Playing the Game: Cheating, Loopholes, and Virtual Identity

Phillip J. Brooke[†] Richard F. Paige[‡] John A. Clark[‡]
Susan Stepney[‡]

[†] School of Computing,
Communications and Electronics
University of Plymouth
Drake Circus, Plymouth
Devon, PL4 8AA, U.K.
phil.brooke@plymouth.ac.uk

[‡] Department of Computer Science
University of York
Heslington, York
YO10 5DD, U.K.
{paige,jac,susan}@cs.york.ac.uk

October 18, 2004

Abstract

A broad range of interactive and distributed systems are essentially virtual worlds; these include examples such as multiplayer games, and even operating systems. They enable the formation and maintenance of virtual societies, which must be *healthy* in order to be prosperous and useful. We describe properties, inspired by writings on law and psychology, that we use to define the notion of *fairness*, which is an essential characteristic of a healthy society. By using multiplayer gaming as a running example, we discuss how a fair virtual society will interact with its real-world counterparts, and outline how one might choose to detect and deal with transgressors who violate rules designed to enable fair interaction and prohibit cheating. This is a conceptual paper, and raises a number of issues and problems that must be considered when designing virtual worlds. Our aim is to develop guidelines for the design of fair virtual societies.

1 Introduction

A virtual world exists purely within the digital domain, yet enables interactions both between and completely within the real or virtual world. The concept of a virtual world is frequently confined to those notions that have captured the public imagination, such as those involving multiplayer games like Half-Life or Quake. However, we can extend the virtual world concept to include auction and trading systems, computer-supported cooperative working (CSCW), and even operating systems when we observe that these systems carry out actions on behalf of, and often with some concept of an identity, for users in the physical world.

The characteristics of a virtual world are therefore: an underlying computing infrastructure (*e.g.*, hardware networking technologies, software, and a communications medium), virtual entities that

enable interaction (*e.g.*, avatars of players in a game), presentation information (*i.e.*, how the world will be shown to users), and rules that constrain the virtual world and govern basic interactions with it and within it.

If time, effort and resources have been spent in providing the computing infrastructure and designing the rules for a virtual world, then we might reasonably hope that it is used as intended, and a virtual *society* will evolve, complete with participants, as well as physical and social rules. This type of evolution requires that people will wish to join and remain engaged in the society. We assert that this requires a *fair* society for an acceptable level of healthiness to be achieved.

Most participants in virtual and non-virtual societies have an informal idea of what it means for the society to be fair, though this is often not stated formally or explicitly. Most commonly, for a society to be considered fair, it must make it difficult to *cheat* – *i.e.*, to break the stated rules that govern interactions – and has an effective and visible process of detecting and punishing cheating. We argue that the notions of fairness and cheating in virtual societies – particularly, but not exclusively multiplayer games – need to be defined precisely in order for more inclusive and robust virtual societies to be deployed. A further reason for more rigorous definitions is that violation of the rules of a virtual society can occur both inside and outside of the virtual world itself, and it is at the border between the real and virtual where the usual definitions of cheating, reputation, privacy and identity start to break down.

2 Motivation

Before we even attempt to precisely define fairness in the context of virtual societies, we ask the question: why do we want fairness? Consider the context of online games, which we use as a running example throughout the paper. In these games individuals will interact in a virtual world to accomplish individual and, often, team goals. The interactions between individuals can, under certain conditions, lead to the development of a virtual society designed to enable fair gameplay. Each individual will control one or more virtual identities, also known as *avatars*. Avatars are the mechanism through which the rules – and thereafter the fairness – of the virtual society are generated and stress-tested.

Fairness in an online game is desirable in order to encourage people to participate: they should want to join and continue to play the game. Games that are unfair – or that are perceived to be unfair – are undesirable and are likely to lose participants. Some best-selling games have developed reputations for online cheating; this reputation anecdotally deters both new and established players [4]. Societies based on such a virtual world cannot be considered healthy or viable in both the medium- and long-term.

