II. Virtual Selves

J. David Velleman

Foundations for Moral Relativism

Publisher: Open Book Publishers
Place of publication: Open Book Publishers
Year of publication: 2013
Published on OpenEdition Books: 1 June 2015
Serie: OBP collection
Electronic ISBN: 9782821854000

http://books.openedition.org

Electronic reference
II. Virtual Selves

Second Life

Most mornings, thousands of computer users log on to a virtual world called Second Life. Their computer screens show scenes of a nonexistent world, peopled by humanlike figures. Each user sees the world from the perspective of one of those figures, which is his avatar in the world and whose movements and utterances he controls through his computer keyboard and mouse. The other figures on his screen are being controlled by other users, all of whom witness one another’s avatars doing and saying whatever their owners make them do and say. Through their avatars, these users converse, buy and sell things, and have all sorts of other humanlike interactions. (You’d be surprised.)

If you saw the virtual world of Second Life on your computer screen without knowing how the images were generated, you would take yourself to be watching an animated cartoon in which human beings, presumably fictional, were portrayed as doing and saying various things. Once you learned about the mechanics of Second Life, you would interpret the doings onscreen very differently. You would attribute them to unknown but real human beings who own and control the avatars that you see. And indeed the typical participant in Second Life attributes to himself the actions apparently performed by his avatar. What a participant causes his avatar to do in the virtual environment, he will report as his doing. “I went up to the professor after class”, he may say, describing an encounter between a student-avatar that he controlled and an instructor-avatar controlled by someone else. In reality, the speaker went nowhere and encountered no one, since he was sitting alone at his computer all along.

DOI: 10.11647/OBP:0029.02
These self-attributions can be startling, given the differences between
avatars and their owners. A young female avatar may belong to an
older man, who may end up remarking, “For last night’s party, I chose a
tight dress to show off my figure.” An able-bodied avatar may belong to a
quadriplegic, who may then report, “I ran all the way.”

The obvious interpretation of such remarks is that they have the status
of make-believe. According to this interpretation, the animated figures
on the speaker’s computer screen are what Kendall L. Walton calls props
in the context of pretend-play.\(^1\) Such props include the dolls that children
rock as if they were babies, the chairs that they drive as if they were
cars, and so on. Just as a child might initiate a game of make-believe by
pointing to a doll and saying, “This is my baby”, the participant in Second
Life may be taken as having pointed to his avatar while saying, “This is
me.”

Obvious though it may be, however, this interpretation makes an
inference that I want to contest. Of course, when a participant says
“I got dressed” or “I ran”, whatever happened was not literally an act of
dressing or running, since the clothes and bodies required for such actions
do not exist. To this extent, the obvious interpretation is correct. But the
interpretation goes on to conclude that the agency of this human participant
is also fictional. When he claims to be the agent of the fictional actions that,
according to the fiction, his avatar can be seen to perform, the obvious
interpretation says that his claim must also be understood as fiction; I will
argue that it is literally true. In my view, the participant literally performs
fictional actions.\(^2\)

---

1 Kendall L. Walton, *Mimesis as Make-Believe: On the Foundations of the Representational Arts*
(Cambridge, MA: Harvard University Press, 1990). I should emphasize that the notion
of a prop is all that I mean to borrow from Walton for the purposes of this chapter. I am
not borrowing his theory of the representational arts.

2 This claim has the consequence that the semantics of our discourse about fiction cannot
be represented by a sentential operator such as ‘fictionally’. The fact that Shakespeare’s
play portrays the prince of Denmark murdering his uncle’s advisor is sometimes
expressed by philosophers of fiction with the statement “Fictionally, Hamlet murders
Polonius.” I will initially rely on this way of speaking merely as a matter of convenience.
In the end, it will turn out to be insufficient to express my claim about virtual worlds.
The claim that a human player performs a fictional action is not a claim to the effect that
something is fictionally true. Nor is it merely the claim that the human player makes
something fictionally true. It is the claim of a relation between an actual person and a
fictional action, a relation that breaches the boundary between the real and the fictional
worlds. Hence it does not consist in any purely literal or purely fictional truths nor in any
combination of the two.
II. Virtual Selves

The problem with the obvious interpretation of virtual worlds is that it exaggerates the similarities between those worlds and make-believe. In order to explore the differences, I will use the label ‘virtual play’ for games such as Second Life, and ‘pretend play’ or ‘make-believe’ for the sort of game typically played by children. Please note, however, that these labels are not meant to be precisely descriptive.

