingly* enhanced, facilitated and multiplied by Information and Communication Technologies (ICTs). And all these “increasingly” explain why it is really only in advanced information societies that we can more readily and frequently witness the occurrence of DM phenomena. ICTs are a most influential enabling factor behind the emergence of DM, working as powerful moral enablers, as I shall explain in some detail in the next section. Individuals are more and more connected and interactive, so that the threshold between online and offline is being



**Fig. 4** % of EU population accessing the internet, away from home or work

gradually erased and DM becomes progressively more important. For example, in 2011, 20.7 % of the European Union population accessed the internet, by a laptop, while being away from both home and the office<sup>7</sup> (see Fig. 4), and that is because our world is becoming our infosphere. We no longer login or logout, we are always *onlife*. Nevertheless, ICTs as moral enablers are not (at least not yet) designed in such a way as to meet the serious challenge posed by the correct management of DM. At the risk of trivialising a much more complex issue by using an elementary illustration, P2P technology, for example, can be used in order to aggregate neutral actions and make them overcome either threshold in both directions, towards evil or towards goodness. More controversially, the debate on network neutrality (Turilli et al. 2012) seems to be a case in which old prejudices against diversification are going to hinder the development of morally good, distributed dynamics.

It might be that some specific technologies will always maintain their dual nature. Web 2.0 applications may be used to aggregate all sort of interests and interactions, even the darkest ones. Very plausibly, at least part of the solution rests in the intelligent synthesis between

1. a more profound and detailed understanding of the logical dynamics of DM and hence new forms of civil education<sup>8</sup>;
2. better design of our technological moral aggregators—as argued for example by Adam (2005) in her discussion of privacy in relation to DM, by Turilli (2007) in

<sup>7</sup> Source: Eurostat—Community survey on ICT usage in Households and by Individuals, [http://scoreboard.lod2.eu/index.php?scenario=2&indicators%5B%5D=i\\_iuport+IND\\_TOTAL+%25\\_ind&countries%5B%5D=EU27#chart](http://scoreboard.lod2.eu/index.php?scenario=2&indicators%5B%5D=i_iuport+IND_TOTAL+%25_ind&countries%5B%5D=EU27#chart).

<sup>8</sup> See for example *LCD'07—Workshop on Logics and Collective Decision Making*, Erasmus International Institute MSH Nord-Pas-de-Calais, March 13–14, 2007, Lille, France.

- terms of ethical protocols design, and by Cavoukian (2009), Pagallo and Bassi (2010), Pagallo (2012, #92), and Pagallo (2012) insofar as privacy might be approached from a design perspective; and
3. improved ethical policies of incentives and disincentives.

Equally plausibly, it seems that part of the solution will also depend on the development of social and technological infrastructures (also known as metatechnologies, more on this in the next section) that will foster the right sort of distributed morality. This is the last point I wish to analyse in this article.

## **Distributed Morality and the Enabling Infraethics**

There is a long and well-established tradition in ethics that seeks to identify, explain and defend moral values, in order to develop and justify, on their basis, universal, normative analyses of morally-loaded actions, and hence support reasonable, if sometimes competing, interpretations of the morally good life and its achievability. One crucial aspect, which seems to have been underemphasised by this tradition, is the analysis, implementation, and furthering of the non-moral factors that can facilitate morality and hinder immorality.

The idea may be quickly introduced by comparing it to a phenomenon well known to economists and political scientists. When one speaks of a ‘failed state’, one refers not only to the failure of a *state-as-a-structure* to fulfil its basic roles, such as exercising control over its borders, collecting taxes, administering justice, providing schooling, and so forth. One also refers to the collapse of a *state-as-an-infrastructure* or environment, which makes possible and fosters the right sort of social interactions; that is, one may be referring to the collapse of (certainties about) the rule of law and, of acceptable ways of dealing with economic transactions, of default expectations about the protection of human rights, of a sense of political community, of civilised dialogue among differently-minded people, of modes of communication to reach peaceful resolutions of ethnic, religious, linguistic, or cultural tensions, and so forth. All these expectations, attitudes, practices, in short such an implicit ‘socio-behavioural infrastructure’, which one may take for granted, provides a vital ingredient for the success of any complex society. It plays a crucial role in socio-political contexts, comparable to the one that we are now accustomed to attributing to physical infrastructures in economics. By analogy, it seems time to acknowledge that the morally good behaviour of a whole population of agents is *also* a matter of ‘ethical infrastructure’ or *infraethics*.<sup>9</sup> This is to be understood not as a kind of second-order ethical discourse or metaethics,<sup>10</sup> but as a first-order framework of implicit expectations, attitudes, and practices that *can* facilitate and promote morally good decisions and actions. Examples include trust, respect,

---

<sup>9</sup> This is related to, but not to be confused with, what Jonsen and Butler (1975) meant by ‘infraethics’, which they understood as a particular level of ethical enquiry concerning public ethics, see Daniels (1996), p. 341.

