Can Visualization of Contributions Support Fairness in Collaboration? Findings from Meters in an Online Game

Ryan Kelly, Leon Watts, Stephen J. Payne
Department of Computer Science
University of Bath, Bath, United Kingdom
{r.m.kelly, l.watts, s.j.payne}@bath.ac.uk

ABSTRACT
In this paper we consider how visualizations might support judgements about fairness in collaborative work. We present a qualitative investigation of *meters*, existing artefacts that enable awareness of contributions in the online game World of Warcraft. Through in-depth interviews with game players, we draw attention to the value of meters as tools for self-reflection and group evaluation. Yet we also describe problematic consequences that arise as a result of meters, distinguishing their usage as in-the-moment awareness tools from instruments used after the fact to apportion credit and blame. We argue that representations like meters may be enough to provoke initial questions about fairness, but are likely to undermine the efforts of collaborators if they fail to combine a set of indices that reflect important aspects of individual work in the context of team activity. We consider broader lessons for the way in which future designs might aim to support fairness in collaborative systems, pointing to multidimensional mechanisms adapted specifically to real-time or retrospective usage.

Author Keywords
Awareness; Fairness; Social Coordination; World of Warcraft

ACM Classification Keywords
H.5.3 Information Interfaces and Presentation (e.g. HCI): Collaborative Computing

INTRODUCTION

Developing systems to support collaboration is a significant research challenge, and part of that challenge involves understanding the properties of group interaction that are relevant to harmonious working. One such issue that has received little explicit attention in the HCI and CSCW literatures is that of *fairness*—in particular, how perceptions about fairness impact a group's work process. Fairness is an important value that guides many social exchanges and feeds into the dynamic coordination of collaborative group work. In particular, perceptions about what is fair can influence the allocation of work among a team, the level of effort that is invested, and, later, the distribution of rewards from a group's effort [23]. Concerns about fairness can present problems for groups; for example, the well-known 'free-rider problem' occurs when "one or more members of a group do not do their fair share of work on a group project" [6]. Prevention of free-riding requires that groups attend to the status of ongoing contributions to ensure that work is completed in line with fairness norms. Such monitoring may, however, be difficult in computer-mediated environments where awareness of one's colleagues is low and knowledge about effort is incomplete [18, 24].

Research has long espoused the importance of awareness for supporting collaborative work (e.g. [13, 15, 19]), and a range of technologies have been developed to support awareness among distributed teams [15, 22, 25]. However, no research has yet explored the question of how awareness mechanisms might specifically support fairness, or the impact that such support might have on the behaviour of collaborative teams. In this paper, we explore these issues through a qualitative study of *meters* in the online game World of Warcraft (WoW). Meters are interface widgets that quantify and visualize individual contributions to the overall team effort. We regard meters as an example of how support for awareness in service of fairness judgements might be implemented in a collaborative system. Our study focuses on how meters impact the work of WoW teams during complex, intrinsically collaborative 'raid' tasks in which players must work together to ensure success.

The present investigation describes how meters benefit groups in WoW, both in terms of enabling immediate assessments about contributions and in supporting retrospective review over the longer term. However, we also highlight a range of tensions and sociotechnical concerns that arise as a result of how meters are interpreted by players. We argue that meter-like representations inevitably undermine valuable contributions to group work, particularly when those contributions are difficult to measure. This research extends previous work on feedback systems for groups (e.g. [1, 4, 11, 12, 29, 36]) to a context beyond the realm of the laboratory, and derives lessons for enabling awareness of contributions and judgements of fairness in other collaborative work applications.

BACKGROUND & RELATED WORK

This study is at the intersection of three areas of research: fairness [6, 17, 23, 30], awareness technologies [13, 15, 19, 22] and collaboration in WoW [2, 3, 33]. Here we consider

Authors' accepted version, to appear in proceedings of CSCW 2016.

how fairness might be supported through visualizations that promote awareness of contributions, before presenting meters as a case in point. We then provide contextual grounding on WoW and the use of meters within WoW teams.

Understanding Fairness in Collaborative Work

There are many rules one can follow in order to determine what is fair in the course of collaborative work. Perhaps the simplest rule is that of equality [23]. That is, people may be expected to provide equal or near-equal input to a project if there is no good reason to behave otherwise. But while the rule of *"what is equal is fair"* may be satisfactory in many situations, it is unlikely to remain sufficient when one begins to consider the differential types of work that collaborative projects typically involve, as well as the range of skills and capacities available to project teams. Even if no differences in capacities exist, there may be a predetermined distribution of work that deviates from equality and yet still qualifies as fair.

Research has identified that judgements about what is fair can be influenced by a range of principles, including perceived needs, rights, and entitlements, as well as claims about merit, status, ability, efforts, or mitigating circumstances [9, 23, 30]. Each of these factors serves to *legitimise* particular patterns of behaviour. In other words, fairness can be characterised in terms of judgements about whether *deviations from equality* are legitimate in one's current situation [38]. In collaborative work, an understanding of job roles, for example, may prescribe a specific distribution of tasks, but task assignments may be adjusted by knowledge about current workload or other relevant factors that legitimise unequal contributions.

Fairness Theory [17] argues that assessments of legitimacy occur through counterfactual thought; in short, comparing what is to what should, could, or might have been in alternate scenarios. Comparisons between imagined and actual states allow one to assess whether current behaviour is justified, and hence permissible, or whether there is evidence to suggest that what is happening violates some normative standard and is thus unfair [17]. The only way to answer questions about what is fair is to collect information pertinent to the issues about which one is concerned. Fairness is thus *inferred* from behavioural information, invocation of relevant fairness rules, and an understanding of how those rules should be applied in one's current social context [30]. However, this makes fairness problematic in that it can be highly subjective; individuals may weigh rules differently, and there may be differences of opinion about what is legitimate in a given situation.

