RESEARCH

Joyful Encounters: Learning to Play Well with Machines

Eleanor Sandry and Gwyneth Peaty
Curtin University, AU
Corresponding author: Dr. Eleanor Sandry (e.sandry@curtin.edu.au)

Human interactions with machines, including computers, consoles, smart devices and robots, are becoming more and more a part of everyday life. However, human–machine relations are often regarded as problematic for people, their interpersonal communication and human society more broadly. This paper provides an analysis of the characteristics that constitute ‘play’ in relation to video games and interactions with robots, arguing it is possible to position time spent on play with machines as valuable in itself, without requiring the outcomes more traditionally regarded as productive. Much of what is valuable in play can be seen to develop from embodied processes of communication within which humans and machines encounter and respond to one another. These encounters are often shaped by stories about the capabilities of machines and humans, while the interactions themselves go on to provoke new narratives. Although human–machine interaction can be theorized as ‘cyborg’ or ‘hybrid’, this paper argues that adopting the idea of the ‘assemblage’ is a better way recognize the flexibility of bringing disparate humans and machines together, whether in relation to playing a game or playing music. In rethinking the value of play, this paper emphasizes how people’s time spent interacting, whether with video games or robots, provides opportunities for them to learn more about themselves and others.

Keywords: human–machine interaction; human–machine communication; play; video games; robotic musicianship; assemblages

Introduction

While human interactions with machines, such as robots, computers and other devices, are becoming commonplace in people’s lives, especially since the widespread uptake of mobile devices including smartphones, human–machine relations are still often seen as at best problematic. At worst, perpetual interaction with devices ‘is linked to depression, accidents and even death’ (Whillans 2018). Anxiety surrounding the impact of machines on human health, relationships and society has a long history in popular culture (Duchaney 2015; Peaty 2018). The time that people spend interacting with machines, especially when this is framed as a form of play, is a source of discomfort. It has been suggested that people need to attend more closely to the importance of face-to-face human–human conversation, setting aside machines, including social robots and smart device screens, to do so (Turkle 2011; Turkle 2015). Public discussions of human–machine interaction consistently reveal concern about the future impact of these encounters, because ‘as machines are made to look and act like us and to insinuate themselves deeply into our lives, they may change how loving or friendly or kind we are—not just in our direct interactions with the machines in question, but in our interactions with one another’ (Christakis 2019). The fear that communicating with and via machines has already replaced and ultimately ‘killed’ face-to-face human interaction is widespread in media discourse on the topic (Morris 2015; Goman 2018; Chan 2019). This paper considers two specific examples of human–machine communication: human interactions with video games and human-robot interactions. While these forms of communication might initially be perceived as quite different, we observe many useful synergies and points of overlap. Significantly, both of these examples are regarded by some as potentially damaging in one way or another to human individuals and human societies more broadly. In this paper we aim to move beyond fear-based analysis to consider the benefits, joys, and contributions offered by these interactions, specifically in the context of playfulness.
Working through an analysis of the characteristics that constitute ‘play’ in relation to video games and interactions with robots makes it possible to position time spent playing as valuable in itself, without always needing to identify outcomes traditionally regarded as productive (although practical effects, such as increased familiarity and dexterity with a computer, device or robotic interface can result). In this paper, much of what is valuable in play is argued to develop from an embodied process of communication, through which both humans and machines encounter and respond to one another, in ways often shaped by and shaping stories about them and their interactions, to enable dynamic and flexible meanings and understandings to emerge. Although such close interactions between humans and machines can be theorized as ‘cyborg’ or ‘hybrid’, this paper argues that thinking of these relations as ‘assemblages’ offers a more productive way to understand the flexibility of drawing together disparate humans and machines in a particular context of play, whether relating to games or improvised music performance. The paper considers how interactions with video games and robots offer ways for people to learn more about themselves, the environment and others; breaking down the boundary of the self in the process through interactions that are not only enjoyable and entertaining, but also have the potential to become collaborations, supported by human–machine communication that coordinates joint action.

**Linking Robots and Video Games as Problematic Technologies**

Fictional depictions of robots often divide them using a simple dichotomy of good or bad. ‘Good’ robots are safe and dependable, acting as benevolent servants and friends, such as the droids R2D2 and C-3PO in *Star Wars*, and *The Jetsons*’ maid Rosie. To develop good robots in the real world, roboticists are working on cute companion robots, such as Jibo and Vector, designed to offer love and friendship in real life, although these companions are often found lacking after time (and are now no longer in production, although Vector may be re-launched in the future). These robots are often described as ‘social robots’ and may be designed to be helpful and friendly but are regarded by some as potentially reducing people’s desire and ability to communicate with other humans (Turkle 2011). In contrast, the more practical robot assistants, such as robotic floor cleaners, may be relatively good at what they do but have nowhere near the capabilities of the household robots of people’s dreams, as idealized in the cases of Rosie the robot from *The Jetsons* and *The Bicentennial Man* (Asimov 1990: 245–290).

Alongside the good robots, bad and dangerous robots, such as those in *Robocop* and *The Terminator*, abound in fiction. Indeed, the first use of the term ‘robot’ in Karel Capek’s 1920 play *R.U.R.* (Rossum’s *Universal Robots*) involved the story of a robot uprising resulting in human extinction. In real life, many news reports focus on a whole range of perceived dangers of robotic technology. From concerns about robots designed to revolutionize the field of war as autonomous killing machines, to robots that will take people’s jobs and cause unemployment, poverty and societal collapse; these and other fears regularly shape how robots, artificial intelligence and other automation technologies are discussed in news reports and other media.

