http://dx.doi.org/10.14236/ewic/hci2014.7

A joint activity theory analysis of body interactions in multiplayer virtual basketball

Divesh Lala and Toyoaki Nishida
Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501 Japan
lala@ii.ist.i.kyoto-u.ac.jp, nishida@i.kyoto-u.ac.jp

Yasser Mohammad
Faculty of Engineering, Assiut University
71515 Assiut, Egypt
yasserm@aun.edu.eg

To create embodied agents which exhibit realistic behaviour, we should first examine how humans behave with each other in the same context. In this paper, we define the context as navigating a virtual environment and using body movement as signals for communication. We undertake a novel experiment in which two humans play virtual basketball as a team in distributed locations, using only their bodies to navigate and execute tasks. Participants interact mainly through moving throughout the virtual world while passing a virtual ball. We propose that joint activity theory concepts are prevalent in virtual world communication, find evidence to support this hypothesis, and generate insights which can be used to create effective agents in the same type of environment. Even with a limited communication channel, it was found that the intention of players was able to be understood, which shows the existence of various joint activity theory concepts.

Joint Activity Theory, Body Interaction, Virtual World, Virtual Basketball, Games, Gesture

1. INTRODUCTION

The current paradigm of embodied agents can be generalised into two main types. First we have those which interact using discrete inputs from the user. An example of this are video game characters, where inputs are the keyboard and/or mouse. These characters exhibit realistic animated behaviour by reacting to these inputs. Another advantage is that they can move around a navigable world during interaction. The drawback is that the user is restricted by the input device as to what they can communicate. The other type are agents which consider human modalities, but the majority of these require the user to stand in front of a screen and face the agent. Interactions in the real world are more varied. When we chat with others we usually don't stand directly in front of them and talk at them. In situations such as team sports, a team-mate may turn away from us or tend to their own individual goals.

Perhaps due to technological limitations, few agents are both dynamic and use the human body as an input signal. We define "dynamic" as both physical dynamism (the ability to move around the world during interaction) and attentive dynamism (the ability to switch focuses of attention over time). The motivation of creating this type of agent is to allow human-computer body interaction in navigable worlds so agents can collaborate effectively with humans in a physical task such as sports. Our long-term goal is to create this agent (which we term a **dynamic interactive agent** or **DIA**) so that its **communicative** behaviour is appropriate for human interaction. We impose some constraints to reduce complexity. The only communication modality is body movement and the task is a simple game situation.

Two features important to DIAs must be addressed. The first is the type of body signals which a DIA must recognise, given that any communicative action can be executed by the user. The second is that DIAs must be able to recognise and express focuses of attention, as this will change throughout interaction.

These features carry more weight for DIAs. Signals for video game agents are not human input, while face-to-face agents focus more on facial expressions, eye gaze and utterances rather than full body movements. For recognizing the focus of attention, video game agents are only capable of doing this through simple discrete inputs from the user. Face-to-face agents make the assumption that behaviours performed by humans are directed towards themselves. DIAs must handle both these issues, which differentiates them from the other agent types.

DIA analysis has not been undertaken due to their scarcity. To address this, we require an agent model which considers the DIA features described above.

Figure 1: *Video game agents react to discrete input (left group), while face-to-face conversational agents are static but react to human signals (centre group). Our target implementation is the dynamic interactive agent (right group), which moves throughout the world and uses a human's full body behaviour as an input signal.*

Fortunately, we have a potential solution in the form of a real-world communication theory. Herbert Clark's seminal **joint activity theory** or **JAT** has been used to describe communicative behaviour between people (Clark 1996). So far this theory is largely conceptual and has not been used to analyse **virtual** world interactions between embodied humans. If this theory presents itself in this environment, we can then use it as a basis for designing DIAs. Our initial aim is for humans to engage with a DIA. We therefore generate the following research questions to be answered:

Q1 How can JAT explain communicative behaviours in virtual environments?

Q2 Can DIAs bring about engagement using only body movement signals?

To answer Q1 we propose an experiment to observe JAT in the virtual world. We focus on body movement, so our experiment will limit the communication channel to only this modality so that we can know for sure which signals are being reacted to. Firstly we create a virtual basketball environment in which DIAs can be located. Secondly we conduct an experiment with human teammates. Our hypothesis for the experiment is as follows:

H1 JAT concepts can explain observed interactions between human team mates in a virtual basketball game where the only communication channel is body movement.

To answer Q2 we analyse the games and generate insights into how DIAs should behave to increase user engagement. Our hypothesis for this is:

H2 Patterns of JAT-based behaviour can be exploited to enable DIAs to bring about engagement.

We present three novel contributions:

1. the construction of a virtual space where humans play basketball with agents using natural body movements (Section 4);
2. an analysis of the virtual world experiment from a joint activity theory perspective (Section 6.1);
3. insights for dynamic interactive agent design based on joint activity theory (Section 6.3).

