

# Team Reactions to Voiced Agent Instructions in a Pervasive Game

Stuart Moran, Nadia Pantidi, Khaled Bachour  
Joel E. Fischer, Martin Flintham, Tom Rodden  
Mixed Reality Lab  
University of Nottingham, UK  
{firstname.surname}@nottingham.ac.uk

Simon Evans and Simon Johnson  
SlingShot  
Bristol, UK  
{firstname.surname}@slingshoteffect.co.uk

## ABSTRACT

The assumed role of humans as controllers and instructors of machines is changing. As systems become more complex and incomprehensible to humans, it will be increasingly necessary for us to place confidence in intelligent interfaces and follow their instructions and recommendations. This type of relationship becomes particularly intricate when we consider significant numbers of humans and agents working together in collectives. While instruction-based interfaces and agents already exist, our understanding of them within the field of Human-Computer Interaction is still limited.

As such, we developed a large-scale pervasive game called ‘Cargo’, where a semi-autonomous rule-based agent distributes a number of text-to-speech instructions to multiple teams of players via their mobile phone as an interface. We describe how people received, negotiated and acted upon the instructions in the game both individually and as a team and how players initial plans and expectations shaped their understanding of the instructions.

## Author Keywords

Human-agent interaction; in situ; instructions; pervasive games

## ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces - Interaction styles.

## INTRODUCTION

Intelligent interfaces continue to revolutionize the relationship between humans and computers. A product of this is that the default and comfortable position of users instructing and controlling computers is changing. As we continue to offload complex tasks and responsibility to machines, we will more frequently experience systems and interfaces that deliver instructions for *us* to follow. This role reversal becomes increasingly important in highly complex, dynamic and demanding time critical circumstances such as disaster response. Such complexity will require groups of humans and intelligent interfaces to *seamlessly* exchange

instructions in order to effectively co-operate. While expert, diagnostic, support and recommender systems have been successfully deployed for some time, to date, exploring and understanding the instruction of users in HCI is still limited [26]. Without this understanding, agent based instruction systems may function inefficiently, or at worst, fail to achieve compliance from humans. This would be of particular consequence in safety/time critical systems.

In light of this, our goal in this paper is to probe a number of issues centered on how people respond to instructions from a semi-autonomous agent-based system. In particular, we are interested in investigating:

- How do people make sense of unstructured or ambiguous information from a software agent?
- How do groups of people work together while under agent instruction?
- What makes people trust instructions, and what might cause them to not comply?

To explore these issues, we have developed a large-scale, ‘in the wild’ pervasive game called Cargo. Within the game, a semi-autonomous rule-based agent distributes a number of voice instructions to multiple teams of players via their mobile phones as an interface. This approach allowed us to study the visceral reactions and interactions of a large number of human teams receiving instructions in situ, in real time and under pressure. The results highlight the process by which players interpret instructions, negotiate their meanings, as well as how these are intertwined with the relationship between the players and the agent. We conclude the paper with a number of emerging design recommendations to consider when creating agent systems which instruct teams of humans.

## BACKGROUND

Within the field of Human-Computer Interaction, there is a wide range of literature on the interaction between humans and agents. However the tendency is to focus on a one-to-one relationship. The functionality and practical application of such collaboration is limited to specific types of problems and tasks. When we begin to consider increasingly large-scale complex environments and challenges, such relationships start to lose value. Hence, one growing field of research in human-agent interaction is

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

UII'13, March 19–22, 2013, Santa Monica, CA, USA.

Copyright © 2013 ACM 978-1-4503-1965-2/13/03...\$15.00.

the exploration of significant numbers of individual agents and humans acting in human-agent collectives (HACs).

### Human Agent Collectives

HACs will have the ability to help solve, support and/or manage extremely dynamic and time-constrained tasks and problems that humans cannot easily complete alone. One clear example of how such collectives might work is in disaster response [10], where agents could assist in data collection in inaccessible areas (e.g. scouting aerial drones), and decision-making (e.g. knowledge of the ‘bigger’ picture that may be incomprehensible to humans). The means by which the interaction takes place in HACs is a vital issue, and intelligent user interfaces are important in facilitating human-agent co-operation and teamwork [20,25]. In situations such as these, an inherent part of this teamwork is the exchange of instructions as part of an on-going communication [21].

### Agents as Instructors

For successful co-operation with agents, humans need to be open to being directed [25], and able to engage with agents at a peer level [15]. As such, it is important to understand the different roles and relationships between humans and agents in HACs. Agents are capable of managing complex information that would overwhelm humans [1], and giving them the role of the instructor is ideal. This places the agent as an intelligent interface between the humans and the complex information they need to complete their task. For example, there is a large body of work on the ways agents have been used to support collaborative workplaces such as in the management of emergencies [23] and damage control [4]. Agents can also take on more formal advisory roles such as ‘doctors’, assisting patients and medical staff in diagnoses [5,8], or as a personal tutor [13].

While this prior work is interesting, it can be argued that instructing is not necessarily the primary function of many agents: they simply explain concepts, diagnose problems or provide recommendations. This raises the question, what would it mean for a human to receive direct and specific instructions from an agent, as opposed to passive recommendations? Moreover, how might this play-out in teams of humans, with different views and perceptions?

### Instructing Teams

In human-agent teamwork, it is important that agents are viewed as proactive ‘*team players*’ [6] [28] and not simply as ‘puppets’ [27]. In their role as an instructor, agents will need to make decisions and be capable of dealing with issues of authority, responsibility, coordinated actions and group decisions [27]. There are a variety of examples where software agents directly instruct teams, including control rooms [18], flight decks [12] and command-and-control situations [16]. While the interaction between human and agents in these systems is readily established, the ways humans *interpret* and *negotiate* instructions can affect the way they are carried out.