In order to define fairness more precisely, it is useful to define the context in which the term will be formalised. We define a *society* – virtual or not – in the terms used by Rawls: it is a self-sufficient association of entities that recognise rules of conduct as binding, and for the most part, they act in accordance with those rules [1]. We assert that a virtual society is an evolution of a virtual world, one in which interactions must take place and where rules governing social interactions have evolved over time in response to interactions and feedback. A virtual society cannot entirely be constructed, whereas a virtual world – underpinning a society – can be designed and implemented.

In general, we want to encourage participation in the virtual society. We would like such a virtual society to be *healthy*, where we may define healthiness as the number of active users (or using some other valuation function, such as how many users renew their subscription in a pay-

to-play game). Healthiness and fairness are inextricably related; however, healthiness is usually an observable property of a society, whereas fairness, in general, is not. Healthiness for a virtual society is defined in terms of the requirements of its stakeholders: investors may want to make money from subscriptions; participants may want to enjoy their experience. There will likely be conflicts between the different stakeholders and their perception of healthiness that must be resolved through negotiation.

Rawls specifies principles of fairness, which essentially say that one has to assign rights and responsibilities to individuals in a society in order to distribute the benefits and burdens that come with participation. All individuals and groups should accept these rights and responsibilities assuming that everyone else is following those principles. Moreover, the social institutions have to do the same. Fundamentally, when we design mechanisms to enable a virtual society, we are trying to develop buy-in from its participants and by potential participants. The rules should advance the good of those taking part, while imposing necessary responsibilities on the members of that society.

Referring again to Rawls's text, there are the principles of fairness, where an entity is required to do their part as defined by the rules of a society when

- the institution is itself fair; and
- each entity has voluntarily accepted the benefits of the arrangement or taken advantage of opportunities to further their interests.

The definition of rules is dependent on the nature of the virtual society concerned. It requires an understanding of the motivation and goals of individual participants and goals of stakeholders who are not participating in societal interactions directly, *e.g.*, financiers. Some imposed goals may encourage or even require cooperation between avatars; others may direct an 'everyone for themselves' goal. There may even be conflicting goals, but it is not the role of the implementation of the virtual society to resolve these conflicts: it is essentially a question of priorities for the individuals who are participating.

An example may help to clarify the issue of balancing rights and responsibilities. One common problem in team-based 'first person shooters' (FPS) is that of the 'team-killer', the user who directs their avatars to attack members of their own team. Team-killers are not discharging their responsibilities to the other members of their team, but are endeavouring to accomplish goals of their own.

We can immediately make a broad classification of the rules of a virtual society into two categories:

Physical Those rules that are non-negotiable, such as the law of gravity in the real world. These are thus the rules of the world itself.

Social The rules that potentially emerge from the behaviour of the entities in the society, or which are explicitly stated beforehand. For example, in a peer-to-peer system, we may require that everyone should contribute content. By analogy, these are the 'rules of the behaviour of people', and as such dictate who will actually be able to fully participate, and how avatar-to-avatar interactions will occur.

The social rules are intended to encourage the common good to all of individuals, teams or groups (some declared, some undeclared), and generally all of the virtual society.

It is worth pointing out that there are both stated and unstated rules; the latter may eventually be formalised if they are shown to be useful in promoting fairness and healthiness. We return to this point later when we discuss the design of virtual worlds.

3 Classification of Virtual Societies

We have broadly classified the rules of virtual societies into 'physical' and 'social' in the previous section. Since many rules depend on the context, *i.e.*, the details of the virtual society concerned and its intended usage, it is useful to classify these systems, with the intention of using this to provide specialised rules for specific types of systems.

Firstly, we may consider whether the system is designed to enable human-to-human interaction, or human-computer interaction. We may then classify the systems by application. Virtual societies can be developed from:

- Immersive worlds with avatars, *e.g.*, multiuser dungeons (MUDs), first person shooters.
- Information exchange systems, *e.g.*, peer-to-peer (P2P), computer supported cooperative work (CSCW), chat rooms, and search engines such as Google.
- Monetary exchange, *e.g.*, auction houses such as Ebay, bartering systems (credit unions and online classifieds listings), or PayPal. These can be subsumed within the previous category if money is taken to be just another kind of information.
- Shared execution environments (operating systems); and, on a wider scale, Grid computing.