<sup>10</sup> For the non-philosopher, metaethics is the branch of philosophy that studies the nature of ethical theories, properties, statements, attitudes, and evaluations.

reliability, privacy, transparency, freedom of expression, openness, fair competition, and so forth. I highlighted 'also' and 'can' above because it is important to understand that such an infraethics is *not* necessarily morally good in itself. Any successful complex society, be it the City of Man or the City of God, has an implicit infraethics. Even a society in which the entire population consisted of angels, that is, perfect moral agents, needs norms for collaboration, coordination, and cooperation. Theoretically, that is, when one assumes that morally good values and the infraethics that promotes them may be kept separate (an abstraction that never occurs in reality but that facilitates our analysis here), a society in which the entire population consisted of Nazi fanatics could rely on high levels of trust, respect, reliability, privacy, transparency, and even freedom of expression, openness, and fair competition. Clearly, what we want is not just the successful mechanism provided by the right infraethics, but also the coherent combination between it and morally good values, such as civil rights. To rely on an analogy: the best pipes may improve the flow but they do not improve the quality of the water, and water of the highest quality is wasted if the pipes are rusty or leaky.

In sociology, economics, politics, and law studies increasing attention has been paid in the last few decades to so-called enablers such as education, health, safety and security, property rights and credit opportunities, clear legislation and reliable implementation of the law, especially in developing countries. The lack of similar studies about the need for, and the nature of, an infraethics is understandable, given the troublesome history of human priorities, but it also seems to be time to redress it.

Within this general context, the specific point I wish to address, in relation to the phenomenon of DM, may be clarified rather simply by means of two questions. Suppose we have a general view of what morally good is and of the sort of distributed morality that might bring it about, then what exactly may facilitate the implementation of the latter? And what exactly may hinder the sort of DM that could bring about the morally evil? Of course, the two questions are as strictly related as the two sides of the same coin. Indeed, they may be further simplified by labelling the referent of the "what" in each of them as an infraethics (understood as the ensemble of moral enablers) and then rephrase them thus: given a dynamic, moral system in which DM plays a significant role, what is the right infraethics that can foster it?

An enquiry into the nature and logic of the right sort of infraethics, its interactions and operations within a dynamic system, and its positive effects on DM does not seek to uncover the morally good and evil, but rather presupposes a satisfactory understanding of both. It addresses a different problem, namely what sort of facilitating framework makes the morally good more likely to occur, and then become more stable and permanent, i.e., to take root; and what sort of hindering framework makes the morally evil more unlikely to occur or, when it has occurred, to remain unstable and more transient, so to wash away more quickly and easily. Now, investigations into ICTs, their personal, social and ethical impact, and hence into ICE issues, have helped us, both historically and theoretically, to unveil cases of moral facilitation and thus identify moral enablers to an unprecedented extent and with a much higher level of clarity. Examples include information availability and accessibility, trust online (Taddeo 2009, 2010), information

transparency (Turilli and Floridi 2009) and openness (as in open source software, Chopra and Dexter 2008), information privacy (Floridi 2005, 2006), and the relation between forgetfulness and forgiveness.

Unsurprisingly, issues of moral facilitation that seem too complex to tackle if we use a first order logic become rather unproblematic once we adopt a second order logic. Trust, for example, becomes very easily treatable if understood it as a second order relation (and hence an enabler), rather than a first order one (Taddeo 2010). However, the temptation of interpreting specific moral enablers, e.g. trust or transparency, in terms of *meta-values*, that is, as if they were values qualifying other values, should be resisted. I argued above that an infraethics is not a metaethics. Logically speaking, a more fruitful way to represent specific moral enablers is by relying on the apparatus of modal semantics, and to treat them as agents in themselves, which operate between possible worlds (PWs). Such enabling agents, when properly designed and regulated, can act as promoters and facilitators of the morally good. At worst, they can prevent, neutralise, or at least limit the paths to evil, that is, undesirable transitions from some PWs to other, morally worse PWs. Or, (in the logical, inclusive sense of the disjunction), at best, they can foster, enhance and consolidate desirable transitions from some PWs to other, morally better PWs, the paths to goodness. The other temptation, to understand specific moral enablers as *infra-values*, i.e. values that underpin other values and make them possible, should also be resisted. On the contrary, moral enablers are better understood as intra-components of the moral system, metaphorically comparable to the lubricant of the moral machinery. They work at the same level as moral values, neither below nor above them, as integral parts of the dynamic moral system, even if they themselves are not moral values.