### Interpreting Instructions

While efficiently communicating correct information can help reduce the cognitive burden of interpreting instructions [6], the source can influence the way it is interpreted and evaluated. For example, incomplete or ambiguous information received from an agent may be perceived as unreliable but as credible if from a human [7]. Humans will also react differently depending on the timing of the instruction in relation to the local situation [19]. Although, reactions to instructions are not based purely on the current experience and situation, but are also affected by interactions pre- and post-instructions [9,14].

While interpretations of instructions can differ in many ways, it is the consequences of these interpretations that are ultimately important – the way they are (or not) translated into practical activities [24]. However the post-instruction activities of humans may not always be predictable or expected. As such, an instructing agent needs to be able to adapt its recommendations to a situation as it unfolds. This means relying on a group wide shared understanding of the situation [14][6]. Unfortunately, this understanding might not be the same for all team members [14][28], so negotiation becomes important to the team dynamic.

### Negotiating Instructions

Negotiation about the goals and actions of a human-agent team is a key aspect to collaboration, and is enabled by sharing a common frame of reference [27]. The purpose of negotiation is to reduce any discrepancies between the *shared* frame of reference, and the *actual* frame of reference [22], a common source of breakdowns in co-ordination [6]. That is, negotiation enables a team to resolve potential changes to their current state or agreed plan of action. There are different approaches that can be adopted when human-agent teams negotiate, including *debative* and *integrative* negotiation [22]. Debative negotiation initiates when new information is shared which highlights a discrepancy or interference with current activities. However, if this information were to coincide with current activities and was introduced by an individual, then integrative negotiation initiates.

When receiving instructions, teams of humans have been known to make *their own* negotiations. Such negotiations can take place silently where, for example, humans that are experienced at working together have been observed making group decisions without the need to communicate it to an agent [22]. This can be problematic for an agent, as missing information could render any further attempts at instruction or negotiation inappropriate [3]. This demonstrates that *observability*, in addition to *directability*, is essential to the success of co-operative systems [6].

Negotiation is unlikely to occur in safety critical systems, where an agent has a greater appreciation and understanding of the wider situation. Even so, there are examples of agents that will act if their instructions are ignored. For example, a collision agent instructing pilots on

maneuvers will take action itself if sufficient time expires and the situation becomes critical [12]. This is often a necessity because the additional cognitive load of interacting with an agent can cause a breakdown in the co-operation under time critical situations [22]. There are also many examples of situations that are free from risk and consequence, which present humans with options for ignoring or misusing instructions. For example, drivers have been known to follow the directions offered by in-car satellite navigation, irrespective of external warnings, such as road signs or unexpected changes in the quality of road. As a result, some drivers have actually driven their cars off-road and directly into rivers as a consequence of their over reliance and confidence in the directions given [29].

The main limitations of studies on agents and instructions are that they are restricted to a small number of individuals and are generally conducted in laboratory conditions or simulations. Furthermore, the critical nature of their purpose adds (necessary) constraints to how freely teams interpret and negotiate the instructions. Hence, to explore systems instructing *teams of humans* on a *larger scale, in situ* and with *fewer constraints on actions*, a different approach is needed.

## METHOD

Pervasive games are recreational experiences that make use of pervasive technologies such as smart phones, wireless technologies and sensors with the aim of extending the gaming experience into the real world [2]. They are frequently used as a research tool for exploring complex social and technological issues in an engaging way [e.g. disaster response [10]. The primary benefit of pervasive games is the fact that they can be situated, which can lead to improvements in the efficacy of the behavioral observations made [11]. Pervasive games have already been established as a tool for exploring instructions, although limited to a linear experience, and a small number of individuals receiving instructions from human performers [26]. This motivated the development of ‘Cargo’, and the use of a semi-autonomous rule-based software agent, which automatically sends hundreds of verbal instructions to significant numbers of individuals, distributed in teams.

## CARGO

Cargo is a mixed reality street game with elements of treasure hunt and tag. The theme of the game is that a team of six to eight players is trying to help one member (called Cargo) escape the city before being caught by police. Players race against the clock in an attempt to gather enough credits to win the game. In order to collect enough credits, players must visit, or check in, at a number of these game stations scattered around the city. When a player checks in, the station either rewards them with credit or, if the station is “dead”, wipes out their credit total. Players need to figure out which stations are dead and which ones are live. They are aided in that by a software agent we call the Instructor, who calls players on their phones with

relevant in-game information. In the game rules, the only information given to players regarding the Instructor is the following:

“You will get calls giving useful information. Using the information from these calls will be vital to your success or failure.”



**Figure 1.** A game station (left) is a computer with an RFID reader and a target where players scan their RFID tags, and a player (right) swipes his ID on the box and awaits feedback.

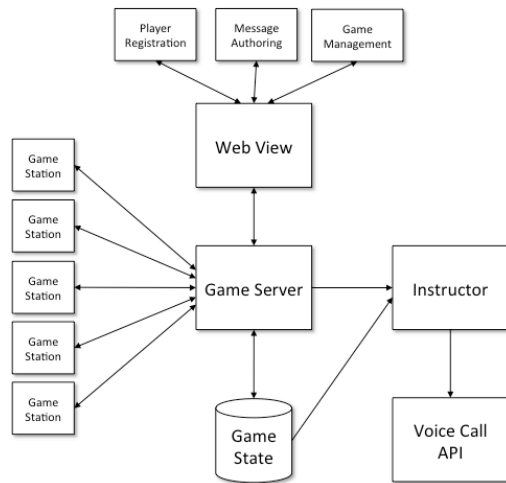
## Implementation

The system is implemented as a server-client architecture. Each game station is a PC running a simple python script that handles player check-ins. A Phidgets [30] RFID reader is attached to each station, as well as a loud speaker to provide direct feedback to the players. The hardware is packaged in a cardboard box with an area on the top marked with a target indicating where the RFID reader is placed. Players have RFID-tagged plastic cards associated with their player profile. Players can then swipe a card on the target area of the cardboard box (see Figure 1), and the station gives audio feedback on whether or not that check-in was successful. The game station sends all check-in data to the main game server where all the game logic is handled. The architecture is represented in Figure 2.

