

## Semantic Frames and Video Games



James Paul Gee  
Mary Lou Fulton Presidential Professor of Literacy Studies  
Arizona State University  
james.gee@asu.edu

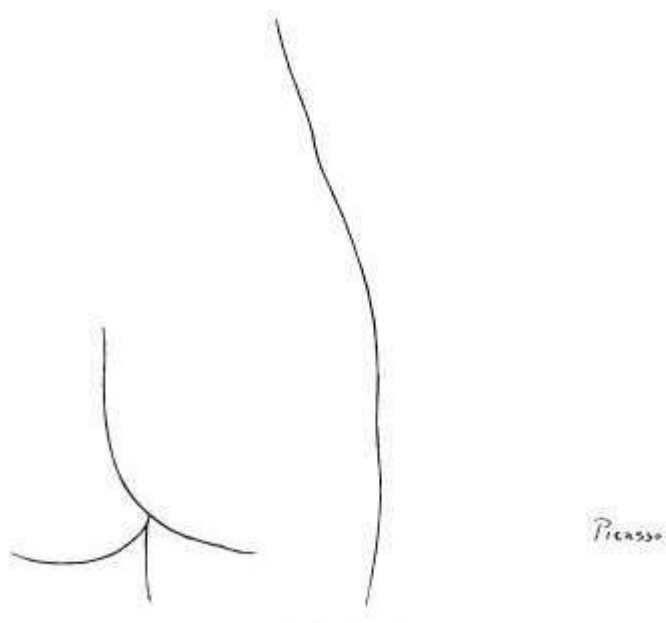
**Note:** *This is a speculative paper meant to generate discussion. Its claims may or may not turn out true. Use at your own risk.*

Visual realism in video games has long been a much discussed topic. We know that games can vary on a scale from low to high realism. And we know that games can be successful without high realism. Indeed, we do not always want high realism. Few people would want to play a First Person Shooter (FSP) where they kill completely realistic looking humans.

The issue of visual realism (some sort of scale towards photo-realism) is interesting, but less important than what I will call “semantic realism”. Semantic realism applies to much more than the visual images of a game, but images are a good place to start.

Consider the two images below. The first is drawing by Picasso. The second is a part of a painting by Carl Samson (the full painting shows a full-length female nude figure in front of Picasso's *Demoiselles D'Avignon*, standing, with a stick in her hand, triumphant over the wounded Minotaur at the bottom of the painting). In terms of visual realism, Picasso's drawing

is low and Samson's painting (the whole painting) is high. But both work visually quite well. So, more than visual realism, in any photo-realistic sense, is relevant here.



<http://www.thisnext.com/tag/fine-arts/items/?start=120>



<http://carlsamson.com/newnews.html>

Why does the Picasso drawing work so well, despite being very low in visual realism of the photo-realistic sort? It works because it triggers what I will call a *semantic frame* in the mind. Let me discuss what semantic frames are and then return to the Picasso drawing.

A semantic frame is a template that captures a paradigmatic meaning. The template can be filled in by users in different ways. For example, humans have a semantic frame in their minds like this: AT SOME TIME AND PLACE an AGENT TRANSFERS POSSESSION OF SOMETHING TO a RECEIVER for SOME REASON. This frame can be filled out in different ways (in different languages) for different cases of this general meaning. It can be filled out as, for instance: “Yesterday, in his office, the Dean granted Professor John Smith a sabbatical to honor his service to the university” or “Tomorrow, at the stadium, the Giants will give Billy Jones a fat new contract in order to keep him on the team”. How we say these sentences can vary in word order (e.g.: “Tomorrow, the Giants will give Billy Jones a fat new contract, at the stadium, to keep him on the team”).

The frame can also be filled out more or less completely. We can say: “The Dean granted Prof. John Smith a sabbatical” and leave out some of the elements the frame allows. If every possible element the frame allows is present in an utterance, I will say the frame is *fully semantically saturated* by the utterance. Semantic frames do not require full saturation. But they do have *a minimally required degree of saturation*. “The Dean granted Prof. John Smith a sabbatical” is not fully saturated, but it is coherently meaningful. “The Dean granted Prof. John Smith” is not fully saturated and it is not even saturated to the required minimal degree. Thus, it is not fully coherent. It is what we will call an *unpragmatic* (misshapen) meaning.

