chapter 11

Online Communities, Virtual Reality, and Artificial Intelligence

LEARNING OBJECTIVES

Upon completing this chapter, you will successfully be able to:

- Describe some key similarities and differences between online communities and traditional communities, and identify and evaluate ethical controversies affecting social networking services (SNSs) such as Facebook,
- Identify and assess some ethical aspects of *virtual environments* (VEs) and *virtual reality* (VR) applications, including massively multiplayer online role-playing games (MMORPGs),
- Explain some implications that ongoing developments in the field of *artificial intelligence* (AI) will continue to have for our sense of self and for what it means to be human,
- Determine whether we need to extend our conventional notion of moral obligation to grant at least some degree of moral consideration to certain kinds of AI entities.

In this chapter, we examine some ethical and social aspects of three diverse, and seemingly unrelated, aspects of cybertechnology: (i) online communities, including social networking services (SNSs); (ii) virtual environments (VEs), including virtual reality (VR) applications; and (iii) artificial intelligence (AI). A unifying theme that brings together this otherwise disparate cluster of topics is the impact they have for our notions of *community* and (*personal*) *identity* in the digital age. Whereas SNSs enable social interactions that challenge our traditional notion of community, some VEs and VR applications allow users to construct new and alternate identities. AI-related developments, on the other hand, invite us to reassess our sense of self and, ultimately, question what it means to be human in a world that we share with non-human "intelligent" agents and entities.

We begin by reflecting on a scenario that briefly illustrates a cluster of issues that we will examine in detail in later sections of this chapter.

► **SCENARIO 11–1:** Ralph's Online Friends and Artificial Companions

Ralph, a 60-year-old bachelor and retired software engineer, moved to a condominium in the Peaceful Manor retirement community approximately one year ago. In many ways, Peaceful Manor is an ideal community for a person of Ralph's age. For example, all of the community's residents are 55 or older.

Furthermore, there is a clubhouse on the premises that includes a wellness center with indoor and outdoor swimming pools, as well as other amenities and activities. In addition, Peaceful Manor provides its residents with three categories or levels of living, ranging from (fully) independent living to "assisted living" to full-time care. (So a resident at Peaceful Manor would never have to relocate because of aging or deteriorating health.) Although Ralph is still fully independent and capable of participating in many of the amenities offered by his retirement community, he has elected not to take advantage of any of them. Instead, he prefers to stay in his condominium (day and night) and to "meet" only with online friends, most of whom he has acquired since joining a social networking service one year ago. Ralph also enjoys interacting with various avatars and (soft)bots that "reside" (only) in online forums and games, and he prefers those interactions over the kind that would be required if he were to engage with fellow humans who live in his retirement community.

In the past year or so, Ralph has not interacted (face-to-face) with any of his fellow residents at Peaceful Manor (and has interacted with people outside that community only when it is necessary for him to do so). Ralph prefers online-only interactions for several reasons. For one thing, he can communicate with his online friends at times that are convenient to him. For another, Ralph does not have to worry about being embarrassed by his speech impediment, since he communicates textually with his online friends. Additionally, Ralph can choose which features about himself to disclose—or, alternatively, to "construct"-such as information about his age, gender, marital status, work background, physical characteristics, and so forth. It turns out that Ralph has constructed a kind of alter ego, or "cyberego," by which he is known (only) to his online friends. Ralph also finds it "liberating" that his digital persona does not require him to disclose information about his speech impediment, whereas this condition would be immediately apparent to anyone with whom he interacts in physical space.

Ralph has already decided that when the time comes for him to move to the next level of care offered by Peaceful Manor, he will request that an "Elderbot" or "Carebot" assist him rather than a human being. He believes that he will be far more comfortable interacting with a sophisticated robot or "artificial companion" than with a human assistant in carrying out routine daily tasks. So Ralph, unlike many people in his age group, desires to interact only with online friends and artificial companions and not with "flesh-and-blood" persons.

What are we to make of Ralph's choices, especially for someone in his age group or category? Because he has freely chosen to interact only with online "friends" rather than fellow residents in his physical community, we can ask whether Ralph's allegiance (and thus his sense of belonging/membership) is to his online community of friends, avatars, bots, and so forth. We can also ask questions about Ralph's (personal) identity—in other words, who is he? For example, is Ralph ultimately the person (or persona) represented in the online world where he spends most of his time? Or is that simply Ralph's alter (or "cyber") ego and thus just one aspect of his overall self (or multiple "selves")? And why would Ralph opt for an artificial companion (or Elderbot) to assist him in the future, if he would be eligible to have a human care for him in that role?

The purpose of Scenario 11–1 was not to answer any of these questions, but rather to get us to begin thinking about a cluster of issues that we will examine in detail in the remainder of this chapter. We start by describing the roles that online communities increasingly play, while also focusing on some key ethical and social challenges that they pose.

11.1 ONLINE COMMUNITIES AND SOCIAL NETWORKING SERVICES

Many people, both young and old, now interact in various forms of online communities. Perhaps the most popular of these are Web-based social networking services (SNSs), such as Facebook, and professional-oriented networking services, such as LinkedIn. Some people also "follow" celebrities and send instantaneous messages in the form of "tweets" via a popular online service called Twitter, while others participate in one or more blogs (Web logs). Many also communicate with one another through digital media services that include video, such as Skype and (Apple's) FaceTime. So we can ask: How has our conventional understanding of "community" evolved, that is, in light of the kinds of social interactions made possible by these various kinds of online forums? We begin our analysis of this question by examining some important similarities and differences between online and traditional communities.

11.1.1 Online Communities vs. Traditional Communities

To better understand what is meant by an online community, we first examine the meaning of "community" in the traditional sense of the term. Webster's New World Dictionary of the American Language defines a community as "people living in the same district, city, etc., under the same laws." Note that this traditional definition stresses the geographical aspects of community by associating it with concepts such as "district" and "city" that have typically constrained community life. So, for the most part, traditional communities are limited by geography.

Cybertechnology has made it possible to extend, or perhaps even ignore, the geographical boundaries of traditional community life. This, in turn, leads us to reexamine the concept of community; individuals physically separated by continents and oceans can now interact regularly in SNSs and other online forums to discuss topics that bind them together as a community. Not surprisingly then, more recent definitions of "community" focus on the common interests of groups rather than on geographical and physical criteria.

Though it may seem surprising to some, online communities are not exactly a new, or even recent, phenomenon. Rheingold (2001), who defines online communities as "computer-mediated social groups," describes his experience in joining the Whole Earth' Lectronic Link (WELL), an early electronic community, in 1985:

The idea of a community accessible only via my computer screen sounded cold to me at first, but . . . [t]he WELL felt like an authentic community to me from the start, because it was grounded in my everyday physical world. WELLites who don't live within driving distance of the San Francisco Bay area are constrained in their ability to participate in the local networks of face-to-face acquaintances. . . . I've attended real-life WELL marriages, WELL births, and even a WELL funeral.¹

Rheingold points out that because of the social contracts and collaborative negotiations that happened when members met online, the WELL became a *community* in that setting. He notes, for example, that in the WELL, norms were "established, challenged, changed, reestablished, rechallenged, in a kind of speeded up social evolution." When the members decided to get together occasionally at physical locations in the greater San Francisco Bay area, the WELL became a "hybrid community," spanning both physical and virtual space. But some "pure" online communities also continue to thrive alongside the hybrid communities. As White (2002) notes, these electronic-only forums also seem like "real communities" because they offer their members "social exchange, emotional support, and learning environments."

Do users now find as much enjoyment and satisfaction in participating in online communities as they do in traditional ones? Mitch Parsell cites a survey (conducted by the U.S.-based Center for the Digital Future) showing that 43% of members of online communities claimed to feel "as strong" about their online communities as their traditional or "real-world" communities. He also believes that this may be due to the enhanced nature of the Web—what some now refer to as "Web 2.0"—which is very different from the early Web, primarily because of the interactive aspects of the experiences it makes possible. Analysts disagree on exactly which criteria differentiate Web 2.0 from the original Web, but most agree that the kinds of services made possible by SNSs and blogging sites have significantly altered the way users interact in online communities. (Recall our description of some key differences between the early Web, or "Web 1.0," and Web 2.0 environments in Chapter 1.)