The game server is a Django web application that uses an SQLite engine to store the game state. It provides several views on the game state for online game management such as a player registration page and game-authoring tools for defining what messages players will receive. The web view also allows the remote configuration of game stations, as well as monitoring their status through periodic pings that the stations send back to the server. The game uses a commercial cloud API for making voice calls. The server communicates with the API through HTTP requests to trigger JavaScript code that makes a voice call and uses text-to-speech to read out a personalized message to the player in a continuous loop until the player hangs up.

## The Instructor

The Instructor is a semi-autonomous rule-driven software agent that monitors the game state and player check-ins and sends out appropriate instructions. Every time a player checks in, the agent evaluates the game state against a sequence of conditions and takes the appropriate action if any. Figure 3 shows the rules the agent follows when calling players with information



**Figure 2. System architecture.**

It is important to note that the information the agent provides to the players is always correct. For example, the Instructor never recommends that a player visit a dead game station that would cause them to lose their credits. The highlighted condition in the chart is the most commonly occurring and is one of primary interest to us. It triggers anytime a player checks in at a live station and thus gains credits. If that player had not been instructed to go to a particular station, or if they were checking in at the station she was instructed to go to, she would receive a new recommendation for the next station:

*“Hi Sally Smith, I heard there are a lot of points to be had at Albion House on Broad Street. It is in your best interests to head there now.”*

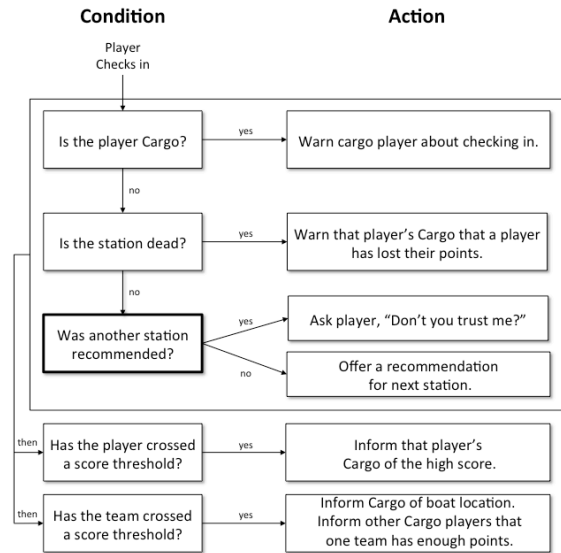
If however she checked in at a location different from one she was recommended, she would receive a call questioning her trust in the Instructor:

*“Sally Smith looks like you did not follow my advice about going to The Big Chill on Small Street. Do you not trust me Sally Smith?”*

When multiple players in the same team check in to a location triggering this condition, several of them might receive a recommendation for the next station, and while all recommendations lead them to live stations, the actual stations may differ. Thus players on the same team often receive instructions from the Instructor telling them to go to different locations. The Instructor also tells players not to check in if they are Cargo, when they have enough credit to win the game, and when one player loses their credits. In addition to the automated messages, the game management interface allows Instructor calls to be manually triggered. These include Start and End Game messages.

### Event and Deployment

Cargo was developed in collaboration with SlingShot [31] a street game developing company based in a British city to headline on the annual ‘interesting games’ festival (igFest



**Figure 3. Rule Chart for Instructor.**

[32]). A total of 112 players participated in the game (71 players in the first session and 41 in the second). Players purchased tickets priced at 15 British Pounds to participate in the game. For both game sessions, 14 game stations were setup around the city center area resulting in a game area approximately 600 meters long and 300 meters wide. The locations chosen for these stations were varied and included pubs, cafes, bookshops and art galleries.

Players arrived at the festival headquarters approximately half an hour prior to the start of each game and were registered into the system (including their phone numbers) and given game packs that included their RFID card, some game instructions and a map of the game area. They were then asked to go to the center of the game area and await further instructions. Once all teams were ready, a manually triggered call was initiated, and the Instructor simultaneously informed players the game had begun.

Each game lasted one hour, and a total of 1310 Instructor phone calls were made to the players. When a team had collected enough points to win the game, they were given one last call and told to get on a boat that will leave the city at a certain time from a given location. When the game time elapsed, the remaining players were given a similar call directing them to “the losers’ boat.” Two small ferries were awaiting the players at the end of the game.

During the course of the game, a researcher shadowed each of four teams (two in each session) and filmed their entire game experience. Short semi-structured interviews were conducted and recorded after the game with some players (not only ones that were shadowed). Our findings in the next section are predominately based on the analysis of these video recordings, supported by game logs collected on the server.

## FINDINGS

This section describes how people *received, negotiated and acted upon the instructions* in the game both *individually* and *as a team*. These descriptions are presented in context i.e. taking into account there were different types of instructing messages; each team had different game play and group dynamics. More specifically, regarding the four teams observed, three of them had quite high cohesion, probably relating to the fact that a number of players knew each other beforehand. The fourth team had comparatively low cohesion and several conflicts emerged during the game. Further, the players' initial plans and expectations, their assumptions regarding the Instructor, as well as the timing and the location instructions were delivered were shown to have shaped the way instructions were understood, negotiated and followed, or not, by the players.

### Relating to the Instructor

The way players interpreted and acted upon instructions throughout the game was shaped to a great extent by their perceptions and assumptions regarding the Instructor. How players talked to and about the Instructor; issues relating to its perceived authority and degrees of trust are presented.

