(PT-AI), held at the University of Oxford in September 2013, and is forthcoming in:
Fundamental Issues of Artificial Intelligence (Synthese Library), Vincent C. Müller (ed.), Berlin: Springer.

Machine art or machine artists?

Dennett, Danto, and the expressive stance

Adam Linson

1 Introduction

As art produced by autonomous machines becomes increasingly common, and as such machines grow increasingly sophisticated, we risk a confusion between art produced by a person but mediated by a machine, and art produced by what might be legitimately considered a machine artist. This distinction will be examined here. In particular, my argument seeks to close a gap between, on one hand, a philosophically grounded theory of art and, on the other hand, theories concerned with behavior, intentionality, expression, and creativity in natural and artificial agents. This latter set of theories in some cases addresses creative behavior in relation to visual art, music, and literature, in the frequently overlapping contexts of philosophy of mind, artificial intelligence, and cognitive science. However, research in these areas does not typically address problems in the philosophy of art as a central line of inquiry. Similarly, the philosophy of art does not typically address issues pertaining to artificial agents.

This paper is framed in relation to Daniel C. Dennett's theory of intentionality and Arthur C. Danto's ontological theory of art. The general structure of my argument is as follows: If, through observation and interaction, we discover an (e.g., artificial) agent's behaviors (and behavioral outcomes) to be similar to those of human agents, this suggests that we may reasonably make similar intentional attributions across agents, understood as intentional systems (Dennett 1987). From this, it follows that, given identical respective outcomes of the behavior of two independent agents (e.g., the production of an artwork by an agent, and of an identical artwork by another agent), we could in principle attribute identical intentions to each agent. However, this equivalent attribution would be problematic in the case of art because two materially identical objects, although indistinguishable in ordinary circumstances, may differ from one another in an important ontological sense, on the basis of their causal origin (Danto 1981). I will argue that for an object to be an artwork, its production must stand in relation to a conscience, which in turn must be based in consciousness. The possibility of consciousness and, ultimately, conscience, is constrained by the design of a cognitive apparatus.

Through a critical examination of Dennett's (1987) idea of the *intentional stance*, applied to the consideration of works of art, I have previously introduced the idea of the *expressive stance* (Linson 2013). According to my account, the intentional stance falls short of an adequate means for understanding art (in relation to artists), and I proposed the expressive stance to address this shortcoming. In the following discussion, I have two primary aims: (1) to further my earlier arguments for the expressive stance with a turn to ontological issues in the philosophy of art, especially those brought forth in Danto (1981); and (2) to tie the ontological issues raised in (1) to a contemporary empirically grounded cognitive neuroscience perspective, in particular, concerning the structure and limits of our broader engagement with the world.

In this paper, I will draw upon ontological issues raised in Danto (1981) to further my arguments for the expressive stance. In doing so, I will draw further contrasts between the expressive and intentional stances, especially with respect to a discussion of machine art. At various turns in this discussion, I will refer to some of Dennett's thought experiments from several of his works to illustrate key points in my argument. Section 2, which follows, provides further background to the subsequent discussion. Section

3 considers the grounds for different interpretations of materially identical artworks on the basis of their origins, which relate to a sociohistorical context, but also to a cognitive architecture. Section 4 considers the relation between an agent's cognitive architecture and the way in which the agent's behavior is interpreted. Section 5 examines the connections between conscience, responsibility, and art.

2 Background

In my initial sketch of the expressive stance (Linson 2013), I argued that an important aspect of the intentional stance (Dennett 1987) could be preserved when considering works of art (and artistic performances), namely, the epistemological constraint that, in our pragmatic engagement with the world, we are confined to an external view of other subjects; we make judgments through an interpretation of our observations and interactions. However, I also argued that for interpreting works of art in relation to artists, the intentional stance's 'lens' of rationality is insufficient to give a full account of how we understand art as such. My position is that we interpret artworks as an expression of the artist, based on a plausible interpretation of the artist's life experience in a sociohistorical context, grounded in externally discoverable evidence (in this paper, I will add cognitive architecture to the pool of relevant evidence). This position stands in contrast to the commonly held view that an artwork expresses an artist's intentions, which, in philosophical terms, implies either that the artwork should be understood as resulting from an artist's purported intrinsic intentionality, or, as Dennett (1987) would have it, in terms of a rationalization of the artist's activity, relative (at least partly) to a domain (see Dennett 2001, p. 319ff.). While I find this latter view tenable with respect to general intentional behavior, it cannot on its own account for the ontological aspects of an artwork identified in Danto (1981).

In an ordinary pragmatic context, outside of philosophical discussion, the theorized ontological status of an object, action, or utterance seems to lack practical significance. This is why an externalist account (such as the intentional stance) is useful for explaining the observation of and interaction with a 'black box', from which we can develop a competence theory that need not be a narrow (i.e., Skinnerian) behaviorism (see Dennett 1987, p. 74). Following the logic of a pragmatic, externalist account, I initially formulated the expressive stance without explicitly addressing the ontology of art (Linson 2013). Nevertheless, the context of art brings certain ontological issues into sharp focus.