In collaborative work, perceptions about fairness are necessarily tied to an understanding of the contributions that are expected of one's collaborators, as well as the extent to which those contributions are being fulfilled. This means that teams need to monitor the status of work to assess whether coworkers are exceeding, adhering to, or failing to fulfil *commitments* [30] relevant to the work situation. However, monitoring the status of work is known to be difficult when information about the actions and efforts of one's colleagues is not easily accessible [13, 21]. This issue was emphasised in early work by Galegher & Kraut [18], who found that groups working face-to-face perceived greater fairness in their work process than other groups working via mediated communication technologies. The implication here is that groups may find it especially difficult to make judgements about fairness when the completion of contributions is obscured by technology.

In summary, we define fairness as a perceived quality that arises through a process of social judgement about what is legitimate in a given situation. In collaborative systems, it may be difficult to perform these judgements without sufficient awareness, defined as "an understanding of the activities of others, which provides a context for your own activity" ([13] p. 107). Following previous research that has demonstrated the value of visual information for building awareness in teams (e.g. [15, 20, 22, 25]), we focus on how contributions can be presented in the form of a visualization. We reason that information about contributions will be of intrinsic benefit to teams by improving collective awareness of individual efforts. This might then allow individuals to engage in information gathering, counterfactual thought, and checking adherence to commitments, all of which are relevant to judging fairness.

Visualizing Contributions to Collaborative Work

Previous research has explored how contributions can be visualized, and although these studies were not designed to investigate the issue of fairness, they are instructive in terms of understanding how visualizations influence group work.

Studies of group mirrors—tools that provide feedback to groups by reflecting specific aspects of their activity—have found that visual information can promote equality of participation [1, 4, 36]. For example, *Second Messenger* [11, 12] is a system that records speech acts in face-to-face discussions and presents these contributions as a histogram on a shared display. The system was found to increase equality of participation by encouraging over-contributors to surrender the floor to their teammates [11, 12]. Turning to text-based communication, Leshed et al. [28, 29] describe *GroupMeter*, a system that allows groups to reflect on their language use through visualizations of word count and the frequency of agreement words. They found that visualizations led to improved levels of agreement compared to situations in which visualizations were not present. The *Participation Tool* described in [27] caused groups of students to increase the length of their utterances, a change that was associated with improved coordination and information sharing. Interestingly, students using the tool felt more able to spot free-riding through monitoring of group activities, implying that information about contributions can indeed be useful for fairness-related assessments.

Other work has examined the role of visual information in promoting awareness of efforts during collaborative writing. Many modern word processing applications provide edit histories that record changes made to a document. An interview study by Birnholtz & Ibara [5] revealed that teams sometimes use these histories to identify, monitor, and evaluate individual contributions. Recent design work by Wang et al. [40] presents *DocuViz*, a tool that utilises the edit history of Google Docs to present a timeline visualization of changes made to a document. The authors highlight how the visual information can help to draw attention to "seismic activity"—patterns of work that appear to have been especially contentious and

thus worthy of further investigation. Similar representations have also been demonstrated as useful for unveiling patterns of collaboration on Wikipedia [39].

While these studies suggest that visualizations of contribution can be of benefit to groups, they also draw attention to challenges that arise as a result of making contributions visible. For example, collaborators may fail to act on a visualization if they believe that it does not accurately represent their own behaviour [12]. Moreover, feedback about effort can lead previously successful groups to adopt detrimental strategies, e.g. by encouraging groups to aim for equality of participation, despite equality leading to poorer performance [11]. Information about contributions may also have negative relational consequences by provoking discontent about how work will be evaluated by others [5]. This may be especially true if a mechanism does not represent all of the contributions that are meaningful to a task [40].

These issues stress that further research is required to understand the impact that visualizations have on group behaviours, both in producing positive outcomes but also in how they may generate disharmony. Other than the research on collaborative writing described above, prior studies have been short-term, experimental, and focused on evaluating interactive prototypes, suggesting a particular need for research in uncontrived contexts. Most importantly for our research, there is a lack of exploration of specific parameters that would be required to support fairness judgements. Studies of the extent to which existing awareness artefacts are, and are not, sufficient for judging fairness would be instructive in this regard.

With these considerations in mind, we turn to the example of meters, a type of mechanism that measures the contributions that players make to groups in the online game World of Warcraft (WoW). Our interest in WoW is motivated by the fact that grouping in the game world is prevalent and often highly regimented, involving structured divisions of labour, tight coupling between individuals, and high coordination requirements. This means that the 'work' required of players is similar in several respects to collaboration in other settings—several authors characterise WoW players as 'working for fun' (e.g. [8, 42]) and studies suggest that findings about game groups can provide lessons that apply to other CSCW settings [2, 3, 33, 34]. WoW grouping also involves concerns relevant to fairness, including the prevention of free-riding [31] and the distribution of rewards from teamwork [26].

In this paper we report findings from an interview study with 10 WoW players concerning their personal experiences with meters. Our aim is to understand how meters support players' inferences about the contributions of their teammates and, by extension, how the information can be used to permit judgements about the fairness of contributions. Furthermore, we investigate how deficiencies of meters can make judgements of fairness problematic. We also probe additional challenges, beyond those reported previously in the literature, that arise as a result of publicising contributions in a meter.

To provide contextual grounding for our study, the following section offers an overview of WoW and the collaborative situations in which meters are used to record contributions. We then outline the qualities of meters that we believe are relevant to fairness, casting each as a more general design parameter that could be leveraged in other tools that aim to promote awareness of contributions.

