

Imagining a non-biological machine as a legal person

David J. Calverley

Received: 10 June 2005 / Accepted: 2 February 2007 / Published online: 13 March 2007
© Springer-Verlag London Limited 2007

Abstract As non-biological machines come to be designed in ways which exhibit characteristics comparable to human mental states, the manner in which the law treats these entities will become increasingly important both to designers and to society at large. The direct question will become whether, given certain attributes, a non-biological machine could ever be viewed as a “legal person.” In order to begin to understand the ramifications of this question, this paper starts by exploring the distinction between the related concepts of “human,” “person,” and “property.” Once it is understood that person in the legal sense can apply to a non-biological entity such as a corporation, the inquiry then goes on to examine the folk psychology view of intentionality and the concept of autonomy. The conclusion reached is that these two attributes can support the view that a non-biological machine, at least in theory, can be viewed as a legal person.

Introduction

Law is a socially constructed, intensely practical evaluative system of rules and institutions that guides and governs human action, that help us live together. It tells citizens what they may, must, and may not do, and what they are entitled to, and it includes institutions to ensure that law is made and enforced (Morse 2004).

D. J. Calverley (✉)
Center for Law, Science and Technology College of Law,
Arizona State University, Tempe, AZ, USA
e-mail: David.Calverley@asu.edu

This definition, on its face, seems to be elegant and concise but, like an iceberg, it is deceptive. My purpose in this article is to look at law in a general sense and then use other fields of study to determine if we can reach some conclusions, which might be helpful in our evaluation of non-biological machines,¹ which exhibit mental attributes now viewed as primarily human. If our inquiry has the potential to produce results that are meaningful to both those who will be faced with deciding how to regulate such an entity and to the designers who are actually making the effort to create such a non-biological machine, then it is worth the effort. As stated by Solum (1992):

First, putting the AI debate in a concrete legal context acts as a pragmatic Occam's razor. By examining positions taken in cognitive science or the philosophy of artificial intelligence as legal arguments, we are forced to see them anew in a relentlessly pragmatic context....

Second, and more controversially, we can view the legal system as a repository of knowledge—a formal accumulation of practical judgments. ... In addition, the law embodies practical knowledge in a form that is subject to public examination and discussion.

As with most endeavors, it is often the questions one asks at the outset which determine the nature of the debate and directs the form of the ultimate outcome. If we want to design a non-biological machine which we will at some later point in time claim is entitled to be treated with the same respect and consideration as a human, we should determine as early as possible in the process whether the result we seek will stand up to scrutiny. One way to do this is to ask if the artifact will ever be capable of becoming a “legal person.” Only in this way will the results be amenable to being evaluated by criteria that are consistent with the way humans govern themselves and view each other. Law offers just this pragmatic, reason driven tool.

Before proceeding, a word of caution is required. It is important to point out that this paper deals with isolated attributes and not overall theories that would fully define a legal person. It is beyond the scope of this short paper to delve into what are the necessary and sufficient conditions to definitively establish that something is a legal person (Solum 1992; Rivaud 1992). I am going to suggest only two aspects of the overall picture. The first is one, which has been addressed both in the context of folk psychology and also in corporate law theory: the concept of “intentionality.” The second is more speculative in this context but is well known to philosophy and that is the idea of “autonomy.”

¹ Briefly, a word about terminology; some use the term “non-biological machine” or “artificial intelligence (AI),” others “artilect,” and still others “artifact.” For ease of use and consistency I will use the term “non-biological machine,” except where quoting directly, but any of the others would suffice.

Humans, persons, and property

In presenting arguments which would tend to support the idea that a non-biological machine can in some way be developed to a point where it, or a guardian acting on its behalf, could make a plausible claim that it is entitled to legal recognition, many factors are implicated. One in particular, which needs initial clarification is the distinction between the related concepts of *human*, *person*, and *property*.

The word “person” is derived from the Latin word “*persona*” which originally referred to a mask worn by a human who was conveying a particular role in a play. In time it took on the sense of describing a guise one took on to express certain characteristics. Only later did the term become coextensive with the actual human who was taking on the *persona*, and thus became interchangeable with the term “human.” Even as this transformation in linguistic meaning was taking place, the concepts of person and human remained distinct. To Greeks such as Aristotle, slaves and women did not possess souls. Consequently, while they were nominally human, they were not capable of fully participating in the civic life of the City and therefore not recognized as persons before the law. Because they were not legal persons, they had none of the rights possessed by full members of Athenian society. Similarly, Roman law, drawing heavily from Greek antecedents, made clear distinctions, drawing lines between property and legal persons, but allowing for gradations in status and in the case of slaves, permitting movement between categories.