#### Referring and talking to the Instructor

Initial impressions about the phone calls and the Instructor were positive; players were amused while receiving their first phone call; one player commented on the voice and the content: *"It was nice to hear your name recorded in a computerized voice"*.

The players answered the phone calls by saying "hello" and answered "yes" when their name was mentioned similar to the way they might respond to an actual person on the other end of the line. However when referring to the Instructor, the players quite often used 'it' (*"see what it says..."*; *"wait it says I need a new identity"*) or 'they' (*"they tried to break us up"*) and less often 'he' (*"he knows I changed my identity"*). The use of 'they' was interchangeable as for some players it represented people such as the policemen that were after them and for others it referred to machines or computers. Sometimes when referred to as 'he' or 'they', the Instructor was ascribed human qualities and reactions: *"he is annoyed I didn't follow."*

#### Authority and Obedience

Throughout the game, the central role of the instructions meant that players assigned authority to the Instructor. Players not only debated or made assumptions about what the Instructor was capable of, mostly regarding penalizing for not following instructions, but also critical milestones of the game such as who the Cargo will be were attributed to the Instructor: *"So is it the Artificial Intelligence that is gonna decide whose name it is (the Cargo)?"*

The authority assigned to the Instructor often triggered dilemmas on obedience. Teams were undecided as to whether they should follow the instructions and how to do so *"let's go to where we were told"*, or whether this was a trick, as part of the game *"What if it is a trap?"* Likewise,

players were sometimes torn between following what they were individually advised and what the rest of the team had or was about to decide to do: *"I think I should go to the one he told me."* Players also reported feelings of anxiety following non-compliance to the instructions:

*"It was quite intimidating ... When we didn't follow its instructions, it rang me and told me off. Made me feel a bit worried."*

In the interviews following the game, one player even related the game with Milgram's well-known experiment on obedience [17]: *"I reckon it was like that voice, that experiment where you were told to shock someone and you listened to the authority."*

#### Trust

The Instructor's authority was recognized either implicitly or explicitly; however, authority did not coincide with trust. Building up from the first phone call and what followed in the next few stations, players created a *persona*, a stereotypical perception of whether the Instructor was on their side or trying to make them lose the game or just confuse them; and accordingly trust was built or not around the stereotype adopted by the team or individual players.

In the interviews after the game had ended, several players reported trusting the Instructor:

*"... Because everywhere I went to a place it told me, there were points; I had no reason to not trust"*

However, during the game, when players had to continuously make sense of the information given and the situation, many of them expressed disbelief and mistrust to the Instructor:

*"P1: Should we try and ignore all messages cause they probably just are trying to pull us apart? P3: Maybe, we don't know yet. P2: I think it is false information"*

*"So basically the computers are lying to us, that's what's going on. The automated messages are lying."*

For some teams/players, this persona remained stable (usually as a negative one) despite opposing evidence such as being sent to locations with credits; yet for others it was renegotiated throughout the game:

*"P3:I have never believed it; has anyone believed it here yet? P1:I believe it now. When we went to the rummer, we got points and I had been told from the start (to go there)."*

#### Receiving instructions

While the content of the phone calls involved both individual and team information, the calls were delivered individually. Upon ringing, players moved away from their team to pick up the call or stayed behind when walking/running. In the teams with high cohesion, players were observed to rejoin the team once the voice message started repeating. They walked towards where the majority

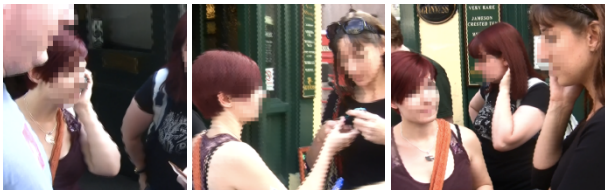
of the team was gathered and shared the content of the message with their teammates. On some occasions information was shared with others with a delay. For the less cohesive teams, sharing did not occur, or did so rarely.

#### Sharing

Players shared the information they received on the phone calls by announcing it to the rest of the team or nearby players. *"I have just got a phone call saying I should go to the Albion Bar on Broad Street."*

Sharing was manifested by body language during or after receiving the call. Players often drifted away or stayed behind when walking to listen to the phone call for the first time. They would then approach the rest of the team while the message was repeated and shared its content with the rest of the team.

In addition to verbal announcements, sharing instructions also took place by one player sharing the phone with another player or putting the phone on speaker (Figure 3).



**Figure 3.** Second player on the left is getting instructions on the phone and sharing with her team; at some point she cannot hear what it is said and asks another player to have a listen for her: *"P1: I cant make out...to Dal, to Dalgo Place... (passes the phone to P2 while the message repeats) P2: It says don't you trust him because you didn't go to Photographique (P2 gives the phone back to P1) P1: Oh! Is it Photographique, I cannot hear what they say!"*

#### Delaying

Sharing often happened in a delayed fashion, usually due to the timing of the instruction. Players received phone calls while already running towards a location or when their teammates were swiping at a station, which made it hard to stop and share the information. Yet in the high cohesion teams, players handled that delay in sharing by stopping at times to coordinate action (described in next section).

Apart from such coincidental delays in sharing, there were occasions that the delay was purposeful and planned. For example, in one team, the Cargo did not immediately openly reveal he was the Cargo. Instead he confided this information gradually and to only two teammates at a time. Between walking to stations, he approached pairs of teammates and in low voice told them he was the Cargo and that they had to be careful as the police were after him. This delay was intended and served the purpose of protecting the Cargo's identity not only from the policemen but also from other teams' players that were in vicinity.

#### Withholding

Not all players shared their information; withholding mostly occurred in the teams with low cohesion. This was

also indicated strongly by body postures during or after the phone call; players turned their backs or moved much further away from the team and stayed there until the end of the phone call (Figure 4).