As already suggested, SNSs are arguably the most popular type of online community. And the most popular of these services (as of May 2015) are Facebook, Twitter, LinkedIn, Pinterest, Google+, Tumblr, and Instagram, respectively. In Section 11.1.3, where we examine some pros and cons of online communities, our focus will be mainly on social/ethical challenges posed by SNSs. Before examining those challenges, however, we briefly examine some controversies arising from a specific kind of online forum: blogs.

11.1.2 Blogs and Some Controversial Aspects of the Bogosphere

What, exactly, is a blog? According to the (online) Merriam Webster Dictionary, a blog (or "Web log") is "a Web site that contains an online personal journal with reflections, comments, and often hyperlinks provided by the writer." Blogs can be maintained by either individuals or organizations. The community of blogs is often referred to as the "blogosphere." Online communities such as myBlogLog and Blog Catalog connect bloggers, whereas search engines such as Bloglines, BlogScope, and Technorati assist users in finding blogs.⁵

While some blogs function as online diaries, others provide commentary on a particular topic or news story. Based on their topics, blogs are often organized into categories such as personal blogs, political blogs, corporate blogs, health blogs, literary blogs, travel blogs, etc. Blogging has become popular because it is an easy way to reach many people; but activities on some blogs have also raised social and ethical concerns. For example, consider the case of a personal blog, illustrated in the following scenario, which raised a cluster of controversial issues.

► **SCENARIO 11–2:** "The Washingtonienne" Blogger

Jessica Cutler, who worked as a staff assistant to a U.S. senator, authored an online diary (on blogger. com) under the "pen name" or pseudonym "The Washingtonienne." In 2004, she was fired when the contents of her diary appeared in Wonkette: The DC Gossip, a popular blog in the Washington, D.C., area. Until her diary was discovered and published in Wonkette, however, Cutler assumed that it had been viewed by only a few of her close friends (who also worked as staff assistants in Washington, D.C.) that she suspected might be interested in reading about the details of her romantic relationships and sexual encounters. In her online diary, Cutler disclosed that although she earned a relatively low salary as a congressional staffer, most of her living expenses were "thankfully subsidized by a few generous older gentlemen." She also described some details of her sexual relationships with these men, one of whom was married and an official in the George W. Bush administration. Cutler did not use the real names of these men but instead referred to them via initials that could easily be linked to their actual identities. However, she was subsequently sued by one of the men implicated in her blog.6

We can see why Cutler's (personal) blog was controversial and why it raised a number of ethical issues. One concern had to do with expectations regarding personal privacy, both for Cutler and the men implicated in her blog. Other concerns included questions about expectations regarding the anonymity and confidentiality of those who contribute to blogs, or who post information on personal blogs that are intended to be shared only with a few close friends. Another question that arose in the Washingtonienne incident was whether some of the controversial content pertaining to the men (indirectly) identified in this blog rose to the level of (online) defamation.

However, it is not only personal blogs that have generated controversy. Consider, for example, some controversies affecting political blogs, which we briefly examined in connection with our discussion of democracy and cybertechnology in Chapter 10. There, we saw that political bloggers have often been responsible for breaking news stories about political scandals and thus influencing public opinion. We also saw that some of these bloggers had political agendas to advance and thus were eager to spread negative stories about politicians whose views they opposed; in many cases, these stories were not only inaccurate but blatantly false.

One question worth mentioning before we conclude this section has to do with whether bloggers, especially those who write and maintain influential blogs, should be held to the same standards of accuracy, accountability, and liability as professional online journalists. Many bloggers claim that they are not journalists and thus should not be held to professional journalistic standards. Critics, however, argue that bloggers have certain "ethical obligations to their readers, the people they write about, and society in general." Unfortunately, an examination of this debate in the detail it deserves is beyond the scope of this chapter.⁸

11.1.3 Some Pros and Cons of SNSs (and Other Online Communities)

Have SNSs in particular, and online communities in general, had an overall positive effect on communication and social interaction? Not surprisingly, arguments have been advanced to support positions on both sides of this question. Those who see these online forums in a favorable light tend to point to the fact that on SNSs such as Facebook, users can make new "friends" and meet prospective college roommates before setting foot on campus; they can also possibly find future romantic partners in online dating services such as Match.com and eHarmony. Supporters also note that users can join and form online medical support groups, as well as participating in various blogs designed to disseminate material to like-minded colleagues. Through these online services and forums, users can communicate with people they might not otherwise communicate with by mail or telephone.

However, SNSs and other online forums have also had some negative effects. In addition to threatening traditional community life, they have:

- Facilitated social polarization
- Threatened our traditional notion of *friendship*
- Facilitated deception

We briefly examine each of these points.

Social Polarization

One positive feature of online forums is that they provide us with choices regarding the kinds of online communities we or may not wish to join; this would seem to contribute positively to human interaction by enabling us to come together with like-minded individuals we otherwise might not meet. However, some online communities, especially those whose focus tends to be on topics and issues that are divisive and narrow, can also contribute to social polarization. Parsell (2008) argues that "extremely narrowly focused" online communities can be dangerous because they "can polarize attitudes and prejudices," which can lead to increased division and "social cleavage." He worries that the narrow focus of many online communities presents us with cause for concern, which he articulates via the following line of reasoning:

- 1. People tend to be attracted to others with like opinions.
- 2. Being exposed to like opinions tends to increase our own prejudices.
- 3. This polarizing of attitudes can occur on socially significant issues. . . .
- Thus, where the possibility of narrowing focus on socially significant issues is available, increased community fracture is likely.9

So, even though online communities can empower individuals by providing them with greater freedom and choice in terms of their social interactions, they can also foster increased polarization in society. In Chapter 10, we briefly examined some political/social polarization issues in the context of democracy and cybertechnology, some of which also overlap with those facilitated by online communities. There, for example, we saw how "personalization filters" used by search engine companies can also influence political and social polarization, especially in ways that undermine "deliberative democracy."

"Friends" in the Context of SNSs

Does it matter that online communication has minimized the kinds of face-to-face interactions that define behavior in traditional communities? Is that necessarily a negative thing? On the one hand, being able to send an e-mail or a text message to someone, or to write on a Facebook user's "wall," or timeline, is far more convenient than having to meet that person face-to-face in physical space to communicate with them. On the other hand, some worry that something is lost possibly some critical interpersonal skills—in excessive online communications, at the expense of avoiding face-to-face interactions. (Recall our example of "Ralph" in Scenario 11-1.) A related, and very important, question that also arises has to do with the implications that online-only communication between individuals may have for our traditional understanding of friendship. In other words, is it possible for people who interact only in virtual (or purely online) contexts to be "real friends"?

To what extent, if any, is physical interaction between individuals necessary for true friendships to develop and flourish? At one time, the notion of "disembodied friends" might have seemed strange. But today, we hear about so-called "friends" who communicate regularly online but have never met in physical space. Cocking and Matthews (2000) argue that the kinds of close friendships we enjoy in physical space are not possible in pure virtual environments (VEs), that is, in contexts that are solely computer mediated. They worry that we miss the kind of interaction that is commonplace in close friendships, because online-only friendship occurs in

a context of communication dominated by voluntary self disclosure, enabling and disposing me to construct a highly chosen and controlled self-presentation and world of interaction.¹⁰

Cocking and Matthews' argument is complex and cannot be analyzed here in the detail that it deserves. But we will consider a few of their key points. The authors argue that it is not possible to realize close friendships in a "virtual world" because purely computer-mediated contexts (i) facilitate voluntary self-disclosure and (ii) enable people to choose and construct a highly controlled "self-presentation" or identity. (Consider again the example of Ralph in Scenario 11-1, who constructed a selective or controlled personal identity for use in his online interactions.) This would not be possible in offline contexts, Cocking and Matthews argue, because there we involuntarily disclose aspects of ourselves through indicators or "cues" in our interactions with others. And because these interactions include acts of "nonvoluntary self-disclosure," one has less control over the way one presents oneself to others. As a result, important aspects of our true personalities are involuntarily revealed, which, the authors believe, makes close friendships possible in offline contexts but not in virtual ones.