Study Setting: World of Warcraft Raid Groups

World of Warcraft is a massively-multiplayer online game (MMOG) in which players create a character and enter a virtual environment populated by various three-dimensional fantasy avatars (orcs, elves, gnomes and the like). Like other MMOGs, WoW emphasises the collaborative and social aspects of gaming [14]. Players are encouraged to communicate through voice- and text-based chat tools in order to form teams and quest together. Grouping is especially prevalent at later stages of the game, where players must band together in teams of 10 or 25 to defeat challenging dungeons known as 'raid' encounters [2, 3]. Raids typically last several hours, and involve a series of battles with increasingly tough 'boss' enemies [2]. Defeating each boss allows players to acquire rewards known as 'loot', yet only a limited number of items are acquired from each encounter, meaning that no individual is guaranteed to be rewarded for their efforts [2, 26].

What makes raids compelling is that they are by no means straightforward, with each requiring a great deal of advance planning and intense coordination among players. Moreover, the risks associated with raiding are high, and each encounter requires teams to follow a specific strategy. The failure of one or more individuals to execute this strategy can result in death for the entire team, an outcome known as a 'raid wipe'. Since wipes are costly, it is critical for players to participate fully in a raid if they want to maximise their chances of success. These issues make collaborator awareness and accountability central concerns in raid activity.

In raids, accountability is partially supported by formalised *social* and *ludic* roles [3], each of which establishes a different type of high-level division of labour among members of a raid team. Social roles are those that are created by the players themselves. For instance, all raids have a leader who acts as a kind of project manager. He or she is typically responsible for creating the group, assigning tasks before the start of the raid, and outlining attack strategies. Other social roles include the 'communications officer', responsible for issuing instructions during battles, and the 'loot master', a player who oversees the distribution of rewards from raid activities. These features make raids akin to real-world work groups.

By contrast, ludic roles are defined by each player's 'character class'. Different classes have different abilities that define individual responsibilities within a raid setting. Some classes, like Warriors and Death Knights, must protect the raid by holding the attention of bosses and other enemies (a practice known as 'tanking'). Classes like Warlocks, Mages, Hunters, and Rogues focus on dealing damage required to defeat enemies. Classes with healing abilities (Druids, Priests, Shamans) are responsible for restoring the health of other players and providing benefactions. Not only do these roles create interdependencies within raid teams (characters responsible for

tanking enemies will not live long without the aid of healers, and neither healers nor tanks can complete a raid without the support of damage dealers), they also establish certain expectations about contributions in advance. For instance, it is expected that healers will keep everyone else alive; thus, if a character dies during a raid, suspicion of fault may initially be directed towards players responsible for healing [3].

Awareness in Raid Groups

Awareness and monitoring of teammates are critical requirements for raids: because the line between success and failure is often very thin, teams can not afford to tolerate non-participation or free-riding from their members. There are at least three ways in which players build awareness in WoW:

- First, players can explicitly communicate about their activities via text-based chat channels. However, because text chat is cumbersome (players need to use their hands to control their character) the majority of groups use third party programs to enable voice communication [8].
- Second, as a mimetic representation of its user, an avatar naturally conveys information useful for awareness as it moves around the virtual environment [19]. This means that simple observation of the physical positioning of one's comrades, as well as the animations portrayed by their avatar, may sometimes be sufficient to assess whether or not they appear to be doing as they should be.
- Finally, WoW has a number of interface elements that support awareness, e.g. a minimap that shows a player's current location relative to others in the environment. These visual elements lower coordination overheads by alleviating the need to explicitly request information [34].

Prior work on awareness in WoW and other virtual environments has identified some deficiencies with the above methods [7, 14, 31, 32, 34] but one issue that has not been investigated is that these channels have proven to be insufficient for monitoring the contributions of raid members. To address this, the majority of raid groups use *meters* as a mechanism for capturing contributions and promoting more detailed awareness of actions. These meters are a type of interface 'add-on' that players install to augment the standard WoW interface, meaning that players are not ordinarily able to inspect the contributions of other raid members without meter support.

Meters as Visualizations of Contribution

Figure 1 displays an example of a meter from WoW. The meter works by parsing data from the system layer of the WoW client. In order to see this data, each member of the raid must have the relevant add-on installed. (Installation of a meter is often a requirement to participate in organised raid teams [37].) Members of the team are represented by coloured bars, with each representing the *damage* an individual has inflicted to the group's enemies. (Meters also exist for other activities like healing, and are qualitatively similar to that of Figure 1.) All damage-related actions are collapsed into a single metric and are automatically presented in aggregated form. Players are ranked in accordance with total damage output, with those at the top of the meter contributing the most damage. The

Figure 1. Screenshot of the ReCount addon, a damage meter used in World of Warcraft.

colour of each bar pertains to the player's character class and is consistent with other aspects of WoW—yellow is associated with Rogue characters, green with Hunters, and so on. The numbers overlaying each bar pertain to each individual's contribution, representing (from left to right in Figure 1) total damage inflicted by the player, their damage per second (known as 'DPS'), and their contribution as a percentage of the overall group effort. These numbers, as well as the position and size of the bars, fluctuate in real time to match current performance. Thus, one player may begin at the top of the meter but can eventually be overtaken by others.

It is important to note that the interpretation of meter data has its subtleties. In Figure 1, for example, the players represented by bars 9 and 11 are healers who are not expected to deal damage. Thus it is immediately possible for other players to make an assessment about teammates—if a healer is high on the damage meter, it would suggest that he or she is not performing their duties in the correct way.

