Ethics Inf Technol (2009) 11:81–90
DOI 10.1007/s10676-009-9183-0

ORIGINAL PAPER

How to be a responsible slave: managing the use of expert information systems

Emma Rooksby

Published online: 12 February 2009

Abstract Computer ethicists have for some years been troubled by the issue of how to assign moral responsibility for disastrous events involving erroneous information generated by expert information systems. Recently, Jeroen van den Hoven has argued that agents working with expert information systems satisfy the conditions for what he calls epistemic enslavement. Epistemically enslaved agents do not, he argues, have moral responsibility for accidents for which they bear causal responsibility. In this article, I develop two objections to van den Hoven's argument for epistemic enslavement of agents working with expert information systems.

Keywords Epistemic dependence · Epistemic enslavement · Expert information systems · Expert systems · Moral responsibility · Moral autonomy

Introduction

Computer ethicists have for some years been troubled by the issue of how to assign moral responsibility for disastrous events involving erroneous information generated by expert information systems.[1] In a nutshell, the problem is that while agents working with expert information systems, such as aircraft control-tower workers, or pilots, may be said to be *causally* responsible for disasters, it is much less clear whether or to what extent they are *morally* responsible for such disasters.

Some writers within computer ethics have recently developed arguments in support of the conclusion that agents working with expert information systems are *not* morally responsible for disasters involving expert system-generated error, even if they were causally responsible for the disasters. These arguments trade on the notion of 'epistemic enslavement', used to describe work situations involving the reliance of human agents on an expert information system. The argument goes roughly as follows: an agent relying on an expert information system to guide her in making critical decisions loses her status as an autonomous moral person, since her work environment prevents her from performing some of those acts that are constitutive of moral reasoning. Such an agent may be said to be *epistemically enslaved*. Such an agent cannot, they argue, be held *morally* responsible for catastrophic events, such as plane or train crashes, because decisions made by that agent are based on expert system-generated errors rather than on their own reasoning. The claim that such agents are epistemically enslaved is particularly associated with Jeroen van den Hoven and it is van den Hoven's argument for epistemic enslavement that I will consider here.

In the first part of the paper, I consider how useful 'epistemic enslavement' is as a moral notion: my position

E. Rooksby (✉)
School of Social and Cultural Studies (English and Cultural Studies/School of Communication Studies), University of Western Australia, Mailbox M202, 35 Stirling Highway, Crawley, WA 6009, Australia
e-mail: Emma.Rooksby@uwa.edu.au

[1] By way of a definition, an expert information system is a system of one or more computers that stores, generates, and retrieves information to be used by human agents in decision-making. Complex expert information systems can also generate advice to humans on how to act; less complex ones just store and provide information to humans.

is that it is not very useful as a guide to ascribing moral responsibility for disasters involving expert information systems. I argue that, although epistemic enslavement is actually very common, epistemic enslavement does not automatically destroy an agent's intellectual or moral autonomy. So, I conclude, epistemic enslavement is both more common, and less interesting, than computer ethicists such as van den Hoven have supposed.

In the second part of the paper I examine van den Hoven's case for assigning prospective responsibilities to agents working with expert information systems, so as best to avoid the occurrence of disasters. Van den Hoven argues that the moral responsibility of individuals working with complex information systems might be preserved, partially if not wholly, if we acknowledge that all individuals have what he calls meta-task responsibilities: namely, prospective responsibilities to ensure that they can fulfil their responsibilities without causing harm. I contest his argument and propose an alternative.

The argument for epistemic enslavement

Van den Hoven's neatest formulation of epistemic enslavement is the following: 'If a user U is epistemically dependent on expert information system S, and U is narrowly embedded in an epistemic niche of which S is part, then U is epistemically enslaved vis-à-vis S'.[2]

To unpack this we need to know more about what van den Hoven means by the terms 'epistemic niche' and 'epistemic dependence'. I'll consider them in turn. Some work environments involving an expert information system prevent workers from evaluating knowledge claims that they employ in their work, in so far as workers are constrained by external circumstances from evaluating the expert information system's judgements about what is the case.

The following four conditions, van den Hoven holds, characterise an epistemic niche:[3]

(i) *Inscrutinizability condition*: it is impossible to monitor what all the computers in an expert information system are doing (inaccessibility); or to keep track of it all (intractability);

(ii) *Pressure condition*: some decisions must be made when there is (a) very little time to make a decision (b) a decision must be made (c) one cannot get extra expertise from outside the epistemic niche;

(iii) *Error condition*: Computers may contain (a) flaws in the specification and world model of the system (b) brittleness (c) bugs and programming errors (d) limits of testing and proof (e) emergent and unpredictable properties of software, resulting from the interconnecting of systems;

(iv) Given i, ii, and iii, information systems are inhospitable to the forms of discursive scrutiny by which we traditionally seek to identify experts and to establish reliability of expert opinions [*Opacity condition*].