Is the argument advanced by Cocking and Matthews convincing? Briggle (2008) disagrees with their conclusion, but he uses different kinds of criteria - one based on "distance" and one on "deliberateness"—to make the case for why friendships in purely virtual contexts can be initiated and can "flourish." First, he points out that communications among friends in offline contexts, which are based largely on "oral exchanges," are not always candid or sincere; consequently, important "dynamics and indicators" that are required to form close friendships can be distorted. But Briggle believes that the distance involved in typical computer-mediated communications can give friends the courage to be more candid with one another than in typical face-to-face interactions.

Second, Briggle points out that online friends depend on written correspondences (as opposed to oral exchanges), and he believes that the deliberateness required in composing those kinds of correspondences can lead to "deeper bonds and greater depth in friendships." He also notes that oral communication in offline contexts, on the contrary, is "often too shallow and hasty to promote deep bonds."11 So, unlike Cocking and Matthews, Briggle concludes that it is possible to form close friendships in purely virtual contexts. He also concedes, however, that some purely virtual relationships can be "shallow."

Deception in SNSs and other Kinds of Online Communities

Some critics believe that online communities reveal a "darker side" of the Internet because people can, under the shield of anonymity, engage in behavior that would not be tolerated in most physical communities. For instance, individuals can use aliases, including pseudonyms and screen names, when they interact in online forums, which makes it easier to deceive others about who actually is communicating with them. We briefly examine a scenario illustrating how online pseudonymity and deception reveal the darker side of online communities.

► SCENARIO 11–3: A Suicide Resulting from Deception on MySpace

In 2006, Megan Meier was a 13-year-old resident of Dardenne Prairie, Missouri, who had an account on MySpace. On that SNS, she received a "friend" request from a user named Josh Evans. Evans, who claimed to be a 16-year-old boy, told Meier that he lived near her and was being homeschooled by his parents. At first, Evans sent flattering e-mails to Meier, which also suggested that he might be romantically interested in her. But Evans's remarks soon turned from compliments to insults, and Evans then informed Meier that he was no longer sure that he wanted to be friends with her because he heard that she "wasn't very nice to her friends." Next, Meier noticed that some highly derogatory posts about her for example, "Megan Meier is a slut" and "Megan Meier is fat"—began to appear on MySpace. Meier, who was reported to have suffered from low self-esteem and depression, became increasingly distressed by the online harassment (cyberbullying) being directed at her—that is, from both the insulting MySpace postings and hurtful e-mail messages she continued to receive from Evans. In October 2006, Meier decided to end her life by hanging herself in her bedroom. An investigation of this incident following Meier's death revealed that Josh Evans was not a teenage boy; she was Lori Drew, the 49-year-old mother of a former friend of Meier's.12

In the period following Meier's death, Lori Drew's true identity was discovered, and she was eventually prosecuted. Although Drew was found guilty of three misdemeanor counts of computer fraud in 2008, the jury was deadlocked on a fourth charge, involving conspiracy. While the Meier incident could be examined from the vantage point of cyberullying, an offense that has affected numerous teenagers interacting in online communities, we consider it here mainly from the perspective of one kind of deceptive behavior that is possible in online communities. But it is also important to note that the Meier incident was not the first case in which pseudonyms and screen names were used in deceptive and devious ways in online communities.

A classic case that also illustrates the darker side of online communities—one which also received wide attention via a popular article by Julian Dibbell that first appeared in *The Village* Voice (in 1993)—involved (an incident that has come to be described as) a "virtual rape" in cyberspace. In an online forum called LambdaMOO – a MOO is a multiuser object-oriented environment—users had "screen names" that represented the virtual characters they portrayed. One character, who used the screen name "Mr. Bungle," had designed a program that enabled him to control the actions of other characters in this MOO. In at least one instance, Bungle used his program to "perform" sexually offensive actions on two of the characters also represented in this online forum.¹³ This incident had a profound effect not only on the members of the LambdaMOO community but also on many participants in other online communities who heard about it.14

In which ways are the controversial incidents involving MySpace and LambdaMOO similar? Both have at least one key feature in common: They reveal a darker side of online communities that is made possible by anonymity and pseudonymity. One relevant difference, however, is that no one was physically harmed in the LambdaMOO incident. But it could be argued that two Lambda members whose characters were "virtually raped" did experience some emotional or psychological harm, and in Section 11.2.2, we will see how emotional harm can occur to the real-life (i.e., "flesh-and-blood") people whose characters are represented in VEs.

A second difference is that LambdaMOO was a "pure" online, or online-only, community, where the virtual characters that participated had screen names and acted out various roles in the game-like or role-playing context of that VE. Also, it was appropriate for members of LambdaMOO to use names other than their actual ones because of the rules defining that Lambda community. In the MySpace incident involving Meier, however, Lori Drew's use of an alias was deceptive by virtue of the rules, or at least the expected norms, on MySpace for setting up an account and initiating a request to be friend another user. There, the expectation was that an individual using that SNS to seek out new "friends" would disclose his or her actual name and not use a false (or pseudo) name to intentionally deceive someone. However, one might also argue that such a rule was not sufficiently explicit at the time, because neither MySpace nor most other SNSs had clear policies when it came to individuals using actual names in setting up an account. (Typically, all that was required then to register for an account on MySpace was a legitimate e-mail address.)

11.2 VIRTUAL ENVIRONMENTS AND VIRTUAL REALITY

We should note that the kinds of online communities examined in the preceding section are also sometimes described as "virtual communities." However, those communities comprise only a subset of what we have referred to in the preceding section as virtual environments (VEs). Before describing what is meant by our broader notion of VE, however, it is useful to understand the meaning of "virtual." This term can be used in three senses. Sometimes, it is contrasted with "real," as in distinguishing virtual objects from real ones. At other times, the term is contrasted with "actual," as when a person says that she has "virtually finished" her project (i.e., she has not actually, or literally, finished it, but she believes that for practical purposes, she has finished it). The term "virtual" can also refer to one's feeling "as if" one were physically present in a space, as when you are conversing with a friend online or on a phone; even though you could be literally thousands of miles away from each other, the sense that you are interacting in real time makes you feel as if you are both in the same room. (Contrast this experience with radio and traditional broadcast communications, where messages can be transmitted in only one direction at a time.) However, it is primarily in the first sense of virtual—contrasted with "real"—that we examine various kinds of VEs in Section 1.2.

How are VEs both similar to and different from the online/virtual communities we examined in Section 11.1? While both kinds of virtual forums are "computer generated," and thus could not exist without cybertechnology, VEs also provide contexts in which users can do more than merely interact with other users. As Søraker and Brey (2015) note, VEs can "visualize imaginary environments" as well as simulate real ones. 15

Not only do VEs subsume virtual (or online) communities, they also subsume (what some authors call) virtual worlds. Søraker and Brey describe the latter as a type of VE in which users typically (i) are "represented by avatars" and (ii) have the "illusion of perceiving a three dimensional world consisting of virtual objects." ¹⁶ Wankel and Malleck (2010) also describe virtual worlds as environments that are "three dimensional." For our purposes, however, a VE need not have a three-dimensional aspect (even if only illusory); as we saw in our analysis of online/virtual communities and forums (in Section 11.1.1), VEs can also include twodimensional environments (such as MOOs). So, in our scheme, a VE can include a wide range of two-dimensional and three-dimensional forums or applications—that is, from online/virtual communities to virtual worlds to virtual reality (VR) applications. Our focus on ethical aspects of VEs in the remainder of Section 1.2, however, will be mainly on VR applications (including video games). So we first need to understand more precisely what is meant by the term VR.