Beyond the surface representation of damage, players can interrogate the meter data in several ways. This makes it possible to explore actions that are not directly related to damage but are nonetheless important to success on raids. For example, some classes in WoW have abilities that they must use at specific points in a fight, e.g. 'Mages' have a spell that turns an enemy into a benign character, preventing its attacks. Players will be assigned these duties in advance, and most meter applications are capable of recording their use. Clicking through submenus allows groups to investigate this data as an additional source of information about the effectiveness of a player's contribution. However, raids are frequently too intense to permit extended interrogation, meaning that this type of investigation typically occurs after the fact (as will be evidenced by the results of our study).

Design Parameters of Meters Relevant to Fairness

We contend that meters have a number of properties that are pertinent to supporting fairness in collaborative work. Here we underscore these properties in terms of two design parameters, positioning each in accordance with our earlier theoretical considerations about fairness.

First, the hierarchical ordering and delineation of players based on contributed work allows for *social comparison* [16] between individual contributions. This is important because fairness is inherently social and people often assess fairness by evaluating their own situation vis-à-vis that of referent others [17]. We believe that allowing for comparison of one person's contribution relative to the next will enable collaborators to answer important questions related to fairness (e.g. *"Am I contributing fairly? Are those other people being fair?"*). Moreover, facilitating social comparison allows collaborators to make judgements about similarity in contributions. This should support fairness by allowing collaborators to observe *deviations from equality*; subjective judgements can then be made about the *legitimacy* of such deviations, and whether or not it appears that collaborators are *adhering to commitments*.

The second parameter pertains to disclosure of *detail about contributions*. That is, damage metrics are shown at a level of detail that allows for social comparisons to be meaningful. Prior research on awareness highlights the inherent tension between the collective requirement to share information and the individual need to avoid potentially damaging disclosures. This is usually described as the problem of maintaining privacy [25]. Some approaches to awareness, including *social translucence* [15], indicate that obfuscating information about actions can allow participants in a collaborative system to be aware of who is around and active, without violating privacy. However, simply knowing that others are 'active' would likely prove insufficient for judging fairness—collaborators may wish to answer more specific questions, such as *who* is doing *what*, as well as *how much* or *how well* work is being done.

One way to think about this parameter is in terms of a continuum between the mere fact that a team member is expending effort through to a complete representation of what that effort entails. If sufficient information about work is disclosed, one might be able to answer questions about whether particular levels of performance are appropriate to the current situation. The type of meter shown in Figure 1 represents a single point along this continuum. Simply eyeballing the mechanism is enough to know who is in the raid and whether or not they are active, especially when combined with visual information available in the standard WoW interface (e.g. which avatars are nearby, and what actions they appear to be performing). Comparisons are then made more precise through the provision of exact figures about each player's damage, thereby indicating their current level of contribution. In this way, the meter not only allows players to compare their contributions but also to reason about the difference between rankings.

It is important to emphasise that we do not regard meters as a panacea for fairness support. Rather, we see meters as a single point within a broader design space for supporting judgements about fairness. However, the nature of this space, as well as the key research challenges that exist within it, are currently not well defined. An initial and obvious design critique of meters is that they are limited in scope, providing only a small amount of information about contributions and roles. Nevertheless, meters do seem sufficiently rich to warrant some degree of evaluation. Previous work by Taylor [37] noted that players in WoW are ambivalent about the use of meters. On the one hand, players find them helpful in monitoring performance and in encouraging friendly competition through observation of the rankings. On the other, players can feel that their contributions are not meaningfully represented because meters do not capture all of the effort that is invested into the team. The potential frustration that might arise from the ambiguity of meter information suggests that future designs could be informed by a careful analysis of player experiences before, during, and after raids. The present study finds that Taylor's observations hold true, but also provides a deeper understanding of meters derived from player narratives, raising new design-relevant problems for contribution awareness mechanisms.

INTERVIEW STUDY: METHOD

Following the precedent set by earlier work on WoW [2, 3, 14], this study was informed on several levels. First, the lead author is a former WoW player with several years' experience of raiding (50+ raids). This participation supports our analysis by providing a grounded understanding of the social context in which our participants' collaboration occurs [41].

Second, we developed an up-to-date understanding of raid behaviour by engaging in observation of players completing current high-level raid content in WoW. Observations were conducted by watching web streams of raid groups, broadcast live via *Twitch.tv*. This phase did not involve collection of data from players; rather, we took notes about aspects of the observed raid encounter, including how the raid was organised, coordinative behaviours performed by players, and the types of interface elements used. These notes, combined with our own experiences of WoW, helped shape the questions used in our third and final stage, where we performed semi-structured interviews with current WoW players. These interviews aimed to capture players' experiences of raiding alongside topics related to the use of damage meters. The findings we report in this paper are derived solely from this latter data source.

Participants

We interviewed a total of 7 male and 3 female WoW players. Interviewees' ages ranged from 18–46 (M = 30). Interviewees had been playing WoW for an average of 6 years (range = 4–8.75 years). All had experience of raid encounters ranging from dungeons in WoW's initial 2004 release through to those in its most recent expansion at the time of study (*Mists of Pandaria*). Our interviewees had experience with a variety of ludic roles in WoW raids, including tanking (protecting other players), dealing damage, and healing. Two players had held the position of raid leader within their groups.

All interviewees had participated in formal raiding guilds. A guild is a persistent collective of anywhere between two to several hundred players [41]. Raid groups are often comprised of players from the same guild—effectively, the formal structure of a guild facilitates persistent grouping [2] and allows certain policies to be implemented that help to manage raids in the longer term, e.g. a 'guild bank' for storing loot. Two of our interviewees described the current guild of their main character as casual (raiding occasionally, perhaps 1–2 times per week), five stated 'medium-core' (2–3 raids every week), and three as

'hardcore' (4 or more nights every week), though during their WoW careers most interviewees had moved between guilds with varying demands of raid commitment. Three informants had held positions of authority within their guilds. Additionally, all respondents had engaged in raids with their guild and in 'pick-up' raid groups (PUGs). PUGs are one-off collectives that form to complete a dungeon and then disband afterwards. Many had experience of PUGs constructed through Blizzard's 'Looking for Raid' (LFR) tool, a system that automates the allocation of players to raids. These demographic aspects suggested to us that the raid experiences we surveyed were broad and were not isolated to either casual or hardcore players.