Together, these four conditions are taken to be jointly sufficient for the existence of an epistemic niche.

And what is epistemic dependence? This notion is developed by John Hardwig, in a paper called 'Epistemic Dependence'. One is epistemically dependent on an expert when one has good reason to believe true a claim held true by the expert, but cannot assess its truth oneself. Hardwig expresses the point as follows: 'A has good reason to believe that B has good reason to believe that *p*.' (Hardwig 1985, p. 338) One of Hardwig's examples is that of his own epistemic dependence on physicists as to the truth of Einstein's theory of relativity. Similarly, someone working with an expert information system is epistemically dependent on the expert information system in the same sense: she cannot evaluate the truth or falsity of *p* herself, has good reason to believe that the expert information system *can* perform the evaluation, and so cannot rationally refuse to defer to the findings of the expert information system.

So to recap van den Hoven's argument, a person is epistemically enslaved vis-à-vis an expert information system when the person is epistemically dependent on the expert information system, and the person inhabits an artificial epistemic niche of which the expert information system is a part. van den Hoven then takes the case further, arguing that people epistemically enslaved to expert information systems should not be held morally responsible for disasters involving their actions.

And how exactly does epistemic enslavement to an expert information system absolve someone of moral responsibility for her actions, when she is instrumental in causing a disaster? Van den Hoven puts the case like this. An epistemically enslaved agent working with an expert information system lacks intellectual autonomy, since she does not know the grounds on which claims she accepts as true, are actually true. And lacking intellectual autonomy, she cannot decide what to do *for herself*, and similarly, cannot fully justify her actions to others. But deciding what to do for oneself, and being able to account for one's actions to others are partially constitutive of moral autonomy. The agent, hence, cannot be ascribed moral responsibility for any disasters that occur as a result of her using the expert information system.

[2] Van den Hoven (1998), p. 100.

[3] Summarised from op. cit. pp. 104–107.

Analysis of the argument

As I see it, there are two main problems with the argument that epistemically enslaved agents are not morally responsible for their actions. One concern is that epistemic enslavement itself is very common, so common that if van den Hoven's argument were correct, people would turn out to have much less moral responsibility for events for which they are causally responsible than we ordinarily assume. Second, and more significantly, epistemic enslavement does not seem to do so much damage as van den Hoven thinks to the enslaved agent's autonomy, either intellectual or moral.

Regarding the first concern, I would suggest that what van den Hoven calls epistemic enslavement, in a more general form than in relation to expert information systems, is very common. Epistemic *dependence* is very common indeed: We are epistemically dependent on many experts—we assume the truth of claims such as that 'e = mc squared', that the building we work in will not fall down and so on, but cannot ascertain them for ourselves. Further, three of the four conditions that van den Hoven thinks make up inhabitation of an artificial epistemic niche also occur in many other cases of epistemic dependence, though in slightly different forms.

The inscrutinizability condition often occurs: we often cannot cross examine the relevant expert because we lack the skill to do so. The same goes for the error condition: the relevant experts may have made mistakes, and one has no way of knowing whether they did (of course the precise details will vary here—human agents are unlike computers in how they may go wrong). The pressure condition (that one has to make a decision very quickly on the spot) is the only condition that is not actually implied directly by the concept of epistemic dependence. In other words, epistemic enslavement boils down to epistemic dependence, plus a pressure condition, circumstances which occur often enough, both in work environments and elsewhere.[4]

This leads us to the conclusion that epistemic enslavement in the specialised forms discussed by computer ethicists such as van den Hoven is only the tip of the iceberg, although it seems to hold a particular interest for the computer ethicist because the 'expert' in question is not human. If we used the term 'epistemic enslavement' of all cases that satisfied van den Hoven's two conditions (inhabiting an epistemic niche and epistemic dependence), then it could be claimed that I am epistemically enslaved to my mechanic, to the people who made the power tools I use at home and even to my own glasses. While this might be a consistent use of the term 'epistemic enslavement', it suggests that the term somewhat overdraws the significance of the issue, which is simply that a reliance on the expert handiwork of others that is broadly justified may not in fact be justified in the minority of cases where an expert is mistaken.