11.2.1 What is Virtual Reality (VR)?

Brey (1999) defines VR as "a three dimensional interactive computer generated environment that incorporates a first person perspective." Notice three important features in Brey's definition of VR:

- Interactivity
- A three-dimensional environment
- A first-person perspective

How do these features distinguish VR from other kinds of computer-mediated environments? First, *interactivity* requires that users be able to navigate and manipulate the represented environment. Because a *three-dimensional environment* is required in VR, neither text-based VEs nor two-dimensional graphic environments will qualify. Brey also points out that a *first-person perspective* requires a single locus from which the environment is perceived and interacted with; the first-person perspective also requires an immersion in the virtual world rather than simply an "experience" of that world as an "object that can be (partially) controlled by the outside."

As noted earlier, we can differentiate VR environments/applications, which are three-dimensional, from our more broadly defined notion of VE. Recall our analysis of online/virtual communities in Section 11.1, which included examples of two-dimensional VEs; LambdaMOO, for instance, satisfies the requirements for a two-dimensional VE, but it would not qualify as a VR application.

Figure 11-1 illustrates some manifestations of virtual environments.

Virtual Reality vs. Augmented Reality

We should note that some authors now tend to conflate VR with *augmented reality* (AR). What, exactly, is the difference between AR and VR, given that both are "computer-mediated" realities, as well as three-dimensional environments? Søraker and Brey (2015) note that in VR applications, a user's "field of vision is substantially replaced by the computer-generated visual output." So VR applications typically provide the user with a "simulated world," thus replacing the real world. AR technology, on the contrary, enhances a user's view of the real world. Via

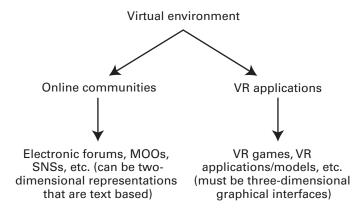


Figure 11-1 Virtual environments.

AR equipment, the user gains either direct or indirect access to the physical (real) world in a way that is enhanced or augmented by computer-generated sensory inputs. These typically include video inputs (e.g., eyewear such as Google glass) and sound inputs. With AR technology, users also interact with their physical environments in real time.

It is perhaps worth noting that some authors build the concept of AR into their definition of "virtual worlds" (which are a subset of what we call VEs). For example, Wankel and Malleck (2010) define virtual worlds as "three-dimensional augmented reality interfaces." However, we limit our discussion of controversies involving VEs in this section to VR applications that either simulate the real world or create completely artificial worlds. These applications are sufficiently distinct from the AR applications that enhance our perception of the real (physical) world we inhabit. For more information on AR, including some potential ethical concerns that this technology raises, see Mann (2013).

In Section 11.1.3, we noted that many actions and behaviors that are considered morally objectionable in the physical world can be performed easily in online communities and forums. We will next see that interactions made possible in VR applications, including some video games, also enable users to engage in morally objectionable acts/behaviors; these include (virtual) prostitution, pornography, pedophilia, torture, mutilation, and murder.

11.2.2 Ethical Aspects of VR Applications

How are ethical issues involving controversial behavior in VR applications different from issues associated with morally controversial acts displayed on television or played out in board games? For example, television programs sometimes display violent acts and some board games allow participants to act out morally controversial roles. But Brey points out that in VR applications, users are actively engaged, whereas television viewers are passive. VR users are not spectators; rather, they are more like actors, as are board game players, who also act out roles in certain board games. This common feature suggests that there might not be much difference between the two kinds of games; however, Brey also notes that VR applications, unlike board games, simulate the world in a way that gives it a much greater appearance of reality. And in VR, the player has a first-person perspective of what it is like to perform certain acts and roles, including some that are criminal or immoral, or both. We next examine some ethical aspects of online video games.

First, we draw a useful distinction between single-player and multiplayer video games; the latter are commonly referred to either as massively multiplayer online games (MMOGs) or massively multiplayer online role-playing games (MMORPGs). An example of a MMOG is Second Life (designed by Linden Lab), which includes members called "Residents" who do not engage in some of the traditional role-playing activities available in many MMORPGs. (Launched in 2003, Second Life had approximately 13 million registered user accounts as of March 2008. Examples of MMORPGs include popular online games such as World of Warcraft (WOW), EVE Online, and Entropia Universe. WOW is perhaps the most popular MMORPG, boasting over 7 million paying subscribers as of May 2015.¹⁹

Many MMORPGs are organized around various "clans" or "guilds" that members join; if they wish, members can also pool their resources.²⁰ Typically, these games are also developed around systems that are either "class based" or "skill points based." In the former scheme, a player's chosen division can determine his or her character's "strengths and weaknesses." So players are encouraged to "interact with others and form teams to balance out strengths and weaknesses."²¹ In Second Life, for example, "Residents" can socialize, participate in group activities, and create trade items (virtual property) and services with one another. The virtual currency used in that MMOG is the Linden Dollar (linden or L\$), which can also be traded for real currencies via a "resident-to-resident marketplace" that is facilitated by Linden Lab. Residents in Second Life, as well as players in other MMOGs and MMORPGs, often select an avatar (a graphical representation that can exhibit human-like features and traits) to represent themselves.

We next consider why some MMORPGs and MMOGs are controversial from an ethical perspective? In doing so, we examine the kinds of morally objectionable behavior they make possible, especially with respect to violence and pornography.

Violent and Sexually Offensive Acts in MMOGs and MMORPGs

Some critics claim that Second Life facilitates child pornography because virtual characters who are adults in real life (RL) can have sex with virtual children, that is, avatars designed to look like children, in that MMOG.²² Cases of virtual prostitution in Second Life have also been reported—that is, where some Residents were "paid to (use their avatar to) perform sex acts or to serve as escorts."23 So, if these reports are correct, there are clearly some forms of sexually offensive acts that take place in MMOGs and MMORPGs, which would not be tolerated outside these gaming environments.

In addition to concerns about sexually offensive behavior in online games, many worry about the kinds of violent acts that are also carried out in these environments. Wonderly (2008) suggests that some forms of violence permitted in online games may be "more morally problematic" than pornography and other kinds of sexually offensive behavior in VEs. She points out, for example, that relatively few video games "permit sexual interaction between characters," and even fewer allow "deviant sexual conduct." But she notes that many popular games permit and that some "even require copious amounts of wanton graphic violence."²⁴

Are violent acts in MMORPGs, including virtual murder, more morally problematic than the pornographic and sexually deviant acts that also have been carried out in these games? Or should violent acts such as virtual murder be tolerated as acceptable behavior? Luck (2009) notes that while most people agree that murder is wrong, they do not seem to be bothered by virtual murder in MMORPGs. He points out, for example, that some might see the virtual murder of a character in a video game as no different from the "taking of a pawn in a chess game." But Luck also notes that people have different intuitions about acts in VEs that involve morally objectionable sexual behavior, such as child pornography and pedophilia. And he worries that the kind of reasoning used to defend virtual murder in games could, unwittingly, be extended to defend virtual pedophilia. For example, Luck suggests that the following line of reasoning, which for our purposes can be expressed in standard argument form, may unintentionally succeed in doing this:

- 1. Allowing acts of virtual murder will not likely increase the number of actual murders.
- Allowing acts of virtual pedophilia may significantly increase the amount of actual pedophilia.
- 3. Therefore, virtual pedophilia is immoral, but virtual murder is not.²⁵

While this kind of argument may appeal to some, Luck points out that it is difficult to defend because of the lack of empirical evidence needed to confirm both (1) and (2). More importantly, however, if (2) could be shown to be false, then virtual pedophilia, like virtual murder, would not be immoral (according to the reasoning used in this argument).