2. RELATED WORK

Related research is framed by our proposed contributions. There are three main domains – the type of embodied agent, communication through body movement and the use of communication theories.

As mentioned in the previous section we can categorise embodied agents in terms of two features – dynamic capabilities and human-centered interactivity. Research on agents who inhabit and move in the 3-d world is weighed towards achieving tasks, such as search and retrieval (Eno et al. 2011), automated interviewing (Hasler et al. 2013), training teachers (Mahon et al. 2010) and many more. For the second feature, research is weighed towards creating a better quality of interactions with humans, measuring factors such as user satisfaction (Novielli et al. 2010) and believability (Demeure et al. 2011).

Research on body movement in agents has also been the focus of many studies, such as those related to affective computing (de Gelder 2009; Karg et al. 2013) or emotion (Castellano et al. 2007). The majority of these focus on improving agent expressions. On the other hand, we are interested in how body movements contribute to social signaling (Vinciarelli et al. 2012). Recently there has been a growing interest in body movement interactions due to the availability of low-cost sensors. Bianchi-Berthouze (2012) showed the importance of body movements in user engagement while Kleinsmith and Bianchi-Berthouze (2013) performed a literature survey of the perception and recognition of affective body expressions.

Real world cognitive theories have been used to model agent behaviour, including neural architectures (Sandamirskaya et al. 2011) and theory of mind (Hoogendoorn and Soumokil 2010). Our aim is to include JAT to these applications because it is well suited to modeling communicative behaviours. JAT for artificial agents has primarily focused on robots (Bradshaw et al. 2009) but little work exists on

	Video game Agent A	**Face-to-face Interactive Agent B**
Execution	Reacts to discrete key signal	Reacts to continuous streaming signal
Identification	Rule-based signal checking	Assumes signals are directed towards themselves
Recognition	Signal meanings mapped to keys	Must infer signal meanings using limited knowledge base
Uptake	Instantaneous explicit feedback	Response time may be delayed

Table 1: *Manifestation of joint action ladder activities in differing agent types*

embodied agent implementation. According to Nova et al. (2008), there has been criticism of JAT from linguists, in particular the concept of common ground. In this work we ignore spoken language altogether and do not attempt to formulate a common ground model but assess co-ordination through non-verbal behaviour.

From our survey there exists two research gaps. The first is the need for a navigable virtual world in which embodied agents use human body movement as an input. The second is the need for an agent model to support this situation. This work takes the first step towards resolving these through the construction of such a world and the use of joint activity theory.

3. JOINT ACTIVITY THEORY

Joint activity theory (JAT) was proposed by Herbert Clark as a method to understand communicative acts (Clark 1996). While there are many aspects of JAT to consider, we will focus on the main features which are relevant to our research. These are signals, joint action ladder activities and common ground. We then discuss how we can convert these theories into concrete implementations.

3.1. Signals

Signals represent the basic building block of communication. In this work we analyse body signals only. Clark distinguishes signaling methods into describing, indicating and demonstrating, but this assumes that both verbal and non-verbal behaviour is present. To simplify to body-only signals, we make a distinction between **explicit** and **implicit** signals.

A typical definition of a signal might be an action that is **intended** to convey some meaning to the receiver. Two properties distinguish explicit and implicit signals. The first is the physical property. Implicit signals arise naturally from regular behaviour. On the other hand, explicit signals can be interpreted as an intentional act of communication even without context. The second is the recognition property. If we imagine a distribution of interpretations of a gesture, an explicit signal would generate some interpretations with higher expected values, while an implicit signal would generate a more uniform distribution. These two distinctions should be kept in mind throughout the paper.

Consider the case where a basketball player wants to receive a pass. An explicit signal for this intention would be raising their arms towards their team mate with the ball. Without context we can understand that this physical movement is a communicative act and some interpretations, such as calling for attention, have a high expected value. An implicit signal would be simply to turn towards the team mate. Without context we have no way of knowing whether this physical movement is intentional communication. The distribution of potential interpretations of this movement is therefore more uniform.

Given that an implicit signal is more ambiguous, why use it over an explicit signal? We propose that it is because implicit signals require less cognitive and physical load for the sender to execute. For the receiver, context changes the signal's interpretation distribution towards one resembling that of an explicit signal. This updated distribution allows the receiver to understand the implicit signal's meaning.

3.2. Joint action ladder activities

Suppose we have the ball and our teammate waves their arm at us. When interpreting this signal, Clark argues that we engage in a joint action. There are four steps which describe this process:

- **Execution:** attendance toward a communicative act;
- **Identification:** identification of a communicative act;
- **Recognition:** recognition of the meaning of a communicative act;
- **Uptake:** acceptance or rejection of a proposed joint project.