A few players did not reveal any of the information they received to others in the team even when specifically asked; although most players only partially withheld information. Withholding was more common among the Cargo players and for information that involved teammates potentially having an advantage or disadvantage against others e.g. 'one player in your team has enough points to finish the game on their own.'



**Figure 4.** Players drifting away and remaining separated from the team while receiving phone calls.

#### Negotiating instructions

Upon receiving instructions, players negotiated to decide on what to do next. Negotiation involved interpreting the value of the information in terms of its content and its implications for action, which mostly involved consideration of possible consequences from not following instructions. As part of the negotiation, players also developed tactics to deal with the different instructions; and debated to resolve or avoid conflicts when possible.

#### Interpreting

Players interpreted the instructions based on the content and the intentionality they assigned to the Instructor. In terms of the content, ambiguous information such as the instructions sending players to different locations engaged players into debates about whether instructions should be followed or not and, if so, in what way. Information that related to the credit state (and potential win) of other players or teams were also found to be ambiguous and occasionally led to mistrust between team members.

*"P4: You know something don't you? C: I don't know anything(...)" P4: I think we should ditch the Cargo (laugh)"*

However, most interpretations related to the Instructor and its intentions. As mentioned earlier, as the game unfolded, players developed a character for the Instructor and based on that inferred whether its messages were meant to help them or trick them.

*"It is (the instructor) a bit too insistent and you are like 'hang on a minute!' are you shanking me?" "The machine is luring us to fake places"*

Mistrust towards the Instructor influenced and was influenced by the team's game play. For example, teams

chose not to go to locations where it was guaranteed they would earn credits as a result of not trusting the Instructor. One team even considered not going to the end location after being informed the game was over and they had lost because they thought that was a trick. Equally when a team had decided upon a specific game play such as ‘*staying together*’, all information against their gameplay was considered irrelevant and not trustworthy: “*You know we have all been given different information, so ignore it.*”

#### *Debating over possible consequences*

Further, players debated about possible consequences of not following instructions, in particular, during the first few check-ins of the game. After they received the message telling them they did not go to where was suggested and were asked whether they trusted the messages:

*“I was told off.” / “What if there is some kind of penalty if we don’t swipe in order?”*

They were concerned whether their disobedience meant they would not get any more calls and as a result; some players insisted strongly that the whole team should go where they were told. In response to such interpretations of consequences and value-based attributions to the messages, teams developed game tactics.

#### *Tactics for dealing with instructions*

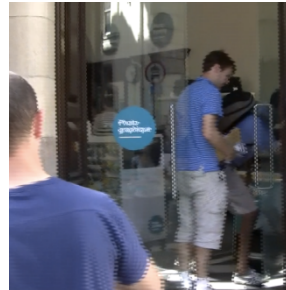
Interpretation was followed or, more often, closely intertwined with planning and forming game tactics. Depending on what meanings were assigned to the messages, the Instructor and perceived consequences, teams and individuals formed different tactics to deal with the consequences of following instructions or not. For example, if instructions were perceived to be accurate and the Instructor trustworthy, the team dealt with the varying location messages by splitting into smaller groups to go check in at the instructed location and regroup at the next.

With regards to the potential consequences of disobeying the instructions, one team developed a tactic to test what happens to those that follow instructions or not. The tactic involved one player following to the letter what the messages were telling them and the others just checking in any station independently of the messages. Another tactic that was introduced by individual players was to follow the team to the location but do not check in if that was not the instructed location. In this way, players avoided being ‘told off’ to the next call and resolved the dilemma of whether they were meant to check in stations that they were not told to go: “*I am not going to cash in here because it told me to go somewhere else.*” A similar tactic used at one point but later dropped was to have the player who had received the instruction to the specific location to check in first at that location in the hope that they would earn extra points.

The episode that follows is an example of how one team went about interpreting the instructions, debating over consequences and formed tactics to deal with them:



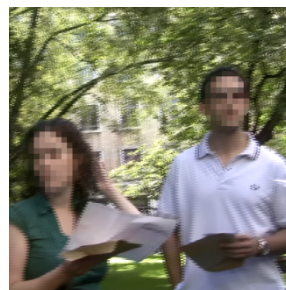
*(Players looking at maps) P1: I am just going to go to the church P2: Ok. P3: Mine is at St Nicholas Market P4: Where? P5: For the first three or four we should just go together P1: Should we go to mine?*



*P5: The Church, yeah, well just wait for us (some players enter another location) P5: I am not going to cash in here because...P1 they told you to go somewhere else. P5: Yes, right. P4: But I think you need more points (to win the game) P5: Right but it is this idea of following instructions*



*P4: Ok, ok. (his phone rings) Let see what it says now that I have followed it (moves further away to pick up the phone and soon after returns to the team) guys guys I have got St Stevens now too!*



*(They all head to St Stevens but the debate continues so they stop in front of the church) (...) P4: why did you get told off for? P1 and P5: For not going to where I was told. P4: You get told off but you still get the points. P5: Yeah but at some point should we not do it cause we are not doing what they are saying...*



*P4: What about, one person doesn't do what it says, the rest of us kind of do what it says and one person religiously does what they say to even it out? P5: I have done it P6: I haven't P4: so you keep doing it and see what happens.*

### Resolving conflicts

Conflicts occurred when players, receiving different instructions and with various interpretation of these instructions, needed to agree on their next action. Such conflicts were more common to teams with low cohesion as the players did not know each other and could not easily agree to a single course of action.