In somewhat similar ways to robots, video games also have an image problem, ‘the alleged dehumanizing potentials of “too much” gaming, often mobilizing adjacent psychologic concerns around addiction and desensitization’ (Taylor & Elam 2018: 246). Part of the reason for this type of response to video games is that, as is the case for robots, people have a limited understanding of how they work, their variety of forms, and what they can help humans achieve. The breadth of experiences and interactions that games and robots make possible are not fully understood or discussed in public forums, beyond niche scholarship and publications dedicated to games or to analyses of human–computer or human–robot interaction. Even among players themselves, the immense range of video games is such that no single person can play them all. As a result, many people question the role video games (or at least the video games they don’t play) can or should hold in society, and are intimidated by them.

In addition, one game can represent a commitment of numerous hours, weeks or months of play. While game-players might regard this as time well spent, in an increasingly time-poor society, adults investing significant time playing can also be regarded suspiciously. Just as some robots, most often those designed as companions or for entertainment, are seen as a waste of time, gamers are often asked what they are achieving through their engagement with video games. Time spent with machines for anything other than obviously productive purposes, in particular for adults, can be regarded as wasteful because it detracts from opportunities for work, social development and human bonding.

Coming to the defence of robots or video games often involves making the case that interacting with them is beneficial in a practical, economic sense. For instance, a *Financial Times* article titled ‘The rise of the robots can help human workers’ argues that collaboration between humans and robots will increase commercial
productivity, since ‘automated systems are far better at dangerous, and repetitive or time-intensive work’, while ‘humans perform well on abstract and creative tasks’, thus ‘putting them together can harness the advantages of both’ (FT 2019). People are also increasingly encouraged to think about how gaming might provide an advantage in the workplace, since ‘there are plenty of soft skills that gamers can utilize in a professional setting, such as teamwork, problem solving and strategic planning’ (Gardner, quoted in Molloy 2019). Playing video games is now a potentially lucrative career for those who compete professionally and those who stream online, so this has become another key rebuttal against the idea that games are a waste of time (Dollinger 2018). Ultimately though, these discussions just serve to reinforce the idea that human–machine interactions must be economically productive or result in concrete practical outcomes in order to have value within society. It is our hope that focusing on play and playful interactions in this paper will help to uncover other important models of value, while disrupting the idea that human–machine communication must always be economically productive and therefore often work-related.

Linking Characteristics of Play with a Framework to Analyse Human–Machine Interaction

Writing before the advent of digital technologies, the mainstream adoption of video games and the mass availability of robots, French sociologist Roger Caillois identifies six core characteristics of play, arguing it is: free, separate, uncertain, unproductive, governed by rules and involves make-believe (1961). The characteristics Caillois pinpoints are particularly relevant to this paper, because they are open enough to be applied to the ways people play with video games and with robots, while raising interesting questions. They can also be integrated with the framework this paper adopts to analyse the human–machine communication that takes place during video game play and in interactions with robots, which focuses on three elements: encounter, addressing the initial and subsequent meetings of humans and machines as communicators that can respond to one another; story, attending to the stories that emerge as humans and machines interact, but also the narratives that arise prior to, around and after those interactions; and dance, acknowledging how communication between humans and machines is an embodied and dynamic process of overlapping interchanges through which meaning emerges (Sandry 2015; Sandry 2018). Beginning with a video game case study, Flower, we then consider the links between robots and video games, before arriving at a deeper discussion of robots themselves, in particular Shimon the robot marimba player.

Playful Encounters With and Within Flower

Caillois (1961) describes play as free, something which people are not forced or obliged to do. People are rarely forced or obliged to play a video game. The choice to play Flower, for example, is as free in Caillois’ terms as any other form of play. The cross-platform availability of Flower means that humans may use a computer, console, smartphone or tablet in order to play. The experience is accessible, requiring little existing knowledge or skill with video games. Flower is also free in the sense that it eschews design features associated with compulsive behaviour. The game does not provide competitive elements such as point-scoring, incremental rewards, social network access or a leader board, which are all associated with ‘addictive’ gaming habits that complicate the notion of free play (Harrigan et al. 2010; Albarrán-Torres 2016). Instead, Flower invites the player on an individual, free-flowing journey motivated by playful interactions within a vibrant and responsive virtual world.

Encounters between human and machine are what bring Flower into existence. As Laura Ermi and Frans Mäyrä point out, ‘the essence of a game is rooted in its interactive nature, and there is no game without a player’ (2005). When playing Flower, and most other video games, these encounters can be thought of as occurring at multiple levels. One level is the player’s interaction with the device through which they are playing the game, using this in ways that may be familiar or new. Another level involves interactions with elements of the game itself. Game studies scholars Mark J. P. Wolf and Bernard Perron have helped define these forms of interaction as diegetic activity (what the player’s avatar does as a result of player activity) and extradiegetic activity (what the player is physically doing to achieve a certain result) (2003: 15). In Flower, the initial diegetic encounter is with a potted plant on a windowsill (mediated by an instruction to click and hold). This begins the game with a video introduction (discussed below in terms of the way it offers a story frame) before the player is then taken to a new environment, where a second encounter with a flower growing in a field of long grass involves picking a first petal that shows the trajectory of the gameplay that follows, catching the breeze and positioning the player as the wind (see Figure 1).