Danto (1981) investigates the ontological difference between materially identical artworks – or between a materially identical artwork and non-artwork – which he relates to their causal origins and interpretive context. For him, the philosophy of art has as a principal concern precisely these ontological investigations. To set up his discussion, Danto (1981, pp. 1-3) imagines a series of apparently identical red squares from different sources, assembled for an art exhibition:

> The catalogue for it, which is in full color, would be monotonous, since everything illustrated looks the same as everything else, even though the reproductions are of paintings that belong to such diverse genres as historical painting, psychological portraiture, landscape, geometrical abstraction, religious art, and still life. It also contains pictures of something from the workshop of Giorgione, as well as of something that is a mere thing, with no pretense whatsoever to the exalted status of art. (Danto 1981, p. 2)

Danto uses this thought experiment to critically examine a number of perspectives on art, concluding that the definition of art must be a philosophical matter, and more specifically, an ontological one.

Although my previous approach takes a different perspective, I presented an example that is essentially similar to Danto's red squares in my sketch of the expressive stance (Linson 2013). My example concerns an improvising musician's choice between two equally valid notes, which I also compare with the case of two improvising musicians playing what appears to be the same note in otherwise identical conditions. I argue that in either case, a note choice has a role beyond its formal or more broadly rational role (e.g., as the final note of a melody, or as a note placed in the service of crafting a memorable tune); the decision to use a given note connects to the context of the performer's life, of which it is ultimately an expression. While this lens of expression stands in contrast to the rational lens of the intentional stance, I nevertheless agree with Dennett's (1987) premise in my view that the interpretation of an artwork as an expression in this sense need not appeal to a notion of intrinsic intentionality.

While Danto (1981, Chap. 7) also holds the artist's expression as part of the basis for understanding the nature of an artwork, Danto and Dennett disagree about what is "inside the head" (for instance, of an artist). Danto, a sententialist, finds that Dennett's notion of the intentional stance poses a woefully inadequate challenge to sententialism (Danto 1988). Despite this disagreement, the expressive stance finds support in both Dennett's theory of intentionality and in Danto's theory of the ontology of art. Or perhaps it is more accurate to say that the expressive stance is a critical reconfiguration of both theories: it borrows from Dennett's (1987) view of intentional systems, but highlights a limitation of Dennett's view with respect to art; this limitation is underscored using Danto's (1981) criteria for an ontological distinction among artworks. But, significantly, my argument also indicates that Danto's theory of art does not depend upon sententialism, even though these are intertwined in his own elaboration.

3 Empirically grounded interpretation

For any artwork, it is uncontroversial to note that some human or nonhuman (e.g., a machine) indeed *physically* produced the work, or at least, through a physical act, offered up a product or process as an artwork (e.g., by submitting it to an art exhibition, holding a public performance, etc.). The expressive stance (following the intentional stance) suggests (*pace* Searle) that we need not posit an originary metaphysically irreducible intentional source for the work, but rather, that our external perspective is all that is needed to form an interpretation of the work. This entails that, given evidence about the physical origins of the work, we can draw an ontological distinction between two materially identical works with different origins, as Danto (1981) points out with his red squares.

To further illustrate the implications of materially identical works with different origins, I will repurpose one of Dennett's typically enjoyable thought experiments featuring Bach and Rudolph the Red-Nosed Reindeer (Dennett 1991, p. 387-388). As Dennett's version goes, if a previously undiscovered work by Bach were found to have an opening sequence identical to the opening of the Rudolph tune, a present-day listener would not be able to hear the work as an 18th-century Leipziger would hear it. This is in part because of the obvious associations that we would inevitably make but that they would not, having never heard the Rudolph tune. I would like to use Dennett's example to point out that, in such a case, Bach could not be said to have been *making a reference* to the Rudolph tune, in the everyday sense of a musical reference, as that tune had not yet been written (we will set aside here a dedicated philosophical discussion of reference, which could no doubt lead us in any number of directions). Bach likely would have had a musical justification for composing his melody. Perhaps Bach would have had a similar justification as the tune's later composer, given their shared human cognitive architecture and relatively similar social experiences. Let us introduce now, to this

scenario, a crude robot that merely randomly generates melodies and lacks any additional cognitive apparatus. If the random-melody generator were to produce the Rudolph tune, we could also say without question that the robot was *not* making a reference to the tune. That is, we know Bach could not be making a reference to the tune due to historical evidence, but we know that our robot could not be making a reference to the tune due to evidence about its cognitive architecture.

More importantly, the melody "by" the robot, though apparently identical to that of the unknown Bach piece and that of the Rudolph tune, is ontologically distinct as a piece of music. Although Danto does not take up the issue of cognitive architecture *per se*, it seems he would agree. Consider his point that a work by Bach differs from a work by a "fugue-writing machine, something that ground fugues out like sausages […]. The person who used it would stand in a very different relationship to the generated fugues from Bach's" (Danto 1981, p. 203). This point acknowledges that an artwork produced by a cognitive architecture similar to our own must be understood differently from an identical work produced by a radically limited collection of mechanisms.