This terminology is slightly different to Clark's, but for clarity will be used in this paper. The executor of the signal, A, goes through the process from uptake to execution when signaling. The intended receiver, B, goes through the opposite process. Clark argues that each level of the ladder should be co-ordinated in order (i.e. from execution to uptake) to advance the joint activity. Assume that A has the ball and is waved at by B. A observes that the gesture is intended for them and attends to it (execution), identifies the waving of their arm (identification), recognises that

the gesture means they want to receive a pass (recognition) and acknowledges the proposal (uptake) through another signal such as passing the ball or moving away.

These activities cannot be physically measured, but we can estimate them by observing behaviour. For example, if A were to pass to B, we assume that A and B have completed a joint project. If A does not react to B, one of the activities was not fulfilled. Perhaps A didn't see B, didn't recognise their signal or did not accept the proposal.

3.3. Common Ground

Finally, we consider common ground. Common ground is "...the sum of their mutual, common, or joint knowledge, beliefs and suppositions." (Clark 1996). It can be already situated or built over the course of an interaction. It is extremely difficult to measure, but we can estimate if it has increased by observing behaviour before and after an event. In this paper we are interested in finding examples of common ground which occur during interactions. This can be in the form of the understanding of signals, behaviours or strategies of gameplay. We also consider how common ground can be represented in agents.

3.4. Benefits of JAT

We propose that JAT is appropriate to use for DIAs. Although JAT is theoretical it can be applied to a broad range of contexts and behaviours where we can identify signals. We can analyse other types of agents from the perspective of JAT concepts, as in Table 1. These agents have benefits and drawbacks according to different JAT concepts. For example, Agent A reacts to discrete signals, which is unsuitable for DIAs.

JAT also allows us to address the DIA-specific features mentioned in Section 1. Recall that these features were the types of signals which a DIA must recognise along with the identification and expression of the focus of attention. Both of these can be understood using the JAT framework. To know which signals to recognise, the identification activity lets us know what constitutes a signal and the recognition activity lets us know its meaning. For the focus of attention, the perception activity tells us whether attention has been gained and the identification activity tells us how to express this attention.

With JAT we are not so concerned with understanding an internal representation of the user, but estimating their state through observable signals. This makes it an appropriate perspective for analysis. We now expand upon our hypothesis **H1** to formulate more detailed sub-hypotheses regarding JAT concepts related to virtual basketball. If JAT is present in a virtual environment, we would expect some evidence of various types of explicit signals as an indication of collaboration:

H1a A relationship exists between collaboration and the frequency of explicit signals.

H1b A relationship exists between collaboration and the variation of explicit signals.

Next, we should investigate the purpose of explicit and implicit signals. We propose that there will be evidence of emotional signals from human players and that implicit signals can be used to determine the attention of a player. Facial expression cannot be communicated, so emotion becomes more explicit. The dynamism of the game means that attention must be signified in some way and updated constantly. Intuitively, implicit signals such as rotation would allow this with the least cognitive load:

H1c Explicit non-task signals are used to express emotional feedback.

H1d Implicit signals are used to indicate the focus of attention.

Finally, we would expect JAT concepts of identification and recognition to also be present. We assume participants understand if a movement constitutes a signal and its meaning, even if they have not played basketball before. We propose that this is due to prior knowledge of how basketball is played and intuition towards abstract gesture meanings:

H1e Implicit signals are identified as such by participants.

H1f The majority of explicit signal meanings are recognised.

We also expand hypothesis **H2**. In this hypothesis we are interested in how the DIA-specific features (signal types and focus of attention) can be implemented in agents to bring about engagement and that this engagement can be seen through increased interactions. We also look for specific features and patterns which an agent can use and propose that these features are understood by humans for co-ordination in the game:

H2a Interaction increases between team mates during the game.

H2b Common physical features exist in explicit body signals used in virtual basketball.

H2c Common physical features exist in implicit body signals used in virtual basketball.

H2d A pattern exists in the focus of attention in virtual basketball.

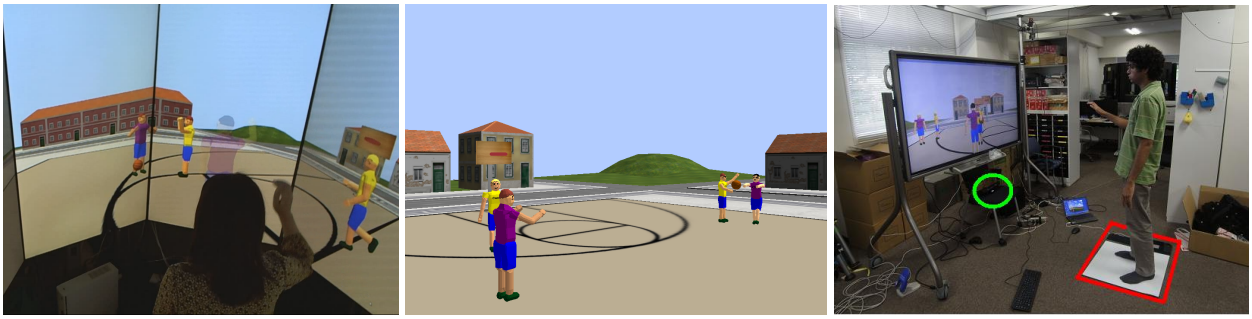


Figure 2: *The virtual basketball environment used in the experiment. Participants stand in an immersive display environment (left) while using a Kinect and pressure pad to interact with the world (right). The centre figure shows the appearance of the virtual world.*