Nor should the fact that the 'expert' is not human be taken to be as worrying as van den Hoven suggests. van den Hoven suggests that cases of epistemic enslavement to expert systems are worrying precisely *because* the 'expert' is not human, and therefore not a suitable subject for moral responsibility: in such cases moral responsibility may disappear altogether. But such cases are, once again, analogous to cases where people rely on the expert handiwork of others, but that handiwork malfunctions: it may or may not to be appropriate to hold the expert responsible for the harm, depending on the circumstances of the case. Like errors made by expert systems, malfunctions in equipment or technology that does not involve expert systems may not be directly ascribable to the designer or the builder of the equipment or technology.

To move onto the second point, which I take to be the more substantial one. This second point is that the implications of epistemic enslavement for the *slave's* moral responsibility are not as decisive as van den Hoven suggests. As I summarised it above, van den Hoven takes his argument to show that people who are epistemically enslaved 'lose their status as autonomous moral persons.'[5] Epistemically enslaved agents become 'unable to think for themselves about what is the right thing to do, and to account to others for what they have done on the basis of their thinking'. So people who are epistemically enslaved cannot be held morally responsible for their actions. But this cannot be right. I will argue that epistemic enslavement does not have such a serious effect on either intellectual autonomy or moral responsibility, taking intellectual autonomy first, and then moral autonomy and responsibility second.

Epistemic enslavement and intellectual autonomy

The nature of epistemic enslavement does not seem to be such that it entirely destroys an agent's intellectual autonomy. Let me explain. Van den Hoven's paradigm cases of epistemic enslavement are ones in which the agent who is in an artificial epistemic niche is also epistemically dependent on information supplied by the expert information system that is part of the epistemic niche.

In fact, what van den Hoven argues for that an agent who enters an epistemic niche is literally compelled to believe whatever information the expert information system in that niche gives to her. Van den Hoven starts from a

[4] I leave aside the fourth condition because, even in van den Hoven's formulation, it is not really separate from the other three.

[5] Op. cit. p. 91.

rejection of doxastic voluntarism, the doctrine that people are completely free to choose what to believe. And he points out that an agent working in an epistemic niche has been trained to take the information system's modelling of the domain of decision to be accurate, and so takes its information to be true.[6] Van den Hoven then takes these two claims to reach the conclusion that a human in an epistemic niche is compelled to believe whatever information she is supplied with by the expert information system, in the sense that the agent literally has no choice but to believe the information.

However, even if doxastic voluntarism *is* false, as is commonly accepted by philosophers today, van den Hoven's conclusion is too strong. The falsity of doxastic responsibility does not entail that those in epistemic niches are always simply *compelled* to believe information generated by an expert system, any more than I am always *compelled* to believe the petrol meter in my car, although of course in many cases I do simply assume, with good reason, that instruments such as petrol meters, are accurate. Similarly, agents working with expert information systems may have good reason to believe those systems, but it's hardly a matter of psychological necessity that they do so.

Consider the case of a worker in a specialized epistemic niche who has some access to obvious contrary evidence to expert system-generated information, even though she has insufficient time to consult relevant experts about whether or not to accept the evidence. For example, an air traffic controller who can view runways and airspace near her control tower may have reason to doubt information provided to her by an information system, *if* visual evidence available to her contradicts the computer's information. Or a market analyst using a programmed decision system may have reason to doubt the recommendation for action given by a decision support system, but still implement that decision on the grounds that the computer has greater expertise than he does. And even if a person in a situation such as these acts on the information provided by the expert information system, she may do so while still failing to *believe* that the information is correct.

The occurrence of such discrepancies can be explained by reference to the distinction between belief, and acceptance, made by L. Jonathan Cohen in *An Essay on Belief and Acceptance*. According to Cohen's (1992) distinction, if someone believes that p she has a strong sense or feeling that p is true, whether or not she affirms or acts upon p. If someone accepts that p, then she decides to affirm p, take p as a premise in her theoretical and practical reasoning, whether or not she feels that p is true. To apply this to my example, a control-tower worker may believe that a runway is not clear, because she has seen a plane headed towards it, and yet have expert system-generated information that the runway *is* clear. Yet she may still *accept* that the runway is clear (on the rational grounds that the system's judgement is generally far more reliable than hers), and act on the judgement that it is clear.

So it is not always the case that information delivered by information systems to people in epistemic niches 'compels belief', even if the pressure of their working conditions means that they act on that information as if they believed it. An individual may have access to strong contrary evidence but, insufficient time to weigh them up against the expert system-generated information, to arrive at a final reasoned judgement about which to accept. Epistemically enslaved agents are not literally forced to believe an expert system, though they may have no reasonable alternative to accepting its information as true.