These first two encounters are required for the game to begin, and therefore the player is guided to complete these tasks, but after this the play is relatively free. You can choose to interact with more flowers and
other game elements, such as windmills, or can simply explore the game space. *Flower* contains no assistive text during play, instead providing lures and suggesting points of interest via visual and sonic cues. Changes in the virtual space are indicated by fluctuating light and shadow, changing colours, and ripples of movement across and through different surfaces. Boundaries, when encountered, are soft; the edges of each map are experienced as a gust of opposing breeze that spins the player around and back into the game space. The result is engaging interaction rather than goal-oriented progression, at least initially.

The environment in the first level of *Flower* is a grassy landscape and, while this may be familiar, the mode of travel is not. Few modes of physical or virtual navigation enable humans to play the role of a gusting wind, disembodied yet impactful. This difference interrupts the player’s sense of familiarity, separating the experience of landscape within *Flower* from everyday life. The player’s movements within the game are fluid and the results are uncertain, depending moment to moment on player engagement and choices. While some video games shape people’s encounters with them in clear-cut ways, another core tenet of play identified by Caillois (1961), the rules, are limited within *Flower* where play is broadly governed only by what is possible within the game space and the affordances of the physical interface chosen. This further accentuates the freedom of play in this game and an uncertainty over its end point.

The sense of freedom, separation and uncertainty that Caillois identifies as key characteristics of play, drawn out in the discussion of *Flower* above, are also key elements of some theories about encounters between self and other. Notably, Emmanuel Levinas’ conception of ‘the face to face’ (1969), for example, emphasizes the separation of self from other that remains even as they are drawn into the proximity of an encounter. Freedom and uncertainty are also a feature of Levinasian encounters, within which the self is positioned as responding to the call of the other without any expectation of continued engagement, the other being free to reciprocate or not (Pinchevski 2005). There is no knowing where the interaction might lead. Levinas himself only considered humans as participants in this type of encounter, but more recently scholars have begun to consider how this idea can be extended to human–machine interactions (Gunkel 2012; Sandry 2015). The idea of otherness extended to robots is discussed later in this paper. In relation to video games, game-players are likely to have a strong sense of self-efficacy in their decision to play, thus it might seem logical to position the human player as the self and the machine, the game and its elements, as other; however, in many ways the reverse is true. Thinking through the example of player–game encounter and interaction in *Flower* highlights the initial choice of the player to play at all, as well as their continued freedom (once they have completed the initial two required interactions with flowers) to engage with the game in whatever way they please. It is the game, the machine, that must respond since it is programmed to do so, while the player can opt out of continued interaction at any time (although enjoyment and the desire to explore further may mean this a difficult decision to make).
Stories Within and About Flower

All games have a trajectory of sorts, whether it is presented in a manner akin to linear narrative or not. We start in one situation and move towards something else through exploration and discovery. 'These things pull us in', notes video game scholar Jesper Juul: 'Video games are like stories, like music, like singing a song: you want to finish the song on the final note' (2010: 4). When it comes to storytelling, games offer a unique opportunity to express and generate meaningful encounters via compelling human–machine interactions. Henry Jenkins famously argued that 'game designers don’t simply tell stories; they design worlds and sculpt spaces' (2004: 121). These spaces are embedded with potential encounters that enable the player to unravel the logic, rules, and possibilities of the game, which operates as an (at times oblique) storytelling partner. Interpreting this 'narrative architecture' is one of the key pleasures of video game play as it unfolds through ongoing interactions between player and machine (2004: 118).

People’s experience of Flower is embedded within a story frame to some extent from the beginning. The introductory video sequence shows a noisy, busy, dark and somewhat forbidding cityscape (see Figure 2), from which the player is transported into the idyllic fields of green grass and flowers. From here the story is felt, heard and seen, rather than told. Each stage of the game draws you back closer to a city, slowly introducing encounters with human structures such as hay bales, electric wires, fences and roads across the grassy fields. As the game’s locations shift from idyllic natural fields into increasingly darkened urban landscapes, the player is given the opportunity to dance their petals across the ruins and return life to these broken structures.

If players read the preview on Flower in Apple’s App Store, they can learn more from the story shared by the game’s designer about the reasons behind Flower’s development. Having moved from the city of Shanghai to California, game designer Jenova Chen was awestruck by the vibrant green spaces with which he found himself surrounded. As Chen explains, ‘I felt like a man growing up in the desert, who travelled to the sea for the first time’ (‘Flower: App Store Preview’ n.d.). He wanted to share his sense of wonder at this new environment with his family and friends back in Shanghai and decided the best way to do this was through the creation of a game. The video and Chen’s explanation could therefore be understood to emphasize the importance of people spending time in natural environments, away from the cities in which they might spend most of their time working and living.

While the introductory video and App Store text frame Flower as being about escaping the city to explore fields of grass and flowers, the game itself has been described as an ‘interactive poem’ (Govan 2009). The game contains no characters or story within itself; instead players take on the role of the wind that collects and then moves an increasing number of flower petals across a wide landscape which slowly transforms in response to the player’s actions. As Caillois (1961) identifies is a feature of all play, people’s engagement with
Flower involves an element of ‘make-believe’ as they become first-person players without bodies. Interacting with Flower immerses people in an idealized natural environment, but this is done in a way that erases the human body, allowing players to become enmeshed in that environment in a way that is not possible in the physical world. The game allows players to imaginatively project themselves beyond their normal corporeal limitations, through an experience of another way of moving through, and being in, a world. Alongside this, considering the interaction between players, the interface and the game as a Levinasian encounter acts as a reminder that even as people feel embedded in the game environment through play, they also remain a part of the physical world on their side of the interface.