4. EXPERIMENTAL DESIGN

In this section, we discuss the actual implementation of the virtual basketball experiment. Its objective is to allow us to test hypotheses **H1a**-**f**, in which we want to see evidence of JAT in the virtual world. We use basketball as a testbed because it contains communication phenomena we will observe:

- situations where explicit signals are used and their intended meanings;
- implicit signals being used to indicate a focus of attention such as a player, ball or goal;
- the signal process used between participants during ball-passing interactions.

From this data we can then find patterns which allow us to address hypotheses **H2a**-**d**. We also conducted post-experiment questionnaires to gauge user responses to interactions with their partner, including rating their collaboration and their interpretation of the intention of their partner.

4.1. Virtual environment

We use our own VISIE environment for the game (Lala 2012). It consists of eight large displays which surround the user and project the virtual environment. For our experiment, two separate installations of these displays were used. The basketball game itself was created using a Java game engine[1]. Separate installations can be connected to each other, enabling the creation of a full multi-player game.

The players manipulated their avatar through body recognition provided by a Kinect sensor. We only captured the top half of the body, due to space limitations and because there was no requirement to capture gestures of the lower body. A third-person perspective was used to aid the user in controlling their representative avatar. To recognise gestures we use the techniques described by Lala et al. (2013) and constructed models for shooting, passing and dribbling. Shooting and passing had to be performed with two hands. Dribbling was performed with the right hand only. Real-time model comparisons allowed the user to manipulate the ball without needing hand-held peripherals. One limitation is that the user had to be facing towards the Kinect for the gestures to be recognised properly.

To navigate through the world a foot pressure pad was used and an algorithm developed by Lala (2012) detected the walking movement and direction of a user. Due to the limitation of a fixed facing direction toward the Kinect, the algorithm was modified so that the user could turn and rotate in the world. In our experiment, walking in place on the pressure pad allowed the user to move their character forward. In order to rotate the user stepped on the extreme left and extreme right edges of the pressure pad to rotate left and right respectively. Faster walking in place equated to faster movement through the virtual space.

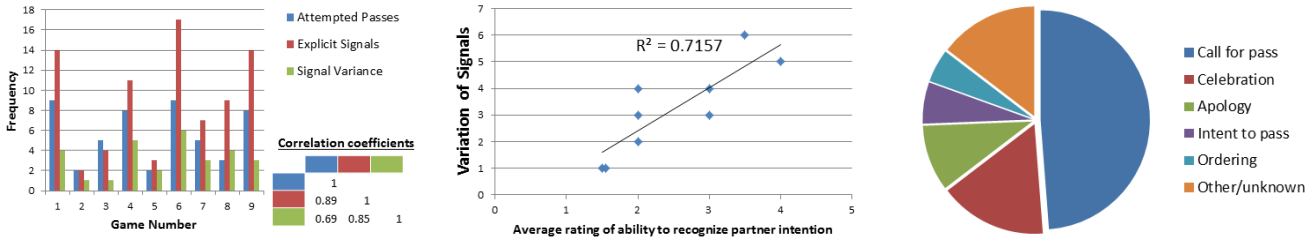
Figure 2 show users participating in the game using VISIE and on on a simple flat-screen. VISIE recorded all virtual world data.

4.2. Gameplay

The rules of virtual basketball are similar to those of pickup basketball with some modifications to improve gameplay. Teams take turns attacking and defending one goal. When possession changes hands the new attacking team returns to the edge of the court before play is restarted. The “traveling” rule (i.e. walking without bouncing the ball) was simplified – the player could only walk a certain distance without dribbling, after which they could not move unless they bounced the ball. To encourage collaboration, if both players touched the ball before scoring the goal would be worth two points rather than one. Players watched an explanatory video before the game began.

Each game was 10 minutes long, with two human players playing against two agent players. Agent

[1]jMonkeyEngine 3.0. `http://jmonkeyengine.org/` (2014)



(a) *Explicit signals, attempted passes and signal variation*

(b) *Recognition of intention (average) vs. signal variation*

(c) *Classification of explicit signals according to meaning*

Figure 3: *Analyses of explicit signal frequency, variation and collaboration in virtual basketball*

opponents had rudimentary behaviour. Before the game each user was given a period of training, in which they could learn how to pass, shoot, dribble and navigate using the pressure sensor. The training phase for each participant was performed in separate installations. Data recorded in-game movements and cameras recorded the movements of the participants in the real world. There were 18 participants, corresponding to nine total games played.