Fairness can be interpreted differently in each kind of system. Before we discuss this, we consider the complication introduced by the relationship between the real world and the virtual one.

4 Real and Virtual Worlds

We illustrate the relationships between both types of world, and the related entities in the metamodel shown in Figure 1. In this figure, players inhabit the real world while controlling one or more avatars in the virtual world. Players may group themselves into zero or more collections, or teams, as may the avatars. There is no necessary relationship between real and virtual world teams; a typical example is when two physically co-located people participate in two different virtual teams in a first-person shooter game such as Firearms. Both players and teams have reputations, which are earned and influenced by behaviour. By participating in a real or virtual society, the players and avatars (respectively) establish other properties. For example, both have a set of capabilities: for the player, these include hardware and network bandwidth; for the avatar, they are a description of its existence in the virtual world (*e.g.*, its strength or appearance). Both players and avatars are expected to obey the social rules pertinent to their environment and will develop a reputation amongst the other entities based on their compliance with those rules and the implemented simulated laws. A hack might be used to manipulated the hardware/software platform on which the world is hosted, in order to unduly influence outcomes. Finally, the avatars' conduct might affect an automated 'scoring' system (*e.g.*, 'alignment' in MUDs). This is an observable characteristic of the avatar's emergent behaviour, and is a calculation of an individual or team's reputation.

