“The Machine Made Me Do It!”: An Exploration of Ascribing Agency and Responsibility to Decision Support Systems

- HANNAH HAVILAND -
Master’s Thesis in Applied Ethics
Centre for Applied Ethics
Linköpings universitet
Presented May 2005

Supervisor: Prof. Göran Collste, Linköpings universitet
"The Machine Made Me Do It!": An Exploration of Ascribing Agency and Responsibility to Decision Support Systems

Abstract
Are agency and responsibility solely ascribable to humans? The advent of artificial intelligence (AI), including the development of so-called “affective computing,” appears to be chipping away at the traditional building blocks of moral agency and responsibility. Spurred by the realization that fully autonomous, self-aware, even rational and emotionally-intelligent computer systems may emerge in the future, professionals in engineering and computer science have historically been the most vocal to warn of the ways in which such systems may alter our understanding of computer ethics. Despite the increasing attention of many philosophers and ethicists to the development of AI, there continues to exist a fair amount of conceptual muddiness on the conditions for assigning agency and responsibility to such systems, from both an ethical and a legal perspective. Moral and legal philosophies may overlap to a high degree, but are neither interchangeable nor identical. This paper attempts to clarify the actual and hypothetical ethical and legal situations governing a very particular type of advanced, or “intelligent,” computer system: medical decision support systems (MDSS) that feature AI in their system design. While it is well-recognized that MDSS can be categorized by type and function, further categorization of their mediating effects on users and patients is needed in order to even begin ascribing some level of moral or legal responsibility. I conclude that various doctrines of Anglo legal systems appear to allow for the possibility of assigning specific types of agency – and thus specific types of legal responsibility – to some types of MDSS. Strong arguments for assigning moral agency and responsibility are still lacking, however.

Nyckelord
computer ethics, agency, responsibility, liability, decision support systems, DSS, technology law, human computer dependency, artificial intelligence, AI
Acknowledgements

I would like to thank my professors and fellow Master’s students at the Centrum för Tillämpad Etik at Linköping University for their feedback on drafts of this paper, and for creating an environment that supports and encourages analysis of practical ethical issues. I have appreciated the solid encouragement, candid advice, and continual patience of my thesis adviser, Prof. Göran Collste, as my ideas have taken shape.

Various other people have lent a helping hand in this work in ways they may not know. They include Prof. Andrew Murray, of the Law Department at the London School of Economics, for helping with legal arguments and finding unusual sources; Prof. Simon Rogerson, of DeMontfort University in Leicester, for pointing out areas of improvement in my arguments; and Mitchell Brown, of Princeton University’s Fine Hall Library, for helping me track down elusive literature. Friends and family in Sweden and abroad have provided support, encouragement, and, when needed, opportunities to focus on things other than research.

Finally, my deepest thanks to the Fulbright Commission, the Stockholm Fulbright Office, and the American and Swedish governments, all of whose generosity and farsightedness in supporting post-graduate academic and cultural exchange has made my time in Sweden, and this thesis, possible.
Table of Contents

Acknowledgements ................................................................................................................... 3
Table of Contents .................................................................................................................... 4
Chapter 1 ................................................................................................................................... 5
Introduction and Rationale: Why DSS Deserve a Closer Look ................................................. 5
  1.1.1 A Cautionary Tale ........................................................................................................ 5
  1.1.2 Intelligent Computer Systems in Everyday Life .......................................................... 6
  1.2 Rationale: Why Focus on Advanced Medical DSS? ........................................................... 6
  1.3 Primary Analytical Questions .......................................................................................... 7
  1.4 Roadmap, Goals, and Methodology .................................................................................. 9
  1.5 Limitations of the Study ................................................................................................. 10
Chapter 2 ................................................................................................................................. 12
Ethical & Legal Perspectives on Agency & Responsibility ....................................................... 12
  2.1 Agency .............................................................................................................................. 12
  2.1.1 Ethical Perspectives on Agency .................................................................................. 13
  2.1.2 Legal Perspectives on Agency .................................................................................... 18
  2.2 Responsibility .................................................................................................................. 22
  2.2.1 Ethical Perspectives on Responsibility ........................................................................ 23
  2.2.2 Legal Perspectives on Responsibility .......................................................................... 24
Chapter 3 ..................................................................................................................................... 31
Advanced DSS in Healthcare: A Shift in the Standard of Care? ................................................ 31
  3.1 Intro to Intelligent MDSS - types, applications, functions ............................................... 31
  3.2 Ethical Status Quo ............................................................................................................ 37
  3.3 Legal Status Quo .............................................................................................................. 37
  3.4 Agency: Non-Traditional Ethical and Legal Perspectives ................................................... 38
  3.5 Responsibility: Non-Traditional Ethical and Legal Perspectives ....................................... 44
  3.6 Summary of Arguments and Counterarguments ................................................................ 48
Chapter 4 ..................................................................................................................................... 55
Conclusions and Guidelines for Best Practice in Law and Healthcare ................................... 55
  4.1 Conclusions and Implications .......................................................................................... 55
  4.2 Suggested Best Practice for Healthcare Professionals ...................................................... 58
  4.4 Larger Implications: Should DSS Usage and AI Research Continue? ............................ 62
Works Cited and Consulted ...................................................................................................... 64
Chapter 1

Introduction and Rationale: Why DSS Deserve a Closer Look

“If machines are developed that behave in much the same way as humans do, in a wide variety of contexts, the issue of whether they are things with moral rights and responsibilities will arise. Consideration would need to be given to how they should be treated. If they behaved like us, would we be justified in treating them differently?” (Weekert 2002: 369).

1.1.1 A Cautionary Tale

In an episode of Krzysztof Kieslowski’s 1988 film, Decalogue I, a father and son in Soviet Poland design a computer program that, among other things, calculates the risk of the ice breaking in the nearby lake. Every day, they enter data from the local weather station. The program performs a simple algorithm to determine the minimum load bearing capacity of the lake that day. Father and son have a curious relationship with the computer, treating it with a degree of respect and reverence usually reserved for an omniscient religious figure – all the more ironic because the father has foresworn religion. Indeed, they seem to treat the computer program’s output as a substitute for the infallible “word of God,” which proves even more tragic when, after running the algorithm one day to calculate the ice’s load bearing capacity, the son falls through the ice.

What is Kieslowski’s intent? A morality tale along the lines of, ‘don’t trust computers/technology but only the word of God,’ perhaps? Certainly, many have interpreted the story from this technical-determinist perspective. There are other films that address similar themes and reach similar conclusions. The story of the human captains, Dave and Frank, versus the intelligent computer, HAL-9000, from Clarke’s 2001: A Space Odyssey comes immediately to mind. But my purpose in recanting such cautionary tales is to highlight the often “blind faith” that many have in a certain type of computer program: the decision support system (DSS). Indeed, the computer program that father and son build in Poland can be seen as a sort of DSS, albeit a very rudimentary one. The user inputs data. The program runs an algorithm and displays a human-readable answer, which in turn enhances the user’s decision-making ability (ultimately, whether or not it is safe to skate on the ice). HAL fills much the same role, although it possesses decidedly more advanced – and more nefarious – capabilities and capacities.
1.1.2 Intelligent Computer Systems in Everyday Life

Blind faith in the advantages and capabilities of a DSS can be just as harmful as irrational fear of the technology. The reality today is that DSS are used in many aspects of modern life. If you should ever require treatment at an advanced hospital, your treatment probably will either be monitored by an advanced medical DSS, or monitored by a healthcare professional who uses such a system to diagnosis your condition or to plan the next step in your treatment. If you buy shares in a mutual fund, the fund manager’s decision when to buy or sell stock is almost certainly informed (if not actioned) by a computer program, which crunches large amounts of market data to identify the optimal conditions – arbitrage situations – for executing a trade. In the U.S., if you apply for private, non-federal health insurance, your medical, financial, genealogical history, even geographical location, may all be analysed by a computer program to calculate the premium you pay, or the amount of the claim you may seek, which themselves are assessed by the calculation of risk that you will draw on the insurance at a certain time, for certain amounts.¹

1.2 Rationale: Why Focus on Advanced Medical DSS?

Of all of these examples of the use of DSS in everyday life, one of the most philosophically interesting is that of medical DSS (MDSS), used in healthcare.² MDSS have been used for decades, although technological advances have improved their quality and expanded their possible applications.

Broadly spoken, MDSS enhance the user’s decision-making capacity, but can do so by different means. As described in more detail in Chapter 3, an MDSS in use today might do any of the following:

- Organize and manage a large body of medical knowledge
- Mine data stored in various patient records
- Train and educate healthcare professionals
- Remind healthcare professionals to perform a task
- Analyse input or stored patient data to form a diagnosis or recommend a treatment plan
- Respond to input patient data to administer or forego a course of treatment

¹ For example, some insurance companies used Personal Injury Evaluation Systems to speed up and simplify the claims process. For empirical data on this, see D. Person, “DSS in Insurance Claims Departments: Personal Injury Evaluation Systems” in International Review of Law, Computers & Technology. 2000, 14(3): 371-83.
² Decision support systems used in healthcare are usually called either medical DSS (MDSS) or computerized clinical decision support systems (CDSS); see Garg et al. (2005). I treat MDSS and CDSS as synonyms for the purposes of this paper.
In the last example, an MDSS may actually replace human action, not just support human decision-making. In Chapter 3 I examine the implications of the often hairline distinction between “supporting” and “replacing” human decision-making and subsequent action.

Many MDSS created today utilize artificial intelligence (AI) in their system design, which in this context denotes them as “intelligent” or “advanced” DSS. Ethicists are increasingly paying attention to the questions of moral agency and responsibility that might arise in intelligent computers systems; however, not much current work analyses the issue using concrete examples of the issues in those sectors in which it is most prevalent.

This paper will explore the ethical and legal implications of human computer dependency, concentrating on the possibility of considering intelligent computer decision support systems (DSS) to be agents in both the moral and legal sense, thus making them partially responsible for their decisions. By referring to real-life scenarios and cases of the widespread use of advanced DSS by professionals in the healthcare industry, I will be able to examine how the concepts of free will, autonomy, and rationality mediate the related but separate concepts of moral agency and responsibility, from both an ethical and a legal perspective. Scenarios ground the research, which is by nature largely hypothetical, in the real world. In the process of highlighting actual and potential (as well as current and future) problems in DSS use, the paper also hopes to make an original contribution to current Best Practices dialogues.

1.3 Primary Analytical Questions

The main question driving this study is this: Are agency and responsibility solely ascribable to humans? To answer that larger question, I set out first to answer smaller, more manageable questions, which can be divided into the general (analysing general ethical and legal concepts) and the specific (pertaining to the use of DSS in healthcare).

General:

1. What are the traditional concepts and types of moral agency? Legal agency? In Ch. 2 I present the most common, mainstream accounts of moral and legal agency, focusing on the mainstream criteria for action and for agency. Legal agency is defined slightly differently and can extend to non-human actors.

As is described more in Ch. 2 and 3, the most significant moral and legal issues that arise with MDSS use are those of responsibility and liability. In this paper I treat moral responsibility separately from legal responsibility (and throughout I use the modifier “ethical” vs. “legal” to ensure the reader knows to which one I am referring), while nonetheless maintaining that the two concepts are inextricably linked.
2. **What are the traditional concepts of moral responsibility? Legal responsibility?**
   Also in Ch. 2, I present the mainstream accounts of moral responsibility, focusing on the criteria for responsibility and its cousin, liability. Legal responsibility almost exclusively refers to legal liability, and thus is usually in the domain of tort law (see Sect. 2.2.2 for more).

Specific:

1. **For what purpose(s) do healthcare professionals use DSS?** As described in Ch. 3, MDSS can be used for education, training, data management, consultation, or a combination of these. My arguments for attributing responsibility to MDSS rests on a careful examination of the type of MDSS in question, focusing on both its role and behaviour in the specific healthcare environment.

2. **How do MDSS mediate, affect, or influence the patient, the doctor-patient relationship, or the doctor/user?** I argue that the use of MDSS undoubtedly mediates a number of dimensions – the patient’s health, the doctor-patient relationship, and the doctor – in a number of ways (as described in Sect. 3.1), which ensures that, when doctors use it (or, refuse to use it), the matter falls under the remit of ethical deliberation.

3. **Do advanced MDSS, by their design, functioning, or usage (or all three) affect traditional concepts of moral agency? Legal agency?** In Ch. 3 I treat moral and legal agency separately, as the two concepts are intertwined but certainly not interchangeable. Here I examine the design of MDSS, the way in which MDSS function, and the ways in which MDSS are operated (by a healthcare professional or patient), to argue that advanced MDSS do affect the mainstream accounts of both moral and legal agency.

4. **Do advanced MDSS, by their design, functioning, or usage (or all three) affect traditional concepts of moral responsibility? Legal responsibility?** In answering this question in Ch. 3, I adopt much the same approach as in the previous question, but focus this time on mainstream accounts of moral and legal responsibility and its cousin, liability.

5. **Is there a Moorian “policy vacuum”? A legal vacuum?** In Ch. 4 I argue that both a policy and a legal vacuum exist (to borrow Moor’s term) when MDSS are used. Matthias (2004) has expressed much the same idea as a “responsibility gap” that arises when not only advanced MDSS but any intelligent, “learning automata” are used (Matthias 2004: 175-6). I characterize the legal situation as a legal “lag.” “Gap,” “lag,” or “vacuum” notwithstanding, this situation is far from ideal, so I propose ways to
rectify this. I hope, in the process, to make an original contribution to Best Practices guidelines for those in the healthcare industry who are affected by MDSS use.

1.4 Roadmap, Goals, and Methodology

Chapter 2 tackles the “general” analytical questions above: the ethical and legal philosophical questions about agency and responsibility. It is intended as a backdrop, or perhaps a framework, against which I will present competing arguments in the following chapter. Chapter 2 also equips the reader with a little understanding of the relevant legal doctrines, which I will then explore in the chapter that follows. Chapter 3 addresses the “specific” analytical questions above, beginning with an overview of MDSS’ design, function, and common usage. Chapter 4 draws together common conclusions from the foregoing analyses, including proposals for possible directions of fruitful, future research. I also address the arguments whether research into, development of, and usage of advanced MDSS should cease or continue, on moral grounds.

My goals can be broadly expressed as to:

1. Give an overview of the disparate legal and ethical approaches to dealing with responsibility and agency, covering the mainstream and traditional approaches, the non-traditional ones, the new approaches, and the older ones. Clarify the meanings of such terms.

2. Expose shortcomings and strengths in existing approaches, as they are applied to MDSS. Show where some arguments go too far, where yet others do not go far enough. Sift through these various accounts to attempt to create a more unified and accurate picture.

3. Contribute to Best Practices for people and institutions that create, use, or are otherwise affected by MDSS.

4. Highlight areas for future research.

In constructing and defending my arguments, I refer to a variety of works of moral and legal philosophy. DSS are a developing and evolving technology; so, too, should be the debates about the technology. For this reason, I have consulted earlier works on DSS (from the 1980s and 1990s) but devote more attention to essays and studies published within the last five years or so. Although I have consulted a variety of sources – edited volumes, journal articles, websites, textbooks, and single-author books – my arguments find their greatest inspiration in the writings of Garg et al. (2005); Chopra & White (2004), Floridi & Saunders (2004), Johnson & Powers (2004), King & Rogerson (2004), Matthias (2004); Weckert (2002); Allen et al.
Haviland (2000), Bainbridge (2000), Collste (2000); van den Hoven (1999); Snapper (1998, 1985); Dennett (1997; 1978), Törnström (1997); Johnson & Nissenbaum, eds. (1995); Lucas (1993); Bechtel (1985), and Gewirth (1978). More works were consulted, of course. This list is intended as a brief – and thus necessarily incomplete – glance at the writings most influential to this paper.

1.5 Limitations of the Study

Before proceeding to the next chapter, I want to demonstrate my recognition of a number of inherent limitations of this study. By presenting them now, I hope to counter in advance some potential objections and critiques of the study, and thus save the reader some time and browbeating.