ICE has cast a powerful light on a less visible side of the ethical discourse, the rich and fertile humus that provides nourishment and strength to the roots of moral interactions. It follows that we have now the opportunity to understand that, in ethics, moral facilitation is a much more influential, macroscopic and perhaps necessary phenomenon—not merely limited to ICE contexts—than we suspected in the past, a phenomenon that lies hidden behind the more visible scenes of many moral interactions. No determinism is involved, but freedom may be exercised more fully and in better directions if the right moral enablers are in place and work properly. Recall the analogy with physical infrastructures: they can help the economy of a country enormously, even if they do not determine the nature of the businesses in question. Once again, this is not an entirely new phenomenon. Within our information societies, moral enablers may often have an ICT nature, hence their study and implementation may be best carried out by an Information Ethics, but they do not need to be only ICT-based. To use the previous example, trust has always been a moral enabler. The fact that only in recent years we have focused so much on its ethical role is largely due to the impact of ICTs, which have worked as a magnifier.

## **Conclusion**

Many more examples could be provided of cases of infraethical phenomena that facilitate the emergence of DM and positive moral behaviours. Consider

fourth-generation bikesharing, for instance. “The advances and shortcomings of previous and existing bikesharing models have contributed to a growing body of knowledge about this shared public transportation mode. Such experiences are making way for an emerging fourth-generation bikesharing model or demand- responsive, multimodal systems. These systems build on the third generation and emphasize (a) flexible, clean docking stations; (b) bicycle redistribution innovations; (c) smart-card integration with other transportation modes, such as public transit and carsharing; and (d) technological advances including Global Positioning System (GPS) tracking, touch screen kiosks, and electric bikes” (Shaheen et al. 2010, pp. 165–166). Clearly, it is a whole ensemble of facilitators that need to be designed, coordinated and implemented, for an infraethics to become possible, and such infraethics can make a difference in terms of DM only if a sufficient number of agents become involved. Bikesharing is a healthy and environmentally good thing and a morally positive trend, but it requires advanced ICT applications, no component of the system in itself would make any difference, and if only a few users were to take advantage of it, the environmental benefits would be virtually null. As stressed above, the risk of misuse and moral hazard are also never entirely absent: such bicycles, for example, could be used to rob a bank or may initially lead to more traffic-related accidents. Yet it seems obvious that the advantages vastly outweigh the risks.

The proper shaping and steering of DM through the design of the right sort of infraethics appear to be an important challenge that will deserve much more intellectual work, education, and political attention. In its scope and influence, DM is a largely unprecedented phenomenon, which characterises advanced information societies, not because it never did or could occur in the past—this would be of course both factually and theoretically wrong—but because ICTs have just begun to make DM a much more common, pragmatically influential, and epistemologically salient phenomenon. Instances of DM that were “too weak” and sporadic in the past to be worth much attention or ethical analysis are now playing an increasingly important role in our lives, and will be more and more influential in the future.

The conclusion is that an information society is a better society if it can implement an array of moral enablers, an infraethics that is, that can support and facilitate the right sort of DM, while preventing the occurrence and strengthening of moral hinderers. Agents (including, most importantly, the State) are better agents insofar as they not only take advantage of, but also foster the right kind of moral facilitation properly geared to the right kind of distributed morality. It is a complicated scenario, but refusing to acknowledge it will not make it go away.