When a frame is fully saturated by a given utterance, all the elements of the frame are explicitly specified by the utterance. When it is not fully saturated, the details that are left out have to be added back in imaginatively by the person who receives the communication. So if I say, “The Dean granted Prof. Smith a sabbatical”, the hearer can seek to imagine or discover the missing elements, elements like where it happened, when, and why. Or, of course, the hearer can just leave these aside or can ask about them.

Semantic frames exist in human cognition for actions, states, happenings, events, and objects in the world. They are based on the experiences people have had in the world. They are types of limited generalizations people have drawn from their experiences in the world. Semantic frames express expectations people have for how things will hang together in the world to be meaningful in certain ways.

Our example of a semantic frame is overly simple and tied to language. Frames need not be filled in with words. In some cases they are filled in with images, which serve as sorts of words when we think in terms of images and experiences and not just words and language.

Let us return to the Picasso drawing. It works because it triggers a semantic frame for “sexually suggestive nude female body”. This frame is a template in the mind that dictates all the sorts of details of a (virtual or real) image and their possible arrangement that could have this meaning. Of course, the words I have used to name this frame do not capture the myriad of elements the frame contains. These elements are all the possible images of parts of a female body and their arrangements that could trigger the meaning “suggestive nude female body”.

Samson’s full painting fills in a great many of the elements that trigger this meaning (especially in the full painting). This is why we say it is close to being photo-realistic. It leaves few of the elements of the semantic frame unspecified; few are left for viewers to fill in for themselves. It “names” (images) the details specifically. Picasso’s drawing fills in few such details, just a bare minimum (no pun intended) of core or fundamental features. It leaves the rest to our imaginations. This is enough to trigger the semantic frame and send us on our imaginative way.

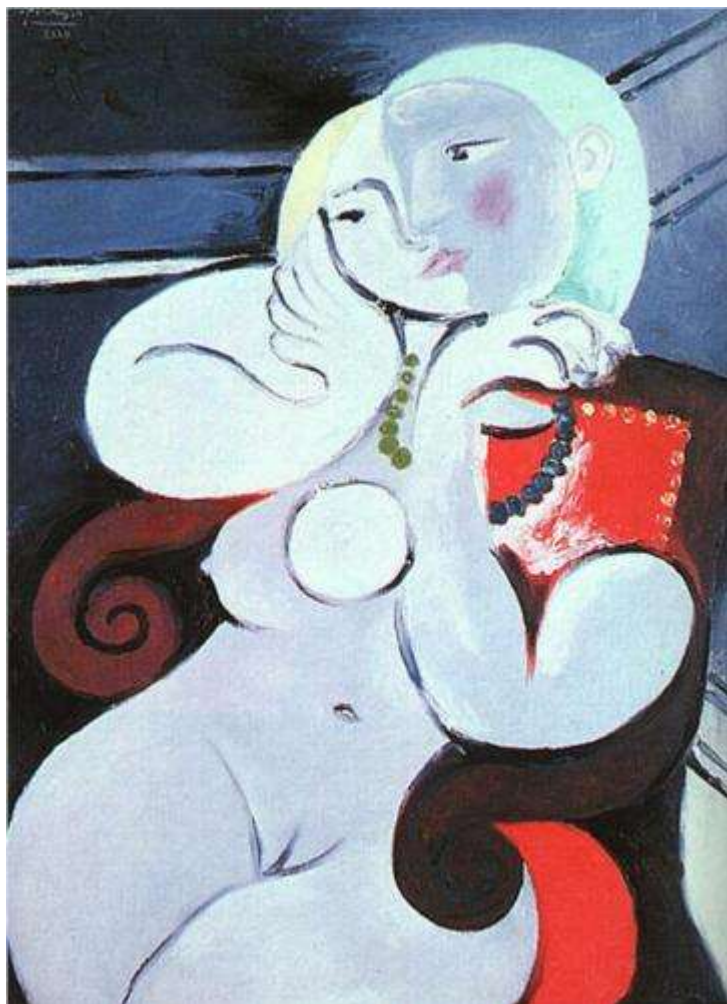
The Samson painting is *semantically saturated*, not fully (after all the woman is standing so that only her front can be seen), but to a great degree. The Picasso painting is not semantically saturated to any great degree. It is quite low in semantic saturation. It offers few specific

details to guide us in filling in the semantic frame. However, the details or features it offers us are so core to and so powerfully associated with the semantic frame that they still trigger it in a powerful way.