Firstly, a dearth of data on the economic and social impacts of MDSS use should make any researcher wary of making claims as to the technology’s efficiency, success, prevalence, acceptance, or even uptake. I have not been successful in locating data for questions such as, “how many MDSS are in use today?” or “is MDSS usage on the rise or decline? Or “which type of MDSS is most prevalent?” Even where data on MDSS are available, poor data quality makes it nearly impossible to venture strong and steady conclusions. This does not seem to have stopped many from claiming that MDSS use is on the rise, or that it increases overall efficiency, or that it increases patients’ well-being. There is little that I can do to rectify these limitations. Where appropriate, however, I am careful to state where I am making an assumption, versus stating verifiable fact. That such claims may be unverifiable at present (or even in the future) does not mean that they cannot be explored, however. One can rephrase the matter in terms of assumptions and potential, plausible implications and consequences (such as, “If an MDSS features $A$ and $B$, then plausible consequences are $Y$ and $Z$”).

Secondly, in the course of researching this paper, I have discovered inherent barriers to developing a full understanding of advanced or “intelligent” MDSS (and I mean barriers beyond my non-expert, and thus limited, technical understanding of the subject), as well as barriers to identifying the applicable law. Many MDSS are developed by companies for commercial use, as opposed to being developed by an academic institution. Commercial interests might brand an MDSS as having artificial intelligence in its system design, but this in reality might amount to little more than a marketing ploy than a true description of the system. Concerns over secrecy mean that technical specifications of commercially-developed MDSS

---

4 See Garg et al. (2005) for the latest (at time of publication) analysis of MDSS use in English-speaking healthcare environments.
are harder to come by. There are more factors that just these two, but the reader likely gets the point: that fully-developed case studies of current implementations of MDSS are harder to undertake that hypothetical studies of MDSS, unless one has access to a research group that is developing such a system. The legal situation suffers from much the same underdevelopment, so most analyses ultimately constitute guesswork – “informed guessing,” yes, but guesswork just the same. I have been unable to locate details on cases involving MDSS in either the state or federal courts, although there are cases in which judges have grappled with analogous issues and have applied potentially transferable principles (a few of which are reported in Miller 1989). Until a judge rules on an issue involving the legal responsibility of MDSS, or until a legislature enacts a law governing it, the legal situation will retain this underdeveloped, uncertain quality. Then again, matters of law are often matters of interpretation.

A third limitation is that this paper relies heavily on secondary sources. Lack of time and resources have prevented me from gathering my own data on MDSS’ impact on the healthcare environment and the players in that environment. That said, a study that analyses previous studies can nevertheless provide a fresh perspective, which, if done well, has value and merit.

Fourthly, I am not a lawyer, so some of the legal analyses and opinions I advance here are (unless otherwise noted) those of an amateur. One does not, however, have to possess a law degree or other certification in order to conduct a thought-provoking analysis of the law’s potential application in a certain area. The view that understanding of one area can be enriched and heightened by insights gained in a different field, can be found in the heart of all interdisciplinary research.

Now that I have laid out the study’s limitations, it is time to move on to the more philosophically interesting bits.

---

5 I have studied law (in assessed university courses and through assessed independent research) over the course of two to three years. Among other things, I focus on IT and tech law in the US and EU, as well as US constitutional and other civil law.
Chapter 2

Ethical & Legal Perspectives on Agency & Responsibility

This chapter presents the relevant ethical and legal perspectives on agency and responsibility. The ethical and legal theories presented here are, necessarily, general, “traditional,” and descriptive, in the sense that they present the basic and mainstream accounts of responsibility and agency, even though they are often applied here to questions of communications and information technology. Ch. 3, on the other hand, discusses how the introduction of artificial intelligence (AI) into medical decision-support systems (MDSS) may change the mainstream or so-called “traditional” accounts of agency and responsibility presented here.

Differences aside, all legal and ethical accounts seem to share one basic assumption: that agency comes before responsibility. Or, to put it another way: agency entails responsibility, provided that certain other conditions are met. For this reason I find it appropriate to discuss moral and legal agency first, before moving on to moral and legal responsibility, which I consider to be natural outgrowths of agency.

2.1 Agency

Sect. 2 provides an overview of common ethical and legal answers to the following analytical questions:

- What is agency?
- What constitutes an agent, from both a legal and a moral perspective?
- What internal (psychological or mental) or external (behavioral) characteristics must something show to qualify as an agent?
- What are the conditions for and implications of “to act”?
- Can philosophy and/or the law attribute agency to non-human actors?

As the reader is surely aware, philosophical minds from Aristotle onwards have wrestled with these questions. While fringe theories still exist, today there appears to be a certain amount of agreement among philosophers, ethicists, and legal practitioners (arguably the three

---

6 Unless otherwise noted, the legal perspectives described are from the U.S. and UK legal systems, not continental (or other) systems, and should not be assumed to apply universally. The scope of this paper precludes a thorough treatment of other legal systems, where the underlying legal philosophy could (and very well may) be different.
groups most immediately concerned with such questions), at least when it comes to practical answers to these questions. Nevertheless, the field is ripe with opportunities for debate and analysis, particularly when computers and other technology are involved.

### 2.1.1 Ethical Perspectives on Agency

#### Common Definitions of an Agent

In order to keep all definitions as theory-independent as possible, I have trimmed them down to their most minimum, basic states. I then show how the definitions function within other theories, and how those theories in turn color, or influence, the definition.

*The Oxford Companion to Philosophy* (1995) does not define “agency” *per se* but rather “agent,” stating:

“A person (or other being) who is the subject when there is *action*. A long history attaches to thinking of the property of being an agent as (i) possessing a capacity to choose between options and (ii) being able to do what one chooses. Agency is then treated as a causal power. Some such treatment is assumed when ‘agent-causation’ is given a prominent role to play in the elucidation of action” (Horn in Honderich, ed., 1995: 18).

There are a number of ways to interpret criteria (i) and (ii) above. “Possessing a capacity to choose between options” implies a number of possible interpretations, including that an agent possess a minimum intelligence to make a decision, or that an agent possess a sort of rationality to do so, or that an agent possess either a minimum intelligence or a type of rationality that allows him/her to make a decision when faced with a choice. “Being able to do what one chooses” implies the commonly-termed “free will” argument, which, as Lucas (1993) has noted, is itself immensely complex. “Free will” can mean that one is free “from compulsion and restraint” (Horn in Honderich 1995: 14), or that one controls one’s actions, or that one has the ability to carry out, or in some way effect, one’s choice.

The very concepts *compulsion* and *constraint* can be interpreted in wide or narrow senses, which in turn will affect understanding of what it means to be free (Ibid 15). A wide interpretation of compulsion and constraint might result in the conclusion that all actors are compelled or constrained in some way (by the “natural law” of society, or by their basic human needs, or even by their basic genetic makeup), and thus that true “free will” is impossible. Run-of-the-mill determinist accounts of agency subscribe to a wide interpretation, which has two major consequences: it radically limits the ascription to an actor of responsibility à la “free
will,” and it implies that all actions are essentially predictable, but come from outside of, and precede, the control of the actor. In contrast, a narrow interpretation might result in the conclusion that actors are only temporarily and limitedly compelled or constrained (by a temporary economic or social situation, for example) and thus have the capacity for an unconstrained, untainted – or “free” – will. A narrow interpretation then allows for ascribing responsibility to an actor, when and where the actor was able to act freely.

**Theory-Neutral Definition of Agent**

In a comparative study of the theories of Gewirth and Hare on moral agency, Törnström (1997) offers a relatively theory-neutral, multi-leveled definition of moral agent as one who possesses the following traits:

“(1) Motivation to act according to moral principles and rules (these principles and rules may or may not be formulated by the agent himself […]

“(2) Capacity to act according to moral principles and rules; […] the agent must have:

(a) Autonomy, which includes:

(I) No coercion or duress.

(II) Full information as regards factual circumstances.

(III) Awareness of other entities that may be affected.

(IV) Knowledge of the likely result from alternative courses of action.

(b) Ability to control his own behaviour.

“(3) Aware of himself as a moral agent, and thereby able to reflect upon his own moral judgements […]

(4) Capacity for rational moral reasoning, which in turn requires […]

(a) Capability to judge different situation [sic] according to the relevant moral rules.

(b) Capability to generalise over different but similar situations.

(c) Capability to pass similar judgements on similar situations.

Most likely, there are some further necessary conditions for these capabilities that must be satisfied at yet one more level:

(I) Command of logic.

(II) Access to a full-fledged language.

(III) Ability to take a disinterested view, in order not to prejudice (unreflected belief) and self-interest distort the judgement.

“(5) Morally appraisable; […] If the actions of an agent with regard to entities that are moral patients or morally considerable entities cannot be morally appraised, then the agent is not a moral agent. […]”

[Törnström 1997: 3-4]

Törnström notes that not all of the above conditions are necessary, in totem, for moral agency; rather, some are sufficient on their own, such as (3). But (3) appears to be a rather circular argument (that a moral agent is moral if he is aware that he is a moral agent and can reflect the morality of his judgements), which Törnström acknowledges (Ibid 3). In Ch. 3 I will
reflect back upon this pared-down framework of criteria to see how aspects of it might be applied to non-human agents.

**Bodily Action, Non-Bodily Action, and Intent**

Earlier in the *Companion*, “action” is generally defined as “…someone’s doing something intentionally” (Horn in Honderich, ed., 1995: 4), although I would draw the reader’s attention to the adverb “generally,” as well as to the semantic difference between “action” and “event.” The presence or absence of intentionality is itself not sufficient to determine the purely physical manifestation of an event, of which intentional actions are a subtype. To illustrate with an example: I bump into a fellow student in the hallway. The physical event (bumping into the student) is still an event, and even an “action,” if we employ the everyday, common sense, colloquial meaning of the term (that some part of my physical body came into contact with some part of the student’s body). But, as Horn also points out, “[a] person may be said to have done something when she keeps perfectly still—when, apparently, no event occurs. In such cases, it seems intuitively right that to say there is an instance of action only if the person intentionally kept still. […] it has to be conceded that there is not always an event when there is an instance of action, and that no fully general link can be made between action and bodily movement” (Ibid). Note that Horn’s explanation relies explicitly on intuition (“it seems intuitively right…”). Questions of intent aside, my example of bumping into a fellow student concerns the realm of bodily action, which Goldman (1970) would argue, ignores the possibility of actions not associated with bodily movements (Horn in Honderich, ed., 1995: 5).

Gewirth (1978) uses the dialectical method to formulate a theory of reason and morality. This method involves making assumptions about the claims and judgments an agent would make, given certain features of an action. Thus Gewirth begins his influential theories on reason and morality by establishing the necessary criteria for action. An action must be voluntary in two senses: (1) it is in control of the agent, who is not constrained or forced (Gewirth 1978: 32) and (2) it is with a goal or intent in mind (Ibid 37). Throughout this part of Gewirth’s argument, there appears to be the following equation: well-being entails having freedom and intent; and yet voluntariness, or unenforced choice, also entails having a purpose or goal in performing the action.

In his 1978 writings, Gewirth explicitly requires that action have both a bodily or physical aspect and a mental aspect (Ibid 42). Presumably the mental aspect leads to the physical or bodily manifestation. Törnström (1997) interprets this as implying “that having a body is a necessary condition for being an agent” (Törnström 1997: 12). Combining this with
the bodily requirement, she then interprets Gewirth’s requirement of non-forced or non-constrained action as required that the agent be alive (Ibid). Törnström then summarizes the remainder of Gewirth’s conditions for action, stating,

“All this leads to a conclusion, that is also a necessary condition for being an agent: only human beings can be moral agents, since (as far as we know) only human beings can fulfil all the conditions. Quite surprisingly, Gewirth at one occasion states this in a formulation that equals a sufficient condition of agency [...]”  

(Ibid 13).

So it appears that, at least by Törnström’s interpretation, Gewirth precludes the application of his definition of action to both bodiless and non-human entities.

**Marginal Agents**

As Törnström points out, Hill (1984) employs a term, “marginal agent,” which is intended to include entities that do not possess the full set of conditions (Gewirth’s or others’) for moral agency, such as animals. Implicit in Hill’s theory is a generally-accepted assumption that moral agency must be possessed in full, in absolute, and not in degrees. Gewirth has a curious remark on the possibility of animals as moral agents, stating,

“’animals other than humans lack for the most part the ability to control their behavior by unforced choice, to have knowledge of relevant circumstances beyond what is present to immediate awareness, and especially to reflect rationally on their purposes’” (Gewirth 1978: 120, as quoted in Törnström 1997: 14).

Note Gewirth’s phrase, “for the most part.” It may sound like common sense to assert that animals lack the ability to control their behavior by unforced, unconstrained choice, lack knowledge of more distant circumstances, and lack the ability to reflect rationally on their purposes. But is this provable? Can we know with certainty? Gewirth provides no logical proof of his assertion—in contrast to the wealth of detailed logical proofs he presents as criteria for moral agency. He assumes, but does not prove, that humans have the capacity for rational thought and moral reflection. He then assumes, but does not prove, that only humans have these capacities. That Törnström does not mount a challenge to this assertion implies that she

---

7 Törnström refers to Gewirth’s statement, “they are ‘human rights’ in that they are rights that all humans have as human agents (and all humans are actual, prospective, or potential agents)” (Gewirth 1978:64, quoted in Törnström 1997:13)

8 The idea that moral agency is an undivisible, undilutable absolute is mentioned, and challenged, by Törnström (Ibid 14)
agrees with it. Of course, and as Gewirth himself acknowledges, his use of the dialectically necessary method lets him sidestep the problems of justifying his assertions by empirical testing (Gewirth 1978: 45). But this does not make the method, or its implications, correct. Ch. 3 will return to this argument and examine its implications in greater detail, drawing upon critiques of anthropomorphic and corporeal (here, exclusively human- and body-based) accounts of agency.

**Requirement of Rationality**

How to define rationality? Gewirth’s method (Ibid 44) requires certain inherent characteristics of the agent, including rationality, for predictive and explanatory power. Gewirth’s account of rationality involves the ability “no only to be aware of the generic features of action, but also understand the conceptual analysis and its implications that, according to Gewirth, necessitate not only the agent’s implicit claim of rights for himself, but also his recognition of the rights of other agents” (Törnström 1997: 12). Törnström then goes on to say something which, I argue, is somewhat dangerous, or which affords enough elbow room for contradictory theories to wiggle in. She writes, “In other words, the agent must fulfil a minimum requirement of being able to think logically, and must thus avoid contradicting himself” (Ibid).

This statement deserves a closer look. One possible reading of this appears to uncover the following equation: one interpretation of such logic-based accounts of rationality, which can be expressed as a trio of necessary, and circular, conditions:

1. not contradicting oneself
   which requires
2. logical thought
   which requires
3. rationality

In Ch. 3, I will analyse how much of this equation can be applied computer systems. In Sect. 2.2.1, I discuss the implications of this equation on Gewirth’s conditions for responsibility.

**Agency Has Aspect**

In most accounts of agency, a moral agent can exhibit aspect, much in the same way that languages do; thus, we can speak of first-person aspect and third-person aspect, or first-
and third-person “perspective.” According to the first-person perspective, a moral agent “pursues personal desires and interests based on his or her beliefs about the world, and morality is a constraint on how those interests can be pursued, especially in light of the interests of others” (Johnson & Powers 2004: 423). According to the third-person perspective, a moral agent can sometimes act to further the wishes or beliefs of another party, called either the “client” or the “principal” (Ibid). When acting from - or in the service of – the third-person point of view, the agent is not free of morality but is “still being constrained […] in the guise of such notions as duty, right, and responsibility,” although the exact content of these notions will be determined by the role the agent plays in this agent-client or agent-principal relationship (Ibid). When humans act as moral agents from this third-person perspective, the type of agency in question is often called “human surrogate agency,” to which I return in Ch. 3.