Nevertheless, it is true people working in epistemic niches may, through lack of time to reflect and reconsider, act against their own beliefs, when confronted with contradictory evidence from an information system. Knowledge that the expert information system is, on balance, more reliable than human judgement, may lead an agent, to discount the contrary evidence, and to *accept* the expert information system's information, even if she doesn't *believe* it, and even though the expert information system is wrong and the agent is right. This point raises interesting moral issues about when to trust one's own judgement, and when to defer to an expert (human or otherwise). These, however, are issues that I do not have space to address here.

So, to summarise this second objection, the pressure condition can push an agent to *accept* expert system-generated information, even if she does not herself *believe* that information. The agent is not enslaved in the strong sense that van den Hoven needs to show that the agent lacks intellectual autonomy. It seems as if epistemic enslavement itself does not completely destroy an agent's intellectual autonomy. Of course, some of the decisions epistemically enslaved agents make may lead to disasters. But the defence open to such agents is that they were rational to defer to the expert information system in the circumstances, not that they were compelled to do so.

Epistemic enslavement and moral autonomy

I'd now like to move on to the topic of the relation between epistemic enslavement and moral autonomy. Even if we granted that epistemically enslaved agents always lack intellectual autonomy (a claim that I have just disputed), it is not clear that such a lack of intellectual autonomy would also deprive the agent concerned of the capacity to act as a morally responsible individual. Situations in which epistemic enslavement is coupled with some degree of

[6] Ibid. p. 103.

discretion in action leave the agent in question responsible for how she responds to inaccurate expert system-generated information, in the sense that she actually has a choice open to her as to how to respond. And this may well be the case for an agent even if she had no choice but to accept that information as true. (Of course, however, in cases where agents have little or no discretion in how they respond, as when they are given orders by an expert information system, there may be no real choice available to the agent, in the sense that there is no reasonable available alternative to the expert information system's order.)

And there are many situations of this sort for agents working with expert information systems. Many expert information systems give the people who use them a choice about how to act, and some give no advice about how to act at all. In these cases, whether or not epistemic enslavement occurs with regard to expert system-generated information, the agent has some discretion as to how to use that information. And when people in such circumstances act on mistaken information, but have discretion about how they employ it, those people may be deemed responsible for how they used the information, even if they are not deemed responsible for their employment of incorrect information.

This point can be clearly put if we consider an appeal to epistemic enslavement as a variety of appeal to ignorance on behalf of the enslaved agent. Following Ronald Milo on the appeal to ignorance: 'In cases where ignorance is appealed to as an excuse, we find that, if the excuse is valid, the agent could not have avoided the act, given his ignorance, and could not reasonably be expected to have avoided the ignorance.'[7] The ignorance in the case of epistemic enslavement would be ignorance that a certain piece of expert system-generated information they have been given is false.

And it is simply not true of all cases of epistemically enslavement, that the agent concerned could not have avoided the act that caused a disaster. It is at least not true for those agents who are epistemically enslaved to an expert information system, but who also have a degree of discretion in how they act on information generated by the expert information system (and where at least one reasonable alternative is available to the agent).

I'll use an example to illustrate this point. Consider the case of the French police officers who mistakenly shot and killed the driver of a car that they believed to be stolen, on the basis of inaccurate computer-provided information (discussed at van den Hoven 1998, p. 98). The officers have information that a car they can see is stolen, but are not given any recommended course of action to follow. We could agree that the officers are epistemically enslaved with regard to the expert system-generated information that the car they can see is stolen, but also hold that the officers' have discretion as to how they respond to that information. The officers have discretion in how they employ the information that the car they saw is stolen, even if their belief that the car is stolen is both false and one that is acquired through epistemic enslavement. And, let us assume, less fatal courses of action are open to them than the one they take. The police officers could, then, arguably have avoided the act that caused the death of an innocent person, even given their ignorance. They can thus be deemed to be morally responsible (in the sense of answerable) for their actions in response to inaccurate expert system-generated information, even if they are not held responsible for their belief in that inaccurate expert system-generated information. In other words, it cannot be true that epistemic enslavement always absolves the enslaved agent from moral responsibility for any disaster in which she had a hand, and in which expert system-generated error also had a hand.