With respect to (2), Levy (2002) has suggested that allowing virtual child pornography may even "reduce the harm to actual children" because it would provide an "acceptable outlet" for pedophiles and would encourage pornographers to seek an "alternative to real children." But it is important to note that Levy does not believe that virtual child pornography should be acceptable; in fact, he opposes it for alternative reasons.

A different kind of rationale for why virtual child pornography should be prohibited has been offered by Sandin (2004), who argues that it can cause significant harm to many people who find it revolting or offensive. But Brey (2008) points out that one problem with Sandin's

argument is that it puts too much emphasis on a kind of harm that is simply "caused by offense." As Brey notes, if we outlaw actions simply because they "offend a large group of people," then (i) what we now take to be "individual rights" would be "drastically curtailed," and (ii) and interracial marriage and homosexual acts would "still be illegal." Hence, none of the arguments considered so far can show why acts that are morally objectionable in physical space either should or should not be allowed in VEs.

Assessing the Nature of "Harm" in VEs and VR Applications

Can a plausible argument be constructed to show why it is wrong to perform acts in VEs in general, and VR applications in particular, that would be considered immoral in RL? We have seen some difficulties with arguments that tried to show that allowing morally objectionable actions in VEs will likely lead to an increase (or decrease) in those actions in the real world. Other arguments have tried to link, or in some cases delink, the kind of harm caused in VEs with the sense of harm one might experience in the real world. For example, some arguments have tried to show that sexually offensive acts in VEs can cause harm to vulnerable groups (such as children and women) in the real world.²⁷ However, the individual premises used to support the conclusions to these arguments typically lack sufficient empirical evidence to establish the various claims being made. On the contrary, some arguments claim that no one is physically harmed in virtual murder or, for that matter, in any act performed only in a VE. But these arguments have also been criticized for lacking sufficient evidence to establish their conclusions.

Should we assume that any harm that one experiences in the virtual realm is not "real harm" but only virtual harm? In our discussion of logical fallacies in Chapter 3, we saw that using this line of reasoning commits the virtuality fallacy; you may wish to revisit that fallacy at this point. Søraker and Brey (2015) have also pointed out the mistaken inference (or logical error) made by some people who assume that because VEs are not real environments, the consequences of one's actions in these environments do not have real-life consequences. And Søraker (2012), who notes that the computer simulation that "underpins" a VE is a physical entity, argues that we need to distinguish between what he calls the "intravirtual" and "extravirtual" consequences of these environments. He then points out that while intravirtual consequences may affect only the "state" of the VE itself, the extra-virtual consequences, which are "triggered by the state of the virtual environment," can also have "potentially dramatic consequences in the real world."

We have already noted that while "virtual" is ambiguous (in at least three ways), it is now typically contrasted with "real." So this has been interpreted by some to mean that a "virtual harm" is equivalent to an "unreal harm." But even though a harm caused in a VE might not result in physical harm to a "flesh-and-blood" person, it doesn't follow that the harm caused is not real. The LambdaMOO incident (described in Section 11.1.3), which involved a "virtual rape," can help us to see why harm in a VE is not itself limited to the virtual characters in that VE. We begin by asking why the "rape" in LambdaMOO was a morally objectionable act. For example, one might argue that it was not a "real rape," and it did not result in physical harm to any "flesh-and-blood" individuals. Brey (1999) believes that we can use two different kinds of arguments to show why it is wrong to engage in immoral acts in VEs:

- **a.** The argument from moral development
- **b.** The argument from psychological harm

To illustrate (a), Brey suggests that we can extend an argument advanced by Immanuel Kant for the treatment of animals to the treatment of virtual characters. Kant argued that even if we have no direct moral obligation to treat animals kindly, we should, because treating animals kindly can influence our development of moral attitudes and behaviors for treating human beings. Similarly, then, the way we treat virtual characters may ultimately affect the

way we treat real persons—raping virtual characters in virtual space, or even viewing such a rape, could desensitize us to the act of rape itself as it affects flesh-and-blood individuals in physical space.

The rationale in (b), the argument from psychological harm, suggests that the way we refer to characters that represent a particular group can cause harm to actual members of the group. Consider a cartoon depicting a woman being raped: Actual (flesh-and-blood) women may suffer psychological harm from seeing, or possibly even knowing about, this cartoon image, even though none of them, as flesh-and-blood individuals, is being raped, either physically or as represented by the cartoon. Extending this analogy to virtual space, it would follow that the "rape" of a virtual woman in a VE, such as a MOO, MMOG, MMORPG, etc., can also cause psychological harm to real (i.e., flesh-and-blood) women.

Virtual Economies and "Gold Farming"

So far, we have examined controversial behavioral issues in VEs in general, and video games in particular, mainly from the perspectives of violence and sexually offensive acts. We also described the kinds of harm that can result from these kinds of behaviors/acts in VEs. However, other behavior-related controversies arise because of the kinds of "virtual economies" made possible by interactions in online games. We noted earlier that some MMORPGs and MMOGs have their own monetary currencies. For example, Second Life uses the Linden Dollar (linden or L\$), which its Residents can both use in in-game transactions and exchange outside the game for real currencies such as the U.S. dollar or the euro. As a result, virtual economies have emerged. Brey (2008) believes that their emergence can also increase the likelihood that moral controversies will arise in these environments—as Brey notes, people will be "more likely to act immorally if money is to be made or if valuable property is to be had."

Virtual property, as in the case of virtual money, can be acquired and exchanged with players in games. It can also be sold and exchanged outside the game to interested parties (in the physical world). In some cases, the virtual property has become so desirable that it has led to violent acts in the real world. Warner and Raiter (2005) describe an incident in China where a person who had stolen someone's virtual sword in a MMOG was murdered in real life (RL) by the "sword's" owner. So virtual economies can have RL implications and can result in physical harm to individuals in the real world. One controversial activity associated with virtual economies in gaming environments is a form of labor and economic exchange called "gold farming."

Kimppa and Bisset (2008) define gold farming as "playing an online computer game for the purpose of gaining items of value within the internal economy of the game and selling these to other players for real money."28 These items can include "desirable items" as well as in-game money (where the rules defining the game's internal economy permit this); they can also include "highly developed" game characters. All of these items can also be sold via online auctions or designated Web sites. Kimppa and Bisset point out that the 2009 "in-game gold market" globally was estimated at 7 billion dollars; they also note that the practice of gold farming is most popular in countries such as China and Mexico that have both low-average income levels and "relatively good access to the Internet."29

Gold farming has also raised concerns about working conditions in the real world. For example, Warner and Raiter describe a situation in rural China where people who participate in World of Warcraft were paid to work 12-hour shifts of gold farming; the workers would acquire "virtual gold" within the game and then sell it outside the game to interested players. The business became profitable, Warner and Reiter point out, because many players who can afford to purchase the virtual gold preferred to buy it rather than to do the work necessary to earn it in the game.³⁰ It was also more advantageous to the "gold farmers" themselves, who could earn more money obtaining and selling virtual gold than they could in traditional agricultural work. But as Brey (2008) notes, this practice has also led to reported cases of Chinese

sweatshop laborers who work "day and night in conditions of practical slavery" to acquire the virtual gold and virtual resources.³¹ Assuming that these reports are true, gold mining raises some serious ethical concerns.

In this section, we have examined some behavioral, or what Brey (1999) also refers to as "interactive," controversies regarding ethical dimensions of VR applications. Other ethical aspects of VEs and VR applications in Brey's model focus on ways in which virtual characters and virtual objects are represented in these environments and applications. For example, these objects can be misrepresented (i.e., in terms of accuracy) or represented in ways that are either biased and indecent (Brey 2008). A discussion of these controversial aspects of VR, however, is beyond the scope of this section.