Pretend Play vs. Virtual Play

One respect in which virtual play differs from typical make-believe is that players cannot make stipulative additions or alterations to the fictional truths of the game. Their play is governed by a single, master fiction, namely, that they are viewing live images of a shared world. This fictional truth is given to the players, not invented by them, and it determines how all the other fictional truths will be generated in the course of their play: whatever is made to appear on the screens of participants will be what happens in the fictional world.

Aspects of determinateness

In pretend play, a child can say, “I’m a pirate, here is my ship, and you are my prisoner.” Five minutes later, the pirate ship can be turned into a spaceship, and the prisoner into an android, by another declaration of the same form. The participants in virtual worlds can make no such stipulations. In their

---

3 What I call virtual play involves some amount of pretending, and its characteristics can be found in games that are not virtual, strictly speaking, in that they do not depend on an information-based ontology. For example, fighting with paintball guns will turn out to be a case of what I call virtual play. In describing virtual play, however, I will confine my attention to the virtual-world participation that is typical of a deeply involved, fully committed player in a game such as Second Life, who spends a significant portion of his week “in world”, under the guise of a single, persisting avatar with whom he identifies (in some sense that remains to be explained). My aim is not to generalize about all participants in virtual worlds of any kind; it is merely to explore what is possible by way of action in virtual worlds, by focusing on the case in which action is most likely to occur. In describing pretend play or make-believe, I will speak of the simplest and most familiar examples of the genre, the spontaneous and unregimented imaginative play of young children. I will use these terms to label opposite ends of what is in fact a continuum of possible games, in which the make-believe and the virtual are variously combined.

4 This statement is not quite true of text-based multiuser domains in which a player makes his avatar act by entering a description of what it is doing. Even here, however, such statements are limited to actions that the player’s avatar is in a position to perform. Other
capacity as human participants in the game, they cannot say anything at all; they can speak only through their avatars. And by doing so, they can make true only the sorts of things that real people can make true by speaking. If a player wants a pirate ship, his avatar must build or buy one; if he wants a prisoner, his avatar must capture one; and he cannot turn his pirate ship into a spaceship unless his avatar carries out the necessary alterations.

A second difference between virtual worlds and the worlds of pretend play is their determinateness in proportion to the knowledge of the participants. What is true in a make-believe world includes only what the players have stipulated or enacted, plus what follows from those overt contributions; what is true in a virtual world is usually far more determinate than the players know or can infer.

Thus, when the children begin playing at pirates, the objects in their environment have no determinate roles in the fictional world of the game, and their characters have no determinate histories. If the children do not assign a fictional role to the coffee table, either explicitly or implicitly, then there is no fact of the matter as to what it stands for in the fiction. Usually, the players are on an equal footing as authors of the fiction, and so the facts of their fictional world are limited to what has been entered into the store of common knowledge among them, since secret stipulations would be pointless in a collaborative game.

By contrast, a virtual world has determinate features that outrun what is known to any of the players. Each player has to explore this world in order to learn what it is like, and he will then encounter others whose knowledge of the world is largely disjoint from his. The need to explore a virtual world interacts with the aforementioned necessity of instrumental action, since a player can explore the virtual world only by making his avatar explore it. He cannot learn about a part of the virtual world unless his avatar goes there. He sees only from the avatar’s perspective, and he cannot see around corners unless the avatar turns to look.5

These differences between virtual and make-believe worlds extend to the nature of a player’s actions. In either context, the behavior of an actual person makes it fictionally true that something is done by his

---

5 These descriptions are subject to a slight but significant qualification. In some virtual worlds, each player occupies a perspective slightly behind and above his avatar, so that the avatar’s body is within his field of view. I think it is not accidental that this perspective corresponds to one that is sometimes experienced in dreams.
II. Virtual Selves

counterpart, but what is made fictionally true by a player in make-believe is less determinate, and more dependent on stipulation, than what is made fictionally true by the player in a virtual world.