Both the Picasso drawing and the Samson painting are *semantically congruent*. By this I mean they trigger the appropriate semantic frame in a person's mind and do so in a way that does not violate the expectations the semantic frame causes us to have of the world. Both images are "realistic" in the sense that they trigger in powerful ways a semantic frame we have gotten from our experiences the world and from media.

Consider for a moment the Samson painting. You see only the top of the painting. So for you only these many elements of the semantic frame "sexually suggestive nude female body" are specified (filled in by the painter). Since you have not seen the whole painting, you can, in your imagination fill in the other elements of the body below the top part you see. After you have done this, you can look at the whole painting and see how you filled in these elements did or did not match how the painter did it. You might be surprised. You might find that Samson filled in the other elements in ways you did not expect. But there are two different surprises you might get. In one case, you will accept that how Samson filled the semantic frame is congruent with your expectations for "sexually suggestive nude female body", just not the way you filled it in in your imagination. In a second case, you will not accept how Samson filled out the frame, it will violate your expectations and be incongruent with your idea of a "sexually suggestive nude female body".

Now consider the image below:



<http://www.wikipaintings.org/en/pablo-picasso/female-nude-sitting-in-red-armchair-1932>

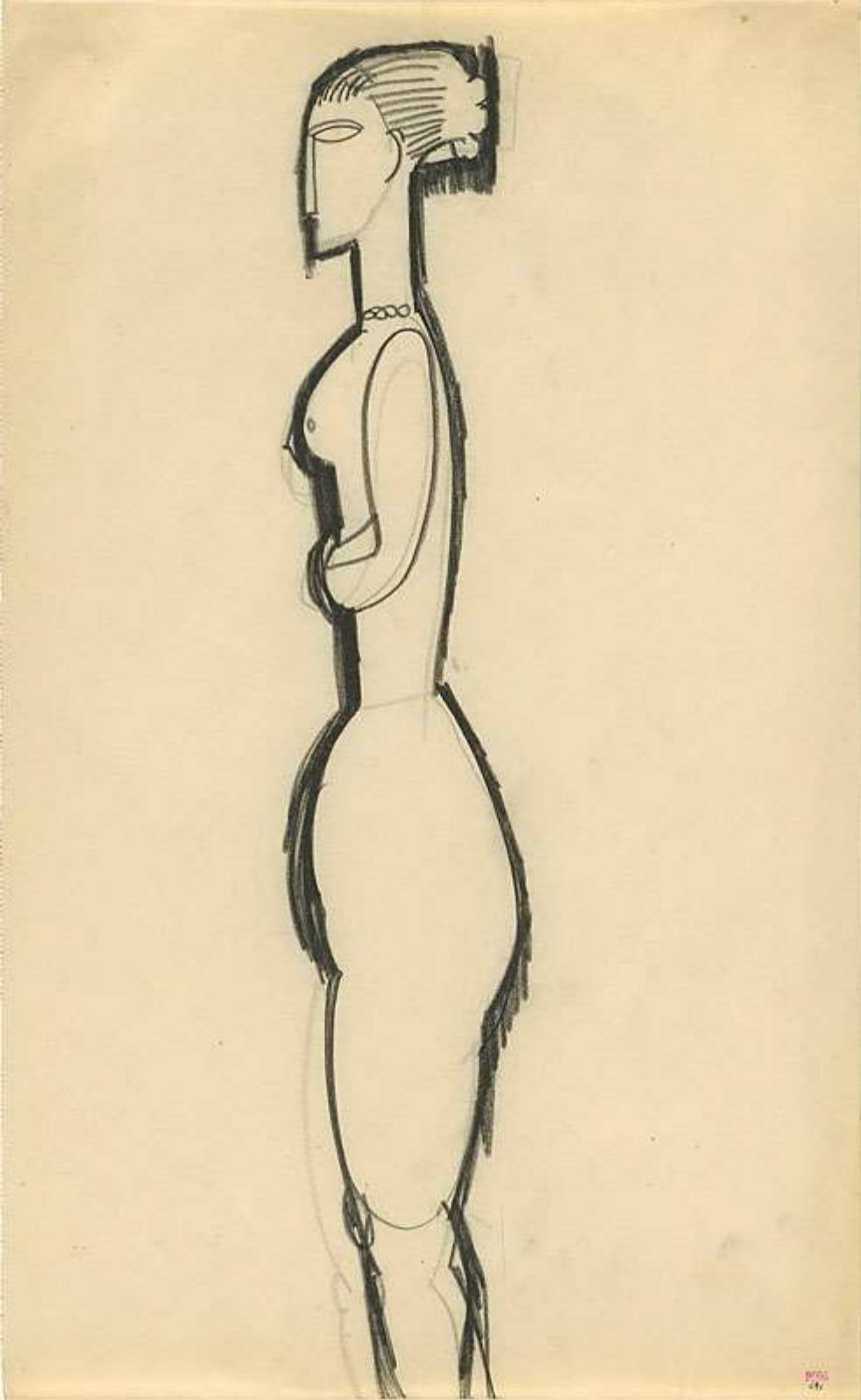
This is a Picasso painting. It is obviously more semantically saturated than the Picasso drawing above, but less semantically saturated than the full Samson painting. Both the Picasso drawing and the Samson painting (even the full version) are semantically congruent. They do not violate or expectations about how elements of the frame can be appropriately filled in. The Picasso painting is not semantically congruent. It violates expectations that our semantic frame for