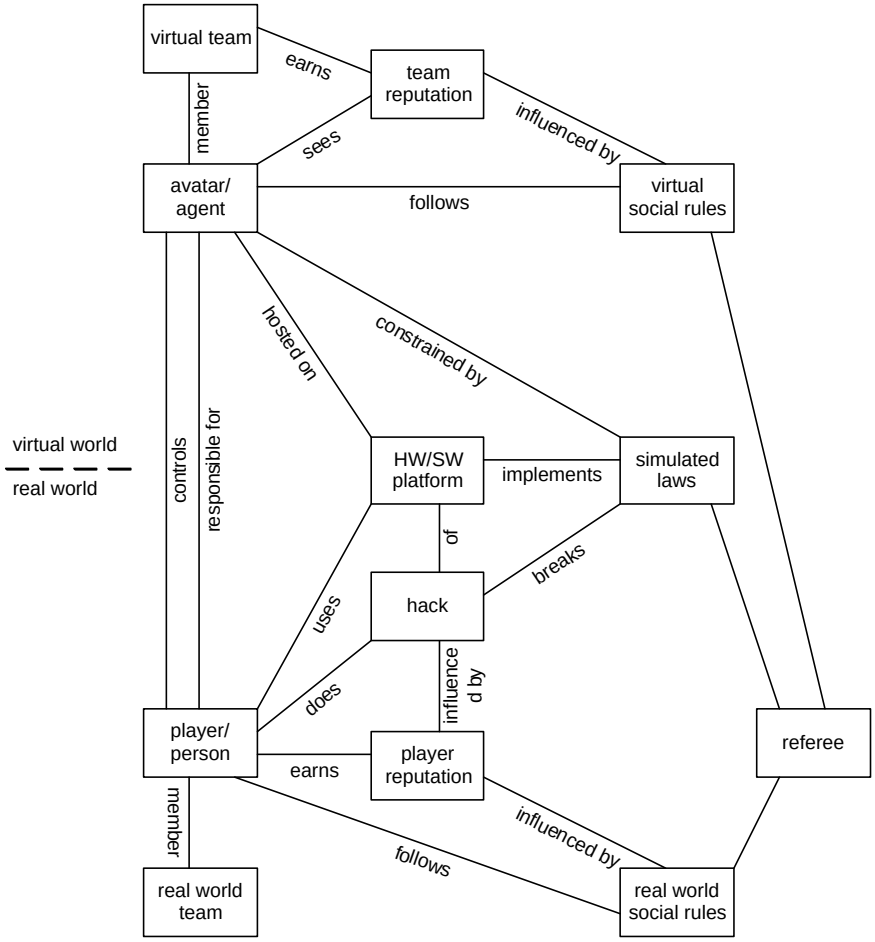

Figure 1: Metamodel of the relationship between real and virtual worlds

A typical problem in virtual societies occurs when entities make claims about their capabilities which differ from their real capabilities. Having a social reputation and a calculated reputation are two different valuations of the trust that entities might place in another. We will address this issue later when we discuss trust further, as calculated reputation can be used as the basis for consistency management and anomalous behaviour detection, as well as a means to reduce the likelihood of virtual identity theft.

5 Notion of Identity

So far, we have discussed real and virtual worlds and societies which include players and avatars as participants. As we attempt to formalise rules (whether physical or social) we increasingly find ourselves attempting to make individuals accountable for their actions so that we can judge fairness (or detect cheating). This requires some notion of identity.

Two definitions of ‘identity’ (drawn from the Collins English Dictionary via WordReference [7]) are

1. the state of having unique identifying characteristics held by no other person or thing

2. the individual characteristics by which a person or thing is recognised

The characteristics or atoms for recognition of identity are dependent on context; in an on-line game it will likely be based on atoms such as physical (virtual) appearance, a simple name, game-play abilities (*e.g.*, speed, strength, skills), and similar features. Identity is lost when one or more of the atoms that allow for recognition are separated from the entity of interest.

The Oxford English Dictionary [3] includes a definition of identity saying "The sameness of a person or thing at all times or in all circumstances; the condition or fact that a person or thing is itself and not something else; individuality, personality." However, the definitions from the Oxford English Dictionary do not reflect the fluid nature of virtual worlds. The definition of identity needs to be modified in the virtual realm in order to take into account multiple virtual personae [8].

An individual player might control multiple avatars, thus having a single (real) physical identity, but multiple virtual identities. We might rebut this by saying that individuals have multiple personae in the real world: an individual's 'work' life may differ from their 'home' life. However, such real world personae cannot be separated. Moreover, one virtual entity may be controlled, at different times, by different persons; consider the case where avatars need to change roles in a team-based game, and players control roles rather than avatars directly. Thus there is a many-to-many relationship between physical and virtual identities.

In general, we use identity to distinguish individuals from others, and to identify them when we encounter them again. It is useful to ask what we use identity for in virtual societies:

Authentication which is the process by which we establish someone's identity in a system.

Accountability of an individual's actions requires the ability to name them (perhaps we want to 'name and shame' transgressors in a system that uses peer pressure to enforce its social rules).

Billing If we wish to charge (in the real world) for access to an interactive system, we need to identify their legal entity in the real world. This might be by a credit card number.

Maintaining relationships We would hope that in a long-running persistent system we can maintain our 'account'. This means that the interactive system must recognise us when we return to it.

A generalised notion of identity in an interactive system might be a tuple comprising

- the user's characteristics
- the owned avatar's characteristics
- the hardware enabling the interactions between the player and the system.

The cyberpunk view considers identity to be fluid, and regards it only as a container for an individual's reputation. The intention is to allow one to transfer reputation from one system to another. However, reputation in one type of application is not necessarily of value in another type (compare FPSs with auction systems). Reputation is also related to trust: an individual might be have a reputation for reneging on deals (so on encountering this individual, we might be more wary of negotiations). One example is when one company purchases another, not only to acquire the company's product line, but also its reputation as a trusted vendor of that product line. Another example is where eBay allows criticism of traders: essentially saying 'this person cheats on deals'.

This suggests that trust and reputation are bound up with identity. These are varying characteristics, which appears at odds with our initial definitions of identity. It is clearly useful to be able to record these characteristics.