In the typical make-believe game of pirates, if one player pretends to stab another, there is no fact as to how much damage has been done until one of them makes the requisite stipulation or takes a relevant action, such as pretending to die. The difference between a graze and a fatal wound is not determined by the physical enactment of the blow. If the players fall to arguing over whether the victim is dead, they cannot examine the action for evidence; even a video replay would not settle the question. The players’ behavior was therefore insufficient to determine whether a killing occurred, and the indeterminacy must be resolved by discussion among them.

This indeterminacy runs in both directions. Not only is it indeterminate what action a player has fictionally performed by means of a particular bodily movement; it is also indeterminate what bodily movement a player must employ in order to perform a particular fictional action. What must a player do in order to climb the rigging of his pirate ship? There is no skill or method of climbing fictional ropes. The bodily means are underdetermined, precisely because so many different movements might be stipulated to constitute the desired action.

In virtual play, however, determinate manipulations of keyboard and mouse are required as a means of causing particular movements on the part of an avatar, and those movements have determinate consequences in the virtual world. In order to bring about what he intends in that world, a player must make his avatar behave in ways that are effective under the “natural” laws governing the world, and he can do so only by providing input that will bring about such behavior, given the design of the user interface.

Role opacity

Yet a third significant difference between virtual and pretend play lies in the relation between the players and their roles. This relation differs in what I will call its opacity or transparency.

In pretend play, the make-believe characters are impersonated by actual children who know one another and see one another playing their roles. What a child chooses to do as a make-believe pirate is attributed both to the pirate, as his action within the fiction, and to the child, as his contribution to the game. The role of pirate is consequently transparent: it allows the player
to show through. The transparency of the role even allows the player to emerge from it completely without any change of venue or medium. When the children start to argue about whether one pirate has killed the other, they implicitly lay down their fictional roles and argue as children: there is no suggestion that the pirates have decided to lay down their swords and “use their words” instead. But the children may be standing in the same places and speaking with the same voices as they did a moment earlier in their roles as pirates.

In virtual worlds, the actual players are usually unknown to one another: they interact only through their avatars. Even if the owners of different avatars know one another’s identities, those identities are not on display in the virtual world; the players don’t see one another’s faces as they would in pretend play. Hence their avatar-identities are opaque. There is no way for players to emerge from behind their avatars to speak or act as their actual selves. They can, of course, communicate with other players whose identities they know, but only in person or by e-mail or instant message or telephone, not in the venue or medium of the game.

Psychological engagement

These differences between virtual and pretend play produce one final difference, which involves the players’ psychological engagement with the fictional world of the game. In make-believe, a player is aware of his power to invent the objects and events of the fictional world, and this awareness affects his attitudes toward them. His cognitive attitudes must conform at any point to the actions and stipulations made thus far, but they are not constrained to the same extent as beliefs would be constrained by reality. Instead of being reality-tested, like beliefs, these cognitive attitudes are tested against the incomplete fiction of the game, into which they can introduce additional details and further developments just by representing them and being voiced as stipulations. Hence these attitudes are only partly like beliefs while also being partly like fantasies. Similarly, the player’s conative attitudes differ from the attitudes that he would have toward real

---

6 Although I noted earlier that paintball games qualify as virtual in my taxonomy, I am unsure whether they resemble online virtual games in this respect. Of course, the actual players are visible, unlike the actual players in a virtual world. But they are unable to set aside their fictional roles as combatants, since there are no “time outs” during which the fiction can be suspended. Hence their roles are transparent in some respects and opaque in others.
II. Virtual Selves

objects and events. A monster that he has made up, and is aware of being able to kill by means of further make-believe, does not frighten him as a real monster would.