It was not until much later, with the rise of liberal individualism in Western societies, that a shift from status based concepts to contract based concepts of individual rights forced legal institutions to begin to clarify the distinctions and tensions between the definitions of “human” and “person.” The idea has been expressed in a similar fashion in the following way:

Only when a legal system has abandoned clan or family responsibility, and individuals are seen as primary agents, does the class of persons coincide with the class of biological individual human beings. In principle, and often in law, they need not. ... The issue of whether the class of persons exactly coincides with the class of biologically defined human being—whether corporations, Venusians, Mongolian idiots, and fetuses are persons—is in part a conceptual question. It is a question about whether the relevant base for the classification of persons requires attention to whether things look like “us,” whether they are made out of stuff like “ours,” or whether it is enough that they function as we take “ourselves” to function. If Venusians and robots come to be thought of as persons, at least part of the argument that will establish them will be that they function as we do: that while they are not the same organisms that we are, they are in the appropriate sense the same type of organism or entity (Rorty 1976).

The distinction between human and person remains controversial. For example, in the sanctity of life debate currently being played out in the United States, serious arguments are addressed to the question whether a human

fetus becomes a person at conception or at a later point of viability. Similar questions attach at the end of life: do humans in a persistent vegetative state lose the status of legal person while still remaining human at the genetic level. Likewise, children and mental incompetents are treated as persons for some purposes but not for all, despite the fact they are clearly human. From this we can conclude that personhood can be defined in a way which gives moral and legal weight to attributes which we ultimately define as relevant without the requirement that the entity either be given the full legal rights of competent adult humans or burden them with the duties those rights imply.

Others have stated “[i]n books of law, as in other books, and in common speech, ‘person’ is often used as meaning a human being, but the technical legal meaning of ‘person’ is a subject of legal rights and duties.” (Gray 1909–1921). However, by qualifying this definition with the caution that it only makes sense to give this appellation to beings, which exhibit “intelligence” and “will,” Gray equated person with human. From this we can infer that the real issue is to determine just what attributes the law is particularly interested in defining so that it can specify the who, or what, to which it applies.

We also see a similar type of development in the concepts of person and property as society developed in the Middle Ages in Western Europe, more particularly in England. There, the concepts of a legal person and property became intertwined. Over time, a person was defined in terms of the status he maintained, and obligations he owed, in relationship to real property. With Locke, these ideas coalesced into the view that a person had a property right in his own labor. By exercising this right he exercised control over things such as land and the goods he produced. It was from this view of a person as property holder that the so-called Fiction Theory of corporate personality initially derived. The validity of this proposition as a basis for legal personhood, as we will see later, is subject to much debate, but at least one generally accepted instance where the fiction has been used can be shown by the comparison between a being of the species *homo sapiens* and the legal concept of the corporation as a person (Note 1987). Because synthetic entities such as corporations were authorized by their state granted charters of organization to own property, they were deemed to be “persons.” In the earliest cases the idea that the corporation was an artificial entity was based solely on this derivative claim. It was only later, following the US Civil War, when courts were forced to respond to arguments based on the anti-slavery amendments to the United States Constitution that the concept of a corporation as the equivalent of a natural person began to be articulated. The answer was that the use of the term “person” in the language of the 14th Amendment to the Bill of Rights,² was broad enough to apply to artificial groupings of people not just humans. This idea, based on the view that corporations are nothing more

² Section 1. All persons born or naturalized in the United States, and subject to the jurisdiction thereof, are citizens of the United States and of the state wherein they reside. No state shall make or enforce any law, which shall abridge the privileges or immunities of citizens of the United States; nor shall any state deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws.

than a grouping of individual persons who have come together for a particular purpose, has come to be known as the Aggregate Theory.

It is beyond the scope of this paper to explore this dichotomy between the views in any detail because if we are correct that an artifact of human making can exhibit relevant characteristics for our purposes, such an artifact will not be an aggregation of humans. True it may be the aggregation of human ideas and handiwork, which led to its creation, but the issues raised by that assertion are better handled by other concepts such as intellectual property rights.