“sexually suggestive female body” gives us. It fills in details of that frame in ways that are not typical of how we would have thought to do this based on our own experience.

Even though the Picasso painting is not semantically congruent, it still works for some people, though of course not for all. Being semantically non-congruent is always a risk. Why does the Picasso painting work for some people despite being non-congruent? It works because it plays with recognizable features of our semantic frame for “sexually suggestive female body” in ways that may suggest new ideas or feelings about this frame. It is important to stress though that such a semantically non-congruent image can work only if the viewer has been “trained” in the non-congruent genre in such a way as to be able to discover new ideas or feelings about the semantic frame or related ones. That is, the viewer must be “trained” to be literate in a new language or style of representation.

Note that all the images we have discussed are “pragmatic” in the sense that they contain at least the minimal elements necessary to trigger the semantic frame (even if some training is required). An unpragmatic image would be one intended to trigger this frame but with too few of the elements necessary to actually trigger it. Consider, then, the image below, a profile study for Modigliani’s only full length standing stone figure in the Australian National Gallery:





<http://www.richardnathanson.co.uk/Modiglianidrawings.htm>

This drawing, at least for me, fails to trigger the semantic frame “sexually suggestive nude female body”. If it was intended to do so, then it is unpragmatic. It fails to communicate this meaning (to trigger this semantic frame) because it does not contain at least the minimum elements necessary to trigger the frame, though it does include lots of other elements, not enough of which together constitute this necessary minimum.

In fact, Modigliani was probably not trying to trigger the semantic frame “sexually suggestive nude female body”, but rather a different set of semantic frames having to do with ancient goddesses. Modigliani’s drawing was probably inspired by ancient Cycladic art (a civilization that flourished on the islands of the Aegean Sea from 3300 until 2000 BCE).

Let us say that language, images, video games, and other media seek to “communicate”, using the word broadly. You can only communicate meaningfully with a human being by triggering semantic frames (whether saturated at a low or high level). Any communication usually triggers more than one frame, though in the examples above we just discussed one. We can ask of all these forms of communication questions like these: How fully does the communication semantically saturate the frames it triggers? How semantically congruent is the communication? If it is not fully congruent, how do people become “literate” in the style of communication giving rise to the non-congruence?

Obviously, in the case of the Picasso drawing not much semantic saturation was needed, but this is only because it used absolutely core or paradigmatic features of the semantic frame.

Had it used more peripheral ones, the image would not have worked, because it would have had too little semantic saturation (been unpragmatic). One moral here is that semantic saturation is killed if core features are not included and that core features can achieve sufficient semantic saturation—however minimal it may be—to work by themselves.