Sloman (1988) makes a related point about cognitive apparatus in response to Dennett (1988; a précis of Dennett 1987). As Sloman states, Dennett "wants the intentional stance to focus entirely on rational behavior and how to predict it, without regard to how the agent is designed, whether by evolution or engineers" (p. 529). One way of understanding this point is to note that, given a certain performance (i.e., behavior or behavioral outcome), we could define a design space of possible models of cognition that would reasonably produce such a performance. The fact that we could go to great lengths to develop an entirely unrelated design that produces an apparently identical performance does not make the unrelated design applicable to understanding the original performance. From this, it follows that we could differentiate between intentional systems in such a way as to not make the same intentional attribution to (e.g.) a human conversationalist as we would to a conversing robot that uses only a look-up table, even if the dialogue were identical in each case.

To further illustrate the idea of an ontological distinction between materially identical artworks, Danto (1981, Chap. 2) turns to Borges' 1939 story, "Pierre Menard, Author of the *Quixote*" (Borges 1998, pp. 88-95), which I will briefly summarize here. In this story, Borges' first-person narrator is a fictional literary critic who reviews the work of a fictional author named Pierre Menard. The world of the narrator and Menard – as in the famous example of Sherlock Holmes' London – intersects with various "non-fictional" aspects of our world, such as the existence of various authors and works, the most central being Cervantes and the original *Don Quixote*. As the narrator tells us,

> Menard did not want to compose *another* Quixote, which surely is easy enough — he wanted to compose *the* Quixote. Nor, surely, need one be obliged to note that his goal was never a mechanical transcription of the original; he had no intention of *copying* it. His admirable intention was to produce a number of pages which coincided — word for word and line for line — with those of Miguel de Cervantes. (Borges 1998, p. 91)

Menard thought of two ways of achieving this delightfully absurd goal: "Initially, Menard's method was to be relatively simple: Learn Spanish, return to Catholicism, fight against the Moor or Turk, forget

the history of Europe from 1602 to 1918 – *be* Miguel de Cervantes" (p. 91).[1] The narrator tells us that, upon reflection, Menard reconsidered his proposed method, having concluded that "of all the impossible ways of bringing it about, this was the least interesting. [...] Being, somehow, Cervantes, and arriving thereby at the Quixote – that looked to Menard less challenging (and therefore less interesting) than continuing to be Pierre Menard and coming to the Quixote *through the experiences of Pierre Menard*" (p. 91, original emphasis).

According to the narrator,

> it is a revelation to compare the *Don Quixote* of Pierre Menard with that of Miguel de Cervantes. Cervantes, for example, wrote the following (Part I, Chapter IX):
>
> … truth, whose mother is history, rival of time, depository of deeds, witness of the past, exemplar and adviser to the present, and the future's counselor.
>
> […] Menard, on the other hand, writes:
>
> … truth, whose mother is history, rival of time, depository of deeds, witness of the past, exemplar and adviser to the present, and the future's counselor. (p. 94)

Comparing these excerpts, the narrator describes the "striking" contrast in styles: "The Cervantes text and the Menard text are verbally identical, but the second is almost infinitely richer" (p. 94).

Danto (1981) uses this story to underscore his point about the materially identical red squares introduced earlier, noting that the conditions under which they were produced – their causal origin – is situated in a sociohistorical context, just as the texts by Cervantes and Menard. His point is that our *interpretation* of artworks must take their sociohistorically situated causal origins into account:

> You can certainly have objects – material counterparts – at any time in which it was technically possible for them to have come into existence; but the works, connected with the material counterparts [...], are referentially so interlocked into their own system of artworks and real things that it is almost impossible to think of what might be the response to the same object inserted in another time and place. (Danto 1981, p. 112)

And while Danto (1981) does not explicitly consider cognitive architecture, I have indicated how this consideration could be relevant to understanding the difference between two works produced under similar sociohistorical circumstances by agents with vastly different cognitive architectures.

---

1 Dennett makes a similar proposal (which he dismisses as unnecessary) for how one might try to experience a Bach cantata as an 18th-century Leipziger would have experienced it: "To put ourselves into the very sequence of experiential states such a person would enjoy […] would require […] forgetting much of what we know, losing associations and habits, acquiring new habits and associations". This would take place in "isolation from our contemporary culture – no listening to the radio, no reading about post-Bach political and social developments, and so forth" (Dennett 1991, p. 441-442).

4 Conduct and context

The relation between behavior and interpretive context will be explored in this section in relation to cognitive architecture, personhood, and computational creativity. As indicated in the previous section, we may consider the relation of cognitive architecture to the interpretation of an agent's behavior. Cognitive architecture, though technically "inside the head" by some accounts, does not necessarily point to intrinsic intentional states. Significantly, an agent's cognitive architecture may be relevant to the determination that the agent is a person, even without an appeal to mental content (cf. Dennett 1981, Chap. 14). An agent may have specific capacities that allow it to engage in humanlike behavior, while nevertheless lacking the capacities that underpin personhood. When artificial agents are designed to exhibit creative behavior, as in some computational creativity research, a specific aim may be to produce artworks. The philosophical dimension of these artworks will be considered below.