There are both practical and theoretical problems affecting the development of a theory of distributed morality, of its moral enablers, and of their correct implementation. One may need to consider, for example, the global nature of information societies and the necessity to negotiate interactions with alternative, pre-existing moral traditions (Floridi 2003). Likewise, consistency and partial-ordering in terms of priority among different instances of DM and several moral enablers are certainly issues of crucial importance, as the debate between defenders of information privacy and defenders of information transparency highlights. Despite these difficulties, however, the study and actual development of DM and the corresponding infraethics are challenges worth tackling. The nature of the ethical

issues facing humanity is increasing in scope, magnitude and seriousness. Big issues call for big agents. We need powerful, multiagent systems that, by aggregating and controlling their distributed actions, can cope ethically well with macroscopic, global moral issues. DM is a new phenomenon whose importance will only grow steadily. The sooner we learn how to harness its power explicitly the better. Infraethical environments in which moral enablers can flourish that support the right sort of MAS and DM will be better equipped to deal with our uncertain future. They may actually play a big role in how we solve some of the most pressing and intractable, ethical problems at a global level.

**Acknowledgments** Previous versions of this article were presented at (a) *CEPE 2007—The Seventh International Computer Ethics Conference* (University of San Diego, 13 July 2007); (b) the International Workshop on *Moral Agency and Technical Artefacts*, organised by the *Netherlands Institute for Advanced Study* in the Humanities and Social Sciences (NIAS, 10-12 May 2007); (c) a research seminar, organised by the *Oxford Uehiro Centre and the Programme on the Ethics of the New Biosciences* (University of Oxford, 26 November 2007); (d) the Internet Ethics Seminar, organised by the *Oxford Internet Institute* (OII, 30 April, 2010); (e) a departmental seminar, organised by the Department of Computer Science, University of Oxford (Oxford, 14 July, 2010). I wish to thank all the participants to those meetings for their feedback and, for their kind invitations and the opportunity to present and discuss this research, Lawrence M. Hinman and Esther Aguilar in relation to (a); Peter Kroes, Henneke Filiz-Piekhaar, Jeroen van den Hoven, Eline van der Ploeg, in relation to (b); Jo Armitage, Miriam Wood, and Julian Savulescu, in relation to (c); Yorick Wilks, Matthew Carlos, and Karen Melham in relation to (d); and Alexandru Baltag in relation to (e). David Davenport, Massimo Durante, Ugo Pagallo, and Judith Simon provided some very useful comments on the penultimate version.

## References

- Adam, A. (2005). Delegating and distributing morality: Can we inscribe privacy protection in a machine? *Ethics and Information Technology*, 7(4), 233–242.
- Allgrove, B. (2004). Legal personality for artificial intellects: Pragmatic solution or science fiction? *SSRN eLibrary*.
- Andrade, F., et al. (2004). Software agents as legal persons. In L. M. Camarinha-Matos (Ed.), *Virtual enterprises and collaborative networks*. Dordrecht: Kluwer Academic Publishers.
- Andrade, F., et al. (2007). Contracting agents: Legal personality and representation. *Artificial Intelligence and Law*, 15(4), 357–373.
- Barfield, W. (2005). Issues of law for software agents within virtual environments. *Presence: Teleoperators and Virtual Environments*, 14(6), 741–748.
- Cavoukian, A. (2009). *Privacy by design*. Ottawa: IPC Publications.
- Chopra, S., & Dexter, S. (2008). *Decoding liberation: The promise of free and open source software*. New York: Routledge Studies in New Media and Cyberculture, Routledge.
- Daniels, N. (1996). *Justice and justification: Reflective equilibrium in theory and practice*. Cambridge: Cambridge Studies in Philosophy and Public Policy, Cambridge University Press.
- Donaldson, T. (1982). *Corporations and morality*. Englewood Cliffs, NJ: Prentice-Hall.
- Ewin, R. E. (1991). The moral status of the corporation. *Journal of Business Ethics*, 10(10), 749–756.
- Fagin, R., et al. (1995). *Reasoning about knowledge*. Cambridge, MA, London: MIT Press.
- Floridi, L. (1999). Information ethics: On the philosophical foundations of computer ethics. *Ethics and Information Technology*, 1(1), 37–56.
- Floridi, L. (2003). On the intrinsic value of information objects and the infosphere. *Ethics and Information Technology*, 4(4), 287–304.
- Floridi, L. (2005). The ontological interpretation of informational privacy. *Ethics and Information Technology*, 7(4), 185–200.
- Floridi, L. (2006). Four challenges for a theory of informational privacy. *Ethics and Information Technology*, 8(3), 109–119.