If, after reading the above discussion on mainstream ethical theories of agency, the reader is still undecided as to exactly what constitutes a moral agent, then I have succeeded in one of my ulterior motives in this chapter. That there is not full agreement among philosophers on the conditions for action, agency, and moral agency (with their adherent requirements of rationality, capability, and autonomy in different amounts), serves handily to advance an argument that I present in Ch. 3: that it might be necessary to reject traditional appeals to rationality and “human-only” mental capabilities in an agent, because (among other reasons soon to be presented) one can no more know with certainty what goes on in the mind of a human “agent” than in the organic or synthetic mind of a non-human one. At the very least, disagreement and uncertainty on the issue gives me some wiggle room to introduce alternate approaches. While it might appear commonsensical to insist that some fundamental and exclusively human mental capacities and thought processes serve to establish a critical difference and distance between humans and non-humans, such an assertion is untestable and unprovable. It is merely an assumption, and as such it is open to attack.

2.1.2 Legal Perspectives on Agency

This paper assumes that the ethical perspectives on agency usually inform the legal ones, and also that there exists a sort of feedback mechanism by which the legal perspectives

---

10 Johnson & Powers (2004) discuss the possibility of extending the human surrogate agency model to computers, noting the similarities between the role morality and role responsibility inherent in each. More on that in Ch. 3.
reinforce (and sometimes influence) the original ethical ones. Some might take issue with this assumption, to which I would respond that the question, ‘which comes first: ethics or the law?’ is a bit like the “chicken or the egg” conundrum, and, ultimately, does not serve to advance any meaningful discussion.

Upon further examination, however, the legal perspective on agents and agency has seemingly little to do with ethical perspectives – at least, not in some important aspects. Although it may seem counterintuitive at first glance, the law actually does not function solely upon humans. For example, non-human entities that may enter into contracts and be both subject and object in a liability case include immaterial things (such as organizations and corporations) and, sometimes, material things (such as ships). Although these things, on their own, cannot perform actions (they can only do so through agents), they are commonly said to have legal “personality” or “personhood.” Potentially confusingly, the term is often used interchangeably with legal “agency.” But I would not want to overstate the amount, type, and occurrence of legal agency currently ascribed to such entities. Legal agency derives its nature, in turn, from the precise nature of a relationship, as I explain below. Indeed, humans, or human-made decisions and actions, are not far removed from either corporations or other material things that are given legal agency, or legal personality. For example, humans create and manage a corporation; thus, one could argue that human brains and brawn are largely behind the actions of a corporation.

In law the terms “agents” and “agency” usually arise only in the context of the relationship between an agent and a principle. Simply put, an agent acts on behalf of the principle. Depending on the relationship, agent and principle can be people or entities, in singular or plural form, or a combination of all of these (such as a publicly-owned company’s board [agent or agents] versus the company shareholders [principle or principles]).

Most relevant to this study is the legal status of non-living, non-human entities, such as digital or electronic “artificial agents.” Today, the commonest approach is to treat such electronic agents as “mere tools” or “mere means of communication” (Chopra & White 2004: 636). This approach creates the possibility that the designer, manufacturer, or vendor could be

---

11 Of course, not all laws concern expressly moral issues, nor do they always concern positive or negative rights; some are simply rules about procedure.
12 The legal definitions I employ here are common to both the US and UK legal systems and can be found in nearly any legal dictionary. The terms are essentially the same, whether in area of state or federal law. For a quick reference, I recommend the on-line legal dictionaries at www.law.com and FindLaw.com
13 Here the term “artificial” does not necessarily have any epistemological reference to a synthetic moral or ethical agency; rather, the term is used throughout the computer science and legal disciplines to refer to any non-human entity that can conclude contracts (but not be a party to them). Chopra & White (2004: 635) give the auction website Ebay as a prime example of an electronic artificial agent that facilitates contracts.
held liable for any damage, injury, or mistakes that result from the agent’s actions, whether in the area of tort law or contract law (see Sect. 2.2.2 below for definitions of those laws). This approach also prevents liability from being assigned directly to the artificial agent; as stated, liability rather bounces back to humans, or the corporations for which they work. Other approaches have met with interest by philosophers and computer scientists, but it is too early to assess their impact on judges’ treatment of artificial agents.

Chopra & White (2004) point out, however, that many seemingly commonsensical notions about moral agency fail to find their complement in legal theory. The ethical perspectives on agency and responsibility (described in Sect. 2.1.1 above, and Sect. 2.2.1 below) often originate in a “grocery list” of criteria required for assigning moral agency and moral responsibility. Often on the list are criteria such as autonomy, the mental capacity to make decisions, the ability to perform cognitive tasks, self-awareness, existence in physical bodily form, and so on (this list is not exhaustive). Not surprisingly, many philosophers adopt a similar approach when refuting the ascription of moral agency and liability to non-human agents: comparing such agents to the preceding grocery list of criteria, and marking where artificial agents fall short. Chopra & White raise the point – and does this paper – that debates on the mental capacities and intelligence of any agent – human and non – inevitably reduce to irresolvable debates about the validity of assumptions about mental and physical processes – assumptions, which cannot be proven. Such discussions can also dissolve into semantic debates (such as on the meaning of “cognitive processes,” and whether the algorithmic and electrical functioning of a computer system can be classified as cognition, even though they are synthetic). Instead, Chopra & White advocate a pragmatic approach (which often finds its purest form in legal theory) that adopts Dennett’s (1977) intentional stance (explained in Ch. 3) and focuses on an agent’s behavior, rather than the mental and physical processes that led to the behavior. They write,

“It is not clear, however, whether a legal system would deny an agent [legal] personality on the basis simply of its internal architecture, as opposed to whether it engaged in the right kinds of behaviour, because its behaviour is what will regulate its social interactions. Note, too, that the distinction made above assumes that we know what the actual features of mentality are” (Chopra & White 2004: 638).

By “the right kinds of behaviour” Chopra & White do not appear to mean “right” in any normative sense, as in “good” or “proper”; rather, they mean “right” in the sense of “the types of behavior that are governed by law.”
Here is a good time to dispel quickly with some oft-heard legal “grocery list” arguments against awarding legal personality to artificial agents:

1. Some argue that **artificial agents are not their own “person,” and so cannot enter into contracts.** In Anglo-American legal systems, however, no contract is necessary to establish a relationship between a principal and agent that will have the force of law.

2. Some argue that **artificial agents do not exhibit the required mental capacity of a human adult** (this argument compares the artificial agent’s internal system design, not behavior, to the assumed mental and psychological makeup of a human). But here again, possessing the mental capacity of a human adult is not a requirement for entering into a contract. Many states rely instead on a “sound mind” for agency, which usually “means that the agent must understand the nature of the act being performed” (Chopra & White 2004: 637), although exactly what is involved in understanding the “nature” of an act is a matter for debate.  

3. Some argue that **some degree of consciousness is necessary for legal agency and responsibility.** But “[t]he [Anglo-American] legal system has not seen consciousness as a necessary or sufficient condition of legal personality. Historically, many categories of fully conscious humans – such as married women, slaves and children – have been denied legal personhood. Conversely, persons in comas or asleep […] are not denied legal personality on that basis […]” (Ibid 638).

   There does exist an important criterion for assigning agents legal personality: their ability to control money. Obviously, this criterion arose out of tort law’s *raison d’être*: to award financial compensation for injury and damage. It also makes only non-human entities that have money (modern corporations, for example) able to compensate for their actions. In the case of computer systems, this means that compensation currently falls to the humans who designed, built, or sold the system, or even to the user who incorrectly used or applied the system. I discuss legal liability in greater detail in Sect. 2.2.2, on legal perspectives on responsibility.

---

14 Does it mean recognizing the internal logic of an act? Or does it mean anticipating the consequences of an act? Or does it mean judging – ethically or morally – the environment and conditions in which the act is to be performed? All of these possible interpretations seem plausible.

15 Although, as will come up again, the corporation itself cannot handle or dispense money; agents must act on its behalf.
Legal Definitions of Agency

The law distinguishes between more types of agency than just the “artificial” versus the “actual” type. **Express agency** exists when the principal, either in writing or by speech, authorizes the agent to act on the principle’s behalf (note that the principle may state the exact nature of the agent’s authority, but does not have to). **Implied agency** exists when the principle, by nature of her actions (incl. the non-verbal and non-written), reasonably signals her intent to forge an agency relationship with the agent. **Special agency** specifies the exact acts that the agent may carry out – such as when, where, and how. **General agency** describes an agency relationship in which the agent carries out actions on behalf of the principle “in all matters in furtherance of a particular business of the principal.”

I discuss various types of liability in Sect. 2.2.2 below.

Legal Definition of Action

The law is generally silent when it comes to defining “action,” as in, “what does it mean for an agent ‘to act’?” To view it another way, the law adopts an open definition, which ensures the law’s applicability to all potential current and future situations, where the definition of “action” might not include any physical or bodily manifestation or produce any physical effect. This way the law can be stretched to cover digital and electronic actions.

Many artificial agents are able to act autonomously and remain within legal boundaries. “Autonomy” is allowed for in many agency relationships, such as general agency (described above). In most legal contexts, “[a] utonomous action takes place without the knowledge (at least contemporaneously and of the specific transaction) of the principal” (Chopra & White 2004: 636). Autonomous action does not necessarily equal unpredictable, erratic action, however. Indeed, unpredictable or unforeseeable behavior could be the limit up to which the autonomy standard holds: if behavior is unpredictable, then the standard could fail. But this is not necessarily the case.

2.2 Responsibility

HLA Hart wrote extensively on responsibility, and his work appears to form the backbone of current Anglo legal philosophy. It can be difficult to divorce the terms “agency”

---

17 Here I refer to both state and federal law; local vs. national law in the UK
18 See, for example, HLA Hart (1994 edition) The Concept of Law. For general background reading on Hart and Dworkin, I have referred to J. Brewer (1998) Kapplingen mellan lag och moral
and “responsibility,” particularly when speaking about legal perspectives, because the law often views agency and responsibility as two sides of the same coin. When a human reaches “legal age,” he or she is usually deemed to possess the conditions for agency. Much of the same argument can be made about moral agency – that it is required for responsibility.

2.2.1 Ethical Perspectives on Responsibility

Ethical perspectives on responsibility largely refer back to the previous section about ethical perspectives on agency, as moral agency usually entails moral responsibility (or, moral responsibility might not be possible without moral agency). I would argue that many ethical accounts of agency are actually sufficiently broad in scope to be accounts of responsibility, insofar as an agent who meets the criteria of agent fulfils many of the requirements for responsibility in the process. Perhaps the term “overlap” is a truer description of the link between agency and responsibility, as one can imagine instances where an agent is causally responsible for an event, but not morally responsible, or morally “blameworthy.” Indeed, whether or not we would withhold or assign blame for an event seems to be the same as whether or not we would assign moral “responsibility.” Moral responsibility entails blame; legal responsibility does not necessarily do so. In some instances, as I described below, responsibility also carries with it a moral requirement of penitence or compensation, which may be separate from any legal requirements of penitence or compensation.

To return to Gewirth’s conditions for action and agency are that the actor be free, exhibit well-being, initiate the “transaction,” control his or her participation in the transaction, and at least be aware of the possible outcomes of the transaction (Gewirth 1978: 129-33). A “transaction” in this sense is any action “that affect[s] persons other than their agents” (Ibid 129). Already couched within these conditions for agency are the beginnings of criteria for assigning moral responsibility, including control, well-being, free will, and awareness of outcomes. Gewirth’s concept of “well-being,” however, does not refer to health, physical or mental – at least, not explicitly. Well-being and freedom form Gewirth’s “generic rules” for responsible action (Ibid 135).

At this point it is necessary to mention an important aspect of Gewirth’s theory of responsibility: the Principle of Generic Consistency (hereinforth PGC). The PGC must be applied to the moral evaluation of any transaction, if one wants connect the agent to responsibility. The PGC states, “act in accordance with the generic rights of your recipients as well as of yourself.” (Ibid). The PGC is “logically necessary,” because “for any agent to deny or violate it is to contradict himself, since he would then be in the position of holding that rights
he claims for himself by virtue of having certain qualities are not possessed by other persons who have those qualities [too]” (Ibid; my own emphasis), which is internally inconsistent and highly “irrational,” according to any account of rationality that is based on logical reasoning. The PGC entails that the recipient of the transaction or action also consent to the action. Consent is synonymous with “unenforced choice” (Ibid 134). To force or coerce a recipient to do something is to do harm or damage, and must be avoided, at the risk of contradicting oneself (Ibid 134-5). In Ch. 3 I explore the idea of assigning moral responsibility to a computer system via the logic-dependent PGC.

2.2.2 Legal Perspectives on Responsibility

HLA Hart categorizes the philosophical underpinnings of legal responsibility into different types, four of which are relevant to this study: role responsibility, causal responsibility, legal liability-responsibility, and capacity responsibility (Hart in Johnson & Nissenbaum, eds. 1995: 515). These types of responsibility are not exclusive but rather can be found in combination and in degrees, whenever one speaks of someone or something “being responsible for” an outcome.

Hart describes role responsibility thus: “[…] whenever a person occupies a distinctive place or office in a social organization, to which specific duties are attached to provide for the welfare of others or to advance in some specific way the aims or purposes of the organization, he is properly said to be responsible for the performance of these duties, or for doing what is necessary to fulfil them” (Ibid). Duties are not the same thing as responsibilities, according to Hart, who refers to the “sphere of responsibility” as constituting both long- and short-term responsibilities. The distinction between duties and responsibilities, for Hart, is a temporal one: duties belong to the short-term. (Ibid 515-16). Role responsibility has, as demonstrated in Sect. 2.2.1, both a moral and a legal dimension.

Causal responsibility is a strictly neutral concept, insofar as one can speak of a cause or product of some action. Non-persons can have causal responsibility for an event, such as when an extreme weather event “causes” the devastation of homes (Ibid). Causal responsibility, in and of itself, does not necessarily imply moral blameworthiness, as moral blameworthiness relies upon a well-defined and strong causal connection between actor and event. Strict liability (described later), is another type of legally-recognized liability that exists regardless of moral blameworthiness, praise, or intent.

Capacity responsibility can arise in morally-neutral situations (where it describes a person’s psychological state) and in morally-charged ones (where it denotes sufficient mental
capacities of “understanding, reasoning, and control of conduct”) (Ibid 524). Although capacity responsibility is often considered when assigning moral responsibility, Hart argues that Anglo law has been too slow to consider it when assigning legal responsibility (Ibid). In order to be capacity-responsible, an actor must demonstrate restraint, which requires that the actor understand the consequences of action and be able to control his or her conduct.

In order to ascribe legal liability-responsibility, the law usually—but not always—considers three general conditions, according to Hart (Ibid 518-520): (1) mental or psychological (capacity to understand one’s legal rights and obligations, capacity to control actions) and mens rea\(^{19}\) (2) causal connection to harm or damage (a measure of the proximity of the action to the harmful outcome), and (3) relationship with the agent who caused the harm (if an agent-principal relationship involved). Legal liability-responsibility, therefore, encapsulates aspects of other types of responsibility, such as causal and capacity responsibilities. Hart says that the same three conditions can be used when determining moral liability-responsibility (Ibid 522).

Hart’s categorization of the types and elements of responsibility serves to clarify an important point: that responsibility does not necessarily entail liability. Often, legal liability is created in situations where there exists some sort of moral blameworthiness. But this is not always the case. The Anglo doctrine of respondeat superior (lit. “may the superior give answer”) arises only in tort law (see Sect. 2.2.2 below for more) and, even then, usually only in employment contexts, where the employee is the agent, and the employer, the principal. In a tort law situation, where the agent’s actions have caused injury or damage, respondeat superior holds that the principal can be legally liable for those actions, regardless of whether they stemmed from the principle’s direct orders (such as in special agency relationships), or whether they are reasonably deemed merely similar to the actions the principle herself would have made (such as in general agency relationships). The concept of respondeat superior falls under the more general term vicarious liability.\(^{20}\)

By the same token, an actor can be causally responsible for an outcome of his actions (such as setting in motion a series of events), without being legally liable, because many doctrines require that the actor fulfil a series of psychological criteria, such as awareness of the law, awareness of consequences of action, and ability to control action. But there are

---

\(^{19}\) Mens rea means lit. “the guilty mind” and refers to the mental and psychological state of the actor when he or she acted. It finds its most common manifestation in the concepts of intent and rational thought. Mens rea is used as a condition in, for example, criminal law, when distinguishing between first- and second-degree murder.