Transparency of these mechanisms is important in building trust and reputation. It is undesirable for the author of a book listed in Amazon to masquerade as an unconnected reviewer; similarly it is undesirable for a reviewer to masquerade as the author of a book listed on Amazon. Furthermore, we would not want to allow identity theft. Including earned reputation as part of the identity contributes further value, and when this is used in combination with a calculated reputation, it can make it more difficult to steal virtual identity.

Ultimately, the definition of virtual identity must be driven by the context in which it is used.

6 Fairness and Cheating

What does 'fairness' actually mean? In a legal context, Rawls has equated fairness with justice. We might view this interpretation as meaning that cheating is not possible. If cheating is possible it is detected either by the institutions or the participants, and punished appropriately and visibly by the institutions in the system. Applying Rawls's concepts, we require that a fair virtual society has a precisely stated set of rules that participants agree with and generally intend to adhere to. Each participant should believe that the others are also adhering to the rules, thus the necessity for visible detection and punishment when the rules are (inevitably) broken. Visibility of punishment is essential in order for participants to believe in the inherent fairness of the system.

As well as explicitly stated rules, we will encounter emergent rules. These evolve with the society and only become apparent with the conventions that develop over time. A real world analogy is this: we might design an estate with planned paths in it; or we might wait and see which routes people wish to walk, and then put the paths matching those routes. The latter paths are 'emergent'. Emergent rules are sometimes made concrete in order to deal with undesired *loopholes*, interactions that were not predicted by the designers of the society and which are therefore not prohibited or captured by any of its rules. Peer pressure often has a significant role to play in deciding whether or not to implement an emergent rule.

Fairness requires visibility in the institutions' application of the rules. Williams discusses the idea that fairness is not directly observable: we can only observe a process and the results of that process, but results alone cannot themselves possibly determine fairness [6]. We must ask questions about the process to determine fairness; for example, in the process of game play we might ask 'Did the players play according to the rules of the game?', 'Did the administrator (referee) apply those rules in an unbiased fashion?', 'Were penalties evenly exacted for infractions?'. Williams continues: in deciding whether a rule is a good rule, in terms of evaluating its contribution towards the healthiness of society, we should not evaluate it in terms of the likely result it will produce for particular entities under specific circumstances, but instead in terms of long-run opportunities produced for a wider group of entities in a wider set of circumstances (the 'greater good').

Physical rules are imposed by design of the virtual world and enforced and implemented by the server(s). These are dependent on the type of and requirements for the virtual world concerned, and are often derived by analogy with the real world. For example, gravity is often constant in a game world — but not always; scenarios exist in games where gravity is lessened or turned off for novelty and to improve gameplay.

Social rules include both intended fixed rules that are implemented with the virtual world, and

which are implemented by the server and the client-server relationship. There are also emergent rules. As a way of more finely identifying types of social rules, we categorise types of cheating we might observe, where cheating is defined as gaining some unfair advantage over other participants. Pritchard discusses the issue of hacking and cheating in online games, and some of this discussion derives from that work [5].

- **Collusion and cooperation:** where several (real world) players physically nearby (or in out-of-system communication) cooperate to exchange information which their avatars could not have found directly. Some examples include: two co-located players exchanging advice while playing in the same virtual world; or exchange of money in wargaming in order to guarantee victory in a battle, *i.e.*, paying to fix the outcome of a wargame.

  More generally, knowledge from outside the system can be exploited. But is that necessarily unfair? For example, in a game which has multiple rounds over the same map, does detailed knowledge and experience of the game map constitute unfairness, or is it simply the mark of an experienced player? It depends on the stated rules. For example, some games state that players should all be of similar experience and capabilities (and as such there are separate novice and expert scenarios). An experienced player engaged in the novice virtual world would lead to unfairness; it would constitute cheating if there were explicit rules set up to prevent this type of behaviour. Both cheating and unfair behaviour of this kind can be addressed by partitioned scoring systems. Some games designed in this way give novice players a substantial bonus when they defeat an experienced player; conversely experienced players gain considerably less for defeating a novice player. Partitioning game scenarios based on experience is analogous to protectionism in business, and as such it is worthwhile to revisit rules enabling partitioning from time to time, to ensure that they are promoting healthiness and fairness.