- Floridi, L. (2008a). Information ethics: A reappraisal. *Ethics and Information Technology*, 10(2–3), 189–204.
- Floridi, L. (2008b). The method of levels of abstraction. *Minds and Machines*, 18(3), 303–329.
- Floridi, L. (2010a). Network ethics: Information and business ethics in a networked society. *Journal of Business Ethics*, 90(4), 649–659.
- Floridi, L. (Ed.). (2010b). *The cambridge handbook of information and computer ethics*. Cambridge: Cambridge University Press.
- Floridi, L. (2011). *The philosophy of information*. Oxford: Oxford University Press.
- Floridi, L. (forthcoming). *Information Ethics*. Oxford: Oxford University Press.
- Floridi, L., & Sanders, J. W. (2001). Artificial evil and the foundation of computer ethics. *Ethics and Information Technology*, 3(1), 55–66.
- Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14(3), 349–379.
- Gips, J. (1995). Towards the ethical robot. In K. Ford, C. Glymour & P. Hayes (Eds.), *Android epistemology* (pp. 243–252). Cambridge, MA: MIT Press.
- Greco, G. M., & Floridi, L. (2004). The tragedy of the digital commons. *Ethics and Information Technology*, 6(2), 73–82.
- Halpern, J. Y., & Moses, Y. (1990). Knowledge and common knowledge in a distributed environment. *Journal of the Association for Computing Machinery*, 37(3), 549–587.
- Hardin, G. (1968). The tragedy of the commons. *Science*, 162(3859), 1243–1248.
- Hardin, G. (1998). Extensions of the tragedy of the commons. *Science*, 280(5364), 682–683.
- Hildebrandt, M. (2008). Ambient intelligence, criminal liability and democracy. *Criminal Law and Philosophy*, 2(2), 163–180.
- Hildebrandt, M. (2011). Criminal liability in a smart environment. In R. A. Duff & S. P. Green (Eds.), *Philosophical foundations of criminal law* (pp. 507–532). Oxford University Press.
- Jonsen, A. R., & Butler, L. H. (1975). Public ethics and policy making. *Hastings Center Report*, 5(4), 19–31.
- Koops, B. J., Hildebrandt, M., & Jaquet-Chiffelle, D. O. (2010). Bridging the accountability gap: Rights for new entities in the information society? *Minnesota Journal of Law, Science & Technology*, 11(2), 497–561.
- Kroes, P. & Verbeek, P. -P. (forthcoming). *Moral agency and technical artefacts*. Springer.
- May, L. (1983). Vicarious agency and corporate responsibility. *Philosophical Studies*, 43(1), 69–82.
- Narveson, J. (2002). Collective responsibility. *The Journal of Ethics*, 6(2), 179–198.
- Pagallo, U. (2012). On the principle of privacy by design and its limits: Technology, ethics, and the rule of law. In S. Gutwirth, et al. (Eds.), *European data protection: In good health?* (pp. 331–346). Berlin: Springer.
- Pagallo, U. & Bassi, E. (2010). The future of eu working parties’ “the future of privacy” and the principle of privacy by design. In *3rd International Seminar on Information Law 2010*.
- Shaheen, S. A., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia: Past, present, and future. *Transportation Research Record: Journal of the Transportation Research Board*, 2143, 159–167.
- Simon, J. (2012). Epistemic responsibility in entangled socio-technical systems. In *Proceedings of AISB/IACAP World Congress 2012*. Birmingham, UK: AISB.
- Taddeo, M. (2009). Defining trust and e-trust: Old theories and new problems. *Journal of Technology and Human Interaction*, 5(2), 23–35.
- Taddeo, M. (2010). Modelling trust in artificial agents, a first step toward the analysis of e-trust. *Minds and Machines*, 20(2), 243–257.
- Turilli, M. (2007). Ethical protocols design. *Ethics and Information Technology*, 9(1), 49–62.
- Turilli, M., & Floridi, L. (2009). The ethics of information transparency. *Ethics and Information Technology*, 11(2), 105–112.
- Turilli, M., Vaccaro, A., & Taddeo, M. (2012). Internet neutrality: Ethical issues in the internet environment. *Philosophy and Technology*, 25(2), 133–151.
- Verbeek, P.-. P. (2011). *Moralizing technology: Understanding and designing the morality of things*. Chicago, IL: University of Chicago Press.
- Wallach, W., & Allen, C. (2009). *Moral machines: Teaching robots right from wrong*. Oxford, New York: Oxford University Press.
- Wooldridge, M. J. (2009). *An introduction to multiagent systems* (2nd ed.). Wiley: Chichester.