From the above discussion we can draw the conclusion that person and human are distinct concepts recognized by the law. Property, in Locke's sense of being the end product of mental or physical effort, is also recognized by the law. However, because a person is able to own property, a person can be defined in a way, which allows us to draw a distinction between property and a legal person. It is now our task, as Rorty (1976) suggested, to find those attributes of the non-biological artifact which functionally indicate that we can make a principled distinction between it and the property it produces, thereby coming to the conclusion that it can support the claim that it is a legal person.

Law and folk psychology

My limited contention in this part of the paper is that if we accept the notion that the definition of person is a cluster concept about which we do not have clearly delineated content, folk psychology gives us a starting point from which to begin our analysis. While it may not be easy to determine whether the various aspects of person we will discuss are necessary and sufficient to meet the minimum requirement of legal personhood, it is possible to get a sense of what would be acceptable to humans in general if they were faced with the question. Certainly under some theories of law, such as positivism (Kelsen 1967), it is logically possible to argue that to the extent law defines what a legal person is, law could simply define a legal person to be anything law chooses it to be, much like Humpty Dumpty in Alice in Wonderland, "nothing more and nothing less." But, this would be a meaningless exercise and intellectually barren. If on the other hand, law, rather than being viewed as a closed system which makes up its own rules and simply applies them to its objects, was in fact viewed as a limited domain which, while it did not necessarily rely upon morality for its validity, drew upon factors outside the law to define its concepts (Schauer 2004), we could articulate the concept of a person by using factors identified by folk psychology which are related more to function. This point will become clearer in the subsequent discussion. The thesis to be examined implies that so long as a non-biological machine had a level of mental activity in areas deemed relevant to law, such as autonomy and intentionality, then it could be a legal person with independent existence separate and apart from its origins as property. Given the wide range of entities and the variety of types of conduct that the law has brought within its scope, we need to identify those aspects of what Leonard Angel called

“functional simimorphy” (Angel 1989). Certainly there is just this type of simimorphy when we look at corporations, and I suggest that nothing we have seen so far requires us to categorically rule out non-biological entities from the equation.

As the quote which began this paper suggests, law is based upon the idea that it has as its target, entities which are susceptible to practical reason. Morse has also written that law “employs the folk psychology model of human action” (Morse 2004a). Consequently, if, for the sake of argument, we accept this formulation of the law, then we can use studies of folk psychology to determine if there are any empirical grounds that will allow us to shed some light on how we can conceive of a legal person. If we do in fact accept the argument that looking at folk psychology is a legitimate exercise in order to ascertain whether the creation of an artificial entity which follows this model can by definition be viewed as a legal person, then we can further speculate on the effects of this attribution. Let us begin by looking at what acceptance of the idea that law follows the folk psychology model might suggest about intentionality.

There are at least two meanings of the word “intentionality,” and those meanings should not be confused because they are not the same. From a philosophical point of view, starting with Franz Brentano (Brentano 1924–1973) and continuing through many later commentators, the philosophical idea of intentionality has referred to the ability of our thoughts to be about something or to represent something. For example, the sentence, “The White House is in Washington, DC,” is a sentence about the location of the White House, and also a statement about a feature of Washington, DC, namely that it is the site of the White House. John Searle has stated that “Intentionality is that feature of the mind by which mental states are directed at, or are about or of, or refer to, or aim at, states of affairs in the world.” (Searle 1999). Searle further notes that while not all intentional states are conscious and not all conscious states are intentional, there is still an essential connection between the two concepts.

Strictly speaking, that is not what is meant by “intentionality” in the vernacular of the law. Nor, as we will see in a moment, is it what is meant by “intentionality” in folk psychology.

The law clearly treats people as intentional agents and not simply as part of the biophysical flotsam and jetsam of the causal universe. ... (L)aw and morality are systems of rules that at the least are meant to guide or influence behavior ... They operate within the domain of practical reason.

...

All things being equal, intentional action or forbearance is the only aspect of the human condition that is fully “up to us,” that is fully within our control, and that can be fully guided by and produced by our reason (Morse 2004a).