So, to summarise my position on epistemic enslavement. Epistemic enslavement is common, and not only with users of expert information systems. Epistemic enslavement does not necessarily destroy an agent's intellectual autonomy, because agents working with expert information systems are not literally compelled to *believe* the information the systems give them, although they may have no reasonable choice but to *accept* it. And epistemic enslavement does not always and necessarily absolve an agent from moral responsibility for a disaster, since in many cases the agent has discretion in how she employs expert system-generated information, and will have a choice among reasonable alternative actions to pursue. Her ignorance about the facts will not be sufficient to excuse her from responsibility, if, despite her ignorance, she could have acted in such a way as to avoid the disaster.

Still, I don't want to deny that in *some* cases at least, agents working with expert information systems are indeed not to be held morally responsible for disasters involving both expert system-generated errors and their own actions. I only want to deny that the notion of epistemic enslavement effectively marks the point at which the moral responsibility of agents working with expert information systems stops.

Assigning moral responsibility for avoiding expert system-generated disasters

Now I want to move on to consider van den Hoven's solution to the problems associated with assigning moral responsibility for avoiding disasters involving expert system-generated errors and human actions. The issue

[7] Milo (1984), p. 223.

addressed by van den Hoven is that of how to assign prospective moral responsibility for avoiding expert system-generated disasters. In this section, I argue that van den Hoven's analysis places too much of the onus on the users of those systems, despite the fact of their being epistemically enslaved to them, while neglecting broader institutional responsibilities.

Van den Hoven adopts Robert Goodin's (1995) consequentialist account of the distribution of prospective responsibilities among the members of a collective. In this account, prospective responsibilities fall into two main types: task responsibilities and negative task responsibilities. A Task Responsibility concerning X is the obligation to see to it that X is brought about and a Negative Task-Responsibility concerning X is the obligation to see to it that no harm is done in seeing to it that X is brought about.

In addition, van den Hoven posits that agents have meta-task responsibilities, which are a variety of supervisory responsibilities that are entailed by task responsibilities[8]:

> A user A has a meta-task responsibility concerning X means that A has an obligation to see to it that (1) conditions are such that it is possible to see to it that X is brought about and (2) conditions are such that it is possible to see to it that no harm is done in seeing to it that X is brought about.

Van den Hoven argues that meta-task responsibilities remove the legitimacy of an employee's appeal to epistemic enslavement to absolve herself of moral responsibility in acting upon an expert system-generated error. The imposition of meta-task responsibilities renders any appeal to epistemic enslavement at best a partial excuse for wrongs done, by making individuals take responsibility for the conditions in which they work. Even if A is epistemically enslaved when making a decision that results in some disaster, she is still responsible for seeing to it that conditions are such that her job can be done, and done without causing harm. And A can still be held responsible (in the sense of being answerable) for any harm that results from her job.

Meta-task responsibilities for an individual acting alone

Van den Hoven assumes that an individual's task-responsibility simply entails a corresponding meta-task responsibility, in the following way: if someone takes up or is assigned a task-responsibility, then she also has the entailed meta-task responsibility to ensure that it is possible to do this task. I will argue against accepting this assumption for the case of an individual participating in a collective action, although it may be acceptable for the case of an individual acting alone.

To do this, I argue that meta-task responsibilities break down into two distinct components, not all of them entailed by individual task-responsibilities. The characteristics of task-responsibilities for members of a group show that *one element* of meta-task responsibility, that pertaining to the actions of others upon whom one relies, is not entailed for each member of a group, but *for the group as a whole*.

Consider first the nature of meta-task responsibility for an individual acting in isolation, someone with the task-responsibility of feeding a dog. Call him the lone dog-feeder. The lone dog-feeder must see to it that the outcome, that his dog is fed, obtains, whatever seeing to it involves. Consider how a meta-task responsibility might be understood in the case of one individual and his responsibility. If the man plans to take the bus to the butcher's each Saturday to pick up dog meat, then to fulfil his meta-task responsibility, he should see to it that it is possible to fulfil his task-responsibility. This would involve ascertaining (or at least presuming) that he is capable of fulfilling it, and, perhaps, ensuring that he is capable of fulfilling it. For example his meta-task responsibility will involve *ensuring* that he has time free on Saturdays, and *ensuring* that he knows which bus to catch, and so on. These responsibilities are self-supervisory responsibilities, entailed by the relevant task-responsibilities; they can be said to be *internal meta-task responsibilities*, i.e. pertaining to the lone dog-feeder's own capacities.