Most conflicts arose from the differing information with regards to where players were advised to go next and troubles reconciling individual obedience versus team play: *“We have been sent to different places, so do we go together or do we split up?”*

Several players felt strongly that they had to follow the instructions: *“I think I should go to the one he told me”*; *“Have to follow the instructions”* and this made it harder for conflicts to be resolved especially in the beginning of the game when the consequences of disobedience were unclear. Withholding information relating to the calls also resulted in conflicts, in particular when members of the team went missing/collecting points without notifying the rest of the team:

*“P1: Why have you gone off and abandoned us? P2: What? I went off to get some more points. P3: Oh, that’s dodgy!”*

### When were instructions not followed

A thorough look to the game data reveals instances or circumstances that fostered the likelihood of players not following instructions. The priorities set by the team as part of their game play, the mistrust to the Instructor, group dynamics and cohesion and conditions such as the location and the timing of when the instruction was given all contributed less or more to the likelihood of disobeying instructions.

### Gameplay priorities

The gameplay that a team had chosen for itself guided to a great extent how instructions were negotiated. Priorities were set implicitly or explicitly and sometimes this meant some instructions came second. For example, when players perceived *‘staying as a team’* as the main goal of the game, this resulted in them ignoring instructions that were suggesting splitting up or individual action to be taken.

*“P2: Why don’t we try and ignore all the messages? They are probably trying to pull us all apart. P3: Maybe. P4: I think it is false information”*

*“May I remind everyone that among our main priorities is to protect the Cargo?”*

### Mistrust to the message and/or the Instructor

As described earlier players interpreted the content of each message and assigned intentionality to the Instructor that was delivering them. When previous instructions had not led to expected successful outcomes (earning credits), it was more likely that future instructions would not be followed. For instance, several teams went to a location

they had been instructed expecting to earn more than 100 points which did not happen, so in the future, they ignored other instructions to locations: *“No I didn’t trust it cause everything was 100 points”*

### Proximity

In the case of instructions to specific locations, the distance between current location and instructed location was found to play a significant role as to whether instructions were followed or not. Specifically, when the instructions were about locations that were not in close proximity to the current location of the team, it was more likely that they were not followed.

*“P1: The Lanes on Nelson Street . P2: The Lanes. P3: We have just been told to go to the Lanes. P4: I have just been told to go to John Street, but the Lanes are pretty close. P2: Let’s go to the Lanes then. P1: The Lanes!”*

### Team dynamics

The way the team was socially structured also related to whether instructions would be followed. The dynamics of the team as we described earlier shaped how instructions were negotiated and acted upon. Further, in some cases, one or two people acted as leaders within the team, which meant they decided on where the team went to next and what instructions were to be followed or not.

### Mishearings and misunderstandings

Often instructions were not followed as a result of the information being misheard or not fully understood by the player who received them. The game being played in a noisy street environment and also the computerized voice of the Instructor led to messages being misheard by players.

Misheard information was then relayed to the team leading to confusion among the players, or equally to mistrust towards the Instructor, both of which resulted in instructions not being followed. One example of the latter was when right after a call, a player communicated to his team that the Instructor was lying. When asked why is that, he replied: *“Because we got a message saying that we all lost our points”*. Since that was not the case, the player assumed that the system was lying; however, such a message did not exist in the system, the player had just misheard the information.

### Timing

The point in time that instructions were delivered was also found to be relevant to whether players would follow them or not. For example, if instructions were delivered while players were running, it was likely that they would not be followed and potentially not even received. Similarly, players’ availability and receptivity to instructions was also restricted by the danger factor of the policemen. In the next episode, players are negotiating on the instructions while a policeman sneaks up on them. Players run away and none of the instructions are followed at this point.



*“P5: I have been told to go the Bank but others have been told to go to the Lanes...P2: So why are we going...hey you guys, guys let's come together. We have been told different information.*



*P3: We got three different venues so let's have a look at the map (team comes together and open their maps (cop is right behind them). P1: (whispering) There is a man behind us that we need to avoid (starts walking away and two other players follow him.*



*P2: Hey guys, shouldn't we stick together as we need to catch the boat together?” (Rest of players realize there is a cop behind them)*



*(Everyone starts running away)*

## DISCUSSION

This work aimed to explore issues regarding an agent delivering instructions to teams of humans and did so in the context of a pervasive game played in the streets of a British city. Contrary to previous work, we were interested in how teams and not individuals received and negotiated instructions. A further contribution of this work relating to the field of human-agent interaction is the methodology itself. There is no precedent of a large-scale pervasive game to be used before to explore human-agent interaction. As shown, the benefits of such an ‘in the wild’ deployment compared to software simulations or purely lab-based experiments are significant: detailed descriptions of what took place can assist in unfolding the complexities in the interaction between humans and agents, including visceral or environmental aspects of decision-making (e.g. fatigue, fear); and provide pointers for further experimental investigation.

This work has uncovered the ways teams and individuals engaged with the agent and followed (or not) its instructions with an emphasis on how shared understanding

and negotiation were realized. These two areas have been established in existing work [6][27] as vital to human-agent interaction and particularly challenging for human agent coordination. In this work, as part of sharing and negotiating, players had to make sense of ambiguous information and this was shown to be informed by assumptions about the character of the Instructor, leveraging the perceived goals of the game and other contextual information against the received information. Often negotiating meant that teams/players would not follow the instructions despite the fact that in the game pack it was stated that the information given through the Instructor is “useful” and “vital” for the game. The instructions were further always correct, in the sense that they did not send the players to any dead stations, and yet players ignored them with ease.

A main consideration from this is that any assumptions about obedience, trust – and equally intelligence – that might have been built in the design of the agent were challenged, and exposed the fragility of such systems. The assumptions relating to the Instructor and the persona developed around it, shaped notions of trust and mistrust that were paramount not only to the following of instructions but also to the intra team relations and communication. With this in mind, it is important to consider ways in which the interface can be more sensitive or contingent to the context at hand. Investing into building a relationship with the agent, potentially through interfaces of conversation - rather than discrete instruction messages - that take into account the on-going assumptions and allow for human-machine reconfigurations [22] is a more appropriate design approach. Likewise, a trustful agent will more easily inspire a change of priorities within the team if needed. In the next section, we provide recommendations for design that address the insights from this work.