Procedure & Analysis

Interviewees were recruited through word of mouth and posts on Facebook. Interviews were conducted one-to-one by the first author. Two interviews were face-to-face and eight were via Skype. Interviews lasted 50–90 minutes, at the discretion of the interviewee. Interviewees were not paid for their participation. Interviewees were told that their responses would be anonymised. Questions (available as auxiliary material to this paper) covered a range of raid-related topics, including how interviewees determined appropriate actions in the context of a raid; the interface elements used to build awareness of teammates; and how players determined whether others were contributing properly on raids. We also asked whether they used meters (which all respondents did) and probed their opinions and experiences with these mechanisms. Interviews ended with a short debrief about the purpose of the study.

All interviews were conducted and transcribed by the first author, allowing for early familiarisation with the dataset and development of initial understanding. We used open and axial coding [35] to categorise responses. The coding process was iterative; transcripts were given two complete readings and initial codes were evaluated and refined through constant comparison, with internal consistency strengthened by scrutinising the data for counter-examples. Codified concepts were then structured into the themes that comprise our results. While it is unclear whether we reached true theoretical saturation, we were seeing repetition in the content of participants' statements by the end of data collection, and thus we felt that we had sufficient data to make a coherent set of claims about the use of meters in raids.

Our analysis resolved on eight themes. We divide these themes into three categories of concern: first, five themes relate to how meters support groups in assessing contributions. One theme describes the limitations of meters for assessing contributions, and two final themes relate to issues that arise as a result of publicising contributions. These themes identify both positive and negative issues from player perspectives. In describing our results we use quotations to illustrate particular points, identifying speakers by participant number, gender, and age.

RESULTS

In analysing our results, we found that awareness is of real concern in the context of raids. Interviewees repeatedly asserted the necessity of maintaining an overall understanding of what is happening at any given time. One participant referred to this as staying 'raid aware', a term that was repeated by other interviewees. Raid awareness is perhaps best likened to collaborative workspace awareness [19] in that it relates to a very general understanding of the group's status and their interaction with various entities in the virtual environment. As one interviewee summarised:

> *"It's important to be aware of everyone around you—this includes healers, tanks, and damage dealers—so you can spot someone who may be standing in the wrong place, or may just be in a bad place in general... you need to focus your eyes on the centre of the screen so you're able to watch the boss and the ground for incoming attacks and keep an eye out for bad positioning. This gives you the ability to tell them to move, which will avoid wipes."* [P2, F, 19]

Additionally, raid awareness allows players to make appropriate actions by observing the behaviour of others (cf. [13]). This was elaborated by the same participant as above:

> *"Being aware of what everyone else is doing allows me to see when I need to use my own abilities, whether defensive or offensive, or to swap to a different position. If I see the rest of my raid move to the other side of the room, I need to move too. But if I wasn't raid aware I wouldn't be paying attention to them, and thus I'd end up probably dying from not moving."* [P2, F, 19]

Thus raid awareness is not only related to the interdependent division of labour among players but also enables players to adapt and react strategically to changes in the high-pressure raid environment.

How Meters Support Groups in Evaluating Contributions

Meters were found to support raid groups in making both real-time and retrospective inferences about contributions. Real-time assessments were required to assess contributions in the midst of raid tasks, whereas retrospective judgements were used to analyse group performance and to identify opportunities for improvement.

Setting and Monitoring Targets for Performance

One role of meters was to provide players with real-time information about whether certain performance standards are being met. This can be of critical importance during tasks where players must achieve, and subsequently maintain, a specific level of performance. Meters provide a resource for gauging whether one is meeting this threshold, and whether there are others who may need to increase their effort. This was important to avoid failure:

> *"I often look at my damage meter to make sure I'm achieving the right amount of DPS, this applies to every raider so that we can kill the boss quickly. If the right amount of DPS is not met, then the boss can enrage after a certain amount of time, causing a wipe."* [P2, 19, F]

The fact that meters permit social comparisons means that players are able to determine what constitutes a legitimate (and thus 'fair') contribution on-the-fly, as evidenced by the following excerpt:

"When I started I was poached from a much weaker raid group... I had to really knuckle down and become a lot better fast. Meters helped because I could see what others were casting, and gauge myself against them. If I was within the warlock group I was doing okay" [P5, 35, M]

Appraising Contributions and Identifying Underperformers

Further to this is the way meters allow groups to address concerns relevant to fairness in participation. Generally this means leveraging qualities of meters to take action against people who appear to be underperforming because of consistently low placement on the meter:

"If you have four people all doing roughly around the same damage, and then two straggling really badly at the bottom, you know it's to do with them, it's not to do with the raid or the boss." [P1, 23, F]

Participants described that teammates are sometimes 'benched' on the basis of meters, i.e. that underperformance gives cause for relegation to a substitute role. However, placement on the meters is not always enough to know *why* an individual is underperforming, and this is important because it reveals how meters are used to interrogate fairness. Player narratives reveal two uses. First, underperformance can be a sign of free-riding, a behaviour that is not uncommon in WoW raids [2] and is difficult to identify because avatars do not accurately convey their user's actual work [31]. Meters allow teams to spot this type of free-riding:

"Everyone must pull their weight to progress through raiding... all it takes is for one person to mess up and it's a wipe, so everyone must do the tactics and conform to their role... meters give raid leaders the information to see who is slacking or failing." [P4, 18, M]

This quote also highlights the use of meters to see who is 'failing'. That is, a player's placement near the bottom of the meter does not necessarily mean they are slacking or free-riding, but could instead be related to ability. This is interesting because it highlights that, while the reasons for underperformance are initially obscure, meters provide an important clue for structuring further inquiry as to the legitimacy of individual contributions. Interviewees described how meters help to identify opportunities for improvement, particularly for raid members in need of guidance. Such improvement can be assumed to support the long-term health of the group, with those towards the lower end of the rankings encouraged to make improvements so that their contributions come into line with other raid members:

"In the raid it is particularly so you can identify weaknesses. It might be that someone's having a real problem with what loot they've got, what armour... so the aim isn't to say, 'oh get out of here we'll get someone new', it's to identify weaknesses and learn from them so you can do it better next time." [P1, 23, F]

"In rare cases it's down to using the wrong rotation[1] *so we'd try to correct that and get them to look further into their class and learn the right rotation."* [P3, 46, F]

Interestingly, presence on the damage meter can be a sign that an individual is *not* making the right contribution. For example, many raid tasks call for players to control roaming enemies using traps or immobilising spells. This entails a drop in damage contribution because the individual in question must divert his or her offensive capabilities away from the primary target. Such 'crowd control' tasks tend to be assigned to specific persons before the start of the raid. If the person in charge of such a task remains present on the meters while additional enemies are around, raid members will be aware that the person in question is not contributing properly:

"In this case if you weren't doing your job then you were high on the meter. It's how we knew the hunter wasn't doing his job, when he started to shoot up the DPS meters because he was just standing still and firing on the boss, instead of killing the [roaming enemy]*."* [P9, 21, M]

In this way, meters allow players to be held accountable for both appropriate *and* inappropriate actions. However, and as will be seen later in our results, these judgements can prove faulty when players rely too heavily on the content of meters and do not take contextual issues into account.

Enabling Equality in Contributions

Responses suggest that a certain degree of equality in contributions may be desirable in raid groups. In line with our earlier contentions, the position of each player's coloured contribution bar within the meter makes assessment of equality relatively straightforward:

"You could see that by a bar on the DPS meter, as long as bars are all aligned then everyone's doing roughly the same, of course there's variations but I don't really mind... it shows everyone's putting in a fair amount of effort." [P1, 23, F]

Equality in contributions appears to be relevant for two reasons. The first is very much related to fairness and the expectation that all members contribute: if contributions appear similar on the meters then group leaders need not devote attention to underperforming individuals. As described above, deviations from expected standards can give cause to raise questions about performance. However, a second reason pertains to the overall risks of raiding, in that it is better to have an equitable distribution of effort in favour of a 'spikier' or more polarised output in which a few individuals contribute significantly more than the rest of the group. This latter scenario is much riskier in a WoW raid setting; if the top contributors disconnect or are removed from the encounter, e.g. due to the death of their character, the raid is at a higher risk of failure due to a larger overall percentage loss of damage output. Concerns such as these may explain why raid groups spend time training new or underperforming players, as it is in the collective interest for contributions to be similar and of a certain (usually high) standard. It seems to us that meters provide raid groups with the information required to assess both of these concerns.

[1] A player's 'rotation' refers to the order in which they activate the offensive capabilities of their character. Certain rotations produce better damage output than others, making rotation selection an important aspect of raid behaviour.

Permitting Analyses After the Fact

An interesting aspect of meter use is that observations based on their data are not confined to in-the-moment monitoring during raids. In fact, some players suggested that, for them, the main benefit of meters lies in their use after the fact. That is, the data that meters gather is saved and used post-hoc to perform deeper interrogations of contribution. According to our interviewees, it has become common for persistent groups to pass the contents of meters to external services such as *World of Logs* for more detailed analyses of player contributions. Four players who raided regularly with guildmates described how they used this information:

> *"I don't know if you've heard of World of Logs... it's become a big thing where you can actually go in and you can see what abilities people were using, how much damage those abilities were doing to the boss, if you're following mechanics of the fight... so they could look at the details and say, "right, you're using this spell so many times, you should've been able to use it maybe ten per cent more". And basically help you out saying, "right, you need to keep an eye on this more, or you need to be using this spell more to keep that ability up", and things like that."* [P6, 34, M]

This practice augments meter use by identifying opportunities for *collective* improvement:

> *"Meters in game are based off combat logs and there's an analysis of the combat logs that can happen offline as well... I mean a meter will tell you, you know, the percentage of things they're doing, how much damage you're doing from all the different spells, who the top three people are, so are you hitting the boss, are you hitting the adds. It'll show you all of that stuff, and you can go back over a fight and look at it. But, World of Logs allows you to look at much greater detail about how all twenty five people are doing and how they're working together. Which is much more difficult to do with things like Recount. So yeah, I mean they wouldn't so much look at the meters, they'd look at different analysis of the data the meters use, yes."* [P7, 39, M]

We reason that this analysis factors into progression, which represents one of the main goals for many persistent raiding groups [2]. Progression refers to the gradual completion of increasingly difficult game content over time; groups that 'progress' are considered to be successful because they are improving in skill, obtaining better rewards, and are able to take on tougher challenges within the context of the game. While meters have some merit for real-time assessment of fairness, an additional benefit independent of their visual design relates to *retrospective* analyses done to gather additional support for claims about the efficacy of player contributions. This information may then be used to ensure players contribute properly in the future.