► 11.3 ARTIFICIAL INTELLIGENCE (AI)

In Sections 11.1 and 11.2, we examined some challenges that online communities and VEs pose for our understanding of community and personal identity in the digital age. We next consider some implications that ongoing developments in the field of artificial intelligence (AI) have for our sense of self and for what it means to be human. AI research has already influenced some philosophers and cognitive scientists to question our conventional understanding of notions such as rationality, intelligence, knowledge, and learning. Some now also question what it will mean to be a human being in an era when some AI entities may exhibit a level of intelligence that exceeds that of ordinary humans. It is mainly from this vantage point that we examine AI-related controversies in Section 11.3. We begin, however, with a very brief description and overview of the field of AI.

11.3.1 What is AI? A Brief Overview

Sullins (2015) defines AI as "the science and technology that seeks to create intelligent computational systems." Sullins notes that AI researchers have aimed at building computer systems that can duplicate, or at least simulate, the kind of intelligent behavior found in humans. The official birth of AI as an academic field is often traced to a conference at Dartmouth College in 1956, which was organized by AI pioneers John McCarthy and Marvin Minsky. Since then, the field has advanced considerably and has also spawned several subfields.

In the 1950s and early 1960s, many AI researchers focused on developing software programs that could play checkers and chess with humans, pass calculus exams designed for undergraduate students, and solve problems that require a high level of human intelligence. The first AI programs were "problem solvers" comprising software code but included little or no hardware; many researchers during this period assumed that just as artificial flight had been successfully developed in ways that bypassed the need to emulate nature's way of flying, so too could AI be achieved without having to copy nature's way of thinking (i.e., via a physical brain). We now refer to the method of AI research conducted in that era, which focused on constructing a kind of "disembodied intelligence," as "classical AI," "symbolic AI," or "good old-fashioned AI" (GOFAI).

The classical AI approach was eventually criticized by researchers in the field who argued that human intelligence cannot be reduced merely to symbolic manipulation (captured in software programs) and that something additional was needed. For example, one school argued that an artificial brain with neural networking (that could "perceive" and "learn" its environment) was also required for a machine to learn and understand the world and thus potentially duplicate the way that humans think. Whereas the latter scheme in AI is often described as a "bottom-up" (or inductive) approach to machine learning, the classical/symbolic AI model is typically viewed as a "top-down" (or deductive) approach.

Another division in the field arose when a group of AI researchers argued that it was not critical to build machines that were as intelligent as humans (or that thought in the same way humans did); rather, they believed that a legitimate goal for AI research would be to develop systems that were "expert" in performing specific tasks that required a high level of intelligence in humans. For example, a system such as an "expert doctor" could be highly competent in diagnosing medical diseases, although it would be unable to perform any tasks outside that very narrow domain. (Recall our brief discussion of expert systems in Chapter 10, in connection with cybertechnology and work.) However, many other AI researchers believed that it was still possible to achieve the original goal of emulating (general) human intelligence in machines. Some of these researchers, including those working on the CYC project, use an approach that builds on classical/symbolic AI by designing software programs that manipulate large databases of factual information. Others, such as "Connectionists," have designed neural networks that aim at modeling the human brain, with its vast number of neurons and arrays of neural pathways, which exhibit varying degrees of "connection strengths." And some AI researchers focus on building full-fledged robots that can include artificial emotions as well.³²

One concern that arose early in AI research, which was more philosophical or sociological than technological in nature, had to do with how we might come to see ourselves in a world where machines would be as intelligent, or possibly even more intelligent, than humans. Weckert (2001) articulates this concern when he asks:

Can we, and do we want to, live with artificial intelligences? We can happily live with fish that swim better than we do, with dogs that hear better, hawks that see and fly better, and so on, but things that can reason better seem to be in a different and altogether more worrying category . . . What would such [developments mean for] our view of what it is to be human?³³

Of course, we can ask whether it is possible, even in principal, to build "machines" (i.e., software programs or artificial entities) that are "genuinely intelligent" and whose intelligence could rival and possibly exceed that of humans. Some critics argue that, at best, AI researchers would be able to build machines that merely simulate rather than replicate human intelligence. A spirited debate about this issue has ensued to this day, and each side has presented a series of arguments and "thought experiments" to defend its position.

11.3.2 The Turing Test and John Searle's "Chinese Room" Argument

In 1950, computer science "pioneer" Alan Turing confidently predicted that by the year 2000 a computing machine would be able to pass a test, which has come to be called "The Turing Test," demonstrating machine intelligence.³⁴ Turing envisioned a scenario in which a person engaged in a conversation with a computer (located in a room that was not visible to the human) was unable to tell-via a series of exchanges on a computer screen-whether he or she was conversing with another human or with a machine. He believed that if the computer was able to answer questions and communicate with the person at the other end in a way that the person there could not be sure whether this entity was a human or a computer, then we would have to attribute some degree of human-like intelligence to the computer.

While most AI researchers would concede that Turing's prophecy has not yet been fully realized, they also point to the significant progress and achievements that have been made in the field so far. For example, in 1997, an IBM computer program called Deep Blue defeated Gary Kasparov, then reigning champion, in the competition for the world chess title. And in 2011, another IBM computer program, called *Watson*, defeated two human opponents in the TV game show Jeopardy in a championship match. (This human-computer competition was viewed by millions of people around the world.)

Watson, like Deep Blue, is a disembodied AI, that is, a highly sophisticated set of computer programs. Unlike Deep Blue, which could be viewed as an "expert system" that is highly skilled at playing chess (but not necessarily competent in other areas), Watson was capable of answering a wide range of questions posed in natural language. Some believe that Watson's skills at least simulate human intelligence in the broad or general sense. But did Watson, in defeating its human challengers, also exhibit the skills necessary to pass the Turing test? And even if Watson could pass the Turing test, would that necessarily show that Watson possessed (humanlike) intelligence?

Some might argue that Watson was merely acting in a manner similar to the individual in John Searle's classic "Chinese room" scenario. In that scenario, a human who is a native English speaker but who understands nothing about the Chinese language is able to perform tasks that require manipulating Chinese symbols to produce answers to questions posed in Chinese. This person, who is not seen by anyone outside the room, receives questions from someone who passes them to him through an opening or slot. The person then consults a series of instructions and rules located on a wall in the room—all of which are written in English—that enable him to substitute the incoming Chinese symbols for other Chinese symbols in such a way as to produce correct answers to the questions asked. Once he has completed the task, he passes the answers back through the slot to a person waiting outside the room. That person might assume that the human who returned the correct answers understood Chinese. However, Searle (1980) argues that it is possible that the person understood nothing about the semantic meaning of the questions he received and the answers he returned; instead, this person had merely followed a set of syntactic rules (written in English) for manipulating symbols that happened to be in Chinese. In fact, the English-speaking person may not even know that the symbols involved are elements of the Chinese language.

Was Watson's behavior in the Jeopardy game analogous to that of the human in Searle's Chinese room—in other words, did Watson actually "understand" the meaning of the symbols (in the questions and answers) involved, or did Watson simply use a series of syntactic rules and cross-checking algorithms to manipulate the information stored in Watson's vast database to get the correct answers? It is not clear to what extent, if any, Watson could be said to have any understanding of natural language. So, even if Watson is capable of passing the Turing test (as originally posed), it would not necessarily follow, using Searle's argument, that Watson possesses human-like intelligence. In fact, some skeptics might argue that Watson is nothing more than a kind of (very broad) expert system, or perhaps some combination of expert systems, that behaves like an advanced version of (Apple's) Siri. Although Siri is capable of responding to many questions with correct answers, it is doubtful that people would be willing to describe Siri as possessing human-like intelligence. In the same way, then, we could argue that we need not ascribe genuine intelligence to Watson. Nevertheless, we can still see why many humans would feel a bit uneasy by the fact that a computer, or AI entity, had defeated two highly intelligent human beings in a championship match, even if it was only in a game show contest.