As a way into this discussion, I will use another of Dennett's thought experiments, which I will briefly summarize here. Dennett also reflects on a story by Borges – in this case, "The Circular Ruins" – to set up his premise: Suppose there is a "novel-writing machine, a *mere* machine, without a shred of consciousness or selfhood" (Hofstadter and Dennett 1981, p. 351). Dennett adds that we can suppose its designers "had no idea what novels it would eventually write". He develops the thought experiment further, coming to the point at which, rather than a simple novel-writing box-like machine, we are faced with a (presumably humanoid) robot that speaks aloud a first-person narrative. Its spoken narrative more or less corresponds to what we can observe about it (e.g., "When it is locked in a closet, it says: '*I* am locked in the closet!'", p. 351). With this setup, Dennett then poses the question: Why should we call this spoken narrative fictional? He answers that we should not, given that this is effectively how human brains work: "Your brain, like the unconscious novel-writing machine, cranks along, doing its physical tasks [...] without a glimmer of what it is up to. [...] *It* doesn't 'know' it is creating you in the process, but there you are, emerging from its frantic activity" (pp. 351-352, original emphasis). If we go along with this, Dennett thinks, we should also be willing to conclude that the robot is a *person*, on the basis that its personhood emerges from its "activity and self-presentation in the world" (p. 351).

There is something very important about interpretation that is missing here, which is surprising, given Dennett's (1987) insightful foregrounding of the role of interpretation in his theory of intentionality. While we can indeed take into account an agent's "activity and self-presentation in the world" for making judgments about personhood, the context for interpreting the agent's behavior is crucial to making the distinction between a fictional entity and a person. Danto (1981, p. 23) recognizes this distinction when he identifies the interpreted difference between an assertion and a *mention of an assertion*, i.e., an assertion in quotation marks; he links this with our ability to identify the conventional context of a theatrical play, for example, which guides our interpretation of actions and words on a stage. Such interpretations must differ from those of actions and words outside of the theater.

We may reasonably go along with the idea that Dennett's robot is a person in one of two ways, both of which seem to undermine the point he seeks to make with his scenario. In the first way, we may take a vantage point from which we have no knowledge that we are dealing with a robot, thus mistaking it for a human and bringing all of our assumptions about humans to bear on the situation. In this case, its activity and self-presentation are already interpreted as relating to personhood, even if this interpretation would be importantly altered with key facts about the robot's cognitive shortcomings. Or, in the second way, we may have foreknowledge that we are dealing with, say, a highly sophisticated robot, wired up like a human and perhaps even having a human-like upbringing, such that by some standards of human equivalence, it already may be considered a person, regardless of its present activity and self-presentation, e.g., if the robot is somehow temporarily deactivated.

This point can be made more clearly using a different example. Imagine a performance artist sitting completely still in a room. With no knowledge about the performer or context, the performer's mere activity and self-presentation would not necessarily lead anyone to believe they are a person. The determination that they are a person could be made either with some brain scans and other medical tests that would satisfactorily prove they are a living human; this would at least imply that they should probably be considered a person ("we normally […] treat humanity as the deciding mark of personhood", although they may be exceptions; see Dennett 1981, p. 267). Or, we could be furnished with the context that we are witnessing a performance in a museum by someone operating within a certain performance tradition, and so forth, in which case we could assume that their motionless sitting *is* in fact their activity and self-presentation in the world. This interpretation requires the broader context, which would reasonably lead us to conclude they are a person. In neither case would a robot known to have crude cognitive machinery, designed for superficial humanlike activity and self-presentation, be considered a person.

While I am in agreement with Dennett (1981) that the conditions of personhood should not be grounded in intentional states, I would, however, respond that its conditions may nevertheless be grounded in capacities such as having experience, making judgments, reflecting, and so on, which certainly some future machine may be capable of, but hardly an unconscious novel-writing machine. Empirical questions about these capacities can be explored with respect to the cognitive architecture that underpins them (see Sloman 1988). Significantly, these capacities relate to our broader engagement with the world, while leaving in tact the idea that they may complement domain-specific cognitive mechanisms for particular specialized activities (e.g., writing a novel).

We must, however, tread carefully in our understanding of the relation between these general capacities and more specific ones, such as creativity. One of the key ideas behind computational creativity is that there are cognitive mechanisms associated with the production of new ideas and, by extension, with the production of new works of art. Boden (1990/2004) distinguishes between what she terms personal or psychological creativity – meaning a new idea arises in an individual cognitive agent – and historical creativity, meaning that an idea is recognized as new and valuable by a community. Research in this area often focuses on psychological creativity though the development of machines with a high degree of autonomy that produce art, especially musical, visual, or literary art. In this context, the distinction can be easily obscured between some of the domain-specific activity relevant to artistic production and the general cognitive activity relevant to personhood. For example, a short story writing machine developed by Bringsjord and Ferrucci (1999, p. 100) has the stated aim that it "holds its own against human authors" (in terms of observable behavior). Their machine's approach "to story generation is based on the assumed limitation of computers to genuinely grasp such things as interestingness" (p. 199). In this case, the ability to identify an interesting story is regarded as a domain-specific capacity rather than a general one.