\(^{20}\) I am not aware of any continental legal systems that recognize vicarious liability of this sort, or vicarious liability of any sort.
exceptions to this, too: if the standard applied is that of strict liability, for example, the actor’s psychological state is immaterial to establishing liability.

**Courts’ Treatment of Liability**

When speaking of legal responsibility, the courts usually refer to legal liability. Liability arises in instances of defect, damage, or harm, and involves a duty to remedy or compensate said defect, damage or harm. In this regard there is an explicit connection between the law and morality – the injurer, who by virtue of his obligation of care to the injured, also has a duty to remedy injury that is the direct and foreseeable consequence of the injurer’s actions. This duty to remedy harm usually manifests itself in economic terms, whereby the liable party compensates the injured party with money or other financial instrument. It can also manifest itself in punitive damages, whereby the liable party is punished (by paying fees, by performing community service, by losing a professional accreditation, etc.).

Liability is also not an absolute quality, insofar as it can be held to various degrees, where some types of law are concerned. The idea that liability can be shared and quantified (i.e., a dollar amount can be assigned to it, which may increase or decrease with respect to the number of parties to the suit, or the degree to which each party is deemed liable) is the primary driver behind the monetary awards given in a liability suit. In some liability doctrines, however, the injured party’s contribution to her own injury is a bar to compensation.

In clinical medicine, liability arises usually in cases where a healthcare professional’s actions have caused injury or harm to a patient. Such liability suits are usually referred to as medical malpractice cases, where “malpractice” implies that there is some generally agreed-upon standard of care in the profession, which has not been upheld. Where this study is concerned, liability suits can arise when the healthcare professional uses (or fails to use) any variety of medical tools, potentially including intelligent computer systems such as MDSS, if the use of MDSS is deemed the standard of care (see Chs. 3 and 4 for more). Many MDSS –

---

21 Hart writes, “[…] though in certain general contexts legal responsibility and legal liability have the same meaning, the say that a man is legally responsible for some act or harm is to state that his connexion with the act or harm is sufficient according to law for liability. Because responsibility and liability are distinguishable in this way, it will make sense to say that because a person is legally responsible for some action he is liable to be punished for it.” (Ibid 521).

22 Harm can be physical or mental, tangible or intangible (such as economic harm), and can affect humans, non-humans, property, and intangibles (such as reputation, or brand).

23 In almost every case, there is a “reasonable standard” applied to the liable party’s ability to anticipate, foresee and prevent injury, as well as to contemplate – reasonably – the groups of people or things to which they might be liable. In most types of negligence, such as UK’s negligent misstatement law, there must be a degree of proximity in the relationship between the injurer and the injured.

24 “Actions” can also include omissions; the failure to act when acting was otherwise reasonable (usually called negligence); as well as intentionally harmful actions.
particularly those used for diagnosis and treatment – are “safety-critical” because they regulate or affect life-critical processes. As will be discussed more in Ch. 3, in the U.S. and U.K., purchasers and users of defective or damaging hardware or software can find remedy through both the civil law (contract and tort law), and, in some cases, the criminal law. As I have already hinted, when it comes to tort law, it is also relevant whether the computer system (here, hardware + software) is a product or a service. In most cases involving computer systems, however, the lines are blurred; most computer systems, particularly those used in clinical medicine, exhibit properties of both products and services. Hardware on its own is a product, whereas software is often classified as a service. Which doctrines of tort law that can apply will depend on which aspect of the system is claimed to be defective.

In order to fall under the protection and control of contract law, the purchaser of the software must also be the aggrieved or harmed party. On the other hand, invoking tort law, which concerns civil wrongs, does not depend on the existence of a contract between parties, and so is more likely to be the appropriate type of law in most cases involving clinical medicine, doctors, and their patients. US and UK tort law includes the law of negligence, negligent misstatement and products liability.

While products and services – both of which are non-human, one of which is intangible – may be deemed faulty and defective, in both everyday speech and legal proceedings, they in and of themselves cannot provide remedy to the injured party. The express purposes of tort law, for example, are to provide remedy to the injured party (usually by offering economic compensation), to ensure the proper application of products and services (by punishing people who incorrectly use a product and thus cause a harm), or to deter the appearance of faulty products and services on the market (by punishing careless or negligent designers, manufacturers or vendors). In most cases, a product, on its own, cannot provide compensation (it has no money, nor can it be punished), but its manufacturer or developer can – and the “manufacturer” may be in the legally-recognized form of a corporation, or it may be a person(s). So, while the courts might in theory locate the direct and proximate cause of an injury in a defective product or service, in practice they will look to human, or to a money-possessing entity, to provide compensation.

A number of legal principles (pragmatically termed “laws”) govern contract and tort law cases involving safety-critical computer systems:

25 Although, if the injured party seeks compensation in the form of ensuring that the product is removed from the market and no longer poses a threat, then the product’s removal could be a form of compensation – albeit intangible, mental compensation. But there again, it is a person or persons who actually remove the product from the market. The product cannot do so itself.
Negligence forms the basis of medical malpractice suits (Miller 1989: 76). It can be invoked only if three conditions are met: (1) that the relationship between the injured party and the defendant involved a duty of care; (2) that the duty of care was breached; (3) and that damage (to persons or property), injury (to persons or property), or other loss (mental, physical, or economic) occurred as the direct and reasonable result of that breach (Bainbridge 2000:184).

In the US, a case of negligence per se (also called negligence res ipsa loquitor, lit. “the thing speaks for itself”) does not require a trier of fact to prove the exact acts of negligence that led to injury, because the circumstances and degree of injury are such that it is obvious that there was negligence (for example, a doctor leaves a medical instrument inside a patient after surgery). Gross negligence occurs when the defendant did not exercise any regard whatsoever for the safety of others; gross negligence can also bring criminal charges. In some instances, contributory negligence (whereby the injured party has to some degree contributed to his or her injury, possibly including implied risk) is a bar to compensation.

The law of negligent misstatement (in UK also called tortuous liability for negligent advice) is of particular relevance to intelligent expert systems, such as MDSS. No contractual relationship is necessary in order to this doctrine to apply; rather, the defendant must simply present himself (and be understood as such) as an expert and give “advice which is intended to be taken seriously and acted upon” (Ibid 188). If the recipient of the expert advice suffers injury, damage or loss as a direct and proximate result of the advice, then he or she may have legal recourse to seek compensation from the designer or manufacturer of the system. At present, UK law has, “[...] the effect of making the persons and organizations responsible for the creation of expert systems liable to the ultimate consumers of the advice generated. The experts who provided the rules and facts used by the system, the knowledge engineers who formalized the knowledge, the programmers and analysts responsible for designing the inferencing and interface programs could find themselves liable if the advice generated by use of the system is incorrect” (Ibid 189).

---

26 Bainbridge (2000: 184) says that UK law of negligence has its origins in Donoghue v. Stevenson [1932] AC 562, which is a generally uncontested assumption. I have yet been unable to determine the exact case law origins of US negligence law.

27 UK law of negligent misstatement has its origins in Hedley Byrne & Co Ltd. v. Heller & Partners Ltd [1964] AC 465, acc. to Bainbridge (Ibid 188). Hedley Byrne has since been stretched to cover negligent provision of a service (Ibid 193), which theoretically brings it into the domain of expert systems based on software per se.
There are at least two defenses possible to this in today’s UK law: there was no duty of care (either the damage was not foreseeable, the relationship between system developer and recipient of information was too attenuated, or the information was used for a purpose unknown to or reasonably unintended by the system designer), or a reasonable disclaimer, exemption clause, or limitation clause existed which limited or absolved the creators of the system of liability (Ibid). That said, no disclaimer can absolve of liability in cases of product liability or malpractice that lead to death or personal injury (Ibid 190).

Products liability allows the consumer to sue the designer, manufacturer, or vendor of a defective product without having to prove negligence and without having to prove a contractual relationship. Of course, products liability can refer to negligence law, but does not have to. In the US, products liability often takes the shape of strict liability (described below). It can also involve claims on warranty of fitness. All products liability cases center on a claim or claims of fault or defect in one (or more) of the following areas: design, manufacturing, or marketing. Locating the defect can be incredibly difficult in cases involving hardware and software.

The most important and effective defense to products liability suits in the UK is that of “state of the art,” whereby the defendant need only prove that “the state of scientific and technical knowledge at the relevant time was not such that a producer of products of the same description of the product in question might be expected to have discovered the defect if it had existed in his products while they were under his control” (Ibid 195). Once a product has left the factory floor, however, “control” by the manufacturer is largely no longer possible.

The doctrine of strict liability (which operates primarily on manufacturers and vendors of products) requires that the injured party demonstrate that the faulty or damaging product is the direct and proximate cause of injury. For strict liability to hold, the causal pathway may not be too attenuated between injured party and faulty product, nor may the injured party have in any way and of own volition contributed to the injury. Not important, however, is whether or not the defendant exercised care in preventing the injury; carefulness does not even enter the picture. Strict liability is, thus, a special type of liability that recognizes causal responsibility, regardless of intent. In other words, someone may be held strictly liable for an event or outcome if it can be shown that that person’s actions directly and causally led to the outcome. The initiator of an accident, or of unintended consequences, can come under this doctrine. As

28 In UK, products negligence law stems from the Consumer Protection Act 1987. In the US, products negligence law stems from various state statutes, some of which are based on the Model Uniform Products Liability Act and the Uniform Commercial Code. No federal products liability statutes exist in the US.

29 See Sect. 4(1) of the UK Consumer Protection Act 1987 for full text of the “state of the art” defense.
mentioned above, the establishment of a clear, limited causal pathway between agent and outcome is sufficient to assign strict liability; but establishing causal or strict liability is quite different from establishing moral responsibility, which sets the bar at a higher level. Should the doctrine be applied often to cases involving harm or damage due to MDSS use, it could effectively curtail research and development of future systems, as manufacturers and vendors would be held legally responsible to create systems of an impossible (and undesirable) standard: the error-free system (Miller 1989: 76). In Ch. 3 I discuss further the full range of potential implications of applying products/strict liability to cases involving MDSS.

Whatever the name – legal “agency,” “personhood,” or “personality” – the concept entails rights and obligations, including the obligation of responsible action. The reader has likely already guessed that the law (both legislative- and case law) lags behind the reality of DSS usage today, as well as in the future. Ch. 3 explores the implications of this legal and moral lag in greater detail.
Chapter 3

Advanced DSS in Healthcare: A Shift in the Standard of Care?

“In contexts where the requirements for attributing responsibility are not met, we are mistaken if we assign responsibility to machines. In such cases we need to make it clear to those designing and employing the machines that responsibility rests on them. Since my argument has the conclusion that currently available computer systems do not meet the conditions for being responsible agents (although some will come close to meeting the conditions), it has an implication that for now issues concerning responsibility must be directed at those designing and using the machines.” (Bechtel 1985:297).

Bechtel wrote the above passage in 1985 – twenty years ago. Have things changed? This chapter explores the possibility that the answer that question now could be affirmative:

“Now it can be shown that there is an increasing class of machine actions, where the traditional ways of responsibility ascription are not compatible with our sense of justice and the moral framework of society because nobody has enough control over the machine’s actions to be able to assume the responsibility for them. These cases constitute what we will call the responsibility gap” (Matthias 2004: 177, author’s own emphasis)

3.1 Intro to Intelligent MDSS - types, applications, functions

Modern healthcare professionals sometimes rely on advanced medical DSS (MDSS) in many aspects of their jobs. At times, they use them on a daily basis. Doctors, trainees, nurses, and other healthcare professionals may rely to various degrees on these advanced MDSS for diagnosis, treatment routes, or education, or a combination of these. For example, GIDEON, Help, and Iliad are just a few of the AI systems in routine use in primary care and point of care. MDSS are used often in the diagnosis and treatment of diabetes (Collste 2000: 181-94). Quick Medical Reference (QMR), which itself grew out of the INTERNIST-1 project, is yet another MDSS, used for consulting and diagnosis (Miller & Masarie 1990: 1).

30 These are a few of the systems listed in the Open Clinical directory, made by Enrico Coiera. Available on-line at http://www.openclinical.org/aisinpractice.html).
Miller and Goodman (in Goodman, ed. 1998) point out that MDSS can perform a variety of different roles—anything from reminder systems, to consultation systems, to education systems, to diagnostic and treatment systems (Miller & Goodman in Goodman 1998: 104-9). In a review of clinical trials of MDSS, Garg et al. (2005) suggest slightly different categories: systems for diagnosis, reminder systems for prevention, systems for disease management, and systems for drug dosing and/or drug prescribing (Garg et al. 2005: 1226). I will argue that it is possible—and necessary—to distinguish types and levels of responsibility assigned to MDSS, depending on whether it is used for reminding, consultation, or education, and whether or not the system merely “supports” the user’s autonomous decision-making, or rather autonomously “replaces” some of the user’s decisions, even actions. As I will argue, the distinction between support and replacement is not always a sharp one, and so requires an analysis of the entire environment in which the MDSS is used.

The design and operation of these systems—which is determined by the type of system—poses a number of potentially problematic issues at the level of the user or operator. These issues influence the ways in which the user may operate and navigate the system (for example, is the user interface graphic?), and how she interprets and applies the information or decisions it provides (for ex., is the recommendation in the form of charts and graphs, or natural language terms? Are recommendations given as option and alternatives, or stand-alone recommendations?). Concern over this framing or “mediating” power of MDSS is one driver for this study.

Miller & Goodman point to many possible mishaps in consultation systems, including the level of medical knowledge (incl. how up-to-date or current its knowledge base is) as well as the system’s appropriateness to the individual patient’s situation and needs. They also consider that the user-system interaction (or, the human-computer interaction) can be fraught with difficulties if the user does not understand what he or she is doing.

Although he is specifically discussing the use of knowledge-based MDSS in the diagnosis of diabetes, Collste (2000) outlines four possible consequences of an expert system. He formulates these consequences, which have ethical aspects, in the shape of four questions to ask when designing and using such a system:

“I. What do kind of information do we want from the system?”

---

31 I discuss the Garg et al. study in greater detail later.
32 Related to Moor’s (1995) “invisible complex calculation factor” and “invisible programming values” discussed later in this chapter.
“2. How will the clinical encounter be affected if the expert system is part of ordinary care?”
“3. How will the expert system affect decision making?
“4. How will the expert system affect the possibilities of taking responsibility?” (Collste 2000: 185)

Collste’s questions – while not forming an exclusive list – are sufficiently general that they are appropriate for the system designer to take into account when designing an expert system. I would argue that they are also appropriate for anyone or any group who seek a rough outline of questions to ask when evaluating MDSS use. Obviously, that group includes committees of hospitals in which MDSS are used, the healthcare professionals themselves.

Having both Collste’s questions and Miller & Goodman’s concerns in mind, I find it helpful to classify the consequences of MDSS into three different “dimensions,” each of which has ethical implications. Some consequences that arise with MDSS are in the dimension of health of the patient; other consequences can be assigned to the doctor-patient relationship dimension; while yet others fall into the doctor-only dimension (or the dimension of the medical professional using the MDSS).