5. RESULTS

After all the game data was collated we observed the games in detail and used our findings to address our hypotheses. We performed the following specific activities:

- identifying explicit signals and their meanings;
- identifying implicit signals;
- inferring the focus of attention;
- observing the ball passing interaction process.

Participants also answered questionnaires regarding collaboration with their partner, however we found almost no meaningful patterns in the responses. This led us to use an observational approach by observing and labeling activities during the game. We felt that these activities were obvious enough to make multiple coders redundant, but acknowledge that this as a limitation of our analysis.

5.1. Explicit signals

To address hypotheses **H1a**-**c** we focus on the explicit signals produced during the games. According to Figure 3a, the frequency and variation of explicit signals generated during the games varied. For example, Game 2 had a much lower frequency and variation of signals than Game 6. In order to measure collaboration we consider the basic interaction unit in basketball, that of passing. This process can be definitively shown to be a collaborative act which must be co-ordinated between two players. We found a positive correlation between explicit signal frequency, variation and ball passing.

Figure 3b provides further analysis of hypothesis **H1b**. From questionnaires we found a slight positive correlation between the variation of signals and the average rating of the ability to recognise the partner's intention. This is displayed in Figure 3b.

To test what kinds of explicit signal were used for hypothesis **H1c**, we identified categories of meanings based on the contexts surrounding the game. This categorisation is displayed in Figure 3c. We can see that nearly half of the explicit signals were used to call for a pass. Other categories included apologies, celebrations and intending to pass. A non-trivial number of explicit signals were unknown.

5.2. Implicit signals

We address hypothesis **H1d** by investigating how participants showed their focus of attention. To do this, we observed the behaviour of the participants during the game and attempted to infer their focus of attention. The actual attention of users could not be reliably known, but we estimated this based on the implicit signals of rotation and movement. An example of the tracked data is provided in Figure 4.

We could identify several focuses of attention which the user could be engaged in during the game. These included objects in the game such as their partner, opponent, the goal and the ball. Opponent avoidance and navigation were also considered to be focuses of attention even though they did not constitute particular objects. We also included an "Unknown" category for situations where we could not interpret the behaviour of the user.

Focuses of attention were able to be identified by implicit signals. These signals were primarily rotation of the body and movement towards an object. For example, a player without the ball (A) would turn towards the player with the ball (B). Even if B moved, A would orient themselves towards them, a signal

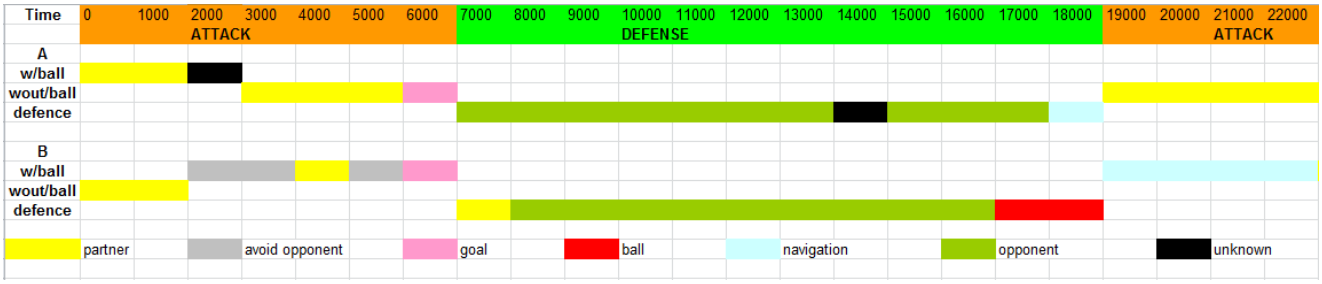


Figure 4: *Example of time series graphs used to display intention states of game participants*

that they were focused on B. This focus was clear throughout the game, considering there was no other indicative signals apart from body movement.

5.3. Identification and Recognition

In order to address hypotheses **H1e** and **H1f**, we have to consider how it could be proven that a participant had identified and recognised a signal. We again consider the passing interaction as a collaborative phenomenon. According to JAT theory, a pass can be considered a joint project. Therefore, if a pass is executed, there must have been some signals used to indicate that it should be thrown. Given this assumption, signals must have been identified and recognised by the participants to the passing action. If we can show that these signals are implicit, then it follows that these are identified (**H1e**). If these signals are explicit and feedback is produced, we can show that these signals have been recognised (**H1f**).

We analysed the process of passing by identifying which signals triggered the interaction. Leading up to the actual passing gesture we identified the initiator of the pass (i.e. passer or receiver) and the signal they used to trigger a passing joint project. This is presented in Table 2. A total of 51 passes were thrown over all games. When the initiator is the passer, the passing interaction is mainly triggered by an implicit signal such as rotation. Conversely, the major trigger to a receiver-initiated interaction is an explicit signal.