Embodiment and the Dance of Interaction With Flower

The smartphone, tablet and console interface for Flower uses the gyroscopic sensor in those devices. Tilting the device controls direction, while pressing any button on the console, or anywhere in the screen, creates forward movement, that is, makes the wind blow. Play on a computer is slightly different, relying on the more broadly familiar computer interface of keyboard and mouse. One of the authors had the opportunity to play on a computer first, before using a console. As might be expected, the console interface, with its gyroscopic control that allowed tilting for a change of direction, led to a more embodied experience of being the wind than key presses. A sense of the rules Caillois (1961) identifies as integral to games are therefore defined by the physical interface used to play, its affordances and limitations, with the computer interface making play feel slightly more restricted than on a console or smart device.

Within the game itself, the idea of rules is less a feature. As discussed above, while playing Flower, the player has no body but the wind. The game therefore allows people to merge with the natural world of the game and identify with its various elements through playful interaction. As players ‘blow’ their collection of petals along, they watch the petals dance through the landscape and explore the virtual space with them. This exploration can be completely free, or players can choose to focus on locating more blossoming flowers, adding petals to their collection throughout their journey (see Figure 3).

As an example of spatial play, Flower draws attention to the way Caillois specifies play as a separate activity distinct from the spaces of everyday life. Building on the work of Henri Lefebvre, Bernadette Flynn argues that:

In games ... it is our encounter with transforming geography that provides the basis for a particular type of spatial pleasure. Rather than being related to plot points, character identification or emergent story lines associated with narrative pleasure, spatial pleasure is grounded in immersive aesthetics, maps, tours, modes of navigation and geometric landscapes (2004: 54).

Figure 3: Screenshot from Flower. Copyright content reproduced here under Fair Use.
Such environments offer new ways of encountering others and spaces, expanding and complicating concepts of self. To play *Flower*, maybe in particular with a smartphone, tablet or console, the player is invited to project their physical presence into the virtual environment. Their body is invisible (it is the wind), with the petals signalling where they are and the direction in which they are moving. The link between the physical and the digital is blurred and the boundary between the human player and the game environment is broken down through the game’s immersive quality.

Although the simplicity of the game’s controls might have been designed to make the game easily accessible to many players, one of this paper’s authors found the motion sickness invoked by the game meant they could only play for a few seconds at a time. The immersive nature of the dipping and swirling dance of motion in *Flower* likely draws other people further into playing the game. Analysed from the Levinasian perspective introduced above, it is evident that the difference between being embodied as the wind inside the game, and being a physical human body outside the game, highlights the separation of players and game, and player and interface, even as they are brought into the proximity of play (or attempts to play). This distance–proximity juxtaposition is part of what makes playing the game so difficult for some and so enjoyable for others, depending on the way their hands, eyes and inner ears coordinate understanding of movement within the game versus movement in the physical world.

The game has no explicitly defined goals, although playing through all of its levels does work towards an end point of lighting up the dark city. In many ways, engagement with this game would seem to be an archetypal example of Caillois’ definition of play as unproductive (1961). Yet, this game could be regarded as having a number of outcomes for players. Through the dance of play, whatever interface is being used, the player learns to interact with their computer or device in ways that may be new to them, or that hone existing skills. *Flower* has been one of the authors’ first interactions with gyroscopic control through a smart device. Although to an extent that particular interaction has been unsuccessful, causing their nauseous response, for other people it might open up a new form of interaction with their device, as well as the virtual environment it contains for this game. Many video games require players to learn the precision control of a game and machine interface, with ‘a growing body of psychological research regarding the “positive” benefits of intensive play’, such as improvements in attention levels, memory, spatial awareness and problem-solving ability (Taylor & Elam 2018: 246). Gameplay through *Flower* reportedly also has broader outcomes, with feedback from people who say that playing ‘the game has helped them push through illness, or come to peace with an emotional time such as bereavement’ (*Flower: App Store Preview* n.d.). More generally, reviews of *Flower* describe playing the game as ‘an experience that is unique and enthralling’ (Govan 2009), ‘a relaxing, calming, and curiously moving experience that has the power to change the way you look at the outside world’ (Michalik 2013). Even without considering the productive outcomes of such gameplay, it therefore seems important to value the potential for relaxation and escape from the everyday, alongside offering a new way to consider the outside world on people’s return to their physical surroundings.