The patient’s health dimension concerns primarily questions of patient well-being, which in turn can be interpreted either as receiving the best possible care or as receiving successful, effective treatment, or both. Within this dimension the success of MDSS use might be judged largely in terms of patient outcomes (more on this in Garg et al. later). The doctor-patient relationship dimension primarily concerns questions about the effects of MDSS on that relationship, which for ex. the Hippocratic Oath requires that the doctor uphold. Within this dimension there is concern less that the MDSS is yet another tool in the doctor’s toolbox, but rather that the time the doctor spends operating and consulting the system draws away from the time he or she could have spent speaking to and consulting the patient, who is the ultimate receiver of the MDSS’s operations. The doctor-only dimension touches most fully upon questions of responsibility for knowledge and decision-making. Whether one views the MDSS primarily as a medical tool, or as a “partner” in decision-making, in most situations the MDSS stores, filters, or even analyses medical knowledge and, in some cases, patient data, all of which serve to reduce the doctor’s knowledge burden (to know the exact data on the patient); yet simultaneously, it serves to increase the doctor’s knowledge burden to understand the basic functioning and operation of the MDSS. The state of the system’s knowledge base is

---

33 I first heard the terms “tool vs partner” in conversations with Simon Rogerson, May 2005.
of primary importance to the doctor. Because the MDSS is a safety-critical system, the doctor needs also to have faith in its analytical and reasoning “abilities” and accuracy of information. To be able to recognize “faulty reasoning” or incorrect analyses that are not immediately obvious (say, incorrect diagnoses that do not stem from incorrect patient data but rather from incorrect analysis), requires that the doctor understand, to some degree, the way the system works. Not doing so results in irresponsible behavior.

Below is an illustration of these dimensions (see figure).

<table>
<thead>
<tr>
<th>Patient’s Health</th>
<th>Doctor-Patient Relationship</th>
<th>Doctor</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Accuracy of info (in the knowledge base), relevance of diagnosis is important for patient’s health</td>
<td>• MDSS decreases the “human element” of diagnosis and treatment</td>
<td>• MDSS removes some of the knowledge burden on the doctor</td>
</tr>
<tr>
<td>• State of medical knowledge must be up-to-date, or else risk harming patient</td>
<td>• The face-time the doctor spends with the patient may be important for patient’s well-being (decreasing his or her anxiety, for ex.) but is also important for doctor’s ability to make a diagnosis which is informed both by test results and the judgment of the naked eye.</td>
<td>• System keeps doctor up-to-date insofar as that knowledge is programmed into the system first</td>
</tr>
<tr>
<td>• Patient has an interest that the doctor correctly implement the diagnosis or treatment recommendations made by the MDSS</td>
<td>• In some cases, the MDSS may actually ”action” a course of treatment, such as a drug treatment, which further removes the human doctor’s hand in treating the patient</td>
<td>• Doctor/user must be able to recognize “faulty reasoning”</td>
</tr>
<tr>
<td></td>
<td>• Doctor may naturally be wary of using system, thus limiting the system’s uptake and development</td>
<td>• Doctor/user must understand, to some degree, how the system reached its decision (involves previous point)</td>
</tr>
</tbody>
</table>

Figure 1. Dimensions in which MDSS use has ethical consequences

Definition of Intelligent System

That the adjectives “advanced” or “intelligent” precede the term MDSS is no coincidence. This paper focuses primarily on such intelligent systems because of their unique system design, which enables greater capacities and wider applications, which, as I will argue later, may accord them a degree of agency and responsibility for their decisions.

Because there are a number of possible ways to define “intelligent” system, it is important to be clear from the outset which definition I employ. Here, “intelligent computer system” does not mean intelligent in the Turing test sense, as it is not really based on overt

---

34 The behavior-driven Turing test measures the outward behavior of a computer program by simply testing the ability or inability of humans to distinguish between human-produced, written output, and computer-produced
behavior; rather, “intelligent” here means that the system features AI in its design. In simplistic terms, AI allows it to assess its environment, to react to it, and to learn from this interaction. AI can itself be divided into categories and degrees of complexity (almost like “degrees of intelligence”). Indeed, the term may sometimes be applied to an MDSS more as a selling point or marketing ploy, than an actual and accurate description of the MDSS. If unsure, one should examine the technical specifications of an MDSS before deciding conclusively whether or not it incorporates AI in its system design.

**Types of AI Systems**

Matthias (2004) categorizes “learning automata” (AI systems) into four types: symbolic systems, connectionism and neural nets, genetic algorithms, and autonomous agents (Ibid 178-181). (I rely on Matthias’s categorization throughout this paper.) Most MDSS are symbolic systems. “A system built along these lines contains long sequences of axioms and derivation rules that are usually expressed in some kind of predicate calculus. The system is either able to derive conclusions from the facts stored in its database [what I call its “knowledge base”], or else to extend this database by adding new rules and facts to it (which is what constitutes learning)” (Ibid 178). This learning process usually occurs when the system interacts with its user/operator or creator, who can usually correct “faulty” or incomplete knowledge, or incorrect “reasoning,” or inappropriate “behavior.”

**AI and DSS: “Weak” vs. “Strong”**

Here we are primarily (but not exclusively) looking at ”weak” AI – that is, AI that helps humans better perform cognitive tasks. The other type of AI – “strong” AI – concerns itself with replicating human behavior as closely as possible, either through replicating human cognition or replicating a similar thought-action pathway. Although many philosophers appear to be drawn more to studying the intersection of “strong” AI and ethics, in this paper I intend to make a case for studying the intersection of “weak” AI and ethics, which has yet to receive the attention it deserves.

The majority of present-day intelligent MDSS fall into the “weak AI” category. Such applications can take many forms, from programs that manage a large body of knowledge and output. Turing devised an experiment in which a person would use a keyboard and computer screen to converse separately with both a computer program and another human. After each iteration, the person was asked whether he was conversing with a computer, or with another person. If, in at least 50% of cases, the person could not tell that he was speaking with the computer program, then the computer should be deemed “intelligent.” I am not aware that any computer programs have yet met the Turing standard.

35 “Knowledge base” is a generic, industry-wide term for the information stored in the system and from which it operates, including not just data but also rules that dictate decision paths and behaviour.
mine data, to programs that train and educate healthcare professionals, to programs that analyse
input patient data to form a diagnosis, to programs that respond to input patient data to
administer or forego a course of treatment. This list is not exhaustive, of course. For all their
differences, however, each of these intelligence systems utilizes AI in some way. Those
knowledge-based, expert systems that to various degrees affect the doctor-patient relationship,
or affect the healthcare professional’s decision-making capacity – such as diagnostic systems,
treatment systems, and data systems – receive the most of my attention.

Here, the verb “affect” necessarily has ambivalent connotations, with positive meanings
(i.e., to improve, to enhance, to support), negative meanings (to limit, to hurt), and neutral ones
(to frame, to shape, to mediate, to set the conditions for). In Sects. 3.4-3.5, I argue that the
scope of this “affecting,” with all its connotations (enhancing, limiting, mediating), is itself
determined by the role the MDSS plays and the MDSS’ design. This, as I will show, in certain
cases requires that we assign a degree and type of responsibility to such intelligent computer
systems.

Above I mentioned that, here at least, the “intelligence” of the intelligent system is
determined less by its outward behavior than by its internal system design. When it comes to
determining agency and responsibility, however, judgment is determined largely by the
system’s outward behavior (and to a lesser extent by its system design, which invariably
determines behavior), because these are inextricably tied to its role in its overall environment.
Indeed, when judging whether or not an agent is morally or legally “responsible,” one is not
always required to assess the agent’s internal (or mental) condition. For example, causal
responsibility, just like strict liability in the law, can be assigned irrespective of mental
condition. A strict interpretation of role responsibility, for example, does not necessarily refer
to the mental or psychological conditions of the agent, either (I return to these points later in
the chapter).

In this regard my analysis is similar to the Turing test, in that it does not consider how
internally complex the computer system is, but rather how it mediates its interaction with its
environment, which here includes humans. Such an analytical approach, like the Turing test, is
in theory able to avoid getting trapped in endless and fruitless debates about an intelligent
system’s ability or inability to replicate human cognition and reasoning, and whether or not
such ability constitutes a capacity for rational thinking and responsible action. Yet, as will later
become apparent, I will try to adopt a “middle road” position, considering both system design
and behavior.
3.2 Ethical Status Quo

As described in Ch. 2, most traditional and contemporary concepts of moral agency and responsibility refer only to human action and human intent. As is discussed later in this chapter, Dennett (1997), Bechtel (1985) and Snapper (1985) consider the possibility of assigning moral agency and responsibility to computer programs; Bechtel concludes, in 1985, that the state of computing is insufficiently advanced to allow us to ascribe agency and responsibility to computer programs. So, too, concludes Dennett, writing in 1997. But Matthias (2004) concludes, almost twenty years after Bechtel’s comments, that the state of intelligent DSS is sufficiently advanced to allow us – indeed, to compel us – to ascribe responsibility to certain types of computer programs. This study builds upon and strengthens the analyses of these four theorists, as well as Dennett (1978) and Johnson (2004).

3.3 Legal Status Quo

As the reader has likely surmised from the discussion in Ch. 2, the law is created by humans and for humans (with little exception). Thus, when discussing MDSS, current law regulates the humans that develop, build, or use the system. While in everyday speech we might refer to an inanimate physical object as an “agent,” in legal terms agency usually applies only to humans. A number of doctrines of legal responsibility are important to this analysis, both in terms of humans and the DSS itself. The legal writings of HLA Hart are arguably the most influential on legal doctrines of responsibility.

Types of Legal Responsibility

As discussed in Ch. 2, MDSS use (or non-use, refusal to use) falls most likely under the various doctrines of tort law, particularly negligence, products liability, and strict liability.

When it comes to considering MDSS as legal agents, with the attendant rights and obligations, the law lags behind. Chopra & White (2004) delineate the conceivable limits of a certain type of agency – artificial agency. Later in this chapter I will make the argument that some types of MDSS may fall within this agency type, thus bringing them a type of legal personhood. If that argument is sound, then nearly all of the legal doctrines on liability discussed in Ch. 2 could be applied to MDSS, their creators, and their operators. Of course, which doctrine to apply depends on the specifics of the case (more on this in Sect. 3.4)

36 “Surrogate agency,” also discussed in this chapter, most often appears in the legal context of living wills, where an ill or otherwise inactive patient might appoint someone to make treatment decisions on his or her behalf, provided the patient has become incapacitated and is unable to make such decisions.
3.4 Agency: Non-Traditional Ethical and Legal Perspectives

How Intelligent Computer Systems May Change Everything

One might argue that nearly all ICT-interested philosophers and practitioners believe that computer technology, no matter how rudimentary or advanced, creates ethical dilemmas, which require ethical reflection. Some of those philosophers would go so far as to say that computer technology often creates unique ethical issues, and that the evolution of computing into cognitive machines is only further enriching, or complicating (depending on one’s view) those issues. Note here that I refer to computer “technology,” which includes much more than simply the hardware and software that constitute a computer; it broadens that definition to include anything from computerized networks to the algorithms that drive software.

It is easy to become bogged down in seemingly endless discussions about the ability of computers to simulate human reasoning and cognition. To venture predictions about the future of AI computing might be enticing, but ultimately not as fruitful as to adopt a behavioural approach, which focuses on the actual role the technology plays, its current functioning and application in today’s healthcare environment. While the behavioural and situational approach also examines the design and operation of these systems, it focuses more on the user: again, the way the user operates and navigates the system, and how she interprets and applies the information or decisions it provides. As will be shown, this approach requires that we abandon, in some cases, traditional ethical and legal doctrines of agency and responsibility.

Recall some of the issues mentioned earlier that might arise when a doctor uses an advanced DSS system. For example, DSS can restrict the doctor’s treatment or diagnostic options to only those options programmed into the system at its conception. In addition, the architecture of the system’s reasoning and logic paths is seldom transparent, which means that the doctor might not be able to understand why the DSS reached its decision. In short, the DSS frames the doctor or healthcare professional’s decision-making process, to various degrees. Does this have any effect on the conditions required for moral agency? Does it affect the criteria for moral or legal responsibility, or both? Non-mainstream ethical and legal perspectives appear to lean to the affirmative.

How to Ascribe Agency to MDSS: Non-Mainstream Ethical Perspectives

Gewirth’s Principle of Generic Consistency (PGC) can be viewed as an account of rationality based on logical consistency. In Ch. 2 I discussed one interpretation of such logic-based accounts of rationality, which can be expressed as a trio of necessary conditions:
(4) not contradicting oneself requires
(5) logical thought, which requires
(6) rationality

Some types of logical thought can be programmed. Can the PGC also be programmed into a computer? Recall from Ch. 2 that the PGC requires “well-being” and “freedom,” whereby the actor must be aware of the rights of other actors. Although Gewirth may have intended it, the PGC does not contain any explicit requirement of empathy (the ability to understand the situation of another). Its requirement of rational thought is a rather narrow one, based on a knowledge of one’s goals and rights, as well as the rights of others. But is it not possible to define such “rights” in terms of rules, which theoretically can be programmed into a computer?

Perhaps the first question to address is this: can a computer system exhibit the generic features of action – freedom and well-being, as Gewirth defines them? Well-being involves the agent trying “to fulfill his purpose and to see to it that, without his consent, other persons do not interfere with his maintaining the conditions that, so far as he is aware, are required for such purpose-fulfillment” (Gewirth 1978: 136).\(^{37}\) One could argue that a system whose freedom and well-being are interfered with is then unable to act according to the generic rights, and the debate becomes irrelevant. The system no longer displays the necessary conditions for agency or action (Ibid). But such an argument rests on the assumption that well-being and freedom are highly definable, rather than amorphous, concepts. If one considers a computer system as an intentional system, (exhibiting desire, beliefs, goals, and the ability to respond to its environment), it may then be possible to argue that is possible to program a “purpose” into it, as well as artificial “moral” codes in the form of rules. A system that requires constant interaction and interface with a human – such as many MDSS in use today – might nonetheless fail to meet the Gewirthian requirements for freedom, as many parts of its functioning are determined or constrained by an external actor. But what about a system that operates largely without the continual input and interaction with humans? While its system design was determined by humans (the computer did not build itself), some aspects of its behavior may – by virtue of its design – be necessarily unpredictable and uncontrollable. Here we may be moving more into the realm of the hypothetical than the currently possible, but the question is nonetheless relevant.

---

\(^{37}\) Maybe HAL-9000 was just defending its Gewirthian “well-being” when exercising self-defense and resisting attempts to be shut down. Its programmed “purpose” was to ensure the completion of the mission to Jupiter, after all: “This mission is too important for me to allow you to jeopardize it,” as he says in Kubrick’s screen adaptation from 1968.
Floridi and Saunders: Artificial Agents

Floridi & Saunders (2004) argue that there is yet another way to locate some computer systems in the realm of morality: through ascribing to them “artificial agency,” as defined in a way different to the artificial legal agency of Chopra & White, described earlier in Sect. 2.1.2. Floridi & Saunder’s type of artificial agency rests upon a distinction between moral agents (“as entities that can perform actions...for good or evil”) and moral patients (“as entities that can be acted upon for good or evil”) (Ibid 349). Essentially the distinction comes down to “sources” (agents) vs. “recipients” (patients) of moral action (Ibid 349-50)

Floridi & Saunders point out that focusing only on the mental and psychological capacities for agency and responsibility – and adopting an anthropomorphic stance which ascribes such capacities solely to humans – is “worryingly dogmatic. Surely more conceptual analysis is needed here: what has happened morally when a child is deemed to enter adulthood, or when an adult is deemed to have lost moral autonomy, or when an animal is deemed to hold it?” (Floridi & Saunders 2004: 374). Underlying these remarks is the concern voiced throughout this paper: that it is likely impossible (not to mention of little practical help) to determine the precise mental and psychological conditions of moral behavior.

Intelligent Computers As “Borrowed Servant Agents” in Healthcare

Snapper (1998; 1985) continues to argue that continued use and future uptake of advanced DSS—indeed, any computer program that saves lives—might be encouraged if some degree of responsibility or accountability were assigned to the intelligent computer system (Chapter 3 in Goodman [ed.] 1998: 45). Although his argument is here referring explicitly to intelligent computer systems, it is not a stretch to conclude that it applies to intelligent MDSS, which themselves are a type of computer system.