  In the case where avatars or users cannot prove themselves to be trustworthy, it may be desirable to revoke levels of privacy, so that other participants can observe them and their actions more freely, thus enabling a form of peer pressure. This could also be generalised to revocation of privilege, *e.g.*, for repeated cheating behaviour.

- Exploitation of **loopholes**, which are usually either implementation faults, ambiguities or omissions in the stated rules, or measures intended to cope with the imperfect real world. Loopholes may be legitimate (rules will not be revised or extended to eliminate interactions to exploit them), or illegitimate; the decision as to legitimacy depends on the stakeholders and a negotiated settlement that is often - but not always - carried out outside the virtual world.

  For example, multiuser dungeons will often put an avatar into a ‘safe’ state when the controlling terminal goes away; this is to cope with network connections becoming faulty. This can be exploited by a player who realises their avatar is about to come to harm: they simply disconnect their client. One solution to this is based on the notion of stability in formal languages. An avatar can only be put into a ‘safe’ state at ‘stable’ points in the execution of the system. The exact definition is difficult, and depends on what the virtual system is simulating. A different solution is used by Xpilot and many other games: a request to pause play does not take immediate effect, but is delayed.

  An example of a legitimate loophole is the Fosbury flop technique in high jump; it was legitimised by use, and is also an example of a loophole emerging due to technology changes, *i.e.*, high jump sandpits being replaced by foam mats. Other examples arise with web-based

sports fantasy leagues, and using reserve roster slots to keep players from being taken by other teams. There are no rules to prevent this, or to condone such use of reserve slots; it is common practice.

Another type of loophole might be exploited to break the virtual physical rules; this refers to the hacks in the metamodel of Fig. 1. However, for the rest of this paper, we will assume that the server and protocol are correctly implemented and are resistant to attack (although this is, in practice, not a reasonable assumption).

Denial of service attacks can plausibly be made against an interactive system server. These could reduce the ability of the server to produce a timely response (or even produce a response at all) for other participants. It is thus very difficult to eliminate loopholes in a rule set; revision and feedback must take place to identify omissions and refine rule sets.

- **Impersonation and identity theft:** pretending to be someone you are not, or attempting to cause confusion about identity, are examples of masquerading and social engineering. Making excess (or inferior) claims can be used to gain advantages (*e.g.*, an avatar pretending to be more capable than it is so as to deter attack): here, the social aspects are exploited by a player to gain advantages for their avatar.

  Virtual identity theft is a variant of impersonation whereby a player in the real world acquires control of avatars by hacking, exploiting loopholes, social engineering, and other means. Its possibility is a particularly dangerous form of unfairness, as it can lead to serious financial and personal repercussions. In an online game the risks of virtual identity theft can be reduced, by making hacking difficult, and by using calculated social reputations – effectively as a form of intrusion detection.

- In the most general case, we cannot distinguish easily between an avatar controlled by a human, and an avatar controlled by a computer. These latter avatars are sometimes called 'non-player characters' (NPCs), compared to the former 'player characters' (PCs). Hybrid-controlled avatars are sometimes considered as unfair behaviour, for example, a human player in a FPS who relies on an 'aimbot' to target the opposition more effectively than the player alone could.

  Although this might be considered to be a technical problem, if it is difficult to distinguish between PCs and NPCs by observing behaviour, then it is also difficult to observe an aimbot in use.

- We might consider unfair conduct to be the most pure form of social unfairness. Such conduct is located entirely in the virtual world and does not actually breach any physical rules or stated social rules at all. The classic example is 'camping', where an FPS avatar controls a single area by waiting for opponents to appear and ambushing them. Some players consider this to be cheating; others consider it to be fair tactics that are a result of exploiting the design of the virtual world.