In a virtual world, however, the players are aware of dealing with objects and events that, however fictional, are still not for them to conjure up or conjure away. These objects and events have the determinateness and recalcitrance characteristic of reality, and so the players tend to have more realistic attitudes toward them. The players’ cognitive attitudes must conform to the truths of a world that is not of their invention, and that world can frustrate or disappoint them as their own fantasies cannot.

The players in make-believe generally invent the attitudes of their characters, fictionalizing about what those characters are thinking and feeling. If a player imagines that “his” pirate is angry or is coveting the treasure, he is not reporting his own feelings. Similarly, what he imagines his pirate to believe about the location of the treasure need not reflect his own beliefs; he may have no belief on the subject, since he may know that the treasure’s fictional location has not been fixed.

In virtual play, by contrast, participants do not generally attribute attitudes to their avatars at all; they simply have thoughts and feelings about the world of the game, and they act on that world through their avatars but under the motivational force of their own attitudes. Players who send their avatars into unknown regions of the virtual world are genuinely curious about what they will find; they do not endow their avatars with a fictional curiosity to motivate their fictional explorations. Players themselves want the virtual items that their avatars buy — want to own them in the virtual world, that is, via their avatars — and they weigh the cost of those items against other uses for which they themselves foresee needing virtual dollars. Players whose avatars get married in the virtual world (and there are indeed virtual marriages) describe themselves as being in love, not as authoring a fictional romance. They do not experience themselves as artists inventing characters; they experience themselves as the characters, behaving in character, under the impetus of their own thoughts and feelings.7

---

7 At this point, one might object that a real person cannot be curious about a merely fictional landscape, nor desire merely fictional property, nor love a merely fictional spouse. Yet participants in virtual worlds insist that they do, and I am inclined to take their avowals at face value. Real curiosity about a fictional landscape strikes me as unproblematic. As I have explained, a virtual world has the determinateness and fixity characteristic of reality: there is a (fictional) fact of the matter as to what it is like in innumerable respects, and one can want to know such (fictional) facts. Desire for fictional things is
Virtual Agency

Consider now the intentions of a player with respect to the actions that result from his curiosity about the virtual world, his desire for some of its goods, or his love for another of its inhabitants. When he first joins a virtual world, the player finds it difficult to control his avatar, not yet having mastered the technique with keyboard and mouse. At this point, he can act with the intention of manipulating the keyboard and mouse in various ways, and with the further intention of thereby causing his avatar to do various things.\(^8\)

As the player gains skill in controlling his avatar, however, manipulations of the keyboard and mouse disappear from his explicit intentions. He still controls the avatar by manipulating his keyboard and mouse, but only in the sense in which he types the word ‘run’ by moving his two index fingers. When he was just a beginner at typing, he still had to intend the movements by which he typed the word, but now those piecemeal movements have been incorporated into skills with which he can perform higher-level actions straightaway. He can simply decide to type ‘run’ without intending the means to that accomplishment, since his typing skills will take care of the means. (Indeed, he may have to type the word, if only in mid-air, in order to remember which fingers he uses to type it.) Similarly, the skilled player in a virtual world does not explicitly intend his manipulations of the input devices.

Even if a skilled player does not have explicit intentions to manipulate his keyboard or mouse, however, the possibility remains that he at least intends to control his avatar — say, to make the avatar walk and talk.

---

Yet I think that the other features of virtual play favor the hypothesis that the player intends, not to make his avatar do things, but rather to do them with his avatar or to do them as his avatar or, more colloquially, simply to do them.

As we have seen, a virtual environment resembles reality in being both determinate and recalcitrant, confronting the player with facts that can be discovered and altered only by way of appropriate steps on the part of his avatar. In general, the player has no access to those facts in propria persona; he must deal with them in the opaque guise of his avatar, which can be neither penetrated nor circumvented by his actual self. Under these circumstances, intentionally manipulating the avatar would entail operating on the virtual world by an awkward remote control. The avatar would persistently stand between the player and the effects he wanted to bring about in the virtual world, like one of those glass-boxed derricks with which players try to pick up prizes in a carnival arcade.