Supporting the Assignment of Rewards

In describing WoW teams we noted that not all players are guaranteed to be rewarded for their efforts [2], raising the issue of distributive fairness as a subject of concern [26]. Two interviewees mentioned that their groups assigned rewards by giving players 'points' after successful raid encounters. These points can be used to 'bid' on loot, with rewards assigned to the highest bidder [26]. The content of meters did not usually factor into this process, but one participant did state that his guild used meters to resolve tied bids:

> *"If there were two people who bidded the same amount, then it'd be decided by the raid leader who would use their information through the meters."* [P4, 18, M]

An alternate approach was to make an attempt at balancing the distribution of rewards with overall benefits to the raid group. For instance, a particular item could be a minor upgrade for a regular player but could be a major upgrade for a weaker member of the team. Meters helped to ensure such decisions were taken fairly:

> *"In one of my guilds we used what we call a 'loot council'... officers in the guild take into account people's current gear and their damage on the meter. If it's a big upgrade for the new player they will award it to them."* [P2, 19, F]

It is, therefore, not the case that being 'top of the meters' equates to a greater chance of reward. One participant recalled the problems that arose through this approach:

> *"A lot of arguments were started over meters, over who could come to raids, who got what equipment. Oh, "person x didnt deserve item y because they did badly"... eventually though we moved to a system of item rewards that gave items to people who would get the best use out of it, rather than person who did most DPS got the best weapon to do more DPS"* [P5, 35, M]

Here we see that meters can provoke disharmony through judgements about what appears to be 'fair' on the basis of a mere ranking. Furthermore, the extent to which meters should be used to assign rewards was called into deeper question by two participants who recognised that people can toy with meter content to achieve contrived outcomes:

> *"Loot off of who's top of the meters is a waste of time because that person's obviously geared to hell and doesn't need the loot. If it should go to anyone, it should be going to the people at the bottom of the list. But there again, you can cheese that by not doing any DPS. Saying, oh yeah I need loot because I'm bottom of the DPS charts."* [P6, 34, M]

This participant's statement raises questions about the reliability of meters for ensuring distributional fairness. Initially, ranking on the meters appears sufficient to demonstrate legitimate claims, yet the participant later asserts that people can 'game' the meters by deliberately withholding contributions. This hints at the potential for meters to be misused, raising questions about their suitability as resources for assessing contributions and, by extension, for making fairness judgements. We now turn to findings that relate to these problems more directly, considering first the specific limitations of meters for assessing contributions and then broader consequences that arise as a result of making contributions available for scrutiny.

The Limitations of Meters for Assessing Contributions

Participants' stories speak to several limitations of meters as measures of collaborative contribution, focusing in particular on the risks of using meters for real-time assessment in the midst of raiding. Participants were especially concerned about the limited scope of meters, in that they do not adequately represent peripheral and supporting contributions that are not related to damage but are nonetheless critical to the group's success. For example, during raids, players sometimes disappear from the meters when performing crowd control (or 'add work') on enemies. Statements by four participants indicate that the failure of meters to represent this work up-front can make it seem as if certain individuals are doing nothing when, in reality, their contribution is critical:

> *"My role was, as a hunter, I was used for a lot of the add work. Misdirecting, which means that you're switching off the main target to control something else and misdirect it... it brings your DPS right down, so there's no way you're going to be top of the meters."* [P3, 46, F]

> *"Some fights you did zero DPS but spend the entire fight banishing something all fight, now folk knew this but if they didn't understand the fight they would think you were not doing what you're supposed to."* [P5, 35, M]

Second, the way in which meters present aggregations of actions can potentially *misrepresent* the value of contributions because information about the *quality* and *relevance* of actions is not immediately available. For example, a person may be performing actions that are captured by the meter but are ineffective for the task at hand. A metered representation does not immediately allow players to detect such behaviour because it provides a collapsed view of performance without accounting for contextual details. One participant described this phenomenon in relation to 'area of effect' (AOE) attacks in WoW. These are abilities that can inflict significant damage to multiple targets at once. This type of action is, however, not always appropriate, and is difficult to detect in a meter:

> *"So you can do a lot of AOE, loads and loads of AOE, and look like they're doing a lot of damage. But actually maybe not be attacking the boss. Now, that kind of DPS padding makes people feel good. And you can't tell that easily."* [P7, 39, M]

Furthermore, meters can cause certain contributions to be *misattributed*. This occurs when one group member's actions have knock-on effects that benefit the group as a whole, but the role of the original contribution is not recognised in meters. For instance, some characters in WoW can bestow others with 'buffs' that boost other players' abilities. However, the provider of the buff receives no credit for this contribution in the damage meter, whereas the output of everyone else is inflated by the strength increase. While it is important to recognise that groups in WoW are typically aware of these peripheral efforts, taking meters out of context could make it appear as though one person has contributed more to the group when, in reality, their contribution was absolutely dependent on the obscured involvement of others. This calls into question the *accuracy* of meters for reflecting contributions.

Lastly, the value of damage meters as measures of contribution can vary according to contextual features of specific tasks. In WoW, this relates to the fact that different characters are more or less effective in different settings. For example, there are some enemies that are more susceptible to ranged damage, and thus players with relevant abilities are at an advantage during these situations. In other cases, these players cannot fully contribute because they must continually move their character and thus cannot participate to the best of their ability:

> *"If you had to move a lot, casters would generally be lower down because they're moving so much whereas if you're tanking or if you're not ranged then it's easier, so there're differences depending on what the fight is."* [P4, 18, M]

The point here is that, as simple aggregations of actions, the contents of meters do not account for contextual variance in task demands. The fact that meters fail to portray individual disadvantages may therefore make it seem as though some players are underperforming when they are in fact making their best effort. One interviewee gave an example where a group leader had berated her for seemingly poor performance. In this case, there was a lack of common ground regarding her character's abilities and the risk of being 'spiked' (damaged) by a raid boss:

> *"The first boss in Icecrown, he did an attack that puts you in a spike if you're too far away... so everybody has to run into the boss, apart from hunters because you've got an area that's a deadzone where we can't attack, so we have to stand outside of the deadzone, but that means we get spiked... And we had one of our officers get rather antsy with me a couple of times for getting spiked and not DPSing... it took our class leader to pull him and say, look, learn the class, you have no idea what you're talking about."* [P3, 46, F]

Taken together, these issues mean that it can sometimes be difficult to get a good sense of individual performance based on meters alone. This is primarily due to their unidimensional scope in terms of measuring overall contribution. Moreover, efforts can be undermined by temporary aspects of a work situation that are not matched to a player's abilities. These issues raise questions about the efficacy of meters for supporting judgements of fairness.