This last type especially illustrates that fairness can be seen as subjective. It depends on the society and its rules, but also on the individual participants. One player's cheating is another's clever tactics. However, this is not helpful for building fair virtual worlds and societies. The key point is that rules for fairness should apply equally to all participants. Consider, as a topical example, spam. From the perspective of evolving fair social rules for e-mail, we have to treat all participants as both senders

and receivers of e-mail, and design the rules that enable everyone to make productive use of limited e-mail resources.

Commerce and negotiation might not work well without some elements of unfairness: each person has a distinct objective and may have access to relevant information of varying quality. The scarcity of a resource drives commerce, and is arguably a source of unfairness. Scarcity can also be linked to monopolies: should a single avatar, player or group have a monopoly over a resource that other players need, such as ammunition, water, or bandages? The changing scarcity of different types of resources means that rules may have varying relevance over time, and need revising and renegotiating with the participants in the society. Furthermore, should players be allowed to trade (in the real world) for artifacts that their avatars or teams possess? At least one online game has problems where virtual artifacts are sold in the real world, *e.g.*, on eBay, allowing players to obtain capabilities beyond their real abilities.

Further, where rules are made, it should be possible for a participant to determine if a particular course of action will break the rules before they embark on that course of action. This suggests that a small number of simple rules (with simple interactions) is required, so that their combined effect can be partially predicted more easily than with complex rules.

7 Deviation from the Rules

In a sufficiently rich environment where individuals can deviate from the rules, some form of detection and sanction is needed to ensure a fair society. It is impossible to violate the law of gravity in the real world. To automatically detect breaches of social rules in a virtual (or indeed real) world is more difficult. There are parallels with intrusion and anomaly detection in computer security here. One could envisage developing profiles of avatar behaviour over time and contrasting these with 'standard' profiles of expected behaviour for particular classes of users (*e.g.*, non-player characters versus real users).

This might partially manage the issue of collusion: a new avatar might be expected to take a particular route through a map. An avatar cheating by following information provided by a colluding player would not conform to that profile by taking less time or using a 'hidden' route. We can use honeypots to develop this detection. Spitzner suggests that honeypots provide a single component of a larger security architecture [2]. We may even have avatars or players acting as *agent provocateurs*, designed to force other avatars to engage in illegal behaviour by following a pre-programmed sequence of behaviour and interaction.

The other form of detection available is that implemented by the participants in the virtual society. Essentially, this relies on the ability of avatars to penalise other avatars for (perceived) infractions of the social rules. The avatars can view each other dependent on the virtual physical rules. Peer pressure can then be used to educate or compel transgressors, and also to guide the production of new rules. This of course generates conflicts with requirements for privacy, and systems will all differ in terms of whether privacy or fairness needs to be emphasised.

The most obvious sanction is to remove an offender from the society completely. But which is the offender? The avatar or the player? (We can complicate this further by considering the case where a single avatar is controlled by multiple players.) This requires a workable notion of identity. It is no good barring an avatar 'name' if players can create new avatars with a different name. (Compare this with the similar issue concerning banning by IP address and ISPs offering dynamic IP addresses.)

Other sanctions might relate to the ability of the avatar to operate in the society. Some systems will use this as an incentive to good behaviour, by only making more useful features (or abilities or skills) available to those with a history of good behaviour. In other words, we are making features contingent on reputation. An example is calculated reputation. A specific example is 'alignment' in a multi-user dungeon, whereby a character's alignment varies on their actions. Killing 'innocent' or 'good' avatars results in a more negative or 'evil' alignment. If reputation changes, then the features available to the avatar may change; moreover, this change will be visible to other player and avatars. An artifact of this approach is that it relates the physical rules to a valuation of the compliance with the social rules. There are no examples of this in the real world, so it is purely a mechanistic interaction of levels of abstraction in an interactive system.

Using reputation as a measure of compliance with social rules also addresses (in part) the issue of spent convictions. This is the concept that after a suitable time of acceptable behaviour, past transgressions are forgotten.