Initiator	Implicit signal	Explicit signal	None	Total
Passer	13	5	3	21
Receiver	3	15	0	18
Both	7	5	0	12
				51

Table 2: *Initiator and signal types used for passing*

5.4. Interaction dynamics

For hypothesis **H2a** we use a Granger causality approach as in Mohammad and Nishida (2011), which takes time-series data of all body movements of all the games and outputs a several causality graphs. These graphs are a collection of nodes and links, representing body dimensions and interactions between participants respectively.

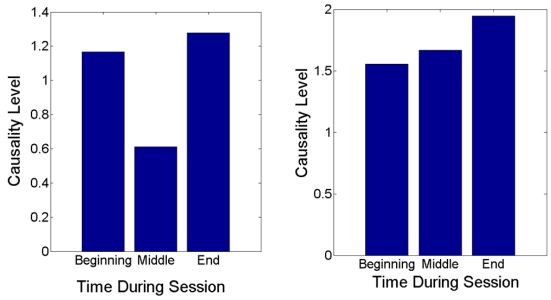


Figure 5: *Average Granger causality levels representing participant interactions. Left and right figures denote attack and defensive states respectively.*

Figure 5 displays the results of analysis over three time periods. We analysed attack and defense states using time series vectors of the arms as inputs. In attacking states, participants decreased their interactions before increasing towards the end of the game. In defensive states there was a gradual increasing trend of interactions. In general we found some evidence for increased interactions. This suggests engagement occurring between participants.

5.5. Patterns of behaviour

We now address our hypotheses **H2b**-**d** by finding commonalities in signals and focus of attention. We first address **H2b** by observing the explicit signals used during the game. The most common feature of all these signals was spatially large arm movements. Participants used this type of movement for all explicit signals. While this feature was common, the signals themselves varied. For example, some participants raised both hands above their head and executed a waving motion. Others stretched their arms to the side. Apology signals tended to be slightly more reserved and used only one arm.

For **H2c**, we identified two implicit signals - rotation and movement. The common feature of these two signals was the facing direction of the user being either at or turning towards its target. There were no cases where the facing direction and the signal target differed (for instance, a “no look” pass).

Finally, we address **H2d** by finding patterns in the focus of attention. For this, we again used the graph of game states in Figure 4. We could see clear patterns emerge depending on the context of the game. For example, an attacking player without the ball tended to focus on their team-mate in possession, while a defending player focused on the nearest opponent.

6. DISCUSSION

We now discuss how our results validate our original hypotheses and can help us understand how to create successful DIAs.

6.1. Joint activity theory in the virtual world

Can JAT concepts explain the observed interactions in our virtual basketball environment? From our results, it would appear to be so. We observed signals and the identification and recognition of them.

We considered a basic collaboration unit, that of passing, and found that relationships existed between this and the frequency and variation of explicit signals (**H1a** and **H1b**). A limitation of this is that passing itself is not a precise empirical measure of collaboration. Furthermore, questionnaires revealed that there was no significant relationship between these variables and a user's self-evaluation of collaboration. Passing was used as a measure in this study because it represented a specific action which fully encapsulated a collaborative process and can be explained through JAT. On the other hand, a user may evaluate collaboration as being tied to success in the game itself. Formulating measures of collaboration is a topic of future research in this environment.

From our categorisation of explicit signals we confirmed that non-task signals were used to show emotional feedback (**H1c**). This was evident in that the second and third most commonly used explicit signals expressed celebration and apologies. Participants certainly did not gain any advantage from these non-task signals but used them anyway. Many participants also used emotional utterances (e.g. shouting "Good shot!") even though there was no way for their partner to receive these signals.

Our analysis of the focus of attention showed that participants used implicit signals to express this (**H1d**). Without sound they relied on rotation and movement towards an object to display their focus. During our analysis it was quite clear what the participant was focused on simply by observing their body orientation. One limitation is that head direction could not be freely moved independently of the body. If this were to happen, we propose that virtual eye gaze is likely to be a better measure of the focus of attention. In any case, this is still an implicit signal.

Our passing analysis revealed that participants were able to identify implicit signals and recognise the meaning of explicit ones (**H1e** and **H1f**). We propose that this is related to the JAT concept of common ground. Participants knew that when their partner turned towards them it constituted an implicit signal. This could not have been due to some learned knowledge picked up during the game. Rather, it was due to prior common knowledge – the communal common ground which we all have as a result of being a human. Similarly, the meanings of explicit signals are also formed prior to the task. Humans know that arm-waving is a signal used to get attention and in a basketball context can assume that it also means a request to receive the ball. This should occur even among people who have never played basketball. The human experience represents a major discrepancy between humans and agents, but is useful for agent creation because it identifies the scope of what must be known in order to recognise the meaning of both implicit and explicit signals.