This mode of operation would be highly disadvantageous. Intending to manipulate one’s avatar so that it does one’s bidding would be (to adopt a different analogy) like intending to maneuver one’s tennis racket so that it hits the ball. And as any tennis player knows, trying to make the racket hit the ball is a surefire way of missing. Given that one must deal with the ball by way of the racket, one does best to treat the racket as under one’s direct control, as if it were an extension of one’s arm. And then one says, “I hit the ball with my racket”, as one might say, “I hit it with my hand”; one does not say, “I made my racket hit the ball.”

The skill of hitting a ball with a tennis racket is a modification of hand-eye coordination, which is a sub-personal mechanism. This mechanism computes and extrapolates the trajectory of a moving object and then guides the hand to intercept it at an angle and velocity that will produce desired results. None of this computation or guidance shows up in the subject’s practical reasoning or intentions; the subject simply decides to catch something or hit something, and his hand-eye coordination takes care of the rest. In acquiring the skill of playing tennis, a player modifies the mechanism of hand-eye coordination to compute the relevant trajectories in relation to the head of his racket rather than his hand, and so he acquires racket-eye coordination, which is also a sub-personal mechanism.

So it is, I suggest, with an avatar. As one gains skill in controlling one’s avatar, one acquires avatar-eye coordination. And then one no longer intends to operate on the virtual world by controlling one’s avatar; one intends to operate with the avatar, as if it were under one’s direct control.
One therefore intends to perform avatar-eye-coordinated actions in the virtual world, not real-world actions of controlling the avatar.

Whereas a tennis racket under one’s direct control serves as an extension of one’s arm, however, an avatar under one’s direct control serves as a proxy for one’s entire body: it is one’s embodiment in the virtual world. Saying “I did it with my avatar” would therefore be like saying “I did it with my body” — something one rarely says, since “with my body” goes without saying whenever one says “I did it” in reference to a bodily action. That’s why a player in the virtual world attributes the actions of his avatar directly to himself, just as he would the movements of his body.  

Combining the foregoing considerations, we arrive at the conclusion that the participant in a virtual world moves his avatar under the impetus of his own beliefs and desires about the virtual world, and he does so with intentions like the ones with which he moves his own body (and its prosthetic extensions) under the impetus of his beliefs and desires. Hence the player’s relation to the avatar, though different from his relation to his own body in many respects, nevertheless resembles it in those respects which are relevant to his status as agent of his bodily movements.

When engaged in virtual play, in other words, a person really has a fictional body. Although the body itself is fictional — it is not really a body or even a real object of any kind — the player’s relation to that fictional body is real, at least in the respects that are most significant for bodily agency, since it is directly controlled by intentions motivated by the player’s beliefs and desires. Hence the player is not speaking fiction when he calls his avatar “me”. He is not strictly identical with the avatar, of course, but his first-person references to it are not meant to imply a strict identity anyway. If a rider in a packed subway car complains, “There’s an elbow in my ribs”, the answer might come back, “Sorry, that’s me” — meaning “That’s my elbow.” Similarly, when a player points to his

---

9 One speaks of doing things “with my body” only when the entire weight or volume of one’s body is involved, as in breaking down a door.

avatar and says “That’s me”, he means “That’s my (fictional) body.” And he is speaking the literal truth.

This equivalence can be restated in the other direction, as follows: Even if you never play video games, you already have an avatar by default; your default avatar is your body.

**Synthetic Agency**

The analogy between a person’s body and an avatar suggests further similarities between virtual and real-world agency. I now want to explore those similarities by focusing on a notable feature of people’s behavior in virtual worlds.