**Linking and Theorising Human Interactions with Video Games and Robots**

Robots and video games are closer bedfellows than you might think. In fact, the last fifty years has seen ‘a synergistic evolution of robotic and video game–like programming environments’ (Lahey et al. 2008). The design of some human–robot interfaces are inspired by video games (McLurkin et al. 2006), while virtual environments created to simulate human–robot communication also make use of video game technologies. ‘Modern computer games share much in common with modern mobile robot simulators’, argues one robotics team that explains it uses ‘computer games technology in three areas of [their] simulator: 3D graphics, physics simulation, and networking’ (Faust et al. 2006). The convergence is not only technological but thematic and physical, extending into more immersive and playful experiences of human–machine communication. Virtual Reality (VR) games are being designed for an array of affordable devices including the HTC Vive, Oculus Rift, PlayStation VR, and Windows Mixed Reality headset. As Lahey, Burleson, Jensen, Freed and Lu highlight, the increasingly physical nature of video games, combined with the reducing cost of social robots, are resulting in significant developments, as ‘concurrent advances are creating new synergies for the advancement of video games and novel human–robot interactions (HRI) through play and learning experiences’ (Lahey et al. 2008). First-person VR games such as *Stormland* (2019) enable the player to identify and act as a robot within a game space that is simultaneously virtual and physical. Other games are being designed that facilitate direct interactions between humans and robots. The *Emotional Robocoaster* is one example: a game played between a human and a robot, ‘geared towards creating a playful human–robot interaction that allows players to explore their emotions through introspection’ (Reynolds-Cuellar &
Breazeal 2017). Play is repeatedly being used to cross the boundary between human and machine, inviting communication that is less about productivity and more about forming engaging connections.

Tim Dant argues that ‘the form of social being that results from the collaboration of human and machine has attracted the term ‘cyborg’’, but ‘the idea of the cyborg tends to fix and reify the human–machine relation (2004: 62), the same being true of the term ‘hybrid’. Cyborgs and hybrids, being permanent combinations, are not good models for human–machine relations, whether they develop as people play video games (discussed above), or when they interact with robots (see below). Instead, in a similar move to Dant who is writing about driver–car relations, it may be better to think of the human–machine relations discussed in this paper as ‘assemblages’ (a term that draws on the work of Deleuze and Guattari), since the humans and machines involved are not only separable, but also can be ‘endlessly re-formed, or re-assembled’ from different components (2004: 62). In these relations, human and machine components do not merge and combine into a single entity, losing their own individual characters and identities; but rather, they mesh, becoming engaged such that they fit with one another and work together in harmony.

In the context of video game play, Taylor and Elam explain that ‘getting good at games involves an intimate choreography between system (input, mechanics, hardware) and player’, involving not only the automation of, for example, ‘muscle memory’ and ‘the direction of attention’, but also the drawing together of ‘human and non-human forces’ to form an assemblage (2018: 244). Players increase their skill in using interfaces and the games they play with those interfaces through hours of practice. Some players reach a level of expertise their opponents feel is not humanly possible, such that they are accused of using, or even of being, bots or software robots (2018: 244). Indeed, the best video game players are often understood to have become machinelike in order to mesh more effectively with the interface and the game, such that ‘discourses of expert players as automatons and/or machines abound’ (Taylor & Elam 2018: 245). More generally, as computers and robots have become more regularly a feature of their everyday lives, many people have developed a propensity to compare themselves to machines and use machine-related terminology to discuss aspects of human life (Turkle 2005).

In contrast, for human–robot interactions it is more likely the onus will be on the machine to become humanlike. For example, Cynthia Breazeal, the creator of one of the first social or ‘sociable’ robots, Kismet, states that interacting with such a robot should be ‘like interacting with another person’, a trait supported by creating robots that are ‘socially intelligent in a humanlike way’ (2002: 1). However, just as Elam and Taylor are not interested in ‘ascribing machinic qualities to human players’, but rather argue that assessments of expert players of this kind are only possible when ‘organic and technical components act in seamless concert’ to result in ‘hyper-efficient gameplay’ (2018: 245), this paper also suggests that robots need not be entirely humanlike in order to support effective human–robot interactions. In contrast to Elam and Taylor the argument here is that human–machine interactions that are playful, interesting and enjoyable to experience are valuable in themselves without needing to be ‘hyper-efficient’. Having begun to argue this already in relation to Flower, this paper now considers examples of human–robot interaction.

**Encountering Shimon, the Marimba-playing Robot**

The second case study analyses people’s interactions with autonomous robot musicians. The focus is mainly on Shimon, the marimba-playing robot, which is not directly under human control, sensing and responding to human musicians for itself. The sense of play in this second example is rather different from that discussed in relation to Flower, as here people are playing music as opposed to a game. Nevertheless, the way in which people perform with Shimon invokes Caillois’ understanding of play (1961), in terms of an interactive co-regulated process that can also be analysed in relation to ideas of encounter, story and dance. Unlike interactions with machines in the form of playing video games—where communication is with and mediated through interfaces such as computers, smart devices, consoles and headsets—interactions with physical robots such as Shimon, and Haile (a drum-playing robot developed at the same robotics laboratory), rely upon various modes of communication between human and machine that take place within a shared physical space.

While playing music might seem to be less about freedom and more about adherence to a score (often one that essentially provides a very strict set of rules), Shimon has been designed to play improvisational jazz alongside human musicians in an ensemble, while Haile takes part in improvised drumming performances. Not only can people choose whether to play with these robots or not, but the musical form that results is free-flowing, emerging from the interaction between humans and the robots. At times, play with Shimon and Haile does use some standard forms of jazz performance, such as call-and-response where
musicians play in turn, one in response to the other, but the exact course much of the performance takes is not precisely planned or programmed, with the exception of a concluding section where all the musicians, humans and robot, play in unison to round things off. Caillois’ suggestion that play is free can therefore be seen in human interactions with these robots (1961). Alongside this idea of free play though, music as a form does incorporate overarching rules of rhythm (so players are in time with one another) and key (to keep the music on track melodically), so this, as well as call-and-response and other specific patterns used in jazz performance, is in line with the idea of governing rules also being essential to the course of play (Caillois 1961).