I have left out one important variable in this discussion so far. People belong to different cultures and different social groups. They have had different experiences in the world (and different sorts of training). They will fill out elements of semantic frames that are left unspecified differently in their imaginations based on their different experiences in the world. In some cases, they may not have enough experience to fill out elements left unspecified. Based on their different experiences, they may also react quite differently to how elements which are specified are specified. They may see them as congruent or not with the semantic frame based on their different experiences, tastes, and values.

Thus, if you look at the whole Samson painting, you may or may not find the female in the painting congruent with your frame (expectations) for the meaning “sexually suggestive nude female body”. If you are a heterosexual, perhaps, you will find her too “masculine” or too muscular or in too war-like a context to fit your frame, though others may not.

Communication and congruence are cultural and social. Thus, design, whether of communication in language, images, video games, movies, or any other media, needs to consider cultural and social factors in terms of response and “recipient design”, that is, how the

person receiving the communication will accept elements specified or fill in ones that are not. This issue becomes crucial, of course, for communications (media) that are meant for large, diverse, or mass audiences.

Video games interfaces—how the player controls a character or other game elements—crucially interact with semantic frames. In any first or third person game, the way the player moves a character triggers semantic frames about movement and action. Movement and action are never, for humans, purely physical. They always have additional meanings in terms of goals and contexts of application. One and the same movement, say an underhand throw of a ball, has similar semantic elements for its physical aspects but different meanings as these are integrated with things like the purpose of the motion and its meaning in context. So, in one case, the throw may be part of a semantic frame for “pitching a softball in order to get a batter out” and in another case, it might be part of a semantic frame for “throwing a small ball to land in a container to test how good a shot I am”.

So if a video game tries to recruit these semantic frames via a game mechanic that is meant to trigger one or the other, the designers have to worry about how much semantic saturation is necessary to achieve an immersive, congruent, and engaging effect. What are core elements of the frame and how many must minimally be present? Should the designer seek for a minimal instantiation of the frame or a more saturated version of the frame? How will these differences affect game play? Are the elements specified for the semantic frame congruent with the

player's expectation of how elements for the frame can be specified? Is the player being coaxed to fill in (imagine) unspecified elements in ways good for the game's success?

For example, consider a game that either via a controller or an embodied controller like the *Kinect* seeks to trigger the semantic frame "throw a baseball to pitch a strike in an effort to strike out a batter". How important to triggering the semantic frame in a congruent way is feeling the heft and solidity of the ball or the strong force of the arm whipping forward?

Obviously these are features that are hard to produce with a traditional controller. The game has to somehow depict them in the avatar's motions and hope that the player "empathizes" with the avatar (in this case a pitcher) in an embodied way that satisfies the semantic frame.

When the Wii first came out, Nintendo was surprised that though in playing Wii tennis or bowling, the player need make only minimal movements to be effective in the game, players used full force motions (some of which broke television sets) in an attempt to fill out the semantic frame in a congruent and satisfying way.

The semantic frames for physical actions (again, not just as physical, but as meaningful types of action) can be triggered well by games with just conventional controllers. The game *Thief: Deadly Shadows* is a good example. The game uses controls and animations in either first person or third that strongly trigger (for me, at least) hiding, sneaking, evasion, and stealth. It is interesting that this works different in first and third person (you can switch between them in

the game) and even switching between the two adds interesting effects that help with triggering semantic frames concerned with stealth, sneaking, evasion, and hiding.

Players can clearly have embodied empathy for their avatar as they combine movements they make with the controller and effects they see on screen. This is so because humans can project their own bodies and minds onto video games characters if they are manipulating their actions in a fine grained way, as they do in most first and third person games, like *Half-Life of Ninja Gaiden*. This effect—that humans feel their bodies extend into virtual space when they control something in a fine grained way—is neurological for humans. Game designers are lucky indeed humans evolved to have this effect, or video games would not work for embodied immersion as well as they do.

This raises the question of what happens when humans use a controller not to manipulate a character in a finely grained way, but simply to push “buttons” to make choices. In a strategy games like *Disgaea* or *Final Fantasy Tactics* players pretty much push buttons to choose attacks for multiple characters. The characters then carry out these attacks with no manipulation required by the player. Such games are rather like chess. Chess triggers semantic frames for strategy and warfare, but not for purposeful embodied movement and action in the world.