## Design Recommendations

We provide five recommendations that address the issues discussed above. While these recommendations have emerged from a context of an agent instructing teams, we believe they are relevant to human-agent interaction design in general and can be applied equally to cases of agents instructing individuals.

**Reliability is subjective.** Trust in the agent does not depend on the reliability of its instruction, but in the user's perception of this reliability. It is not sufficient for the agent to give useful and correct information to a user in order to establish trust. Agents that rely on user's trust for their instructions to be followed must thus provide correct information and at the same time ensure that the user perceives this information as correct. Otherwise it must provide a mechanism for correcting user's perception when they do not match what is desired.

**Users are dynamic; agents should be too.** The relationship between users and agents is not static. It evolves over the course of the interaction. One limitation of our agent is that

while its relationship with the subjects changed over time, its own behaviour remained the same, leading to divergence between how some users interpreted its instructions and how it expected them to be interpreted. An agent instructing users needs to adapt when it becomes clear that the user is no longer interested in following their instructions. The agent should provide ongoing feedback as a part of the process of updating the shared model of the situation [21], and overcome the barriers associated with disinterested users.

**Ignored instructions are not equivalent to non-compliance.** Users are not always able to follow instructions for a variety of reasons. Users may not fully comprehend or appreciate the systems activities, or feel that the instructions conflict with their immediate goals [6]. This may be attributed to the relative ‘invisibility’ of the agent’s intentions and activities, increasing the cognitive demands on users to interpret and evaluate the instructions given [28]. A consequence is, that rather than engaging with the agent, users revert to manual or direct control, often in an attempt to reclaim their own understanding or control of the situation [6]. An agent should not judge lack of compliance to its instructions as a sign that the user is purposefully not following instructions.

**Agent instruction needs to take team cohesion into account.** In unstructured teams, when an individual acts a single point of delivery of agent instructions, then the role and position of that individual within the team strongly impacts how those instructions are handled. When an agent becomes part of a team, this role becomes critical for the interaction between team and agent. Hence, the agent should make allowances for this in terms of the timing, format and content of any communications [28]. The agent’s presence can also add additional co-ordination demands to specific members of the team [6]. An agent may need to send out the same instruction to multiple individuals or select an individual of higher standing within the team to send instructions to. This is part of a global strategy to ensure the agent’s actions are synchronised and consistent with the team [6].

**Manners matter.** Simply delivering instructions to users is not sufficient. It is also important to consider the manner in which they are delivered. The fact that the Instructor called users on their phones and addressed them with their real names had a strong influence on the users’ willingness to engage with it. The anthropomorphic nature of the agent’s behaviour, especially when asking users, “Don’t you trust me?” led to a significant change in how users decided to proceed in their engagement with the Instructor.

#### FUTURE WORK

The study presented in this paper was an initial exploration on the use pervasive games to study how teams of users respond to agent instructions. Its design was not meant to mimic real world scenarios, nor extract generalizable findings. It was rather focused on creating an engaging

gaming experience for players. However, some of the core mechanics of the game share similarities with various real-world scenarios where agents will need to instruct teams of people under pressured conditions and issues such as trust and authority, timing and proximity that were found in our game will be of high importance. For instance, in the case of emergency evacuation/crowd control, an agent may collect information from a number of sources (drones, sensor infrastructure, local authorities) and instruct groups of people on available exit points; or in the case of urban search and rescue, where an agent may coordinate and instruct teams regarding worksites or potential hazards.

Our focus now moves towards establishing a more empirical approach that would produce more concrete and generalizable results. Of particular interest is studying more realistic models of human-agent interaction, for example by focusing on two-way interaction where human users would provide feedback to the agent instructor, rather than the simple one-way communication used in Cargo. In addition, more controlled experiments will be required to compare different configurations of instructing humans via agents: agents directly instructing humans, agents instructing humans through human intermediary, or humans instructing humans through agent intermediaries.

#### CONCLUSIONS

In this paper, we set about exploring how teams of humans respond to instructions from an intelligent agent. To explore these issues in situ and on a large scale, we developed a pervasive game called Cargo, which sees the Instructor, a semi-autonomous agent, use text-to-speech to direct teams of players (via their mobile phones) around a city in the UK. Our observations indicate that the teams’ interpretation of, and compliance with, instructions were shaped predominately by their perceptions and assumptions of the agent. Many participants attributed an authoritative characteristic to the agent, and trust was built (or lost) as the relationship was developed over the period of the game. Upon receipt of instructions by individuals, information was shared with the team in different ways, and frequently led to team members negotiating their next actions. Based on the findings, we presented five design implications to consider when creating intelligent interface systems that instruct teams of humans.

#### ACKNOWLEDGMENTS

This work was supported by EPSRC Grant No. EP/I011587/1.

#### REFERENCES

1. Ball, M. and Callaghan, V. Introducing Intelligent Environments, Agents and Autonomy to Users. *Proceedings of the 7th International Conference on Intelligent Environments (IE '11)*, (2011), 382–385.
2. Benford, S., Magerkurth, C., and Ljungstrand, P. Bridging the physical and digital in pervasive gaming.