Neither Shimon nor Haile use a voice interface and, while both communicate through the music played, Shimon in particular has been developed to use nonverbal communication. Shimon communicates with musicians and audiences through gaze-direction and body movements, as well as the music itself. These modes of nonverbal communication are a feature of all music ensembles, including those that consist only of human performers. Although Shimon does have some bodily characteristics that can be compared with that of a human, such as arms, a neck and ‘socially expressive head’ (Hoffman & Weinberg 2010: 3099), this robot is still overtly non-humanlike (see Figure 4).

Shimon’s body is very closely integrated with its marimba. Robot and instrument appear inseparable and are always encountered together by people in a way that might emphasize the robot’s non-humanness. In addition, unlike humans who often hold four mallets, two in each of their hands, Shimon has four arms each controlling a pair of mallets, making eight in total (Hoffman & Weinberg 2010: 3098–3099). This allows it to play the marimba in a uniquely nonhuman way that is designed to demonstrate accomplished ‘robotic musicianship’ (Hoffman & Weinberg 2010: 3098). Although Shimon’s appearance is machinelike, its head and neck are constructed to allow a ‘unique organic movement’ (Hoffman & Weinberg 2010: 3099). The robot’s head has a single video-camera eye with a shutter that opens and closes to help convey emotional state and liveliness throughout a performance (Hoffman & Weinberg 2010: 3099). In Caillois’ terms, playing music with Shimon (and even being in an audience watching a performance with Shimon) involves an element of ‘make-believe’ (1961). The robot is not alive in anything like the same sense as a human musician, and yet its body movements can be interpreted as lively and attentive when it is taking an active role in a music ensemble.

People’s initial encounter with Shimon is therefore with a machinelike robot that is clearly different from them, but with behaviours that nonetheless support a level of familiarity and understanding when engaged in playing music. Even as people are brought into proximity with the robot to play ‘face to face’ (Levinas 1969), the separation between human and non-human musician remains clear, as does the experience of playing with a robot as opposed to another person.

Figure 4: Screenshot of Shimon from Robot Music video on YouTube (The Kennedy Center 2015). Copyright content reproduced here under Fair Use.
How Narratives Shape, and are Shaped by, Understandings of Shimon

Before working with robots, Weinberg experimented with pushing his musical ability by developing software to generate music computationally, but he realized software that improvises was lacking something (PBA30 2015). He argues that developing robots to play acoustic instruments—which produce a rich, emotional and expressive sound that is difficult to create computationally—is more challenging and satisfying (PBA30 2015). Although Shimon is encountered physically as fully integrated with the marimba it plays, with the player and instrument potentially seen as one entity, Weinberg’s explanation of the decision to create robots situates Shimon as a musician separate from the marimba it plays.

The idea of proximity and separation in relations with Shimon is further emphasized by the way Weinberg says his core idea in creating robotic musicians is to ‘create robots that listen like humans but improvise like machines’ (Weinberg 2015). Shimon listens like a human in order to match ‘the human’s playing style, tempo, and harmony in real time’, but when it responds the robot extends what the human has played to contribute ‘its own musical phrases and ideas’, possible to play with its unique non-humanlike capabilities, to which the human goes on to respond in their turn (Hoffman & Weinberg 2010: 3099). The result is ‘a back-and-forth inspiration between the human and robot’ (Hoffman & Weinberg 2010: 3099), from which an improvised piece of jazz music emerges. When talking, Weinberg’s statements emphasize Shimon’s agency as a robotic musician that does not just ‘come up with its own idea’ of where the music could go next, but must also ‘think about its own body’, ‘its own capabilities’ and what it can play (PBA30 2015).

The success of Weinberg’s idea is evident in comments made by Greg Hendler, one of the musicians who has played with Shimon. Hendler explains that in many ways the robot reacts as you might expect a human would (PBA30 2015). Shimon’s behaviour communicates that it ‘can hear what you’re doing’, such that human musicians feel they can ‘bounce’ around musical ideas with the robot (PBA30 2015). However, Hendler also notes that ‘some things are more rigid’, since it is impossible to talk to the robot ‘other than through music or changing the programming’ (PBA30 2015). This experience means that, as encounters with Shimon develop over time, people become aware they are playing with a machine that is humanlike in some ways and non-humanlike in others. In part, as Hendler’s comments suggest, the experience of communicating with this robot as a musician involves an understanding of music as a language, the flow of rhythm, melody and harmony operating as a narrative that both supports and emerges from human–robot interaction. However, embodied communication is also essential in enabling the ensemble to play and perform, a concept this paper analyses in more detail in a later section.

Weinberg explains that people who hear about or see Shimon in action don’t always understand his goal. They ask whether he is ‘evil’, since he is taking ‘the one thing that is so human, music, and even that you are going to bring to robots and make them even better than us’ (PBA30 2015). This means that Weinberg regularly needs to explain that his robots never play by themselves, instead they improvise in their uniquely machine-like ways ‘based on a seed’ provided by a person (PBA30 2015). Rather than being created to replace human musicians, Shimon is designed to inspire humans ‘to play music and think about music in new ways’ (PBA30 2015). For Weinberg, the design and development of robot musicians is a creative outlet; a way to explore new ways to develop and perform music.