The lone dog-feeder's meta-task responsibilities may also include seeing to it that people, and social regularities upon which he plans to rely for seeing to the feeding do not obstruct his plans; these could be called *external meta-task responsibilities*. For the individual agent, these involve two separate tasks, of *ascertaining whether* external conditions are such that it is possible to fulfil his task-responsibilities, and of *ensuring that* external conditions are such that it is possible to fulfil his task-responsibilities. These two tasks are linked in the activity of planning. Alternative courses of action will be open to the lone dog-feeder; he will have to ascertain what they are, and what external obstacles he will need to negotiate for each. If at least one course presents no external obstacles, then it is possible for him to fulfil his task-responsibilities *simpliciter*. If all do, or if he chooses one that includes external obstacles, he will have to plan additional actions to the tasks involved in his plan for fulfilling his task-responsibilities. What additional actions his external meta-task responsibilities will require, depends on which course of action he plans to take.

Once he has settled on a course of action, the lone dog-feeder is committed to negotiating the obstacles involved in following that course of action. Thus, if he takes a course of action that involves his *ensuring* that the shop has

[8] Op. cit. p. 103.

chopped chicken livers available, he acquires a conditional responsibility to ensure that the shop does indeed have them.

In the case of an individual with a task-responsibility, then, the relevant meta-task responsibilities can be specified in terms of two components. First, there are self-supervisory meta-task responsibilities, or *internal meta-task responsibilities*, involving ascertaining and ensuring that one has the relevant capacities to fulfil one's task-responsibilities. Second, there are *external meta-task responsibilities*, firstly to ascertain whether external circumstances permit one to fulfil one's task-responsibilities (on any of the available action plans), and to ensure that they permit one to fulfil them (given one's choice of action plan). Even in the case of an individual acting alone, external meta-task responsibilities seem rather far-fetched. We would at least need to insert some qualification into the account of meta-task responsibilities limiting the requirement on the agent to that of only 'taking reasonable steps' to ensure that etc. Even more may need to be done. But let us turn to meta-task responsibilities in relation to an individual acting as a member of a group. The problems here are rather more serious.

Meta-task responsibilities for an individual acting as a member of a group

What do meta-task responsibilities look like in relation to an individual acting as a member of a group? The self-supervisory responsibilities that I characterized as internal meta-task responsibilities will apply to each individual, and they will take the same form, since they still concern only the individual's capacities.

For external task-responsibilities the picture looks somewhat different. Van den Hoven's formulation of meta-task responsibilities applies directly to each individual in a collective; so it gives each member of a collective the responsibility to see to it that those collective members upon whom the fulfilment of her task-responsibilities depends, fulfil their task-responsibilities. However, I argue that, for typical cases of collective action, an individual member's external meta-task responsibilities are a subset of the internal meta-task responsibilities of the group.

Consider what might be asked of an individual member of a group, as external meta-task responsibilities. Assume that the group uses an expert information system in some way. They cover rather more than the 'checking of information systems' that van den Hoven cashes out as the meta-task responsibility for an end user of an information system, covering several individual action-types.[9] These include, but are not limited to, seeing to it that the tools and equipment (including information systems) one uses are functioning appropriately; seeing to it that other members (both junior and senior) of the group on whom one relies in fulfilling one's task-responsibilities are fulfilling their task-responsibilities, and without causing harm (this extends to cover all those on whom the other members depend too); and so on.

These meta-task responsibilities are a subset of the meta-task responsibilities of an organization as a whole. That is to say, no individual member of a group has an external meta-task responsibility that is not a member of the set of internal meta-task responsibilities held by the members of the group.

To see this, consider a typical case of collective action by agreement, an action to which more than one person contributes.[10] Imagine a group with a task or set of tasks to perform. These tasks are divided up into part tasks; each member is given task-responsibility for a part task of the larger task of the collective as a whole. Each individual fulfils her task-responsibility *as her part* of the larger task-responsibility of the collective, while also intending that the members of the collective together fulfil the larger task-responsibility (call it L). The fulfilment of L consists of the fulfilment of L's parts ($P_i \ldots P_n$) by the group's members, where the members fulfil those parts as parts of collectively fulfilling L. In such circumstances, 'each participant has accepted an obligation towards the others to do his part and correspondingly has the right to demand that others do their parts.'[11]

If one member of a collective intending to L is not fulfilling her part- task (P_m) of L, the other members are responsible for seeing to it that the part task is fulfilled, since the fulfilment of the part task is a condition of the fulfilment of L, to which all members of the collective are committed. In cases where one person can't do their bit, the others, if they are indeed committed to the fulfilment of L, 'will try to help him in carrying out his part and to bring about a successful joint action.'[12] But the extra things to be done are not required of any single member of the group, *but of the rest of the group as a whole*. The responsibility to 'take up the slack' does not fall on any single member of the group until the group members determine where it should fall.