Yet, as Sharon Sytsma (1998) points out in a review of Goodman’s 1998 book Ethics, Computing, and Medicine: Informatics and the Transformation of Healthcare, in which Snapper’s article appears, Snapper’s recent argument suffers from at least one logical blunder and one serious omission (Untitled review in Quarterly Review of Biology 73(4): 481). Sytsma argues that something crucial is missing from Snapper’s argument: namely, a thorough analysis of the concept of responsibility. Here at least, Sytsma does not think that Snapper takes up important philosophical questions like, ‘what constitutes a responsible agent?’ or, ‘does a DSS have the capacity for responsible action?’ Sytsma appears to imply that responsibility for actions can only be attributed entities that act with a certain degree of rationality and self-awareness. The logical blunder, according to Sytsma, is that of False Bifurcation, which here
results in assigning responsibility entirely either to the computer program or to the user, but not to both.

Sytsma’s critiques deserve a closer look. In earlier writings (see Snapper 1985), Snapper does not put forth the argument that responsibility should be “bifurcated;” there is no ‘either/or’ distinction between assigning responsibility to either the user or the computer. He does, however, seem to focus more on the area of legal responsibility, as opposed to moral responsibility, which is likely the impetus behind Sytsma’s critique. Indeed, I would argue that Snapper’s argument is weakened by his inability, possibly reluctance, to distinguish between types of responsibility. While he is careful to distinguish between types of intelligent computer systems in healthcare (such as machines used for diagnosis and treatment suggestion versus machines used for dispensing medicine or regulating vital body functions), Snapper forgets, or refuses, to treat legal and moral responsibility as separate entities. He appears, furthermore, to use interchangeably the terms “responsibility” and “accountability” for performance and judgements, such as when he writes, “It is important to limit the doctor’s accountability for machine judgment if we wish to encourage use of intelligent machines” (Ibid 50). It is essential to keep in mind, when reading Snapper, that he is almost always referring to legal responsibility, and rarely moral responsibility.

Nonetheless, one should be careful not to misinterpret his reasoning. Snapper’s 1998 argument implies that the spectrum of legal responsibility should be lengthened to include “intelligent machines,” and then only in some instances and only where appropriate. Snapper writes,

“…the machine itself could be liable for harms, analogously to cases where liability passes from the doctor to other human medical assistants. Although this option is often overlooked because of a popular misconception that legal liability can only attach to humans, there is really nothing conceptually problematic about it.” (Snapper in Goodman [ed.] 1998: 47).

Here Snapper is espousing the view of the “MDSS as partner” in decision-making, where the system is not merely a tool or means of communication. He continues, pointing out that non-human entities such as corporations and, sometimes, material things, can be legally held liable for harms. In the “agent approach,” these intelligent computer systems often act essentially as agents of a “principal” (here, the physician user), and, at least in Anglo legal systems, the doctrine of respondeat superior holds that principals may be held liable for the harms caused by their agents. Snapper does issue the caveat that it is dangerous to equate a machine agent’s actions with a human agent’s actions (Ibid 48), but then points out that, “Much
of the process for assessing [legal] responsibility does not depend on the human features of the agent, only on the nature of judgment” including, presumably, the appropriateness and correctness of that judgment (Ibid).

Perhaps in order to avoid many of the proclaimed dangers of the agent approach, Snapper prefers the “borrowed servant” analogy, in which the computer agent acts as a servant or agent “borrowed” from the hospital, much in the same way as human healthcare support professionals (namely, nurses) act as borrowed servants of the attending doctor only during those periods in which the doctor—and not the hospital—has direct control over the professional (Ibid 49). Legally, direct control over decisions translates to certain types of accountability and responsibility (although, for example, the doctor cannot necessarily be held causally responsible for the borrowed servant’s actions). In select cases, as some duties previously carried out only by humans can now be carried out by computers, a degree accountability for those duties “may be passed from the physician to the machine” (Ibid 51), although the current law does not explicitly reflect such a passage. When a DSS is used by a physician to diagnose a condition and propose treatment suggestions—duties which are normally carried out by human experts—Snapper likens the DSS to a “consulting physician,” who can never possess full responsibility for the attending physician’s judgment and yet may retain a certain amount of accountability for that physician’s judgment, which has been influenced by the expert.

Note that Snapper does not actually argue that the DSS, acting as a sort of diagnosis and treatment consultant, actually assumes accountability for the ultimate action, judgment, or outcome. He argues that the physician ultimately assigns the “level of accountability” to the expert’s advice, based on how much it guides his decision (Ibid 54). This has important implications. If the format, navigation, and functioning of the computer system is such that it appears to provide, for example, a limited range of diagnoses, without qualifying or justifying its judgment, and the physician is (by custom, practice, or for efficiency) required to rely on such a system, then Snapper would argue that some accountability, some duty of diagnosis, has transferred from the physician to the machine.

Snapper’s consequentialist-based arguments for responsibility are primarily for role responsibility, as they are behavior-driven and focus on the necessary role the technology plays in modern healthcare.38 He often speaks in terms of dependence and reliance upon such

38 Snapper writes, “Indeed, I generally dislike utilitarian analyses of accountability and prefer to lay liability at the feet of those who justly deserve it. And yet the utilitarian approach seems reasonable in special contexts.” (Snapper 1998: 55)
systems (see Ibid 55 for example). His reliance on an intuitive but ultimately unverifiable assumption – that, if society does not ascribe some legal responsibility to the computer system, uptake of MDSS will suffer, and welfare will decrease – does not serve to strengthen his argument, however. There are other utilitarianist implications of Snapper’s argument. Assuming that the law could stretch to ascribe some responsibility to an MDSS, Snapper seems to imply that doctors should be prepared to consent to this legal sharing of responsibility, even if they harbour moral objections to doing so. There is likely a stronger moral argument to be made.

Computers As Surrogate Agents

Johnson & Powers (2004) argue that advanced MDSS are not mere “tools” but rather are able to mediate and influence the decisions that operator makes. “To think that only human designers are subject to morality is to fail to recognize that technology and computer systems shape what humans do” (Johnson & Powers 2004: 434). Johnson & Powers argue that this shaping power of computer systems is indicative of their possessing “a kind of intentionality” (a lower-order, not higher-order intentionality) (Ibid). Like Dennett’s intentional system framework (described below), surrogate agency is a good model for explaining and predicting the effects of certain types of computer system’s shaping or mediating power. Not all computer systems, nor all behaviors of those systems, fit the analogy to surrogate agency. “The best computer surrogate agents, then, are likely to be expert systems, or perhaps even ‘artificially’ intelligent computers, that can advise clients or users through a maze of complex rules, laws, and guidelines.” (Ibid 432).

In human surrogate agency, the agent acts as “an agent of a third party – a client. The surrogate agent acts from the point of view that can be characterized as a ‘third-person perspective.’ In acting, the surrogate agent considers not what he or she wants, but what the client wants. […] The morality constraining surrogate agents is role morality, a system of conventions and expectations associated with a role.” (Ibid 423). To act upon the wishes of the client is the exhibit third-person agency (Ibid 427), which at present is the only type of agency a computer system can express. But, “while computer systems do not have first-order desires and interests, they are designed in ways that represent interests, interests about the interests of users” (Ibid). Users make known their interests through their interaction with the system, requesting it to perform certain tasks. In this regard, users have an expectation that the system will accurately and correctly perform the tasks requested of it. The system, assuming it is
designed to fulfil requests (as opposed to ignore them, act duplicitously, or otherwise “misbehave”), then adopts the user’s interests as its own.

**How to Acribe Agency to MDSS: Non-Traditional Legal Perspectives**

At present I am unaware of any cases in the state or federal courts that involve MDSS and questions of agency; this section is necessarily hypothetical. But, adopting the preceding theories – surrogate agency, artificial agency – it would seem possible legally to apply agency to computer systems. In case the application is not clear: the computer system acts as an agent of the principal, who is the user of the system. One might make the argument, in a some healthcare situations, that the principal can also be the patient, if he or she is the ultimate recipient of the MDSS’ action.

**3.5 Responsibility: Non-Traditional Ethical and Legal Perspectives**

**Non-Traditional Ethical Perspectives**

**Dennett and Bechtel – Computers As Intentional Systems**

Dennett (1978), whose argument is extended by Bechtel (1985), has – perhaps unintentionally\(^{39}\) – identified an intriguing possible way to tie computers to accountability and responsibility: via his “intentional system” theory. From the outset it must be noted that Dennett’s account of intention and rationality do not correspond to the everyday, colloquial accounts. As Bechtel also points out, Dennett’s original theory did not explicitly foresee its application to assigning some version of responsibility or accountability to a computer program, as the theory actual refers no more to a computer system than to an alien life form in another galaxy. Nonetheless, Bechtel’s application and extension of the concept to moral responsibility theory has opened new doors, some of which I would like to walk through and examine a bit closer, now that Bechtel’s writings are two decades old.

We can apply the intentional system label under some conditions to something so that we can predict its behavior. According to Dennett, an intentional system exhibits three primary features: (1) it has beliefs, (2) it has desires or goals, and (3) it responds to its environment. Dennett writes, “One predicts behavior in such a case by ascribing to the system the possession of certain information, and supposing it to be directed by certain goals, and then by working

\(^{39}\) Pardon the pun.
out the most reasonable or appropriate action on the basis of these ascriptions and suppositions. It is a small step to calling the information possessed the computer’s beliefs, its goals and subgoals its desires” (Dennett 1978: 6; author’s original emphasis). Here, possessing information or misinformation means not simply storing the info, but also actually using and acting upon that info (Ibid 7). From an epistemological perspective, it is not important to know (or be able to prove) that the computer actually “believes” in the same sense as, say, a human might; rather, the important point is that we, once having adopted the intentional stance, can predict or explain its behavior by ascribing beliefs to it (Ibid). The strength of the theory lies in its post facto explanatory and pre facto predictive powers. The intentional approach, as Dennett himself later admits, does suffer from a mild case of mentalism (Ibid 60), but Dennett has ready a response to critiques of this sort:

“All that has been claimed is that on occasion, a purely physical system can be so complex, and yet so organized, that we find it convenient, explanatory, pragmatically necessary for prediction, to treat it as if it had beliefs and desires and was rational. The chess-playing computer is just that, a machine for playing chess, which no man or animal is; and hence its ‘rationality’ is pinched and artificial” (Ibid 7-8).

To critiques that the intentional stance serves to anthropomorphize its subject, Dennett counters that we only ascribe to the system a limited account of rationality, which here corresponds to its having justified beliefs, or beliefs in logical truths. This account of rationality does not exclusively apply to humans, as other animals may exhibit logical behavior (Ibid 9).

**Bechtel and Moor**

Bechtel (1985) agrees with Snapper that any attribution of responsibility to computer systems must be shared with humans; thus, his argument does not suffer from False Bifurcation. He also suggests that one reason to consider partially attributing responsibility to computers is that humans are unable to bear the responsibility that currently rests on them when they design or use the computers (Bechtel 1985:297).

At this point it is helpful to leave Bechtel for a moment and refer to another philosopher’s writings on the inability of humans to bear full responsibility when they use computers. Moor (in D. Johnson and H. Nissenbaum [eds.] 1995) provides some terms that can be helpful to use when analysing a computer system, its embedded values, and its effects—all of which can be largely invisible to the user and, sometimes, even to the programmer. There exist varieties of
this “invisibility factor” (Ibid 14-15). Most important to this study are two varieties: “invisible programming values” and “invisible complex calculation,” the existence of both of which seems to require that we take away some degree of responsibility from the computer user. Moor identifies invisible programming values as, for example, what the programmer decides to include and to omit from the program’s specifications, which in turn will mediate the entire user experience. Invisible complex calculation occurs when a computer program makes calculations that humans may be unable to comprehend. Moor writes:

“The invisibility factor presents us with a dilemma. We are happy in one sense that the operations of a computer are invisible. We don’t want to inspect every computer calculation. In terms of efficiency the invisibility factor is a blessing. But it is just this invisibility that makes us vulnerable. We are open to invisible abuse or invisible programming of inappropriate values or invisible miscalculation. […] We must decide when to trust computers and when not to trust them. This is another reason why computer ethics is so important” (Ibid 15).

Do not these invisibility factors—which are an unavoidable aspect of computing (i.e., we would severely retard computing capacity if one did away with complex calculation, or required that it be “visible”)—require that the moral responsibility of the user be reduced? M.J. van den Hoven (1999) and Matthias (2004) argue that it does. Van den Hoven argues that users of MDSS are subjected to “epistemic enslavement” because they are not able to control the acquisition of their beliefs (here, the beliefs they form as a result of the MDSS’ recommendations). A plausible counterargument to this view is to assert that the user is always able to choose not to believe the MDSS’ decision. Van den Hoven responds that, once the user has seen the decision, “once the user has given in, she is unable to provide good reasons to opt out [to not believe]…If a user is epistemically enslaved vis-à-vis system S, then non-compliance with the system’s output constitutes a form of moral risk taking [which] the user cannot justify, at the moment of non-compliance” (van den Hoven 1999: 6)

If one removes some degree of responsibility from the system’s creator or operator, to what would one add the now “unassigned” moral responsibility? Bechtel (1985) would argue that, under limited conditions, one may assign it to the computer system. He first establishes his conditions for assigning responsibility, which stem from Aristotle: “that the decision stem from the agent and be within the agent’s control” (Bechtel 1985:298). But to claim that control stems from the ability of the agent to do otherwise, which is the “free will” or “free choice” argument, and “is thought to require
freedom from causal determination,” would imply that one must show that computers also possess a sort of free will or free choice (Ibid). Bechtel critiques the freedom from causal determination argument, saying that probabilistic and random events also appear to be free from causal determination. Instead, moral responsibility rests on whether or not the outcome was caused “in the proper way” (Ibid). Even though computers (usually) perform what they are programmed to perform and thus do not possess a “free will,” the criterion for assigning moral responsibility is not dependent upon a “violation of causal determination of events,” and thus a non-issue (Ibid). Bechtel then looks for a type of causal determination “that is compatible with a decision stemming from an agent and being within the agent’s control,” for which he finds argumentative evidence in Dennett’s idea of intentional systems.

The requirement that “decisions stem from the agent and be under the agent’s control” is met if one, like Dennett, views a computer system from the intentional perspective, in which it possess beliefs, desires, and the ability to respond and adapt to its environment (even, Bechtel says, to have beliefs and desires about its environment [Ibid 304]).

Note that Bechtel’s argument rests, at least in part, on a simplified understanding of the internal system design of the computer system (that it has beliefs, desires, can respond to its environment, etc.). Nonetheless, his argument still rests primarily on an analysis of the external environment. He writes, “an intentional system must both have the right kind of relationship to the environment and have an internal configuration that permits such a relationship” (Ibid 300).

**Non-Mainstream Legal Perspectives**

Chopra & White (2004) are correct to point out that the law does not yet preclude the possibility of applying artificial agency to non-human actors. As explained in Ch. 2, legal agency and legal personhood do not rely necessarily on consciousness, adult mental capacity, or the presence of a contract between agent and principal, all of which are oft-voiced barriers to granting legal personhood.

If one considers a system that reaches decisions or initiates action independently of human intervention, prompting, or other control, one could call such a system “autonomous.” The argument can be made that such a system is functioning as an autonomous agent – albeit an “artificial” (non-human) one – and thus may have legal personhood, with its attendant rights and obligations.
3.6 Summary of Arguments and Counterarguments

This paper has now covered a lot of ground, in terms of laying out the ethical and legal foundations for agency and responsibility, and in terms of presenting arguments in support of assigning agency and responsibility to some MDSS. At this point, I want to summarize the arguments (and potential counterarguments) presented:

Argument for Artificial Moral Agency: (Floridi & Saunders 2004) Ignoring the fact that non-humans (such as corporations but also computers) can be both the sources and receivers of immoral action results in a view of morality which is both narrow and absolute, as opposed to recognizing the often dual or shared nature of morality. An artificial moral agency approach is more appropriate. Anglo law recognizes such artificial legal agency, in certain circumstances, and grants such agents legal rights and obligations.