A notable form of social rule breaking is civil disobedience. A rule considered to be unfair or inappropriate may be protested (or simply ignored) by avatars breaking that rule. This may make a case for revising that particular rule; more generally, social rules evolve and emerge and need periodic review. Jurisdictions have some mechanism for creating new rules which may themselves dispose of, or interact with older rules. Rules will of course have to be modified over time, whether due to changing patterns of behaviour, or bug-fixes, or simply to make the virtual world seem like a new, different, and more exciting one – *i.e.*, to encourage players to continue to spend their money on the game. This is already the case with expansion packs or new scenarios for online games.

8 Conclusions

We have described the relationship between a virtual world and its real-world counterpart. We have discussed some properties inspired by the concept of fairness as a means to ensuring both the initial and ongoing healthiness of the virtual world.

Furthermore, we have noted that laws can be placed in one of two broad categories, specifically 'physical' (*e.g.*, the law of gravity) and 'social' (*e.g.*, don't kill other virtual players). While physical laws cannot be broken (except through error of, or attack on the computing infrastructure supporting the virtual world), social laws can be easily transgressed. Responses to these may be automated, but in the case of the many rules which are ambiguous, subjective or simply difficult to detect automatically, the society itself may choose to deal with those problems. The analogy with intrusion detection is an apt one here.

So now, we suggest some guidelines for those developing virtual worlds and virtual societies:

- Clearly specify the goals of the system in terms of a classification of the actors or archetypes who will want to use it.
- Clearly define the virtual physical rules. Implement the server as correctly as possible. Where relevant, draw analogies of virtual physical rules to the real world.
- Define a small number of initial social rules to drive the desired goals. While it may be tempting to leave all social rules (and some physical rules) unwritten, so as to determine them via emergent behaviour, it is necessary to constrain social interaction to a degree to promote initial subscriptions and short-term health.

- Both the design of the system and the behaviour and desires of the participants determine the emergent rules.
- A small number of simple rules (with simple interactions) is better than a large body of laws.
- Enable the participants (particularly the avatars) to make, review, revise and enforce the social rules.
- Provide mechanisms for modelling, measuring and viewing reputation as a part of observing the process of evolving rules.

An issue that remains unresolved is the relationship between the fairness of the virtual society, and the fairness of any business models that guide how participants might be charged for making use of the virtual society's infrastructure. Clearly, fairness in an online game is not the same as fairness in the gaming company's charge and cost model for play. But the two must be related, and it remains to be seen how to best engage feedback on unfairness that arises from the virtual society in any model of fairness in any business model. This will be of importance for providers who wish to generate revenue from selling and maintaining the infrastructure for virtual societies. More generally, there will be conflicting goals in the design of virtual worlds and societies from the different stakeholders. By being precise in specifying rules, and by keeping the rule set small, we can more easily identify conflicts and thereafter determine where problems might arise in the rule implementations.

There are inevitably different *cultural* views of what constitutes fairness. The Western, predominantly capitalist and individualist view is at odds with many others. An intriguing issue will be the design of truly international fair virtual worlds and societies, where fairness and rules can be negotiated within the setting of a virtual world.

References

[1] J. Rawls. *A Theory of Justice*. Oxford University Press, 1971.

[2] L. Spitzner. *Honeypots: Tracking Hackers*. Addison-Wesley, 2002.

[3] Oxford English Dictionary. http://www.oed.com/

[4] A. Park. Battle.net cracks down on cheating…again. *Gamespot News*, `http://www.gamespot.com/pc/rpg/diablo2/news_2871872.html`, June 2002.

[5] M. Pritchard. *How to Hurt the Hackers: The scoop on Internet cheating and how you can combat it*. Information and Security Bulletin, February 2001.

[6] W. Williams. *Fairness: Results Versus Process*. Ideas on Liberty, October 1998.

[7] Word Reference. http://www.wordreference.com/

[8] S. Turkle. *Life on Screen: identity in the age of the Internet*. Simon and Schuster, 1997.