Given that the idea of an unconscious novel-writing machine is clearly not as far-fetched as it might sound to some, in this context, I would like to consider a distinction between a work produced by an expert system and a work produced by a person. Following points by Danto (1981) and Sloman (1988) introduced above, we may say that materially identical works, one produced by an expert system and the other by a person, must have a different ontological status as artworks, because of the way the work relates to the conditions under which it was produced. These conditions must be taken into account for a defensible interpretation of the work. As Danto states,

> It is not just that appreciation [of an artwork] is a function of the cognitive location of the aesthete, but that the aesthetic qualities of the work are a function of their own historical identity, so that one may have to revise utterly one's assessment of a work in the light of what one comes to know about it; it may not even be the work one thought it was in the light of wrong historical information. (Danto 1981, p. 111)

This relates to the present discussion in that, if we are moved by (e.g.) a novel and think it captures something about human experience, but we then discover that it was produced by a robot with a cognitive apparatus that is limited in certain key respects, this new information demands of us that we modify our original interpretation. We may still agree that the story moved us, but not because the robot shares an idea about human experience.

In fact, in this example, it is the robot's designers who share an idea about human experience. They may have designed, for instance, a largely autonomous system that generates stories and decides that they are complete and ready for the public. In doing so, however, the designers have made a decision of their own that significantly impacts what the robot does. Namely, they have decided that their purpose-built robot is itself complete and ready for the public. The designers' judgment that their art-producing robot is ready to be 'released' is part of their judgment that its output objects can, from that point forward, be treated as artworks (novels, paintings, performances, etc.). Even if such a system has an internal 'critic' to evaluate its own output, the criteria for the critic have likewise been established by the designers (though computational implementations of aesthetic evaluation may be "almost comically faulty"; see Thomson 2011, p. 60). The designers thus bear aesthetic responsibility for the machine output, even if they do not know the precise objects it will produce.

My position is that any such work, despite the designers' lack of knowledge about future system output, is nevertheless an expression of the designers, rather than of the machine. If the works are regarded as artworks, they must be understood as works by human artists, mediated by autonomous machine production. This is importantly different from an actual machine artist, which would only be possible if the machine were a person (as some future machine may be, but, as far as we know, no current machine is). Thus, mine is not among the familiar positions that art could only be art if made by a human, or that a machine could never be creative, have emotions, etc., which often seem to be the main positions defended against in this context by philosophers including Dennett and Boden.

Returning to Dennett's robot, let us assume it is highly sophisticated, beyond all current technology, and that we could have a conversation with it that is apparently about its childhood. Even if it could pass a Turing test for intelligence, there remain two important facts for us to know about the robot before we ought to be tempted to consider it a person. The first fact concerns its location in history: Was it powered up today for the first time, revealing its designer's remarkable generator of childhood stories? Or, did it actually spend time as a less-developed robot with a humanlike upbringing, analogous to a human child? While this may seem an obvious piece of discoverable evidence, it remains outside of the Turing test, as the machine could always give answers as if the latter were the case, even if the former were true. The second fact concerns its cognitive architecture: Does it use a cognitive architecture similar to our own, capable of experience, judgment, reflection, etc., or does it merely use a crude look-up table or similarly simplistic mechanism? As Sloman (1988, p. 530) points out, there are "*design* requirements for various kinds of intentional abilities" and certain philosophical considerations are mistakenly based on "an oversimple view of the space of possible designs" (see also Shieber, in submission).

While an external view of a 'black box' may suffice for understanding a limited practical engagement (typical of our daily encounters with others), the answers to the above questions would be evidence that gives us more or less justification to treat the robot as significantly equivalent to ourselves – perhaps even as a person – regardless of its inherent biological difference. Assuming we do find it to have autonomously developed over time, as part of our society, accruing experience, exercising a faculty of judgment, undergoing reflection, etc., if it *then*, under these circumstances, were to produce an artwork (write a novel, paint a painting, improvise a musical solo, etc.), we would be in a position to reasonably interpret its work as that of a machine artist, despite some prior role of a robot designer. Such a robot would not be regarded as a traditional domain-specific expert system, but more like what we would ordinarily regard as a person. Its artistic output would be ontologically different than randomly-generated output or output generated because its designers settled upon domain-specific mechanisms to produce even largely unforeseeable works.