People’s responses to Shimon also draw attention to whether research into robotic musicianship can and should be productive or practically useful. A comment below a video of one of Weinberg’s talks (2015) on YouTube encapsulates the problem, noting Weinberg’s projects seem ‘fun’, but immediately questioning the source of funding (and suggesting they want their money back if public funding was used). It can therefore be argued that the concept of play in robotic musicianship does not just relate to the playing of music itself but also to questions raised in relation to the idea of playing with robots without very clear practical goals in mind. In a similar way to play with video games, play with robots, even play that creates music, can be regarded as unproductive, an idea that Caillois (1961) embraces in his conception of play, without judging this lack of productivity to be a waste of time or money. The results of jamming with Shimon are also uncertain, like play is more generally (Caillois 1961), and to some extent whether you regard it as productive or unproductive may well depend, at least in part, on your like or dislike for jazz ensembles that play improvised music.

Other writing about Shimon also suggests that discussion on the difficulty of justifying money spent on research into robotic musicianship is more widespread than in comments on YouTube videos. Bretan and Weinberg reinforce the practical gains to be made from this type of research by suggesting that breakthroughs in ‘timing, anticipation, expression, mechanical dexterity, and social interaction’ that are important when creating a robot musician also ‘have numerous other functions in science’ (2016: 100). In particular,
the development of human–robot interaction (HRI) algorithms ‘that enable anticipation or synchronization in a musical context’ is ‘useful for other HRI scenarios where accurate timing is required’ (Bretan & Weinberg 2016: 102). As well as assisting in the development of robots for activities beyond music, they also argue that ‘by building and designing robotic musicians, scholars can better understand the sophisticated interactions between the cognitive and physical processes in human music making’ (Bretan & Weinberg 2016: 102), an idea that is sometimes extended to suggest the use of robots to better understand how humans operate in the world more generally.

Alongside this, Bretan and Weinberg also emphasize the benefits of building a robot to play ‘music that humans could never create by themselves’, inspiring people ‘to explore new and creative musical experiences, invent new genres, expand virtuosity, and bring musical expression and creativity to uncharted domains’ (2016: 102). As might be expected, these potential gains rely on the machinelike nature of robotic musicians, able to use ‘compositional and improvisational algorithms that humans cannot process in timely manner’ as well as allowing designers to explore ‘mechanical sound production capabilities that humans do not possess (from speed to timbre control)’ (Bretan & Weinberg 2016: 102).

Although these ways of justifying work with robot musicians focus on locating productive economic and scientific outcomes for such research, videos of ensembles with Haile and Shimon clearly show how much enjoyment human musicians gain from playing music with these robots. In this paper, we argue that this pleasure is something that should be valued in itself, so in the next section we consider the embodied communication between human and robot where this pleasure and enjoyment in the process of making music together becomes most evident.

Ensemble–playing as a Dance of Interaction

As Weinberg explains, most of his experience as a jazz musician is of playing in improvisational groups with other people, able to alter what he does to build on the play of other musicians in real time (Weinberg 2015). This interplay between humans is enabled through the exchange of auditory cues, the music, but also visual cues, body language and gaze direction, something that Weinberg has built into interactions with Shimon. People’s jazz improvisations with this robot are not only enabled by its ability to play the marimba (Hoffman & Weinberg 2011: 233), but also its ability to communicate ‘gesturally’ (Hoffman & Ju 2014: 98). In particular, Shimon’s head ‘bobs’ to signal its ‘internal beat’, a movement that helps the ensemble’s timing. Shimon responds to the beat set by the music humans contribute, but can also decide to change the tempo, indicating this by a change in its head movement as well as the music it plays. Shimon’s body also allows it to turn to the marimba while taking the lead and focusing on its own play, turning towards a person ‘to signal that it expects the musician to play next’ or take the lead (Hoffman & Ju 2014: pp. 102–103). Shimon’s musicianship is therefore reliant on a ‘choreography of movements’ in support of its improvisational performances with human musicians (Hoffman and Weinberg 2011: 234).

Alongside the language of the music itself and the frameworks that drive particular jazz improvisations, such as call-and-response, playing in an ensemble with Shimon clearly relies on ‘embodied communication’, a type of communication Donna Haraway suggests ‘is more like a dance than a word’ (2008: 26). As Stuart Shanker and Barbara King note, communication framed by the ‘dance metaphor’ involves ‘co-regulated interactions and the emergence of creative communicative behaviors’ within a ‘dynamic system’ (2002: 605): a system of just the sort encountered in an improvisational music ensemble, where a constant exchange occurs between musicians as they play.

Not only does the idea of embodied communication as a dance fit well with the understanding of playing music with Shimon discussed in this paper, but it also brings to mind Andrew Pickering’s conception of ‘tuning in goal-oriented practice,’ which he argues takes the form of ‘a dance of agency’ (1995: 21). Although Pickering develops this idea in an effort to embed machine agency alongside human agency in scientific practice, as François Cooren notes it seems relevant to ask questions about agency, in particular the agency of non-human others, more broadly (2010: 21). Cooren argues that extending the attribution of agency to non-humans certainly ‘does not force us to abandon the differences between its various forms’ (Cooren, 2010: 4), thus recognising the continued presence of difference between human and non-human agency. Instead, he suggests that ‘it just invites us to pay attention to what is active (agissant in French) in a given situation’ (Cooren, 2010: 4). From this perspective, whenever one can identify someone who or something that makes a difference, whether in terms of activity or performance, there is action and agency (2010: 20).