Argument for Surrogate Agency and Beyond: Advanced MDSS are not mere “tools” but rather are able to mediate and influence the decisions that operator makes, which bestows on them a sort of role morality. Consultative and diagnostic MDSS can assume this role responsibility via an extension of human surrogate agency to the computer system. Johnson & Powers’ application of the surrogate agent model to MDSS, with its third-person interests and desires, is appropriate for some types of MDSS but it is not future-proof, as advances in AI may lead to computer systems with first-order preferences and intentions. Such advances would go beyond true surrogacy (Ibid: 432). Also, the human surrogate agent model, when applied to computers, does not shed much light on the computer’s liability for its actions in carrying out the client’s wishes (Ibid 433)

Counterargument 1: MDSS is simply a tool which supports human decision making and action. In that supporting role, and being a computing technology, the design of the MDSS will necessarily result in some loss of user control, some limiting of user’s scope for decision making. For example, of the treatment options suggested by the MDSS, there will also be options left out. Any tool, however, mediates the user’s experience, so that, in and of itself, simply pushes the matter into the realm of ethical deliberation. It does not, however, necessarily mean that the system should also assume responsibility.

Counterargument 2: Most counterarguments to this focus not on the feasibility of extending the surrogate agent model but rather on any attempt to ascribe agency and
moral responsibility to non-human actors, regardless of the type of agency or responsibility. As Ch. 2 described in detail the current mainstream accounts of action, actors, and agency – particularly their requirements of anthropomorphic and corporeal action, which do not allow room for digital actions. Most accounts also require the requirement of intentionality. The type required for moral deliberation - what Dennett calls “higher-order intentionality” (Dennett 1997:354) - is itself an outgrowth of empathy, perceptual input, and complex interaction with one’s environment. As of 1997, Dennett argued that such higher-order intentionality was not yet present in any AI systems, let alone MDSS (Ibid).

**Argument for (Legal) Liability on Utilitarian-Consequentialist Grounds:** MDSS perform a necessary role both in the immediate doctor-patient interaction and in the overall healthcare environment. Advanced MDSS are able to perform tasks “better” than, or at least faster than, the humans who operate them. In order to provide the best possible care, healthcare professionals must then use them, which constitutes a new standard of care.

**Counterargument 1:** Not all MDSS play such a role. Consultation MDSS are intended (by their creators) to be used for just that – consultation – but nothing more. The attending physician or healthcare professional may consider the recommendations, but nonetheless retains full autonomy (and thus full responsibility) in reaching decisions.

The counterargument above, however, ignores the mediating power of the technology.

**Counterargument 2:** Where is the data? Again, the Garg et al. study indicated that data on practitioner performance only demonstrated overall improved practitioner performance in 64% of the cases studied (Garg et al. 2005:1223). Of the 52 trials which assessed at least one patient outcome, only seven reported an improvement in outcomes (Ibid).

**Argument: Uncertainty in Law Hinders Uptake** The uptake of MDSS will be hindered by uncertainty in the moral and legal arenas. Some doctors will morally object to relinquishing some decision-making control to a computer system. Other doctors will be reluctant to use MDSS because they are unsure of the legal consequences that could come to bear on their use.

---

40 “Overall”: of the 97 studies (out of 100) that assessed practitioner performance, across system types.
Counterargument: This argument is more of a cluster of assumptions than an argument, in the sense that the statements are largely unverifiable (how, say, a social scientists would go about empirically testing that “uncertainty” will negatively impact the future use of MDSS, is far from clear). That said, it is not entirely unthinkable, and appeals to one’s intuition.

As for uncertainty in the legal situation: this uncertainty is a constant feature (some might call it an asset) in any legal system in which doctrine can be set (and broken) by precedent, or case law. While it can be valuable to venture predictions about which legal doctrines would plausibly apply in a particular case, in the end the only definitive answers come from either a legislature’s lawmaking or a judge’s ruling.

Argument for Shared Responsibility Due to Unpredictability: (Matthias 2004; aspects of Bechtel 1985) Advanced MDSS perform not only invisible complex calculation but also unpredictable, uncontrollable complex calculation. It is not morally defensible to hold either the system designer or the operator responsible for behavior they cannot conceivably control (Matthias 2004: 175-183.

Counterargument 1: This depends on the type of AI used by the MDSS. The system operator of a symbolic system-based MDSS can almost always edit the system’s knowledge base to update facts or correct errors, so a great deal of control is possible.

Counterargument 2: Matthias (2004) has not defined “responsible.” The term is very nuanced. Responsibility – particularly causal responsibility – is hardly confined only to the area of blameworthiness. It also involves praise and recognition, as in “the aid agency was responsible for saving the lives of thousands of refugees.” Here it can mean that the agency has a role responsibility, to “see to it” that lives are saved. It can also mean that the agency’s actions (or omissions) are causally responsible for saving lives.

Argument: MDSS Constitutes New Standard of Care: Advanced MDSS often form the de facto standard of care, bestowing a moral responsibility on healthcare professionals to use them. When used properly, expert systems may lead to improved health, increased efficiency, cost savings, and overall increased welfare for both healthcare professionals and patients.

Counterargument 1: Again, where is the data to back up such an assumption of improved patient outcomes and increased efficiency, not to mention the claim to improved “well-being,” however defined? Hard data are hard to come by. Garg et al. (2005) conclude, in their recent review of 100 clinical trials of MDSS, that the majority
MDSS do improve practitioner performance, but that the current lack and low quality of data preclude the conclusion that MDSS use also improves patient outcomes. Success rates varied significantly across system application areas (diagnosis v. drug prescribing, for example).

**Counterargument 2:** MDSS negatively impact the doctor-patient relationship, by decreasing “face time” when the doctor is updating the system, operating it, or considering its decision. This argument is of the more general “technology encroachment” type.

**Counterargument 3:** Even if this assumption could be verified independently, it still smacks of utilitarianism. Just because a technology is used now, does not mean that it *should* be used. One must move from the “is” to the “ought.” To that end, in Ch. 4 I outline guidelines for when MDSS should, and should not, be used.

**Potential Legal Scenarios**

The use of MDSS raises legal questions in a number of areas. While this paper has primarily focused on the possibility of assigning some level of legal responsibility to advanced MDSS, the paper has also noted that current liability and negligence law already acts upon users of MDSS (usually, the medical professional) and objects of MDSS use (the “objects” are usually patients, but can sometimes the medical professionals, too, if they are using the system for increasing their knowledge).

As described in Chs. 2 and 3, the legal issues that arise with the use of MDSS most often involve questions of liability for harm or damage caused. The law of negligence – as it is invoked in medical malpractice suits – turns in part on the standard of care due the patient. Recall Goodman’s quote that opened this chapter: “the future of the health professions is computational.” This view implies that MDSS uptake will increase until it becomes the norm. Goodman writes:

>“Suppose the use of medical computers evolves to the point that half of all practitioners employ them to good effect. Are those who do not use computers then providing substandard care? […] It will not be obvious when we cross such a Rubicon. Some medical liability insurers now offer modest premium discounts for physicians who adopt certain computer systems for clinical records, drug interactions, and patient education (Borzo 1995), and it has been suggested that computer use in medical offices is good for risk management (Bartlett 1994). Other computerized tools are said to reduce unacceptable levels of human errors (Boon
and Kok 1993; Mango 1994). Is all this to be understood as evidence of a shift in the standard of care?” (Goodman 1998: 7)

“Standard of care” is a future-proof legal concept, insofar as it will change in step with developments in technology. The standard is always evaluated with reference to the field in question, and so will differ from field to field, country to country. Assuming we cross that Rubicon, at some point it will no longer be possible for medical professionals – and the hospitals that employ them – to avoid using MDSS; indeed, they will have a duty to use them, both legally (to uphold the standard of care and thus avoid malpractice) and ethically (to uphold the principle of beneficence, to provide the best possible care to patients).\footnote{One could potentially go so far, invoking King and Rogerson’s (2004) study of ethical principles, to argue that the patient’s right to equality (here, equal care and treatment) requires that MDSS be used, once they are believed to provide the best possible care (Ibid 518).} Once this happens, and assuming that intelligent systems continue to develop “intelligent” capabilities, one could even argue that advanced MDSS move from being mere “tools” in decision making, to being “partners” in decision making, perhaps “partners” in healthcare.\footnote{This terminology of “tool” vs “partner” draws upon the writings of King, Garibaldi & Rogerson (2002), “Intelligent Medical Systems: Partner or Tool?” in \textit{Proceedings of ETHICOMP 2002}} From this perspective it would appear imperative that some degree of legal liability be assigned to the intelligent system.

There are additional important, potential – although not necessarily likely – legal scenarios involving MDSS use. I refer the reader back to Ch. 2 for definitions and explanations of the legal terms used below.

\textbf{Strict Liability scenario 1}: Should strict liability be applied often to cases involving harm or damage due to MDSS use, it could effectively curtail research and development of future systems, as manufacturers and vendors would be held legally responsible to create systems of an impossible (and undesirable) standard: the error-free system (Miller 1989: 76). Because of the potentially exorbitant costs to physicians and hospitals to purchase products liability insurance for using MDSS, Miller argues that judges are not likely to apply strict liability except in the clearest and most extreme of cases.

\textbf{Strict Liability scenario 2}: What Miller did not consider in 1989, however, is that today’s systems that feature AI in their system design are unable to function entirely error-free, as committing mistakes is part of the way in which they “learn.” This makes strict liability an even more prohibitive doctrine to apply (as it would stifle development), and more difficult legally to justify (as it relies on a clear connection between human error in design, manufacturing, or marketing, and damage or injury to the defendant).
Products liability scenario 1: the system designer of a controllable system that aids or replaces human decision-making can be held liable (under products liability) for the damage the system creates only if the product has functioned or performed incorrectly due to its design, manufacturing or marketing, all of which are set prior to the product leaving the factory floor. For products liability, the type of MDSS is of little issue, so long as its incorrect or damage-causing behavior can be traced back to human error.

Products liability scenario 2: the system designer of a system that performs outside of human control (outside of human predictability) is likely not liable for the above situation, under products liability, because the incorrect or damage-causing behavior cannot necessarily be traced back to human error. To do so would effectively end the development of all “learning automata,” as Matthias (2004) calls such AI systems.

Negligence scenario 1: If a healthcare professional uses system that functions as aiding, but not replacing, decision-making, the operator retains ultimate control of, and legal liability for, the application of the MDSS’ advice, under negligence law. If a patient is injured as a result of the operator’s decision (even though it was partially informed by the MDSS), and not as well by the patient’s own accord,\footnote{I.e., the patient has not also contributed to his own injury. Contributory negligence is a bar to recovery in some states.} then the patient has right to seek recovery via negligence law.

Negligence scenario 2: (as described earlier) If MDSS use can be proven to form the standard of care,\footnote{Recall that judges will refer to the standards generally agreed upon by like-positioned experts in that area of medicine. Thus, standard of care can differ not only by time but also by area of medicine and geographical location.} then a healthcare professional’s decision not to use it could bring negligence claims. In legal systems that recognize vicarious liability, the employer of the healthcare professional (usually a hospital) is potentially also open to negligence claims.

Negligent misstatement scenario 1: Insofar as an MDSS is considered an “expert system” by users, the person or persons who create, edit, or update the MDSS’s knowledge base are likely to be held responsible for the accuracy of that information, as they have a duty of care to the receiver of such expertise.

Negligent misstatement scenario 2: But where the MDSS “learns” from iterated interaction with the user, and in the process “makes mistakes and adjusts behavior, it is less clear that a person or persons would be held liable for the accuracy of such statements, as it is far less clear who holds the duty of care. Here one runs into an argument much like before: to hold the system designer or operator liable for the to-some-degree-unpredictable statements of
a computer system, would serve to stifle development in the industry. An analogous situation, with roots in the real world, is this: licensed physicians graduate from medical school, where they receive much of their ground education, which is later supplemented by interning at medical facility. If a physician’s actions cause damage or injury to a patient, neither the medical school nor the medical facility where the physician received his or her practical training, is open to liability under negligent misstatement, although it could be argued that the physician bases his or her actions on the “statements” or expertise offered by, and gained at, the medical school and the medical facility. One could thus draw an analogy with the duty-of-care status of the person(s) who supply and edit the MDSS’ knowledge base.

There is one important theoretical reason why negligence (including negligent misstatement) claims might still operate only on humans, as opposed to the system. In theory, where a healthcare professional owns and operates an expert system, he or she could be held partially liable for the misapplication or misinterpretation of the expert system’s advice. This is because “[p]ossible existence of errors in communication may make it difficult to establish the product as the direct cause of the patient’s injury” (Miller 1989: 79). Errors in communication could stem from the user interpreting commands and prompts differently than that system (or, the system designer) intended. Unfortunately, the user would necessarily discover communication errors post facto, after acting upon information provided by the MDSS, and then noticing harm or damage to the patient (because, before that point, the communication cannot be deemed to be incorrect or faulty, but must be considered correct). I discuss the implications of this legal uncertainty in Ch. 4.
Chapter 4

Conclusions and Guidelines for Best Practice in Law and Healthcare

“The future of the health professions is computational”
(Goodman 1998: 1)

MDSS are likely here to stay. This chapter aims to provide general best practices – ethical and legal guidelines – to those people and groups affected by MDSS, from the programmers who design the system, to the hospital board that approves its use, to the physician who uses it, to the patient whose treatment is mediated by it, even in the smallest way. I conclude with a brief consideration of a recurring debate concerning the broader theme of intelligent computing: whether or not it is ethical and responsible to continue research, development, and uptake in the field, given the current ethical and legal situation.

4.1 Conclusions and Implications

This paper has identified essentially two ways of viewing DSS:

1. as a black box\(^{45}\), where “users are unable to monitor what is going on in the system as a matter of fact (inaccessibility) and secondly, even if they could monitor what was going on, they would not be able to keep track of it (intractability)” (van den Hoven 1999: 2)

2. complex internal system

To determine which view is adopt, one can focus on the purpose of the MDSS. Is it used (or to be used) as an aid or a substitute for human decision-making?\(^{46}\) In the last sections of Ch. 3 I summarized the arguments and counterarguments for assigning a least three types of agency to computer systems, all of which are dependent upon the classification of a DSS into an aid/supporting system, or a replacement/substitute for human decision-making. Nearly all such arguments eventually rest upon role responsibility: surrogate agency, borrowed-servant agency, artificial legal agency (a computer system can theoretically fulfill the legal

---

\(^{45}\) I am not aware if the term “black box” is widely-used. The instance of it that I have come across in my research, and which was not cited to an earlier source, is from van den Hoven, 2001. Proceedings of Fifth ETHICOMP Conference. Available only on CD-ROM of conference proceedings. Page 2.

\(^{46}\) These points, in this form, were raised by Simon Rogerson during conversations in May 2005.
conditions for artificial agency à la Chopra & White), and “artificial” moral agency (à la Floridi & Saunders 2004). Where an MDSS actually actions a course of treatment, causal responsibility can be assigned. The causal pathway is less clear for an MDSS that is used in a primarily support position.

A description of the relevant legal doctrines of tort law indicates that the easiest way to assign agency and responsibility is to do so via the law. In this respect, there appears to be a significant divergence between moral and legal philosophy, which usually overlap to a very high degree. If one argues that an MDSS fulfills the conditions for legal personhood via artificial legal agency, this comes with rights and obligations. If an MDSS can be classified as an agent, then it can be held liable under both negligence and strict liability law (although which doctrine(s) would apply will depend entirely on the specifics of the case (what sort of MDSS, put to what uses, how much human-computer interaction).

Throughout this paper I have discussed the implications of ascribing legal responsibility to MDSS. One that I have yet to mention, however, is that, if the MDSS were held responsible in a tort law context, then the injured party would have a difficult time recovering for damage, as someone or something (i.e. a corporation) with pockets must pay. One reply to that, however, is that the law does not technically require that there be a “guilty party” in order for someone to be compensated for damages. Ways that a society provides compensation without pointing a finger a guilty party include through insurance schemes and taxes, with which an aggrieved person can be compensated (Matthias 2004: 176, see his footnote 4). Another reply is to question the absolute, divisible approach to responsibility that such an approach espouses. Moral responsibility, as well some types of legal liability (such as joint and several liability) allows for sharing of responsibility among actors; such is the essence of dual responsibility. The complement is found in Anglo law, as well. In most legal liability cases, liability is quantified and ascribed to parties in percentages, who then “pay out” a monetary amount in compensation that is analogous to their allotted liability percentage.