From this we see that the legal definition of intentionality differs from the philosophical. Legal intentionality is concerned more with the concept that people act for reasons which they themselves control. According to the folk

psychology conception, the direct cause of an “intentional action” is an “intention.” An intention is action directed. It is based, in a hierarchical relationship, upon two inter-related but different concepts, a desire for an outcome, and a belief about the consequences of the act before it takes place. Both components are necessary to form an intention, so in this sense the intention is derived from the presence of desire and belief, but they are not sufficient for us to take the next step and ascribe intentionality on the strength of the existence of an intention alone. Similarly, desire alone, or belief alone is not sufficient to give rise to the intention. In order for an action to be performed intentionally, the intention has to be present as we have noted, but more is required. In order for there to be intentionality, there must be an intention coupled with action and accompanied by the skill to perform the act and awareness that the act is being performed. The awareness component specifies the agent’s state of mind at the time of acting (knowing what he or she is doing), and the skill component refers to the agent’s ability and skill to perform the action he or she intends (Malle and Knobe 1997; Malle and Nelson 2003).

It is important to note that desire differs from intention in that intention represents a mental state, which leads to an action. It is based on reasoning and involves commitment to act whereas desire is not based on reasoning or commitment. We see from this that it is the intention that is relevant for the legal system because it is the point at which acting begins.

We need to be careful in this regard, because, as Malle and Nelson (2003), strongly suggest, the folk definition actually differs from the legal definition in large part because there is no single clearly understood legal definition, which is consistent. In fact some of the legal definitions are often at odds with the folk sense of intentionality. For our purposes this simply means that we cannot apply legal definitions of intentionality taken from statutes or cases without further understanding the components which make them up and the instances to which they refer. To do so runs the risk of creating further confusion in the attempt to delineate a course of action which we can rely upon to lay out the parameters of what it would take for a non-biological machine to become a legal person. However, in the broad theoretical sense we can still use the folk psychology definition to argue in favor of our claim that intentionality can serve as a component of legal personality. Again, a word of caution is in order. Recent studies by Knobe (2003) and Nadelhoffer (2005) seem to point to the fact that a person’s view of the morality of an action also has significant impact on the determination of whether it was intentional or not. This has interesting implications for any position which is based on a purely positivist view of law, and may indicate that a natural rights view is more pervasive at a folk level. If this is the case, designers of non-biological machines may have to take this moral component into account as well.

Returning then to our analysis, if we can argue that a non-biological machine can form an intention derived from a desire and a belief, and can act intentionally based upon that intention, there seems again to be no theoretical reason why it could not be viewed as a legal person by the average lay person

applying common sense. But our intuition seems to tell us that more may be required, so our inquiry cannot stop at this point.

As previously noted there has been considerable debate concerning the nature of the corporation as a legal entity. First it was viewed as a mere fiction, then justified on the basis that in reality it was an aggregation of individuals acting in concert. There is another suggestion, which I will now examine to see if it can help us to determine if intentionality has any bearing on how a non-biological machine might be treated before the law. If we set aside the distinction between a legal person and a moral person for the time being and focus more on the underlying conditions required for there to be any sort of “intentionality,” we see that the topic has been the subject of some debate in the literature. Peter A. French is perhaps the person most noted for advocating the idea that a corporation is something more than a mere legal fiction, or an aggregation of human employees or shareholders. His view argues that the corporation has natural rights and should be treated as a moral person, in part because it can act intentionally. In this context, French uses the term “intentionally” in virtually the same sense that folk psychology does. Thus, it offers some meaningful basis for comparison to the situation we are considering.

French’s premise is that “...to be a moral person is to be both an intentional actor and an entity with the capacity or ability to intentionally modify its behavioral patterns, habits, or modus operandi after it has learned that untoward or valued events (defined in legal, moral or even prudential terms) were caused by its past unintentional behavior” (French 1984).

Needless to say, French is not without his critics. Donaldson (1982) argues from an Aggregate Theory stance that the corporation cannot have a single unified intention to act. He then goes on to argue that simply having intention is not enough to make the claim that the actor has moral agency, a position at odds with most of the animal rights movement. Werhane (1985) carries this point further and, using the example of a computer system, argues that the appearance of intentionality does not necessarily mean that it acts out of real desires or beliefs. In other words intentionality does not imply that it is also free and autonomous. While I recognize Werhane’s point, I disagree that such a system is impossible to construct. One example of a theory which could lead to just such a functional artificial agent is set forth in Pollock (2006). Further, drawing on Daniel Dennett’s ideas concerning intentional systems, one can certainly argue that Werhane’s position requires one to accept the premise that only phenomenological intentionality counts for moral, and perhaps legal purposes, but that does not appear to be supported by intuition. Functional intentionality is probably enough in a folk psychology sense to convince people that a non-biological system is acting intentionally. Solum (1992) suggests as much in the following language:

How would the legal system deal with the objection that the AI does not really have “intentionality” despite its seemingly intentional behaviors?