The same holds for meta task-responsibilities for a group. Relative to a single group member, meta-task responsibilities concerning the performance of other group

[9] Remember that external meta-task responsibilities include both tasks of *ascertaining that* and tasks of *ensuring that*.

[10] This is a very rough definition of collective action. For instance it does not specify whether the collective agreement in question must be explicit, or whether it may be tacit.

[11] Tuomela (1995), p. 76.

[12] Ibid. p. 95.

members are *external*. But relative to the *group taken as a whole*, meta-task responsibilities to see to it that all parts of the task can be fulfilled without harm are *internal*, concerning the capacities and organization of the group as a whole. How the group sees to it that it is possible for its members to fulfil its various task responsibilities is up to the group as a whole, at least in the first instance. For example a group might institute supervisory responsibilities, by nominating some, or even all, group members as managers or as monitors. Alternatively, the group might commission an independent assessment of its structure and functions. If the organization as a whole sees to it that L can be performed without harm (by its members collectively delegating this meta-task responsibility to some particular members), then surely there is no need for individual members to see to it separately.

This position seems to me fairly plausible, at least for the case of a group of equals deciding together how they will fulfil a group task-responsibility.[13] It may, of course, leave us with no obvious candidate to hold answerable for a disaster. But that, surely, is a different sort of problem.

Effects of giving each member external meta-task responsibilities

Should managers assign external meta-task responsibilities to each individual employee? Two problems suggest that such a step should not be taken. These problems are, first, the impracticability of so doing, and second, the political burden of meta-task responsibilities when placed on the shoulders of subordinate employees.

To the first problem, which I call the Demandingness Problem. The error condition that is one of the conditions for any employee working in an artificial epistemic niche makes it difficult for the individual public administration employee, a non-specialist in information systems, to see to it that information systems are free from error. The implication is that individual employees will, in almost all cases, have to rely on the word of an expert in assessing computer systems. As we have already seen, in the error condition, even the designers of expert information systems cannot guarantee that such systems are error-free.

Clearly, most employees will not be able to find errors in information systems on their own. This can be confirmed by reviewing the four types of condition that together make up the error condition. The first subcondition, that of flaws in the specification and world model of the system, is the one where the individual worker is most likely to be able to find fault. But consider the other four: (b) brittleness; (c) bugs and programming errors; (d) limits of testing and proof; and (e) emergent and unpredictable properties of software that emerge as a result of the merging and interconnecting of systems. Even an expert investigation of the information system prior to working with it is unlikely to turn up errors of any of these kinds. And further problems are associated with employing an expert. First, the employee is not in a position to assess the expert's assessment of the system. Second, it is unclear what difference it would make for each employee to commission their own checks, rather than relying on periodic checks commissioned by the group as a whole (i.e. by assigning the relevant external meta-task responsibilities to supervisors or some other member of the organization).

To the second problem, which I call the Source of the Error Problem. One feature of information systems makes fulfilling external meta-task responsibilities especially difficult *in political terms* for subordinate staff members. The relevant feature is that information systems take both expert system-generated and human inputs. An information system is both an information-transferring *medium* and an information-generating *expert*. Thus there will be cases in which individuals are epistemically enslaved, but working with both expert system-generated and human inputs. In such cases, the epistemically enslaved worker will be unable to assess the origin of the information to which she has access. The *source of the error* will be opaque to the enslaved worker.

Now, if a subordinate staff member has a meta-task responsibility for seeing to it that it is possible for her to fulfil her positive and negative task responsibilities, then that agent may find herself responsible for seeing to it that her superiors give her accurate information. The Source of the Error Problem thus makes the imposition of external meta-task responsibilities especially onerous in the case of junior staff members. It is one thing to be responsible for seeking out errors in expert systems (that is the Demandingness Problem), and quite another to be meta-task responsible for seeking out deliberate distortion or falsehood by one's superiors.

It must be noted, then, that as a result of the Demandingness Problem and the Source of the Error Problem, individuals will sometimes *simply not be in a position* to see to it that they will be able to fulfil their positive and negative task responsibilities. Hence, such individuals cannot be held automatically accountable-responsible for failing to fulfil their meta-task responsibilities, as it would be asking the impossible of them. This, unfortunately,

[13] It may be less plausible for cases in which decisions about how a group will fulfil a group task-responsibility are made by only some members of the group; it may also be less plausible for groups in which task-responsibilities are allocated, not to individuals, but to *functional place-holders,* which could be filled by any individual who happened to be assigned that function (for example, by becoming employed by the organization to fill that place-holder). In the next section I suggest that these complexities can be used to provide some conditions under which individuals, even those with subordinate positions in an organization, have meta-task responsibilities.

limits the promise held by meta-task responsibilities in apportioning moral responsibility and answerability for disasters involving human agency and expert system-generated error.