We have shown that our initial hypotheses regarding JAT concepts were somewhat correct. Signals were a combination of explicit and implicit expressions and represented some kind of collaboration involving not only task-based signals but also explicit non-task signals such as celebrating. It was shown that the focus of attention could be deduced from implicit signaling. Identification and recognition in JAT was shown through passing behaviours, where participants made use of communal common ground to identify and recognise various signal types.

6.2. Modeling human behaviour

What body movement patterns exist which can be exploited to further user engagement? We have some evidence from the Granger causality analysis (**H2a**) that participants increased interaction activities over time. This is important, as it indicates engagement and participants becoming familiar with each other. We also concluded that humans understood explicit signals given to them. These had a common feature of being spatially large arm movements (**H2b**). We can use this property to both generate explicit signals and recognise them from human players.

A similar case emerges for implicit signals. We have found that in this environment these are rotation and movement (**H2c**). Recognition of implicit signals will pose an additional challenge because we must determine what constitutes a signal. Unlike explicit signals we cannot use a gesture recognition model as a discriminator. In section 6.3 we discuss how this may be achieved.

DIAs must consider switching focuses of attention. We have shown that this changes according to game

states (**H2d**). For example, when a human teammate has the ball, the agent should look to focus on the human. If the ball is loose and the human is seen moving towards it, the agent should not focus on trying to retrieve the ball, but navigate to a more appropriate position. Implicit signals are used to achieve this. We propose that signaling a clear focus of attention helps co-ordinate players.

6.3. Insights for DIA creation

What insights did the basketball game give us for creating DIAs? We have shown that JAT in the virtual world operates in much the same way as the real world. We now use this knowledge to gain some further insights.

6.3.1. Identify implicit signals through cumulative evidence

We have seen that implicit signals serve two purposes. The first is as a communicative act to a receiver, such as a trigger for a passing action. The second is to indicate a focus of attention. Furthermore, it was shown that humans are able to distinguish intentional implicit signals from similar movements which lack meaning. DIAs should also address this problem.

One strategy is to provide agents with a classifier model which can recognise implicit signals. We propose a cumulative evidence model, which is a sum of weighted factors which contribute to the evidence that a movement is actually a signal. Clark also suggests that evidence is an important component of identifying and recognising signals. Consider an agent with the ball observing a human rotating their body. The agent must recognise whether this is a signal with the meaning “pass me the ball”. The agent weighs up factors over a fixed time period and uses these as evidence for an intended signal. Factors include the facing direction of the human, the distance between team-mates and the location of opponents. If the evidence is greater than a threshold, the assumption is that a signal has been produced

$$\epsilon = \sum_{i=1}^{n} w_i f_i \qquad \epsilon^* = \sum_{j=1}^{t} \epsilon_t \qquad \epsilon^* > thr \rightarrow \sigma$$

where ϵ is evidence for a point in time, ϵ^* is the cumulative evidence, thr is the threshold level and σ indicates that an implicit signal has been identified. There are n factors of f which each have a weight of w. This model can be implemented between each pair of players. Factors include primitive behaviours such as the orientation or movement towards a player as these were found to be used as implicit signals.

Recognising the focus of attention is simply a generalisation of this cumulative evidence model by using the same factors to gather evidence on **salient features** in the environment such as the goal, the ball or a subtle facial movement. The model then becomes a classifier for a feature. We do not explicitly state any specific action for the agent to take as this will depend on their individual goals and beliefs.

6.3.2. Recognise explicit signal meanings through context

It was shown that explicit signals were recognised by participants despite no prior interaction with their partner. Furthermore, these explicit signals varied among participants. A recognition model for a specific gesture is not appropriate. A better strategy is to recognise features that indicate an explicit signal has been executed. From our experiment, we know this feature is a spatially large arm gesture. Once we detect this, we can make the assumption an explicit signal has been executed.

The next step is to determine the meaning of the explicit signal. Two pieces of information become important. The first is the aforementioned implicit signal, from which we can determine the “target” of the explicit signal. The second is the context of the game, which allows us to infer the explicit signal’s meaning. In our experiment there are a limited number of signals and these can be differentiated by the game context. A limitation of this approach is extending it to situations where contexts cannot be as reasonably determined as basketball, where the state of the team and the player are clear demarcations.

6.3.3. Express common ground

Common ground ties the JAT concepts together. Implicit signals are only identified and explicit signals recognised because of our communal common ground of being human. A confounding issue is that it is impossible for us to measure or quantify common ground. A vast amount of communal common ground with other humans is already present and representing all this knowledge is infeasible. Our basketball game has a limited communication channel so this decreases the common ground required to play the game.