The case against real intentionality could begin with the observation that

behaving as if you know something is not the same as really knowing it.
 ... My suspicion is that judges and juries would be rather impatient with
 the metaphysical argument that AIs cannot really have intentionality. ...

If the complexity of the artifact's behavior did not exceed that of a thermostat, then it is not likely that anyone would be convinced that artifacts really possess intentional states—that they really believe things or know things. But if interaction with artifacts exhibiting symptoms of complex intentionality (of a human quality) were an everyday occurrence, the presumption might be overcome. We can see therefore, that we cannot rule out the possibility that a non-biological machine can be developed which would meet the criteria of intentionality set forth above. I also suggest that Werhane's criticism concerning lack of autonomy could also be met by a properly designed machine. That is the topic to which I turn next.

Responsibility and autonomy

If asked whether humans are different from animals most people would say "yes." When pressed to describe what that implies in the context of legal rules, many people would respond that it means we have free will, that our actions are not predetermined. Note, however, that Morse (2004) argues that this is a mistake in that free will is not necessarily a criterion for responsibility in a legal sense. From the perspective of moral philosophy the debate can be couched in slightly different terms. In the view of the "incompatibilist," in order for people to be held responsible for their acts they must have freedom to choose amongst various alternatives. Without alternatives there can be no free will (van Inwagen 1983; Kane 1996). The incompatibilist position has been strongly attacked by Harry Frankfurt, who called their argument the "principle of alternate possibilities" (Frankfurt 1988a). Frankfurt has argued that it is possible to reconcile free will with determinism in his view of "personhood." His conclusion is that people, as opposed to animals or other lower order beings, possess first and second order desires as well as first- and second-order volitions. If a person has a second order desire it means that she cares about her first-order desires. To the extent that this second-order desire is motivated by a second-order volition, that is, wanting the second-order desire to be effective in controlling the first-order desire, the person is viewed as being autonomous so long as she is satisfied with the desire. The conclusion is that in such a case the person is autonomous (Frankfurt 1988b).

It should be noted that in this context Frankfurt is using the term "person" as the equivalent of "human." Others would argue that "person" is a broader term and more inclusive, drawing a clear distinction between person and human (Strawson 1959; Ayer 1963). As is clear from the previous sections, my preference is to use the term "human" to apply to *homo sapiens* and the term "person" to conscious beings irrespective of species boundaries.

It is helpful in this regard to compare Frankfurt's position with Kant's belief that autonomy is viewed as obedience to the rational dictates of the moral law (Herman 2002). Kant's idea that autonomy is rational also differs from that of David Hume who argued that emotions are the driving force behind moral judgments. Hume seems to be an antecedent of Frankfurt's concept of "satisfaction" if the latter's essay on love is understood correctly (Frankfurt 1999). Transposing these contrasting positions into the language used earlier to describe law, I suggest that it is possible to equate this sense of autonomy with the concept of responsibility. As discussed above with regard to intentionality, humans are believed to be freely capable of desiring to choose and actually choosing a course of action. Humans are believed to be capable of changing desires through the sheer force of mental effort applied in a self-reflexive way. Humans are therefore, as practical reasoners, capable of being subject to law so long as they act in an autonomous way.

"Autonomy" has, however, a number of potential other meanings in the context of machine intelligence. Consequently, we need to look at this more closely if we are to determine whether the above discussion has any validity in aiding the design of a non-biological machine.

Hexmoor et al. (2003) draw a number of distinctions between the different types of interactions relevant to systems design and artificial intelligence. First there is human to agent interaction where the agent is expected to acquire and conform to the preferences set by the human operator. In their words, "(a) device is autonomous when the device faithfully carries the human's preferences and performs actions accordingly." Another sense is where the referent point is another agent rather than a human. In this sense the agents are considered relative to each other and essentially negotiate to accomplish tasks. In this view "(t)he agent is supposed to use its knowledge, its intelligence, and its ability, and to exert a degree of discretion." In a third sense there is the idea mentioned before that the agent can be viewed as manipulating "...its own internal capabilities, its own liberties and what it allows itself to experience about the outside world as a whole." Margaret Boden, in a similar vein, writes about the capacity of the agent to be original, unguided by outside sources (Boden 1996). It is in this third sense where I suggest that the term autonomy comes closest to what the law views as crucial to its sense of responsibility. However, before exploring that point further, I believe that a digression to show how the first two of the alternative senses of "autonomy" mentioned can readily be accepted and dealt with by the legal system.