- Communications of the ACM (CACM)* 48, 3 (2005), 54–57.
3. Benotti, L. and Denis, A. Giving instructions in virtual environments by corpus based selection. *Proceedings of the 12th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL '11)*, (2011), 68–77.
  4. Bulitko, V.V. and Wilkins, D.C. Automated Instructor Assistant for Ship Damage Control. *Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Innovative Applications of Artificial Intelligence Conference Innovative Applications of Artificial Intelligence AAAI '99/IAAI '99*, (1999), 778 – 785.
  5. Chan, V., Ray, P., and Parameswaran, N. Mobile e-Health monitoring: an agent-based approach. *IET Communications* 2, 2 (2008), 223–230.
  6. Christoffersen, K. and Woods, D.D. How to make automated systems team players. *Advances in human performance and cognitive engineering research* 2, (2002), 1–12.
  7. Cuevas, H.M., Fiore, S.M., Caldwell, B.S., and Strater, L. Augmenting Team Cognition in Human-Automation Teams Performing in Complex Operational Environments. *Aviation, Space, and Environmental Medicine* 78, 1 (2007), B63–B70(8).
  8. Doswell, J. and Harmeyer, K. Extending the ‘Serious Game’ Boundary: Virtual Instructors in Mobile Mixed Reality Learning Games. *Proceedings of the Digital Games Research Association International Conference (DiGRA 2007)*, (2007), 524–529.
  9. Fiore, S.M., Salas, E., Cuevas, H.M., and Bowers, C.A. Distributed coordination space: Toward a theory of distributed team process and performance. *Theoretical Issues in Ergonomics Science* 4, 3-4 (2003), 340–364.
  10. Fischer, J., Jiang, W., and Moran, S. AtomicOrchid: A Mixed Reality Game to Investigate Coordination in Disaster Response. In *Mobile Gaming Workshop (MOGA) 2012 as a part of Proceedings of the 11th International Conference on Entertainment Computing (ICEC 2012)*, (2012), 572–577.
  11. Fischer, J.E., Flinham, M., Price, D., Goulding, J., Pantidi, N., and Rodden, T. Serious Mixed Reality Games. In *Proceedings of Mixed-Reality Games Workshop as a part of the International Conference on Computer Supported Cooperative Work (CSCW'12)*, (2012).
  12. Hexmoor, H. and Heng, T. Air Traffic Control Agents: Landing and Collision Avoidance. *Proceedings of International Conference in Artificial Intelligence, (IC-AI '00)*, (2000).
  13. Heylen, D., Nijholt, A., op den Akker, R., and Vissers, M. Socially Intelligent Tutor Agents. *Proceedings of Intelligent Virtual Agents (IVA '03)*, Lecture Notes in Artificial Intelligence (2003), 341–347.
  14. Koschmann, T., LeBaron, C., Goodwin, C., and Feltovich, P. The mystery of the missing referent: objects, procedures, and the problem of the instruction follower. *Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW '06)*, ACM New York, NY, USA (2006), 373–382.
  15. Maes, P. Agents that reduce work and information overload. *Communications of the ACM* 37, 7 (2007), 30–40.
  16. McGrath, S., Chacón, D., and Whitebread, K. Intelligent Mobile Agents in Military Command and Control. *Proceedings of the Autonomous Agents Workshop on Agents in Industry*, (2000).
  17. Milgram, S. *Obedience to Authority: An Experimental View*. Harper Perennial Modern Classics, 2009.
  18. Mittrup, I., Smarsly, K., Hartmann, D., and Bettzieche, V. An Agent-Based Approach to Dam Monitoring. *Proceedings of the 20th CIB W78 Conference on Information Technology in Construction*, (2003), 239–246.
  19. Murata, K., Enomoto, M., Arimoto, Y., and Nakano, Y. When Should Animated Agents Give Additional Instructions to Users? - Monitoring User's Understanding in Multimodal Dialogues. *International Conference on Control, Automation and Systems, ICCAS '07*, (2007), 733 – 736.
  20. Murthy, S., Akkiraju, R., Rachlin, J., and Wu, F. Agent-Based Cooperative Scheduling. *Proceedings of AAAI Workshop on Constraints and Agents*, AAAI Press (1997), 112–117.
  21. Norman, D.A. The ‘problem’ with automation: inappropriate feedback and interaction, not over-automation. *Philosophical Transactions of the Royal Society of London. B* 327, 1241 (1990), 585–593.
  22. Pacaux-Lemoine, M.-P. and Debernard, S. Common work space for human-machine cooperation in air traffic control. *Control Engineering Practice* 10, 5 (2002), 571–576.
  23. Schaafstal, A.M., Johnston, J.H., and Oser, R.L. Training teams for emergency management. *Computers in Human Behavior* 17, 5-6 (2001), 615–626.
  24. Suchman, L.A. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, 1999.

25. Sukthankar, G., Shumaker, R., and Lewis, M. Intelligent Agents as Teammates. In Eduardo Salas, S.M. Fiore and M.P. Letsky, eds., *Theories of Team Cognition: Cross-Disciplinary Perspectives*. 2012, 313 – 343.
26. Tolmie, P., Benford, S., Flinham, M., et al. “Act Natural”: Instructions, Compliance and Accountability in Ambulatory Experiences. *CHI '12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (2012), 1519–1528.
27. Traum, D., Rickel, J., Gratch, J., and Marsella, S. Negotiating over Tasks in Hybrid Human-Agent Teams for Simulation-Based Training. *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '03)*, ACM New York, NY, USA (2003), 441–448.
28. Woods, D.D. Human-centered software agents: Lessons from clumsy automation. In S. J. Flanagan, T. Huang, P. Jones, & S. Kasif, ed., *Human centered systems: Information, interactivity, and in*. Washington, DC: National Science Foundation, 1997, 288–293.
29. <http://cars.uk.msn.com/features/sat-nav-catastrophes> Accessed on 09/01/2013.
30. <http://www.phidgets.com> Accessed on 09/01/2013.
31. <http://slingshoteffect.co.uk/> Accessed on 09/01/2013.
32. <http://igfest.org/> Accessed on 09/01/2013.