5 Conscience and consciousness

As I have argued, artificially intelligent art-producing machines can be regarded as a class of expert systems, but works of art as such (as opposed to objects identical to artworks) cannot be adequately understood as mere exercises of specialized skills. Rather, such skills, even when local to the production of a specific artwork or performance, fundamentally relate to a broader engagement with the world. Thus, the critical difference between a domain-specific expert system and a person, as viewed from the expressive stance, can in some sense be understood as a reassertion of Dreyfus' (in)famous Heideggerian critique of artificial reason (Dreyfus 1992; see also Dreyfus 2008), but narrowly focused on art.

With respect to art and intentionality, in Linson (2013) I explicitly draw on Dreyfus' (1993) Heideggerian critique of Searle's two senses of the phenomenological "Background". In short, the critique holds that our actions are not only relative to our physical bodies and cultural circumstances, but also to our broader engagement with the world. Roland Barthes makes a similar point when he argues that it is not only one's body and sociohistorical circumstances that give an author (or artist) a unique perspective, but also a decision to have a particular take on the world (Barthes 1968). For Barthes, this take is linked to the idea of conscience.[2] Dennett (1981, p. 297) also accords an important role for something like conscience when he partly locates our moral responsibility in the fact that we decide when to terminate our deliberation process for executing a given action.

Conscience is not an especially well-defined term in philosophy, and some view it as an inner voice that distinguishes between right and wrong. But this view is too narrow for what I am trying to capture here. I am instead suggesting that conscience is relevant even when not faced with a straightforwardly moral question. In some sense, any social act has an inherently ethical dimension, and our decisions to act in certain ways relate to our general sensibilities about how we ought to act; these sensibilities are partly arrived at through socialisation and partly arrived at through self-examination and reflection. Making art of any kind (not only overtly political art) is not the type of activity that is usually held up as a moral act, but it may be viewed as the outcome of an artist's deep convictions about humanity. As philosopher of law Larry May (1983, p. 66) states, conscience motivates us "not to view our own selves as the end to be served, but to view humanity *in* our own and other persons as the end to be served" (original emphasis). An artist's convictions of this sort pertain not only to the production (or performance) of a specific work, but also pertain to the more general decision to produce art for the public, in other words, the decision to be an artist.[3]

2 It is interesting to note that the French word, *conscience*, may be used to mean either conscience or consciousness.

Before addressing conscience further, we may note that Dennett and Barthes agree that one's consciousness (in the general sense of subjectivity) is affected by the society and historical period in which it developed. To take one example, Barthes (1968) points out a difference between the writing styles of Balzac and Flaubert that he attributes to the fact that their lives are separated by the events of revolutionary Paris in 1848. The societies and traditions of these literary figures are, on one hand, remarkably similar, but, on the other hand, their respective worlds differ significantly. It seems Dennett would agree, on the basis of a related example:

> There are probably no significant biological differences between us today and German Lutherans of the eighteenth century […] But, because of the tremendous influence of culture […] our psychological world is quite different from theirs, in ways that would have a noticeable impact on our respective experiences when hearing a Bach cantata for the first time. (Dennett 1991, p. 387)

As Dennett suggests, when we hear Bach's chorales today rather than in Bach's time, "we hear them with different ears. If we want to imagine what it was like to be a Leipzig Bach-hearer, it is not enough for us to hear the same tones on the same instruments in the same order; we must also prepare ourselves somehow to respond to those tones with the same heartaches, thrills, and waves of nostalgia" (Dennett 1991, p. 387).

Generally speaking, whether we are concerned with short stories or novels, musical compositions or improvisations, performance art or any other artistic media, we regard artworks as a result of decisions by one or more artists – at the very least, deciding when a work is finished and ready for the public, though we may also include decisions about formal and structural aspects of the work and, at a more "global" level, decisions such as the adherence to or flouting of traditions, etc. From the vantage point of the expressive stance, we can read such works as expressing something – whether their author intended them to or not – about the times and society in which the author lived, and the experiences the author underwent in these contexts. In this sense, we may say the works are an expression of the artist's life and, in particular, of the artist's consciousness. But if the decision-making apparatus of an individual artist is relevant to the interpretation of an artwork, how can we adopt an external vantage point that disregards the traditional notion of authorial intentions "inside the head"? And, assuming we disregard such intentions, how might this relate to the theoretical notion of the "death of the author"?[4]

It was in fact Barthes himself who, in a 1967 essay, "The Death of the Author", introduced its titular notion in the contemporary sense (Barthes 1977). Barthes used the phrase in response to an earlier generation of literary critics who believed that there must be one ultimate meaning of a given literary work (to which we may also add, any artwork). This meaning was assumed to be deducible from the author's intentions, conscious or unconscious, public or hidden, which could, for example, be partly uncovered in the author's journals, biography, and so on (a version of this view is known by literary critics as the "intentional fallacy"; see also Dennett 1987, p. 319). Barthes argued that a multiplicity of interpretations should be possible, because the meaning is partly constituted by the reader, who brings his or her thoughts and experiences to bear on the interpretation of the material. We can make sense of this view in relation to other artistic practices as well: "responding to a painting complements the

3 Conscience also plays a role in Danto's (1981) contention that artists must be morally responsible in their decisions about what to portray and how to portray it, which he explores in relation to the concept of the "psychic distance" an aesthetic attitude has from a practical one (p. 21-24).