**What the Available Data Say – and Do Not Say**

Already in Ch. 1 I indicated that a relative dearth of data across a large population and time period, makes it difficult for anyone to verify claims as to the prevalence, use and effectiveness of MDSS. As previously noted, such claims often find their way into utilitarian-based arguments for assigning responsibility to MDSS on both moral and legal grounds, such as the “standard of care” argument. Some data do exist, however.
The Garg et al. study analysed the results of nearly 100 trials that measured “improvements” in either practitioner performance (such as helping a doctor detect high blood pressure) or patient outcomes (such as achieving lower or better-controlled blood pressure) from the use of MDSS in English-speaking healthcare environments.\textsuperscript{47} Trials on the latter were far more likely to be inconclusive, or to report less or no improvement (Ibid 1226-27).

<table>
<thead>
<tr>
<th>Type of MDSS</th>
<th>Total no. of trials analysed</th>
<th>% of trials reporting improved practitioner performance (% of x trials)</th>
<th>% of trials reporting improved patient outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic systems</td>
<td>10</td>
<td>40% of 10 trials</td>
<td>0% of 5 trials</td>
</tr>
<tr>
<td>Reminder systems for prevention</td>
<td>21</td>
<td>76% of 21 trials</td>
<td>inconclusive</td>
</tr>
<tr>
<td>Systems for disease management</td>
<td>40</td>
<td>62% of 37 trials</td>
<td>18% of 27 trials</td>
</tr>
<tr>
<td>Systems for drug dosing and prescribing</td>
<td>29</td>
<td>Single dosing: 62% of 24 trials Multi-drug dosing: 80% trials of 5</td>
<td>Single: 11% of 18 trials Multi: 0% of 4 trials</td>
</tr>
</tbody>
</table>

(Data [reformatted] were obtained from: Garg et al. 2005: 1223-38).

Table 1: Recent Data on Improvements from MDSS Use

Judging from the results above, one should think twice before making statements that MDSS improves patient’s well-being. Such statements can be made of course, but not without further qualification on what the speaker means by “well-being,” and what he or she means by type of MDSS.

Responsibility Differs Between Types of MDSS

Recall the three “pillars” of dimensions in which MDSS use has an ethical impact (Fig. 1) from Ch. 3. When considering whether or not to assign some amount of moral and legal responsibility to the intelligent MDSS, most attention is given to the “doctor-focussed” pillar above, where mostly questions of agency and autonomy appear. Although M&G are correct to identify the different roles that MDSS occupy – or their different purposes in the healthcare industry – these roles are largely at the application level. A system used to diagnose diabetes on the basis of input blood tests, is still a consultation system, insofar as the attending physician

\textsuperscript{47} Thus one may assume that only a subgroup of MDSS in use today were studied, as there undoubtedly are MDSS developed for, and used by, non-English speaking healthcare professionals.
consults the system for a diagnosis. Although the argument can be made that the physician relies on the system for providing him or her with a correct diagnosis, in theory nothing prevents the physician from revisiting the data and making the diagnosis on his or her own.  

The autonomy of the patient is also an important issue, and is often tied to the nature of the doctor-patient relationship. Miller and Goodman raise the possibility of requiring the patient’s consent to be treated by a physician who uses MDSS. The use of MDSS undoubtedly mediates the doctor-patient relationship (in ways described briefly in the middle pillar, above), which may in turn require the patient’s “informed consent.” There is a legal dimension, as well as moral one, to the concept of informed consent. Miller and Goodman write,

“Should patients be told the accuracy rate of decision machines – when they [the patients] never were given comparable data for humans? Would knowledge of such rates maximize the validity of consent, or constitute another befuddling ration that inspires doubt more than it informs rationality?” (Miller & Goodman in Miller, ed. 1998: 112).

Not surprisingly, Miller and Goodman conclude that MDSS are not appropriate in all cases. For example, a doctor would be wasting time and money to use it in a case of a regular headache or sprained ankle. “Not all such uses will be appropriate; one is duty-bound to select the correct tool for a job, and even appropriate uses must be evaluated against a complex web of reasonable expectations, publicly defensible standards, and rigorous evaluation metrics. It is possible to err here, and err badly” (Ibid).

The authors call for standards in educating users (both doctor and patient), in designing MDSS, and in “appropriate tool use” (Ibid 113), so as to combat legal and ethical uncertainty, as well as to combat a degeneration of respect in the doctor-patient relationship.

4.2 Suggested Best Practice for Healthcare Professionals

When to Use MDSS

Beauchamp and Childress formulated in 1979 arguably the best-known code of ethical guidelines for healthcare professionals in their successive editions of Principles of Biomedical Ethics. Their guidelines, which they call “action-guides” often fall somewhere on the spectrum

---

48 Although here I would like to point out that it depends on the task. Some MDSS may crunch large amounts of data and perform very complex calculations to reach an answer. A human might not be able to replicate this ability, or at least not replicate it within the same timeframe.

49 “Informed consent” requires that the patient understands – to some undefined degree – how the intelligent system works, and why it is to be used in their case. Note that the standard of sufficient information for the patient’s “informed” consent is lower than that of the physician; in other words, the patient need not understand the system as well as the physician does.
between the theoretical and the practical, as principles exist somewhere between the level of particular cases and the level of ethical theory (Beauchamp & Childress 1989: 6-7). Many of the action-guides, or principles, are deontological, as they imply a duty or obligation. The principle of respect for autonomy, for example, involves the duty to respect autonomy – the doctor’s and the patient’s. While their mid-level principles are not the only beacons of ethical behaviour in the medical profession, they have proven highly influential to others who have attempted to formulate a set of common guidelines. For example, the International Medical Informatics Association (IMIA) has developed a set of principles for professionals who work with health informatics, which incorporates and expands upon Beauchamp and Childress’ original list, to include the Principles of Beneficence, Non-Malfeasance, Autonomy, Justice, Equality, Impossibility, and Integrity (King and Rogerson 2004: 508-9).

It is one thing to formulate principles; it is another, however, to apply them. Differences appear to arise most often at the level of application. For example, King and Rogerson (2004) have conducted a series of interviews with clinicians at a Viennese research hospital. The researchers’ intent is to evaluate the clinicians’ willingness to use, and actual usage of, expert systems, against the backdrop of the IMIA’s expanded code of ethical principles (listed above). These clinicians not only use different types of expert systems (including MDSS) but also invent and develop such systems to use in their respective specialties. Some of these expert systems make “‘decisions’ about when a patient is ready to be weaned from artificial ventilation, […] [some] examine blood test results in order to detect the presence of Hepatitis, […] [some aid] Primary Care Physicians in making diagnoses of unusual Rheumatological cases, [and some monitored] the spread of infections within that hospital” (Ibid 508). King and Rogerson find that the Viennese clinicians are almost always able to uphold the ethical principles while designing and using MDSS, except when to do so “demanded opposing action. These clashes principally occurred when the clinician’s duty to provide the best care possible for patients was at odds with the clinician’s right to autonomy” (Ibid 518).

There is a well-known caveat to any attempt at principles or codes of conduct (although codes of conduct tend to be more specific than principles, which are general guidelines): that people may end up focusing narrowly on applying the codes only to those areas that the codes specifically cover, at the expense of examining other areas.

---

50 Clinicians’ autonomy here refers to their right to reach their own informed decisions for the good of their patients, as opposed to relinquishing some of this right to the “omniscient” computer program. Patients’ autonomy is a separate ethical issue, and includes patients’ increased awareness and knowledge of medical standards and treatment.
Clearly Formulated and Communicated Guidelines from Hospital

Hospital boards must draft clearly worded policies that guide medical professionals on the use of MDSS, covering when and how to use them. Guidelines are necessary to clarify both the legal and ethical situation – i.e., to help demarcate the lines of liability between the hospital and the professional who uses the MDSS, as well as to encourage physician’s to use MDSS, when and where appropriate. Hospitals and, in turn, medical professionals, differ in the sorts of MDSS they use (and can afford), the types of treatment they provide, and even the legal systems in which they operate and by whose laws they must abide. Given these differences, it is, therefore, not unreasonable to suggest that guidelines be determined on a hospital-by-hospital basis. Below I provide an example of possible guidelines, which could be read as a list of minimum criteria that must be fulfilled:

The use of MDSS is generally\(^51\) recommended:

- when the MDSS helps the medical professional to give better care, whether by increasing her efficiency or knowledge, or by decreasing the intrusion or harm caused to the patient\(^52\)
- when the use of MDSS fulfils the standard of care in a particular area\(^53\)
- when all medical professionals who use the system are fully educated in the MDSS’ correct application, general functioning; and when the hospital can verify that such training has occurred
- when the patient has given his or her informed consent to the use of the safety-critical system in his or her diagnosis or treatment - where and when consent is possible to obtain, and where there is risk involved in the use (or non-use) of the system\(^54\)
- when the senior medical professional responsible for the patient is fully informed of the legal responsibility (i.e., partial or full responsibility) he or she undertakes when using MDSS to inform medical decisions, including diagnoses and treatment paths\(^55\)

---

\(^{51}\) “Generally” recommended because the specifics of the case at hand – which are likely unique – should always serve as the ultimate litmus test.

\(^{52}\) Collste (2000: 192-3) notes that “efficiency” in one area may entail inefficiency in others. For example, an MDSS may help the doctor to diagnose a condition faster, but may in the process require that the doctor spend less “face time” with the patient, which might have a negative effect on the doctor-patient relationship.

\(^{53}\) Recall that “standard of care” has both an ethical and legal dimension. Within the legal dimension, the focus is a standard upon which professionals in the field in question would generally agree; thus, standard of care is a fluid concept, changing with the emergence of new understanding and new technology.

\(^{54}\) Beauchamp & Childress in *Principles of Biomedical Ethics* (1979 and subsequent eds.) recognize that informed consent is needed only where risk is present.

\(^{55}\) The amount of legal responsibility will depend on the state legal code, and the particular legal question at hand (such as a negligence suit brought on the hospital or medical professional by a patient who suffered after undergoing treatment proposed by a MDSS). While this paper has argued that a sharing of responsibility should
Van den Hoven (2001) formulates a very flexible principle to guide users of MDSS:

“(4) End-users out to endorse (or act upon) the output of Information Systems they are epistemically dependent upon, and with which they know they will be working under conditions of narrow embeddedness, only after an inquiry of acceptability of the system, the cost of which is proportional to the cost that could reasonably be expected if what is endorsed and acted upon should prove in any sense to be inadequate” (van den Hoven 2001: 6-7).

“Narrow embeddedness” describes environment in which the system is used, where the system has invisible complex calculation ability (earlier called “inscrutinizability” and is a factor in all modern computing), the where the user has a limited timeframe for making a decision, and where the system is relied upon (and interpreted as) as a reliable expert. 56

**In-Depth and Continuous Training**

In order to use MDSS responsibly, the healthcare professional must be educated in – and understand – at least three aspects of the system: its capabilities, intended application areas, and general system architecture. Effectively and responsibly educating users in these areas requires collaboration with system designers, and so cannot just happen in the hospital, where it might be carried out only by hospital staff.

Education of users in these three aspects would aim to answer questions such as:

- What is the system able to do? Does it just calculate a diagnosis based on data the user inputs, or can it also regulate treatment by controlling, say, medicine dosage?
- If the latter, is the system checked regularly for accuracy and fault-free operation?
- How many of the system’s capabilities will be used on average when treating a patient?
- What is the system’s general architecture? What, generally speaking, are its programmed reasoning paths?
- Who is responsible for updating and/or editing the MDSS’s knowledge base?

---

56 van den Hoven defines “expert” very narrowly and neutrally: “…information systems…are inhospitable to forms of discursive scrutiny by which we traditionally seek to identify experts and try to establish reliability of expert solutions. In situ the room for critical questioning and evaluation is severely restricted” (Ibid 4).
• As the system is upgraded, are all targeted users retrained and educated about the new or altered functions and capabilities of the system?
• What are the explicitly intended application areas of the system, according to both the system designer(s) and the institution or body which approves or requires MDSS use?

Reponsible System Design

Van den Hoven has also developed a flexible principle for system designers:
“(5) Information Technology professionals and System Designers ought to allow users to work with systems in such a way as not to make it impossible for them to live up to their obligations as users, specified by (4) [see van den Hoven above]” (van den Hoven 2001: 7)

Questions that system designers might ask themselves include:
• Has quality assessment accompanied the entire product cycle – from conception, to design, to production?
• Has the system been tested for quality control?
• If the system in some way replaces physician action (through administering or prescribing drug dosages, for example), are the safety mechanisms hardware- or software-based?
• Is the interface as user-friendly as possible, given its intended or target user group?
• Has the system been developed in partnership with the target users?
• If the system has been developed with little target user input, is the training program adequate to fully educate users in the system’s intended and appropriate uses?

4.4 Larger Implications: Should DSS Usage and AI Research Continue?

Of course, the issues explored in this paper have a broader and wider scope that the immediate world of healthcare. Healthcare happens to be an industry in which the technology has gained a firm foothold, and in which the use of DSS is often safety-critical (which it is not often in, for example, the finance industry).

Recall the quote that opened this paper, from Weckert: “…If they [AI machines] behaved like us, would we be justified in treating them differently?” (Weckert 2002: 369). This paper would argues that the law might answer, ‘no.’ But a deeper concern underlies Weckert’s comments: should AI research continue?
In answering this question, Weckert and others appeal to calm restraint, and advocate that one recognize the differences between pure research and technological development (Ibid 370). Pure research results in knowledge, which itself has either mental or physical consequences, or both. A physical consequence of knowledge could be the technology into which it is developed, whereas a mental consequence of knowledge is not put to use, but may nonetheless profoundly influence its possessor (Ibid 371). But arguments for the external control of pure research often suffer from the following flaws: they rely on unreliable predictions about the future course of some technology; they ignore the fact that research can have both negative and positive consequences; and repression of research leads to a slippery slope (alternatively, the “opens the floodgates” argument) where other areas of society become repressed (Ibid 172).

Thus the potentially negative mental or physical consequences of pure research require society morally to restrict some types. Weckert proposes a guiding principle: “there is a presumption in favour of freedom until such time as a prima facie case is made that the research is dangerous. The burden of proof then shifts from those opposing the research to those supporting it” (Ibid 373).

Some might argue that, if MDSS development continues to head in the direction of emulating human behavior and decision-making, then the need to build an “artificial morality” into computers will grow. Allen et al. (2000), in an article exploring the possibility of programming some version of moral agency (albeit artificial, inorganic) into computers, write: “[t]he risks posed by autonomous machines ignorantly or deliberately harming people and other sentient beings are great. The development of machines with enough intelligence to assess the effects of their actions on sentient beings and act accordingly may ultimately be the most important task faced by the designers of artificially intelligent automata.” (Allen et al. 2000:1). Most accounts of moral responsibility that rely on anthropomorphic criteria would fall into this group. But there are accounts of moral and legal responsibility that do not rely on such anthropomorphic and corporeal criteria. Such accounts have already found a plausible way to tie MDSS to the moral and legal realms, of which human actions are not the sole domain. Even if one disagrees with attempts to assign moral agency and responsibility to today’s MDSS, one cannot ignore that egal agency and responsibility may already be assignable, even though the application of AI to MDSS is in its infancy. Weckert’s comments are more current than one might think.
Works Cited and Consulted


Open Clinical directory of DSS in healthcare. Built by Enrico Coiera.
Available on-line at:
http://www.openclinical.org/aisinpractice.html. Last accessed
01-02-2005.

Metaphilosophy, 16: 289-295.


the Theories of Alan Gewirth and R.M. Hare. D-uppsats i Praktisk Filosofi, TEMA
institutionen. Linköping: UniTryck AB.

of the Fifth ETHICOMP Conference. Leicester: DeMontfort University CD-ROM
only.


Johnson and H. Nissenbaum (eds.) Computers, Ethics &