Participants in virtual worlds report that when acting with their avatars, they act in character. Rather than acting in their own characteristic ways, they act in ways characteristic of people like their avatars, who may differ from them in gender, age, race, physiognomy, and physique. Weaklings create muscle-bound avatars with which they swagger; wallflowers create ravishing avatars with which they seduce. If a woman’s avatar is a ponytailed guy with a pack of cigarettes tucked in his sleeve and a guitar around his neck, then she acts like a jazz musician, even if she is a Wall Street banker. If her avatar looks like a Wall Street banker, then she behaves accordingly, no matter who she is. Indeed, participants report that the major attraction of living a “second life” is that, having adopted avatars different from themselves, they find themselves behaving like those different people rather than their real-world selves.

What explains this feature of virtual-world behavior? I believe that the explanation can be found by comparing virtual action to a kind of agency that is thoroughly artificial.

As long as an avatar is standing idle, it is indistinguishable from what is called a non-player character, or NPC — that is, a graphical figure whose behavior is controlled by software rather than by a human player. If the software behind an NPC is sufficiently sophisticated, it can generate behavior similar enough to that of a player-controlled character that other players may be unable to tell the difference. In Second Life, NPCs perform tasks of user support, for example, by answering routine questions from newcomers to the world. NPCs are examples of what might be called synthetic agency.
There is a literature on synthetic agents, divided into two segments. One segment discusses software programs that their designers describe as autonomous; I will describe these synthetic agents as rationally independent, so as to leave open the question of their autonomy in the philosophical sense of the term. The other segment of the literature on synthetic agents discusses what have come to be called believable agents, which are believable in that they give the impression of behaving like persons, even if they take nonhuman forms.

When a synthetic agent is rationally independent, it can carry out tasks without human direction or assistance. Like any software application, of course, this agent must be given instructions that “tell” it how to perform its function. But the function that its preprogrammed instructions tell it how to perform is the higher-order function of carrying out first-order tasks of some open-ended kind, for which precise steps are not specified in advance. Performing those tasks will require figuring out how to perform them, by adopting and prioritizing goals, generating and testing strategies, devising and revising plans, and so on.\(^1\)

Rationally independent software agents can be fairly smart, giving the impression that they are not just calculating but also evaluating, strategizing, and learning. Hence the designer’s description of them as autonomous is not entirely inapt. But they tend to come across as autonomous automata — smart and independent machines in which there appears to be nobody home.

Believability is at a premium in synthetic agents that must interact with real people. Consider, for example, a system designed by computer scientists at the University of Memphis to do the job of a Navy “detailer”, who negotiates with sailors about where they will be posted at the end of their current assignment.\(^2\) As the time for reassignment approaches, a sailor must email the detailer to learn about available openings, and the two of them carry on a correspondence with the aim of finding a good fit for the

---

\(^1\) One model for creating independent software agents is called the BDI model, whose initials stand for Belief/Desire/Intention. See Michael J. Wooldridge, *Reasoning About Rational Agents* (Cambridge, MA: The MIT Press, 2000). This model was in fact developed with the help of Michael Bratman’s classic work *Intention, Plans, and Practical Reason*, but even models developed without reference to the philosophical literature resemble the BDI model in their focus on goals, deliberation, and planning.

sailor’s skills, preferences, and family needs. In order to fill the detailer’s shoes, the software needs an impressive degree of intelligence, including the ability to process natural language and the ability to optimize multiple parameters at once. But the detailer must also perform the very human task of negotiation — advising, cajoling, bullying, and ultimately persuading the sailor to accept an assignment. The Navy therefore wanted the system to seem like a human detailer, so that the sailor would forget that the party at the other end of the correspondence was a computer. In short, the Navy wanted a software agent that was not just rationally independent but also believable.