A key aspect to note, which is particularly well illustrated in discussions of video game play as well as when humans play music with Shimon, is that ‘attributing agency to materiality and machine does not amount to
dispossessing humans from their strategies, their intentionality, their goal-oriented practices, or even their wit’ (Cooren 2010: 21). Instead, the recognition ‘that action and agency are not human beings’ privileges’ allows analyses to be decentered, to ‘show that people are acted upon as much as they act’ (Cooren, 2010: 22). This draws attention to the importance of all sides of the relation in their different ways, whether one is considering video game play, and the player–interface–game assemblage, or the human–machine assemblage that forms in improvisation ensembles where Shimon plays music with humans.

Videos of human musicians and Shimon interacting demonstrate how embodied communication allows human and musician to mesh into an assemblage, such that they can improvise in response to one another and play closely together. What is also evident is the enjoyment felt in this play, as well as a recognition of the skill and abilities of the robot, of its activity in the world and therefore a sense of its agency. For example, Weinberg points out the moment when Guy Hoffman, playing piano alongside Shimon, looks at the robot as if ‘his kid did something nice’ when it responds particularly well to the dynamics and expression of the piano’s last phrase (2015). Similar responses are also seen when professional darbuka drum players, presumably unused to playing drums with a robot, are pleasantly surprised by Haile’s drumming, smiling and nodding as they appreciate how the robot’s response to their beat builds a new rhythm (Weinberg 2015).

These momentary expressions in video footage are direct responses to the robot’s actions, rather than a part of keeping the ensemble coordinated, and are easy to overlook. However they are reminders of the pure enjoyment that many people gain from playing music with others, whether human or robot, judging time taken for rehearsal and performance as time well spent. They also highlight how surprising these machines can be, giving a sense of how well the machine’s agency can be read by humans in interaction, and how important that machine’s activity can be in shaping the course of the interaction as it develops.

**Conclusion**

This paper, through its analysis of interactions with a video game and two robotic musicians, helps to break down generic ideas of these types of machine as powerfully frightening and dangerous with the potential to harm human lives. In addition, rather than seeing time spent in interaction with machine-others as lost time, distracting people from more important human–human contact, we argue that human–machine interactions are valuable and rewarding in their own right, in part because of the differences between the humans and machines involved.

Shimon has not been created to replace a human musician, but rather to show the unique possibilities of a robot musician, including the ways that such a machine can spur people on to make novel musical decisions in improvisation. Play and communication with other social robots could be regarded in a similar way, as providing new interaction possibilities, not as replacing human–human interaction and taking time from the pursuit of human social relations. Furthermore, as demonstrated in play with Shimon and Haile above, the pleasure that arises when playing music with robotic musicians, as well as when interacting with social robots more broadly, should also not be ignored. The playing of video games can also provide many positive outcomes for people. Time spent playing need not be time lost to other more productive pursuits; rather, it can enable people to learn new coordination and interface skills. More importantly, playing games allows people to escape for a time from their worries, gaining satisfaction from gameplay, and relaxing, the latter effect likely being particularly evident when playing low-intensity games such as *Flower*.

Although human–machine relations have often been characterized as cyborg or hybrid, we suggest that the idea of human–machine assemblages is more useful, whether those assemblages consist of players, interfaces and games, or humans and robots. There is no need for humans and machines to merge in ways that belie or erase specific differences in order to drive successful interaction. Instead, even as they are brought together, whether in video game play or to play music, humans and machines remain separate from one another, with the potential to bring their own particular attributes to the relation. The human–machine assemblages that emerge in any given spatiotemporal context are unique, with the potential to support joyful boundary-crossing experiences unattainable in other ways. As opposed to emphasising the need for humans to become more machinelike in their abilities (seen in discourse around expertise in video game play) or the need for robots to be humanlike in order to support effective interactions with people (seen in discourse around the creation of social robots), this paper argues that it is the relation that builds between humans and machines, even as their differences are acknowledged and retained, that is valuable.

The paper analysed human interactions with video games and with robots in terms of initial encounters, stories around and within the interactions, and communication as an embodied dance, alongside a broadly defined understanding of play. This supported a consideration of the complexities of people’s interactions
with games and with robots. In particular, human–machine interactions involve not only encounters with machines and the worlds they contain or inhabit, but also the stories that are told around them. In addition, human–machine interactions are rarely wholly language based, whether using text or speech; but rather, are embodied. Communication occurs with and through physical interfaces for games and also through co-regulated movement in physical spaces with robots.

Any encounter with a machine that is mediated by playfulness and curiosity is likely to have a very different trajectory from an encounter dominated by fear and anxiety. Taking the idea of freedom more broadly though, playing freely is a source of joy, surprise, amusement, and exhilaration; sensations that engage us and have the potential to bypass fear, prejudice, and other boundaries that might prevent interaction and connection. Play, whether with video games or with robotic musicians, has the potential to broaden our horizons, while an extension of arguments about the value of difference in human–machine interactions may also be a useful dynamic to consider in pursuing joyful human–human encounters across cultural or racial boundaries.

**Competing Interests**
The authors acknowledge the financial assistance of Tencent Research in the preparation of this article.

**Author Information**

Author Profile for Sandry: http://zigzaggery.edublogs.org/.

**References**


PBA30. 2015. Shimon the robot (and friends). *This is Atlanta*. Available at https://www.youtube.com/watch?v=0dOn-EvSPUs [Last accessed 29 January 2020].


Taylor, N and Elam, J. 2018. 'People are robots, too': Expert gaming as autoplay. Journal of Gaming & Virtual Worlds, 10: 243–260. DOI: https://doi.org/10.1386/jgvw.10.3.243_1