Unfortunately, an extended discussion of key questions involving both Watson and the Turing test, as well as an in-depth discussion of the history of AI itself, is beyond the scope of this chapter. AI's history, though relatively brief, is fascinating, and several excellent resources are available; so, fortunately, there is no need to replicate that discussion here.³⁵ We next examine some AI-related social and ethical issues affecting human-machine relationships, especially as they arise in the context of cyborgs.

11.3.3 Cyborgs and Human–Machine Relationships

So far, we have considered whether machines could, in principle at least, possess human-like intelligence. We have also considered how our answer to this question can affect our sense of what it means to be human. Next, we see how the development of cyborgs and the concerns it raises about human-machine relationships may also have a similar effect on us. We approach these concerns from the perspective of two distinct, but related, questions: (i) Are humans becoming more computer-like? (ii) Are computers becoming more human-like? We begin with (i).

Cyborgs and (AI-Induced) Bionic Chip Implants: Are We Becoming More Computer-like?

Many humans now receive nonhuman body parts, in the form of computerized chips, in implants. As we are implanted with more and more (AI-induced) bionic parts, are we becoming cyborgs? With so many bionic parts becoming available, some worry that humans and machines could soon begin to merge. Kurzweil (2000) has suggested that in the near future, the distinction between machines and humans may no longer be useful. And Moor (2005) believes the question we must continually reevaluate is "not whether we should become cyborgs," but rather "what sort of cyborgs should we become?" Clark (2003) suggests that we already are cyborgs because of our dependency on technologies. Are these views plausible, and why should we be concerned about the role that (AI-induced) chimp implants can play in questions about the human-cyborg dispute?

Our first line of response might be to note that implant technologies are by no means new. However, Weckert (2001) points out that while "conventional" implants in the form of devices designed to "correct" deficiencies have been around and used for some time, their purpose has been to assist patients in their goal of achieving "normal" states of vision, hearing, heartbeat, and so forth. Whereas these are all examples of "therapeutic implants," future chip plants in the form of "enhancement implants" could make a normal person "superhuman." Anticipating the kinds of concerns that enhancement implants will likely raise in the near future, Weckert asks:

Do we want to be "superhuman" relative to our current abilities, with implants that enhance our senses, our memory and our reasoning ability? What would such implants do to our view of what it is to be human?37

Some suggest that the current controversy involving implants can be framed in terms of an "enhancement vs. therapy debate." Moor notes that because the human body has "natural functions," some will argue that implanting chips in a body is acceptable as long as these implants "maintain and restore the body's natural functions." Moor also suggests that a policy framed along the lines of a therapy-enhancement distinction will appeal to many because it would "endorse the use of a chip that reduced dyslexia but would forbid the implanting of a deep blue chip for superior chess play."38 Such a policy would also permit a chip implant that would restore eyesight for a person going blind, but would not license implanting a chip for X-ray vision for a person with normal eyesight. But Moor also suggests that the therapy enhancement distinction might easily become blurred or confused if the only chip available to restore "normal" vision also happened (even if unintentionally) to enhance the person's night vision beyond the "normal" range.

Even if it turns out that a clear therapy-enhancement distinction regarding implants is possible, another important question remains: Who will be responsible for framing and enforcing policies affecting those implants? Without clear policies in place for the use of bionic implants, it may be difficult to prevent, or even to discourage, people from receiving enhancement implants that enable them to become either cyborg-like or superhuman? In Chapter 12, we examine some proposals for policies and ethical frameworks to guide developments in emerging technologies, which would also apply to bionic chip implants.

Our main concern in this section has been with the question of what it will mean to be human as more and more people elect to be fitted with bionic parts (and thus potentially also become more machine-like). Next, we consider the flip side of this cyborg-related question: What will happen as the AI entities we develop become more human-like?

Distinguishing AI Entities from Humans: Are Computers Becoming More Human-like?

Consider that some AI entities (e.g., "bots") in the form of avatars (typically graphical representations which, as we saw in Section 11.2.2, can exhibit human-like features and traits) already assist users in organizing their work schedules, reminding them of important scheduled meetings, arranging travel, and so forth. Also consider that personal digital assistants, and smart phones using voice-recognition programs (such as iPhone's Siri), now interact with humans on a daily basis. Even though they are merely virtual entities, some appear to exhibit human-like qualities when viewed on screens or when heard on electronic devices. Also consider that some avatars (and AI bots), which now act on our behalf, exhibit characteristics and stereotypic traits associated with humans in certain professions. For example, an avatar in the form of an AI "agent" designed to interact with other AI agents as well as with humans, such as a "negotiation agent," may look like and have the persona of a (human) broker.

After interacting with your human-like agent (bot or avatar) over a long period of time, is it possible that you might begin to act as if you are conversing with a real person? We can also ask whether it is possible that you might, after an extended period of time, begin to confuse some virtual entities (with whom you interact) with actual flesh-and-blood characters that those entities represent. For example, as virtual entities become increasingly more human-like in appearance, it may also become more difficult to distinguish between our interactions with some (physical) person's screen avatar and with an actual human represented by that avatar or virtual entity.

This confusion in interacting with artificial entities will likely become more exacerbated as we move from our interactions with virtual entities on screens (of computers and devices) to interacting more regularly with physical AI entities—namely, robots that appear to be sentient as well as intelligent. Consider that sophisticated robots of the near future will not only look more human-like but may also exhibit sentient characteristics; that is, these robots, like humans and animals, would (arguably, at least) be capable of simulating the experiences of sensation, feeling, and emotion. Robots and other kinds of AI entities of the not-too-distant future may also exhibit, or appear to exhibit, consciousness. Many AI researchers have questioned the nature of consciousness; for example, cognitive scientists and philosophers ask whether consciousness is a uniquely human attribute. Some also question whether it might be an emergent property—that is, a property capable of "emerging" (under the right conditions) in nonhuman entities, such as advanced AI systems.

Stanley Kubrick's classic 1967 film 2001: A Space Odyssey portrays a computer named HAL with higher-order thinking functions resembling human consciousness. In addition to performing ordinary computational tasks, HAL engages in sophisticated conversations with members of the spaceship's crew, plays chess, and even critiques the art (drawings) of one crew member. To take control of the spaceship from HAL, the sole surviving member of the crew removes the logic components of HAL's higher-order ("mental") functions: HAL is forced to undergo a sort of virtual "lobotomy." Of course, HAL is merely science fiction, but consider how far developments in AI have progressed since HAL's film debut in 1967.

Today, some researchers working with highly advanced AI systems claim to be on the verge of modeling a form of higher-order thinking in these systems, which might be viewed as also bordering on consciousness. Suppose that we reach general agreement that human consciousness can be understood as an emergent property and that conscious activity—similar to that of humans—emerges whenever a sufficient number of neurons (or "connection strengths" in neural networks) are present. Would this, once again, be something that causes us to reassess our conception of what it means to be human? Would it also mean that we should consider the question of whether to grant at least some level of moral consideration to sophisticated AI entities?

▶ 11.4 EXTENDING MORAL CONSIDERATION TO AI ENTITIES

We have seen that even if AI entities do not achieve full consciousness, many could still be capable of exhibiting (or at least simulating) other human-like characteristics such as rationality and sentience. For example, HAL has been described as an AI entity that was able to simulate human consciousness. Yet HAL was not sentient, since "he" had no body (and thus could also be said to have a form of "disembodied intelligence"). So it is possible that some AI entities (like HAL) may exhibit (or at least simulate) rationality and consciousness but not sentience, while other AIs may exhibit (or simulate) rationality and sentience but not consciousness. In either case, these AI entities would exhibit or simulate at least some human-like characteristics. An important question that arises is whether we are prepared to meet the kinds of social challenges these entities will likely pose and also whether we may need to grant these entities at least some degree of moral consideration.

The following scenario illustrates one way in which questions about extending our sphere of moral consideration can arise as we develop more human-like AI entities.