The pioneering work on believable agents was done by a group of computer scientists at Carnegie Mellon University, in what was known as the Oz Project. To find the secret of creating synthetic agents that were believable, they looked to the “character-based” arts, such as acting and, more to the point, cinematic animation as developed in the studios of Walt Disney and Warner Brothers. A. Bryan Loyall, whose doctoral dissertation was the first extended treatment of the subject, found several recurrent themes in the reflections of these “character” artists. The artists seemed to agree that the first two requirements of believability are the expression of a personality and the expression of emotion. The notion of personality here includes traits of the kind that social psychologists list under that heading, such as extroversion or introversion, but it also includes distinctive styles of speech and movement, specific predilections and tastes, and other characteristics that endow each person with what we call his individuality. As for the expression of emotion, it is now widely recognized as a necessity by designers of believable agents, including the ones who designed the automated Navy detailer. That system was equipped not only with models of memory and consciousness but also with a model of the emotions, which were manifested in its behavior. For example, the automated detailer was programmed to be impatient with sailors who contacted it at the last moment before needing a new assignment.

The third requirement of believability, after the expression of personality and emotion, is what Loyall terms “self-motivation”, defined

as the agent’s acting “of his own accord” rather than merely responding to stimuli. Loyall says that self-motivation is achieved when behavior “is the product of the agent’s own internal drives and desires”,14 but the example he cites does not bear out this gloss. The example comes from Disney animators Frank Thomas and Ollie Johnston, who describe self-motivation in more colloquial terms as “really appear[ing] to think” — a description that is even less informative:15

Prior to 1930, none of the [Disney] characters showed any real thought process. [...] The only thinking done was in reaction to something that had happened. Mickey would see [something], react, realize that he had to get a counter idea in a hurry, look around and see his answer, quickly convert it into something that fit his predicament, then pull the gag by using it successfully.

Of course the potential for having a character really appear to think had always been there [...], but no one knew how to accomplish such an effect. [...] That all changed in one day when a scene was animated of a dog who looked into the camera and snorted. Miraculously, he had come to life!

Surely, what made this dog “really appear to think” was not that he manifested “internal drives and desires” or the results of deliberation. Indeed, deliberation in the service of desires is precisely what was manifested in the behavior attributed here to Mickey Mouse as an illustration of not yet appearing to think. The sense in which the dog “really appeared to think” is that he did not just manifest his internal states; he appeared to be aware of them and to be expressing that self-awareness. Indeed, he appeared to be expressing it to the audience, hence attempting to communicate.

Loyall lists several additional requirements of believability, but I will focus on only one, which subsumes and integrates the requirements mentioned thus far. Loyall calls it “consistency of expression”:16

Every character or agent has many avenues of expression depending on the medium in which it is expressed, for example an actor has facial expression, body posture, movement, voice intonation, etc. To be believable at every moment all of those avenues of expression must work together to convey the unified message that is appropriate for the personality, feelings, situation, thinking etc. of the character. Breaking this consistency, even for a moment, causes the suspension of disbelief to be lost.

14 Loyall, Believable Agents, 20.
16 Loyall, Believable Agents, 22.
Thus, the believable agent must produce behavior that not only expresses his personality, thoughts, emotions, and self-awareness but also does so coherently, in the sense that the features expressed and the behaviors expressing them fit together into what Loyall calls a “unified message”.

**Real-World Believability**

The need for a unified message would explain why participants in a virtual world act in the character of their avatars. Acting in character helps to make their avatars believable, by unifying the avatars’ behavioral “message” with the message conveyed by their appearance.

But why is unification necessary for believability?

Think of it this way. When participating in a virtual world, a player undergoes an updated Turing Test. Turing imagined having a subject communicate via teletype with an unseen interlocutor who was either a second person or a computer.\(^\text{17}\) Turing said that if the computer could fool the subject into thinking that he was communicating with another person, it would qualify as intelligent. As it happens, a similar test confronts the computer that controls non-player characters in a virtual world. Ideally, NPCs would behave in ways indistinguishable from the actions of avatar-embodied persons — though as of yet, NPCs are far from ideal.

What is usually overlooked about the Turing Test is that it tests intelligence only indirectly, by testing for the appearance of personhood, and that it can serve in both respects as a test for human beings as well as for machines. The performance of the machine is judged by being compared with that which would be expected of a person, and there is no reason why the performance of a human cannot be judged similarly.