In tactical role playing strategy games like *Disgaea* or *Final Fantasy Tactics* meaning cannot be made by triggering semantic frames for purposeful embodied action on the player’s part. Thus, there is a danger there will be a dearth of meaning to what the characters are doing as they

engage in combat and why they are doing. In such games, a narrative, as well as goals and achievements set by the designers, triggers these aspects of meaning, rendering action that a player tends to watch, rather than directly produce, meaningful and lucid.

*World of Warcraft* constitutes a very interesting case here. The battle interface for *WoW* is a series of boxes one clicks on to initiate a wide variety of attacks, defenses, buffs, and magic. It is a complex and intricate system with different casting times and cool downs for some of the boxes and interactions among the various choices one makes. On the face of it, it seems that this interface will fail to trigger semantic frames for combat, beyond what happens in a game like *Disgaea* or *Final Fantasy Tactics*. But yet *WoW* feels much less like chess than does the game play in tactical role playing strategy games.

It is my view that the interface for *WoW* does not trigger semantic frames for embodied action, the way games like *Ninja Gaiden* or *Metal Gear Solid* do. Rather it triggers—and in a very deep way—semantic frames for cognition, in this case for complex real time decision making. This certainly does not sound sexy, but it works well. It is also why, perhaps, so many *WoW* players devote so much time off line engaged in “theory crafting” where they seek to understand all the complex interactions among the choices and their statistical underpinnings. *WoW* captures not action, but the planning, cognition, and reflection in and on action behind goal based action. For these semantic frames, *WoW* achieves a masterful amount of semantic saturation while still leaving a good deal to the imaginative play of the mind as it quickly makes choices and gets feedback from the game as to their success.

From the beginning there have been players in *WoW* who have claimed that it is not “skill based” enough and allows players that are not highly skilled to succeed too well. I think what this criticism reflects is that these players are applying a standard of games which trigger semantic frames for embodied action to a game which triggers semantic frames for cognition (planning, decision making, reflection in action based on feedback, etc.). Of course, such cognition is always involved in playing games, since they are forms of sequestered problem solving. But *WoW* makes decision making into an interface that triggers semantic frames for how we humans think (consciously and unconsciously) in order to solve problems in action and interaction. This is why I think the battle system in *WoW* feels less like chess and more like playing a real time strategy game (perhaps, in part, because of *WoW* origins in real time strategy games).

There is today a movement to move to more directly embodied gaming with motion sensitive devices, motion capture, and the use of the body itself as a controller. The goal is to get a more direct feel for action. But there is a real dilemma here. The human brain is built with an innate ability to mirror action that people observe. We can “re-enact” or “mimic” in our minds actions that we are observing “as if” we are doing it ourselves. This mirroring process is actually crucial to the comprehension of speech and meaning in language. Thus, in games like *God of War*, where sometimes the player has to quickly push a button label flashed on the screen to trigger a (usually dramatic) action on the part of Kratos, the player feels “as if” he or she is doing the



action, thanks the mirroring process and the fact that responding quickly to the button flashes mimics a part of the semantic frame for quick decisive movement in battle.

I pointed out above that this mirroring process becomes even deeper when the player is allowed to finely control the movements of the character. Such fine grained control of action makes the mind think one's own body has done the action. This gives rise to the affect of feeling that one's body has stretched into the game and become shared with the avatar's body.

Games that use the player's actions directly, like Wii tennis or any interactive media where a person can, for example, use his or hand body directly to accomplish an action, like making a overhand throwing motion to pitch a baseball or an underhand throwing motion to toss a ball into a bucket, could work less well than a traditional controller games, because such directly embodied games may very well not trigger the mirroring process. The player's mind could become confused between the parts of the semantic frame for throwing, for example, that he or she is contributing directly as a natural action and the images that the game is supplying.

It will be easy in such designs to create semantic non-congruence or even unpragmatic "communications". It may be, for instance, that feeling the solidity and heft of the ball or feeling a tight connection between my movement and the ball's trajectory (things which are hard to do in games and interactive media) will matter more in such direct embodied games or interactive media, than they do in more traditional controller games which strongly trigger the mirroring process. This is to say, that traditional games, on the one hand, and new directly

embodied games and interactive media, on the other hand, may trigger semantic frames in different ways. No one at this point really knows and much research remains to be done.