To conclude, it is difficult to assess in the abstract the effectiveness of giving individuals working with expert information systems external meta-task responsibilities, independent of the question of whether such meta-task responsibilities fall automatically on their shoulders upon their acquisition of certain task responsibilities. Something can be said, however, in hypothetical terms. It *may* indeed turn out to be best, in a given case, to give *each* agent his or her own meta-task responsibility. But this would be something to be determined for each individual case.

Some preliminary suggestions for avoiding expert system-generated disasters

Meta-task responsibilities for group members, whether distributed as I have argued they are, or for each individual, as van den Hoven has assumed, do have some role to play in complex organizations. I believe that they are best treated as supervisory responsibilities, falling on the group as a whole in the first instance, and distributed according to the two major considerations of goodness of outcome, and fairness of distribution. Here I provide a few suggestions on the distribution of meta-task responsibilities in work-places that use expert information systems. First, workplaces should give employees extra time to fulfill any supervisory and meta-task responsibilities they may be allotted, however these are distributed within the organisation. Likewise, giving staff working with expert information systems greater training in operating them will allow staff to make the best use of what discretion is available to them when working in artificial epistemic niches.

Second, it is important to take advantage of the fact that most employees working with expert information systems do not inhabit specialized epistemic niches for all of their working hours. No matter what their working conditions, it will almost always be possible for employees to take the time to reflect on the conditions in which they work, and to consider, as far as they are able, the workings of the expert information system that they use. This allows them, for instance, to become aware of institutional pressures that generate unnecessary pressure conditions in the work place, and of oddities in the expert information systems' functioning, and to report these to the appropriate authorities.

Third, to enable employees to exercise meta-task responsibility in cases where they become aware that the group as a whole has not done so, institutional innovations that allow that exercise, such as 'internal ombudsmen', 'open door policies' and anonymous complaints centers, should be explored. Mark Bovens describes these and other options in *The Quest for Responsibility*. Bovens and Stavros Zouridis (1998), in a paper on IT in public administration, make further suggestions along these lines. For example, they recommend an 'informatization review' when new executive rules are introduced.[14]

Conclusions

To conclude, then, the condition of epistemic enslavement is not as relevant to the question of agent's moral responsibility for disasters involving expert information systems as van den Hoven has argued. It does not entail a complete loss of either intellectual autonomy or of moral autonomy for agents working with expert information systems. And in some cases, agents working with expert information systems will actually be, to some degree at least, morally responsible for disasters in which they have had a hand. Whether such agents are also to be blamed for such disasters, in part or whole, is a further question to be addressed.

Attributing meta-task responsibilities to people working with information systems is one way of being able to assign responsibility for information-system disasters. However, it will only be reasonable to the extent that individuals working with information systems *are* actually capable of ensuring that an information system will not prevent them from performing their positive and negative task-responsibilities. This may be reasonable in the case of internal meta-task responsibilities, concerning the individual's own capacities and plans. It is less reasonable concerning the circumstances in which the individual operates. Given the error condition specified by van den Hoven, the Demandingness Problem, and the political difficulties associated with the Source of the Error Problem, it seems unwise to be too optimistic about individuals' abilities to fulfil external meta-task responsibilities.

References

Bovens, M. (1998). *The quest for responsibility: Accountability and citizenship in complex organizations*. Cambridge: Cambridge University Press.

Bovens, M., & Zouridis, S. (2001). From street level to system level bureaucracies: how ict is transforming administrative discretion and constitutional control edited version of a paper presented at the 2001 PAT-NET conference. Leiden University, The Netherlands, 21 and 22 June.

[14] Bovens and Zouridis (2001), p. 18.

Cohen, L. J. (1992). *An essay on belief and acceptance*. Oxford: Clarendon Press.
Goodin, R. (1995). *Utilitarianism as a public philosophy*. Cambridge: Cambridge University Press.
Hardwig, J. (1985). Epistemic dependence. *The Journal of philosophy*, 82(1), 335–349.
Milo, R. (1984). *Immorality*. Princeton NJ: Princeton University Press.
Tuomela, R. (1995). *The importance of us: A philosophical study of basic social notions*. Stanford CA: California University Press
van den Hoven, J. (1998). Moral responsibility, public office and information technology. In I. Th. M. Snellen & W. B. H. J. van de Donk (Eds.), *Public administration in an information age: A handbook* (pp. 97–112). Amsterdam: IOS Press.