4 This concept was mentioned briefly in Linson (2013) but, due to space limitations, was not addressed in depth.

making of one, and spectator stands to artist as reader to writer in a kind of spontaneous collaboration" (Danto 1981, p. 119).

The idea of the death of the author – much like Dennett's (1987) critique of intrinsic intentionality – is to delink the (outwardly observable) work from some imagined definitive "theological" key in the author's mind that holds the ultimate answer to the intentions behind the work. According to Barthes' view – which is also similar to Dennett's (1981; 1987; 1991) critical view of introspection – the author's intentions, which were held by prior critics as constituting the originary source of an artwork, should not be assumed to be consciously furnishable by the author, say, in an interview, nor should they be assumed to be discoverable by the critic researching the author's memoirs. Rather, we should not understand any definitive intentions as being ultimately expressed in or by the work. Coming to an understanding of an artwork – or hermeneutically engaging with its meaning, without a notion of a final point of arrival – is a process of interpretation that is grounded by evidence. There may be irresolvable conflicts among competing interpretations, but this situation does not fundamentally differ from that of the sciences: As Danto (1981, p. 113) notes, there is a "slogan in the philosophy of science that there are no observations without theories; so in the philosophy of art there is no appreciation without interpretation".

The mode of interpretation proposed by Barthes (1977) is very close to Dennett's position on intentional interpretation, where we may encounter a number of plausible reasons for someone's decision leading to an action, reasons which can never be definitively proven, although a better or worse case can be made (see Dennett 1987, Chap. 4). For Dennett, this scenario exemplifies taking the intentional stance: What is "inside the head" is not pragmatically relevant to the interpretation of an intentional system's activity or output. Rather, we interpret such activity and output on the basis of a community of shared meaning. Thus, Barthes' "death of the author" is already to some extent a preliminary version of the intentional stance epistemology for the arts. But, importantly, this interpretation-centric view is not a pejoratively construed "anything goes" version of postmodern theory. A reasonable basis for interpretation is important, just as Dennett (1987, p. 100) points out for interpreting Jones' likely delusional rationale for the terrible-looking extension to his house. As with scientific observation, there may be facts that undermine certain interpretations (recall Danto's (1981) point that "one may have to revise utterly one's assessment of a work in the light of what one comes to know about it", p. 111).

We have already indicated the importance of evidence about an artwork's origins for determining its ontological status. We have also pointed out that those origins at once relate to sociohistorical location, cognitive architecture, and, at the intersection of both, the artist's consciousness – and, more specifically, the artist's conscience, an important aspect of personhood. A person – natural or artificial – can be said to produce an artwork through a process of making decisions, not only about the form and content of the work, but about when it is ultimately ready for the public. Dennett (1981, p. 297) makes a related point concerning the basis of our responsibility for moral decisions: "In many cases our ultimate decision as to which way to act" is importantly connected to "prior decisions affecting our deliberation process itself: the decision, for instance, not to consider any further, to terminate deliberation; or the decision to ignore certain lines of inquiry". He continues:

> These prior and subsidiary decisions contribute […] to our sense of ourselves as responsible free agents, roughly in the following way: I am faced with an important decision to make, and after a certain amount of deliberation, I say to myself: "That's enough. I've considered this matter enough and now I'm going to act," in the full knowledge that I could

> have considered further […] but with the acceptance of responsibility in any case. (Dennett 1981, p. 297)

When an artist finally decides to declare a work as ready, this constitutes the termination of a cycle of on-going action and judgment and the taking of responsibility for the production of the work. Here, it is true that, from the intentional stance, we can understand the work in one way by an appeal to reason: the painter decided to add no more, so the canvas would not be overworked; the poet decided to stop editing the poem, so its initial impulse would not be obscured. But apart from an explicit or attributable rationale, the decision to conclude with the production of a work is shaped by the artist's sensibility, a sensibility that arises from the life experience of the artist and, ultimately, from the artist's conscience, formed in the course of experience. This is not to deny that the life experience of artists is in many ways due to factors beyond their control, including sociohistorical circumstances of geography and culture, contingencies about their physical bodies, and the evolutionary biological inheritance that amounts to a cognitive apparatus. Within these constraints, however, a conscience is formed that allows one – an artist or indeed any person – to take responsibility for one's decisions, such as those pertaining to an artwork.

6 Conclusion

We are now in a position to account for the question posed in the title: Machine art or machine artists? As I have argued, we may identify how an artwork relates to a specific historical period and culture, a specific artist's body and cognitive apparatus, and, subject to these constraints, an artist's conscience. A conscience in this sense is certainly historically situated and embodied, but also relates to a person's unique accumulation of experience. One develops such a conscience and it affects one's decisions about how to carry out certain activities, such as producing an artwork and determining that it is ready for the public. The development of a conscience is only possible given particular facts about our cognitive apparatus that facilitate our accumulation of experience, our faculty of judgment, our capacity for reflection, and so on. Thus, my argument entails that for a machine-produced artwork to be regarded as a contribution by a machine artist – rather than by the machine's designers – the machine's cognitive architecture must make possible experience, conscience, and the closely related broad engagement with the world. Without such a cognitive architecture, machine-produced art must instead be understood as the work of a human artist, mediated by a machine.