► **SCENARIO 11–4:** Artificial Children

The 2001 movie *A.I.*, directed by Steven Spielberg, is a story of an artificial boy, named David, who is physically indistinguishable from human boys and who appears to be capable of experiencing human-like emotion as well as displaying human-like intelligence. David also appears to need the love of human parents and "he" displays this need in a way that might seem genuine and convincing (to many humans). The human parents who adopt this artificial boy later decide to abandon him.³⁹

Does this "boy" deserve at least some kind of moral consideration? Do David's human parents have any clear moral obligations to their "adopted child"? Or do they have the right to discard "him" at their convenience, in the same way that they might discard a desktop computer that no longer serves their needs and thus is no longer useful to them? Of course, we could ask whether such an artificial child should have been developed in the first place. But given the likelihood that artificial entities similar to David will be developed, we need to seriously consider questions about which kinds of moral consideration, if any, those artificial entities may warrant.

11.4.1 Determining Which Kinds of Beings/Entities Deserve Moral Consideration

One reason why many humans believe they are morally significant, and thus deserve moral consideration, is based on the assumption that humans, unlike all other creatures and entities, are rational beings that possess intelligence.⁴⁰ In recent years, however, the (traditional) assumption that only humans have intelligence has been challenged on two distinct fronts. On the one hand, research in animal behavior suggests that many primates, dolphins, and whales exhibit skills that suggest some level of intelligence (while, on the contrary, some humans either cannot or are no longer able to exhibit these skills). On the other hand, and more importantly for our purposes, we have seen that developments in the field of AI suggest that some highly sophisticated computers can exhibit forms of intelligence and rational decision making, which were previously thought to be possible only in humans.

If some AI entities are capable of exhibiting (or simulating) rationality and intelligence (and possibly even consciousness)—characteristics that traditionally have been reserved to describe only humans—it would not seem unreasonable to ask whether these entities might also warrant moral status. And if some of these entities can exhibit (or simulate) human-like emotion and needs, as in the case of the artificial boy in the movie AI, would that also be a

relevant factor to consider in understanding and addressing concerns about moral consideration for AI entities? An important question, then, is whether we will need to expand the conventional realm of moral consideration to include these entities.

Prior to the twentieth century, many ethicists and most lay persons in the Western world generally assumed that only human beings deserved moral consideration; all other entities animals, trees, natural objects, etc. - were viewed merely as resources for humans to use (and misuse/abuse) as they saw fit. In other words, humans saw these "resources" simply as something to be used and disposed of as they wished, because they believed that these resources had no moral standing and that we had no obligations toward them.

By the mid-twentieth century, the assumption that moral consideration should be granted only to humans had been challenged on two distinct (though not altogether unrelated) fronts. One challenge came from animal rights groups, who argue that animals, like humans, are sentient creatures and thus capable of feeling pleasure and pain. Based on this comparison, proponents for animal rights have argued that we should also grant ethical consideration to animals, in which case it would be morally wrong for humans to abuse animals or to treat them simply as resources.

On a second front, some environmentalists made an even bolder claim, arguing that we should extend ethical consideration to include new "objects," or entities. Jonas (1984) argued that because modern technologies involving atomic and nuclear power have presented us with tools of destruction that could devastate our planet on a scale never before imaginable, we needed to expand our sphere of moral obligation to include "new objects of moral consideration." These "objects" included natural objects such as trees, land, and the environment itself, as well as abstract objects such as "future generations of humans" that will inherit the planet.

In the past 50 years or so, our thinking about both who and what should be included in the sphere of moral consideration has evolved significantly. We have moved from a traditional moral system that granted consideration only to human beings to one that at least debates whether animals, land, and the entire biosphere deserve moral consideration as well. Do we once again need to expand our sphere of moral consideration to include "new objects"—that is, nonnatural or "artificial objects" such as sophisticated AI entities?

Floridi (2002) has suggested that we need to grant some level of moral consideration to at least certain kinds of informational objects or entities. Initially, one might find Floridi's assertion strange, perhaps even preposterous, but we have seen that some sophisticated AI entities already exhibit a form of rationality that parallels that of humans. The question that concerns us here is whether these artificial entities merit moral consideration because they, like humans, have rational abilities. If our primary justification for granting moral consideration to humans is based on the premise that humans are rational entities, and if certain artificial entities also qualify as "rational entities," then we can make a compelling case for granting at least some moral consideration to them. For example, even if they do not qualify as full-blown moral agents (as typical adult humans do), they may nevertheless meet the threshold of what Floridi calls "moral patients."

11.4.2 Moral Patients vs. Moral Agents

In Floridi's scheme, moral patients are "receivers of moral action," while moral agents are the "sources of moral action" (capable of causing moral harm or moral good). Like moral agents, moral patients enjoy moral consideration and thus have at least some moral standing; unlike moral agents, however, moral patients cannot be held morally responsible for their actions.⁴¹ While animals may not be moral agents (i.e., morally accountable for what they do or fail to do), they can nevertheless qualify as moral patients that deserve moral consideration by humans.

One reason for viewing them as moral patients is because of their ability, like humans, to feel pain and suffer emotionally. (Consider that many pet owners relate to their pets in ways that suggest they are moral entities, not only when they try to protect their pets from harm and suffering but also when they reward and punish their pets' behavior.) So if we extend this analogy, it would seem to follow that AI entities that exhibit sentience, independent of whether they also happen to exhibit a high level of rationality, could qualify as moral patients and thus warrant some moral consideration.

Consider again the example of the artificial boy examined in Scenario 11-4. Would it be plausible to grant "him" at least some degree of moral consideration — that is, as a moral patient because of his "emotions" and "needs," even if this "boy" failed to satisfy a high threshold for rationality? We have noted that animal rights proponents argue that animals deserve moral consideration because of their ability to suffer pain, irrespective of any rational capacity they may or may not also have. So it would seem reasonable to ask if we can extend that analogy to apply to sentient AI entities such as the artificial boy. In this scheme, then, AI entities that exhibited (or simulated) either rationality or sentience, or both, could qualify for moral consideration.

A more interesting question, however—and one which we cannot answer here—is whether AI entities that (at least appear to) exhibit consciousness could also qualify as full, or even "partial," moral agents. We briefly discuss the concept of moral agency in Chapter 12, in connection with our analysis of autonomous machines.

We conclude this section by acknowledging that more questions have been raised than answered. The critical question of who/what (in addition to humans) deserves moral consideration (and if so, to what extent) is extremely complicated, as well as very controversial. While criteria such as rationality and sentience (in connection with an agent's ability to act in the universe) have played key roles in answering this question in the past, other characteristics such as autonomy and free will are also generally considered relevant criteria for possessing full moral agency. Note that we have not argued that any current AI entities have either free will or autonomy.

In Chapter 12, we examine a cluster of AI-related questions from a very different perspective. There, we consider whether it is possible to construct "moral machines" or AI entities that are capable of making moral decisions. For example, can those machines be embedded with (software) code that will enable them to make what Wallach and Allen (2009) call "good moral decisions"? We will see that this is an important "practical question," as we develop autonomous machines that act more and more independently of human oversight. Our focus on AI issues in this chapter, however, has been on questions that can be viewed as more theoretical and philosophical in nature—especially given our focus on AI's impact so far for the question of what it means to be human.

► 11.5 CHAPTER SUMMARY

In this chapter, we have examined three diverse technologies that pose challenges for our conventional understanding of community, personal identity, and what it means to be human in the digital era. In particular, we have seen how SNSs have affected our traditional notion of community, and we considered some arguments for the pros and cons of online communities. We then looked at some VR applications, including online video games and MMORPGS, and considered some of their ethical implications as well as for our understanding of personal identity in the digital era. Next, we examined the impact that key developments in the field of AI have had so far for our sense of self and for our sense of what it means to be human. Finally, we considered whether we may need to expand our conventional framework of moral obligation to include at least some AI entities.