In fact, you have probably taken a test just like Turing’s. If you have exchanged instant messages with someone over the Internet, then you have used Turing’s setup. In order to use it successfully, you had to send messages that your interlocutor would interpret as coming from a person rather than from a “zombie” computer churning out spam or a virus commandeering his machine. And if you have participated in a virtual world, then you have faced the task of acting with your avatar in ways that the other participants would interpret as the actions of an avatar-embodied person rather than an NPC.

---

\(^{17}\) Alan Turing, “Computing Machinery and Intelligence”, *Mind* 59, no. 236 (1950): 433–460.
In order to pass the Turing Test of instant messaging, you have to send unified messages — that is, messages containing intelligible discourse that expresses consequent thoughts and coherent feelings. You send such messages so that they will be understood. But even a spam-bot sends intelligible messages: what makes you more believable than a spam-bot?

What distinguishes you from a spam-bot is that in trying to make yourself understood, you also betray an awareness of participating in a project of mutual understanding. You give your interlocutor to understand how you have interpreted what he has said, and you adapt what you say not only to what he has said but also to what it indicates about his interpretation of what you said before. By such means, you engage in a subtle form of social interaction in which the interactants adjust their messages so as to communicate successfully.

That’s what the animated dog seemed to be doing when he snorted. The self-awareness that he appeared to express included the awareness of being seen by viewers who would interpret his snort as an expression of disdain. He looked as if he was communicating disdain, not just expressing it — as if he was expressing it, that is, with the intention of being so understood, hence as if he could be asked, “What do you mean by that?” His believability was thus achieved by more than a unified message; it was achieved by the appearance of sending a message with an awareness of how it might be received. His believability was achieved, in other words, by the appearance of sociality.

In face-to-face interaction, the messages sent and received are visual as well as verbal. What people do and say is interpreted in the context of how they look, and incongruities create misunderstandings. When a down-and-out musician asks about the Dow Jones average, we wonder whether he is putting us on. If he uses the jazz idiom ‘bad’ while dressed as a banker, he is sure to be misinterpreted. That’s why players in virtual worlds unify their behavior and appearance: they are engaged in self-presentation for the purpose of social interaction. And because they are clearly prepared to suit their behavior to that purpose, they are believable.

But avatars are just the virtual bodies of real people, who act with them as they act with their real-world bodies. Does the similarity end there? Do people need to be believable only when acting virtually? After all, people unify their behavior with their appearance in the real world as well. Cut the musician’s hair, dress him in a suit, give him a briefcase, and he will
II. Virtual Selves

begin to act less like a musician and more like a banker. Give the banker a ponytail and he will begin, as we say, to let down his hair.

What's more, the dressed-up musician won't just act like a banker; he will begin to think and feel like a banker too. We call some people suits not because they wear suits and not just because they act like people who wear suits; they grow into their suits and thereby become “suits”.

What follows is that the participant in Second Life, wearing his avatar like a suit, should come to have the thoughts, feelings, and, yes, personality of his avatar. Having the body of someone who can coherently feel confident in being aggressive — or coherently feel seductive or argumentative or whatever — he develops the corresponding traits, and then he animates his avatar with them and with the appropriate thoughts and feelings. This philosophical inference is confirmed by players in Second Life. They don’t say, “In Second Life, I look like a nebbish and I act as if I am shy”; they say, “In Second Life, I am a nebbish and I am shy.”

A character in Second Life is thus a chimerical creature in which a fictional, virtual-world body is joined to a literal, real-world mind. That real mind holds a self-conception of the hybrid creature to which it belongs, a creature whose personality, thoughts, and feelings it can know introspectively, unify among themselves and with his appearance, and communicate directly through its fictional body. Of course, the same mind holds a self-conception of a real-world human being to whom it belongs, but that self-conception is different: it is the conception of a different self. Two distinct creatures, one wholly real and one partly fictional, can be literally animated by one and the same mind, for which they help to constitute different selves.