Danto (1981) has confined his account of an artist's relevant cognitive processes to a sentential account of the mind, which, following my above arguments, should be viewed as entirely unnecessary to his ontology of art. An empirically based neurobiological account of our cognitive architecture, for instance, could plausibly describe how the mind might call upon experience to guide action without appealing to a sentential account. By dispensing with sententialism and taking cognitive architecture into account, the expressive stance can facilitate the ontological distinction between artworks and non-artworks that Danto envisions, while preserving Dennett's insight that intentional states should not be understood as intrinsic, as I have argued with respect to an artist's intentions.

Along the lines envisioned by the Turing test, we may say that, using the best available evidence during strictly external observation and interaction – that is, without an appeal to what is "inside the head" – we may potentially detect no difference between a human and a machine. If an unknown entity is determined to be sufficiently adequate at conversing, playing chess, etc., then we are reasonably entitled to make certain assumptions about it (e.g., that it can think). For example, during an ordinary form of interaction with a neighbor at the local store, we may set aside the question as to whether this neighbor is indeed just like us, or whether they are perhaps a robot (or a zombie, etc.). The neighbor,

like us, is an intentional system, that is, a system to which we can consistently attribute intentional behavior.

However, a look inside an intentional system – at its cognitive apparatus, not at some metaphysical notion of mental content – could reveal a relatively crude apparatus, like a look-up table. This discovery may render false our prior assumptions about the agent and thereby lead us to draw different conclusions about its personhood, even if we would ordinarily grant it personhood on the basis of our external interactions and observations alone. Assuming equivalent external performances, an agent with a cognitive apparatus similar to our own should be regarded as more deserving of the ascription of consciousness than a look-up-table-based agent. If consciousness can be reasonably attributed to the agent, we may then inquire into what mechanisms structure the agent's decision-making, and whether or not these can be said to ultimately relate to a conscience that underlies the agent's ability to take responsibility (e.g.) for producing an artwork. To determine whether an artwork is the expression of a machine or its designer, we must recognize the fundamental relationship between an artwork and a conscience. Ontologically speaking, an object can only be an artwork because of this relationship.

References

1. Barthes, Roland, *Writing degree zero*. Hill and Wang, New York (1968)
2. Barthes, Roland, *Image, music, text*. Fontana Press, London (1977)
3. Boden, M., *The Creative Mind*. Routledge, London (2004 [1990])
4. Borges, J. L., *Collected fictions*. Penguin Books, New York (1998)
5. Bringsjord, S., and Ferrucci, D., *Artificial intelligence and literary creativity*. Psychology Press, Hove (1999)
6. Danto, Arthur C., *The transfiguration of the commonplace*. Harvard University Press, Cambridge (1981)
7. Danto, Arthur C., The notional world of D. C. Dennett, *Behavioral and Brain Sciences*, 11, 509-511 (1988)
8. Dennett, D. C., *Brainstorms*. MIT Press, Cambridge (1981)
9. Dennett, D. C., *The intentional stance*. MIT Press, Cambridge (1987)
10. Dennett, D. C., Précis of *The intentional stance*, *Behavioral and Brain Sciences*, 11, 495-505 (1988)
11. Dennett, D. C., *Consciousness explained*. Little, Brown and Co., Boston (1991)
12. Dennett, D. C., The evolution of culture, *The Monist*, *84*(3), 305-324 (2001)
13. Dreyfus, H. L., *What computers still can't do*. MIT Press, Cambridge (1992)
14. Dreyfus, H. L., Heidegger's critique of the Husserl/Searle account of intentionality, *Social Research*, 60(1), 17-38 (1993)
15. Dreyfus, H. L., Why Heideggerian AI failed and how fixing it would require making it more Heideggerian, in *The mechanical mind in history*, eds. Husbands, P., Holland, O., and Wheeler, M., MIT Press, Cambridge, 331-371 (2008)
16. Hofstadter, D. R., and Dennett, D. C., *The mind's I*. Basic Books, New York (1981)
17. Linson, Adam, The expressive stance: Intentionality, expression, and machine art, *International Journal of Machine Consciousness*, 5(2), 195-216 (2013)
18. May, Larry, On conscience, *American Philosophical Quarterly*, 20(1), 57-67 (1983)
19. Shieber, Stuart M., There can be no Turing-test-passing memorizing machines (in submission)
20. Sloman, Aaron, Why philosophers should be designers, *Behavioral and Brain Sciences*, 11, 529-530 (1988)

21. Thomson, Iain D., *Heidegger, art, and postmodernity*. Cambridge University Press, Cambridge (2011)