In each of the first two senses of "autonomy" discussed above, there appears to be a referent to which the agent always defers in making its decision. In the first sense it is the human operator. Similarly in the second sense, it is the weighting or value placed upon the various decisions, weighting which is determined, not by the agent, but by the operator who is setting the conditions for the agent's interactions with other agents. In each of these situations there appears to be a "controlling" entity, which is setting the parameters of action. From a philosophical and legal sense this would strongly imply that the agent is not the competent causal agent of a consequence that has legal significance.

Let's look at this more closely. Assume for example that I program an agent so that it enters virtual space, say the Internet, to perform a task I set for it such as locating a particular set of documents. I give it various search criteria, i.e., set the search parameters, then leave it to its own devices in determining how best to accomplish the task and fulfill its duty. Assume further that the agent in fact proceeds to do as directed, but in the process commits what you and I and the world would view as an egregious harm. Perhaps in order to get the document it has to fraudulently represent to another agent that it is authorized to access a particular computer. Perhaps it determines that the best way to obtain the document requested is to copy it from a site where there is a charge for access. In order to avoid this fee, it manipulates another computer to access a third person's bank account to make the payment. Because the initial directions did not explicitly rule out these courses of action, the artifact is not constrained from following them.

In each of these cases the law would have little difficulty in ignoring the "autonomy" of the agent and ascribing legal responsibility to the person who programmed the computer. As explained in Heckman (1999), the law, using various well-established rules such as strict products liability, would have little difficulty in determining that the real actor in this scenario is the person who sets the chain of action into motion.

As a further aside, space considerations preclude analysis of the meaning of the term "agent" as it is used in philosophy and in law. Suffice it to say that law has a technical understanding of the term "agent" which implies that the agent is directed and controlled by a principal which may, if one is not careful, predispose one to conclude that the first two senses of "autonomy" are the only legally relevant ones (Restatement (Third) of Agency 2006). Another topic beyond the scope of this paper, which provides a basis for speculation, is whether a philosophical agent can act morally without exhibiting free will, mental states or responsibility (Floridi and Sanders 2004).

In the third sense of "autonomy" mentioned above, the answer is not so straightforward. Change the above scenario slightly and assume that our initial point of departure is merely a stated desire we have to read a particular document. Our "friend," a conscious machine, hears our expression, and motivated by friendliness and social convention, decides to get the document for us as a birthday gift. Acting upon this determination the machine consciousness then proceeds to commit similar proscribed acts as mentioned above. Note that this sequence of activities would fulfill our intentionality criteria discussed previously. Here, I suggest, something more has happened, something more human like. If, in this scenario, the agent is autonomous in the sense described by Frankfurt, what I call the strong sense of autonomy, then it is conceivable that the law could directly affect the question of how we effectively evaluate such an artifact. If we adopt the strong definition of autonomy, and argue that if it is achieved in a machine, as it would be in the above example, then at least from a functional viewpoint, we could assert the machine is the equivalent of a human in terms of its being held responsible. As noted earlier, one would expect to be met with the objection that such a

conclusion simply begs the question about whether the artifact is phenomenally conscious (Werhane 1985; Adams 2004). But once again, in the limited area we are examining we can put this argument to one side. For law, and for the idea of a legal person we are examining, it simply may not matter. Functional results are probably enough.

In our example, if one can conceive of a second order volition, the desire to be a good friend and to comply with social convention, and can as a result affect a first-order action, the obtaining of the document, constrained only by the idea that one is satisfied by that result, does that not imply a functionally simimorphic characteristic of human action? Going the next step, we can then argue that law acts at the level of this second-order volition. It sets parameters, which, as society has determined, outline the limits of an accepted range of responses within the circumscribed field which it addresses, say contract law or tort law or criminal law. This would imply that law acts in an exclusionary fashion in that it inhibits particular first order desires and takes them out of the range of acceptable alternatives for action (Green 1988; Raz 1975). Note that this does not mean to imply that these are the only possible responses or even the best responses the actor could make. To the extent that the subject to which the law is directed (the citizen within the control of the sovereign in Austin's terms; Austin 1832) has access to law as normative information, she can order her desires or actions in accordance with law or not. This would mean, to borrow the terminology of Antonio Damasio (1994) that the law sets the somatic markers by which future actions will be governed. By acting in a manner where its intentionality is informed by such constraints, and doing so in an autonomous fashion as just described, the artifact appears to be acting in a way which is functionally equivalent to the way we expect humans to act. I suggest that this does not require that the artifact have a universal, comprehensive understanding of the law any more than the average human does. Heuristics, or perhaps concepts of bounded rationality, could provide the basis for making decisions, which are "good enough" (Clark 2003). Similar arguments have been advanced on the role of emotion in the development of a machine consciousness (Slovan and Croucher 1981; Arbib and Fellous 2004; Wallach 2004). Perhaps, in light of work being done in how humans make decisions (Kahneman et al. 1982; Lakoff 1987; Pollock 2006), more pointed analysis is required to fully articulate the claim concerning law's normative role within the context of autonomous behavior. One further caution, even though I suggest that accepting law as a guide to a second order volition does not diminish the actor's autonomy, this proposition can be challenged by some theories such as anarchism (Wolff 1970/1998).