So far we have considered common ground as “What do we know?”, but ignored “How do we show what we know?”. In the experiment the participants were aware that their team mate was a human. Based on prior common ground they can infer that their signals would be understood. An agent may not be afforded this assumption so must prove their common ground and actively attempt to engage in communicative acts with the user. This separates agents as tools from agents as intelligent entities. Although we “use” the former to win a basketball game, we may not believe that it “knows” anything.

7. CONCLUSION AND FUTURE WORK

In this paper our main task was to show that joint activity theory was applicable to virtual worlds. We created a virtual basketball environment which can be used as a tool to study dynamic interactive agents, found evidence of joint activity theory concepts within these experiments and performed an analysis from this perspective. We also generated three insights into creating dynamic interactive agents: identify implicit signals through cumulative evidence, recognise explicit signal meanings through context and express common ground. Future work is the creation of effective agents based on these findings.

Acknowledgments

This research is (partially) supported by the Center of Innovation Program from Japan Science and Technology Agency, JST.
AFOSR/AOARD Grant No. FA2386-14-1-0005

REFERENCES

Bianchi-Berthouze, N. (2012). Understanding the role of body movement in player engagement. *Hum-Comp. Interact. 28*, 1–36.

Bradshaw, J. M., P. J. Feltovich, M. Johnson, M. R. Breedy, L. Bunch, T. C. Eskridge, H. Jung, J. Lott, A. Uszok, and J. van Diggelen (2009). From tools to teammates: Joint activity in human-agent-robot teams. In M. Kurosu (Ed.), *HCI (10)*, Volume 5619 of *Lect. Notes Comput. Sc.*, pp. 935–944.

Castellano, G., S. D. Villalba, and A. Camurri (2007). Recognising human emotions from body movement and gesture dynamics. In *Proc. 2nd Intl. Conf. Aff. Comput. and Intell. Interact.*, ACII '07, pp. 71–82.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Phil. Trans. Royal Soc. B: Bio. Sci. 364*, 3475–3484.

Demeure, V., R. Niewiadomski, and C. Pelachaud (2011). How is believability of a virtual agent related to warmth, competence, personification, and embodiment? *Presence: Teleoper. Virt. Env. 20*(5), 431–448.

Eno, J., S. Gauch, and C. Thompson (2011). Agent-based search and retrieval in virtual world environments. In A. Soro, E. Vargiu, G. Armano, and G. Paddeu (Eds.), *Inf. Retrieval and Mining in Dist. Env.*, Volume 324 of *Studies in Computational Intelligence*, pp. 125–143. Springer Berlin Heidelberg.

Hasler, B. S., P. Tuchman, and D. Friedman (2013). Virtual research assistants: Replacing human interviewers by automated avatars in virtual worlds. *Comput. in Human Beh. 29*(4), 1608 – 1616.

Hoogendoorn, M. and J. Soumokil (2010). Evaluation of virtual agents utilizing theory of mind in a real time action game. In *Proc. 9th Intl. Conf. on Auton. Agents and Multiagent Sys.*, AAMAS '10, pp. 59–66.

Karg, M., A. Samadani, R. Gorbet, K. Kuhnlenz, J. Hoey, and D. Kulic (2013). Body movements for affective expression: A survey of automatic recognition and generation. *IEEE Trans. on Aff. Comp. PP*, 1–1.

Kleinsmith, A. and N. Bianchi-Berthouze (2013). Affective body expression perception and recognition: A survey. *IEEE Trans. Affect. Comput. 4*(1), 15–33.

Lala, D. (2012). VISIE: A spatially immersive environment for capturing and analyzing body expression in virtual worlds. Masters thesis, Kyoto University.

Lala, D., Y. Mohammad, and T. Nishida (2013). Unsupervised gesture recognition system for learning manipulative actions in virtual basketball. In *Proc. 1st Intl. Conf. on Hum-Agent Interact.*

Mahon, J., B. Bryant, B. Brown, and M. Kim (2010). Using second life to enhance classroom management practice in teacher education. *Educ. Media Intl. 47*(2), 121–134.

Mohammad, Y. and T. Nishida (2011). Discovering causal change relationships between processes in complex systems. In *2011 IEEE/SICE Int. Symp. on Sys. Integration*, pp. 12–17.

Nova, N., M. Sangin, and P. Dillenbourg (2008). Reconsidering Clark's Theory in CSCW. In *8th Int. Conf. on the Design of Coop. Sys. (COOP'08)*.

Novielli, N., F. de Rosis, and I. Mazzotta (2010). User attitude towards an embodied conversational agent: Effects of the interaction mode. *J. Pragmatics 42*(9), 2385 – 2397.

Sandamirskaya, Y., M. Richter, and G. Schoner (2011). A neural-dynamic architecture for behavioral organization of an embodied agent. In *2011 IEEE Intl. Conf. on Development and Learning*, Volume 2, pp. 1–7.

Vinciarelli, A., M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D'Errico, and M. Schroeder (2012). Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Trans. Affect. Comput. 3*(1), 69–87.