Ultimately the question comes down to whether this type of second order volition can be instantiated in a non-biological machine or whether it is exclusively the realm of biological species. Searle (1980) and others would say that it can only be accomplished in biological systems. Others, however, have started to look at theoretical possibilities where just this type of activity can occur in non-biological systems (Covigaru and Lindsay 1991; Clark 2003;

Holland 2003). In actively moving toward a complex non-biological machine it is possible that this question will be answered empirically.

Conclusion

Our inquiry began with the question of whether we could envision a non-biological machine, which embodies attributes relevant to the way we define a legal person. It was initially suggested that a relevant distinction could be made between the related concepts of human, person and property. In making these distinctions we found that “person” can, and does refer to more than humans. Particularly in the legal sphere, person is a concept quite able to embrace entities, which are non-biological in nature. Next, we saw that folk psychology, in its view of intentionality, provides guidance for us by showing that intentionality, based on a formed intention derived from desire and belief, could carry the weight, at least in a functional sense, of acting in a way sufficient to allow us to say that it is an attribute of a legal person. Finally, we saw that philosophy supplies a further argument in support of our thesis in that the concept of autonomy, if equated with responsibility, allows us to argue that a non-biological machine exhibiting this action could be viewed as a legal person.

The task therefore, is for the designers of non-biological systems to convincingly instantiate these ideas into an actual system. The hope would be to design a set of criteria which, if embodied in an artifact, would support the contention that such an artifact is a legal person. Nothing we have seen in the theory we have examined would suggest that the idea of a non-biological machine constituting a legal person is automatically ruled out. We can therefore conclude that the possibility of a non-biological machine having legal independence is theoretically possible.

Acknowledgments This article is derived from prior papers delivered at a number of conferences held in 2005. The primary papers have appeared in print in *Proceedings of the Symposium on Next Generation Approaches to Machine Consciousness* sponsored by The Society for the Study of Artificial Intelligence and the Simulation of Behavior (SSAISB), and in the AAI Fall Symposium 2005, *Machine Ethics*, sponsored by the American Association for Artificial Intelligence. Permission to use those papers is appreciated. Likewise, the comments and suggestions from attendees at those conferences have been helpful in refining the ideas.

References

- Adams W (2004) Machine consciousness: plausible idea or semantic distortion? *J Conscious Stud* 11(9):46–56
- Angel L (1989) *How to build a conscious machine*. Westview Press, Boulder
- Arbib M, Fellous J (2004) Emotions: from brain to robot. *Trends Cogn Sci* 8(12):554
- Austin J (1832) *The province of jurisprudence determined*, 1955 edn. Weidenfeld, Nicholson, London
- Ayer AJ (1963) *The concept of a person*. St. Martin's Press, New York
- Boden M (1996) Autonomy and artificiality. In: Boden M (ed) *The philosophy of artificial life*. Oxford University Press, Oxford

- Brentano F (1924–1973) *Psychology from an empirical standpoint* (Rancurello AC, Terrell DB, McAlister L Transs.). Routledge, Kegan Paul, London
- Clark A (2003) Artificial intelligence and the many faces of reason. In: Stich S, Warfield T (eds) *The Blackwell guide to philosophy of mind*. Blackwell, Malden, MA
- Covigaru A, Lindsay R (1991) Deterministic autonomous systems. *AI Mag Fall*, pp 110–117
- Damasio A (1994) *Descartes' error*. Harper Collins, New York
- Donaldson T (1982) *Corporations and morality*. Prentice-Hall, Englewood Cliffs, NJ
- Floridi L, Sanders JW (2004) On the morality of artificial agents. *Mind Mach* 3(14):349
- Frankfurt H (1988a) Alternate possibilities and moral responsibility. In: Frankfurt H (ed) *The importance of what we care about*. Cambridge University Press, Cambridge (Original work published 1969)
- Frankfurt H (1988b) Freedom of the will and the concept of a person. In: Frankfurt H (ed) *The importance of what we care about*. Cambridge University Press, Cambridge (Original work published 1971)
- Frankfurt H (1999) Autonomy, necessity and love. In: Frankfurt H (ed) *Necessity, volition and love*. Cambridge University Press, Cambridge (Original work published 1994)
- French P (1984) *Collective and corporate responsibility*. Columbia University Press, New York
- Gray JC (1909–1921) *The nature and sources of the law*. Macmillan, New York
- Green L (1988) *The authority of the state*. Clarendon Press, Oxford
- Heckman C (1999) *Liability for autonomous agent design. Autonomous agents and multi-agent systems*. Kluwer, The Netherlands
- Herman B (2002) Bootstrapping. In: Buss S, Overton L (eds) *Contours of agency*. MIT, Cambridge, MA
- Hexmoor H, Castelfranchi C, Falcone R (2003) A prospectus on agent autonomy. In: Hexmoor H (ed) *Agent autonomy*. Kluwer, Boston
- Holland O (2003) Machine consciousness. Special issue of *J Conscious Stud* 10(4/5)
- Kahneman D, Slovic P, Tversky A (eds) (1982) *Judgment under uncertainty: heuristics and biases*. Cambridge University Press, Cambridge
- Kane R (1996) *The significance of free will*. Oxford University Press, New York
- Kelsen H (1967) *Pure theory of law* (Knight M, Trans.). University of California Press, Berkeley
- Knobe J (2003) Intentional action in folk psychology: an experimental investigation. *Philos Psychol* 16(2):309–324
- Lakoff G (1987) *Women, fire and dangerous things: what categories reveal about the mind*. University of Chicago Press, Chicago
- Malle B, Knobe J (1997) The folk concept of intentionality. *J Exp Soc Psychol* 33:101–121
- Malle B, Nelson S (2003) Judging mens rea: the tension between folk concepts and legal concepts of intentionality. *Behav Sci Law* 21:563–580
- Morse S (2004) *New neuroscience, old problems. Neuroscience and the Law*. Dana Press, New York
- Morse S (2004a) Reason, results, and criminal responsibility. *Univ Ill Law Rev* 2004(2):363–444
- Nadelhoffer T (2005) Skill, luck, control, and intentional action. *Philos Psychol* 18(3):341–352
- Note (1987) The personification of the business corporation in American law. *Univ Chic Law Rev* 54:1441
- Pollock J (2006) *Thinking about acting: logical foundations for rational decision making*. Oxford University Press, New York
- Raz J (1975) *Practical reason and norms*. Hutchinson, London
- Restatement (Third) of Agency (2006) American Law Institute, St. Paul, MN
- Rivaud M (1992) Comment: toward a general theory of constitutional personhood: a theory of constitutional personhood for transgenic humanoid species. *UCLA Law Rev* 39:1425
- Rorty A (1976) *The identity of persons*. University of California Press, Berkeley
- Schauer F (2004) The limited domain of law. *Va Law Rev* 90:1909–1955
- Searle J (1980) *Minds, brains and programs*. *Behav Brain Sci* 3:417
- Searle J (1999) *Mind, language and Society*. Basic Books, New York
- Slooman A, Croucher M (1981) Why robots will have emotions. In: *Proceedings of the IJCAI, Vancouver*
- Solum L (1992) *Legal personhood for artificial intelligences*. *North Carol Law Rev* 70:1231
- Strawson P (1959) *Individuals*. Methuen, London
- van Inwagen P (1983) *An essay on free will*. Oxford University Press, Oxford

- Wallach W (2004) Artificial morality: bounded rationality, bounded morality and emotions. In: Smit I, Lasker G (eds) Cognitive, emotive and ethical aspects of decision making in humans and artificial intelligence, vol I. IIAS, Windsor, Canada
- Werhane P (1985) Persons, rights and corporations. Prentice Hall, Englewood Cliffs
- Wolff RP (1998) In defense of anarchism. University of California Press, Berkeley, Los Angeles (Original work published 1970)