Additional Consequences of Publicising Contributions

Further to these concerns about the design of meters and players' interpretation of contributions, responses indicate that the general phenomenon of publicising contributions has a number of corollaries that impact harmonious collaboration.

Meters can Promote Competition over Coordination

Since damage meters rank players hierarchically in accordance with overall contribution, players have come to see topping this ranking as an accolade carrying a certain degree of prestige [37]. Topping the meter implies that the individual in question has not only made the greatest contribution but is also the most skilled. In some cases this can be positive, with one interviewee describing how being 'top of the meters' became a motivating target for her group:

"Someone's belting away a load of damage and there might be some friendly egging on in raid chat, like, oh look at me, see who can beat me. And then everyone else tries harder to beat that." [P1, 23, F]

However, this type of competition can be problematic when individuals become overly focused on topping the meters. As mentioned by three interviewees, this can damage team effectiveness by causing individuals to neglect basic coordination:

"Sometimes they'll go away with blaring DPS, doing as much as they can to show off, and come top of the DPS meter on everyone's screen. But they don't care about the tactics for the boss, or care what their specific role should be, they just want to show off and be top of the DPS meter. Which should be a secondary goal, not a main goal." [P1, 23, F]

This type of "over-nuking" behaviour was also mentioned by Taylor in her early discussion of meters [37]. The main issue at hand is that meters can shift the motivational focus of individuals away from harmonious work towards intra-group conflict. The resulting tensions can be undesirable or even dangerous for the group. In our analysis, we see an additional corollary that we believe arises as a result of focusing too intensely on competition; that is, meters become the yardstick by which other players are assessed, irrespective of whether these assessments are appropriate. Four people stated that this leads to bullying or other abusive behaviours:

"The dps meters starting draining the fun out of things when used as batons to beat people with about performance." [P5, 35, M]

"People do use it as a kind of bullying method. If they can see what everyone's doing. If there's anyone that's slightly not as high as everyone else, they'd be doing less damage usually, so they'd be the main targets for, "oh my God we've just wiped and it's your fault because you're not doing your job". So it does draw attention to the more vulnerable players who are trying but just aren't there yet." [P1, 23, F]

Meters can Amplify Evaluation Apprehension

Evaluation apprehension refers to the fear that contributions will be viewed negatively by other members of a group, causing individuals to withhold their efforts to avoid criticism [10]. With regard to meters, it is possible that the broadcasting of contributions, which is then held as a measure of 'performance', may cause individuals to become nervous about the way in which their efforts are interpreted by others. Such an experience was recounted by one interviewee who, as a new raid member, had been worried about performing poorly and being evaluated in a negative light. (Raids often require new members to undergo a trial period where performance is analysed and a decision is made on whether the individual can stay.) Her response hinted at an almost self-fulfilling prophecy, where nerves about evaluation caused her performance to deteriorate below her usual standard:

"It made me feel anxious... like I'm being scrutinised. I was worried about how people would judge my actions, and that made me worse in some ways." [P1, 23, F]

Evaluation apprehension is likely true of group participation irrespective of the work context, yet meters may amplify the effect by publicising in detail a limited perspective on an individual's performance. Any negative behavioural outcomes may therefore be especially problematic in group settings where each individual must meet an acceptable level of performance, as was the case in WoW raid groups.

DISCUSSION

Our primary motivation for the present work was to explore how fairness judgements might be enabled through visualization of contributions to collaborative work. Taking WoW meters as an example of how contributions might be visualized, we sought to understand how these meters support group work, as well as the extent to which they permit inferences about fairness. We found that meters allow individuals to reflect on their own performance, to identify underperformers, and to monitor the status of collective efforts. Meters are also used to perform retrospective evaluations of teammates in order to inform future collaborative episodes. However, player narratives speak to several shortcomings of meters as reflections of collaborative work. In particular, meters do not provide immediate evidence of important peripheral contributions, display little information about the quality of work, and do not account for contextual variation between tasks. Our study also found that the public nature of metering can encourage unhealthy competition and induce apprehension. Here we reach for broader lessons based on these findings.

Meters as Resources for Fairness Judgements

We have shown that meters provide collaborators with an initial basis for asking questions about fairness in their group. That is, meters offer cues that provoke deeper investigation by prompting people to check up on the actual behaviour of their teammates, such as when checking underperformance. Without meters, these coarse preliminary judgements would be difficult or need explicit clarification, which may be problematic for team members during tightly coupled, time-pressured scenarios like WoW raids [34]. Meters give groups a complementary source of information about each person's contribution to the group effort, beyond that which is possible through simple observation of an avatar's animations or physical positioning. This emphasises that groups can benefit from an additional source of information about contributions that would ordinarily be unavailable, an observation that aligns with previous research on the value of visualizations for supporting group processes (e.g. [1, 4, 11, 12, 29, 36]).

However, unlike previous research, our study of real-world teams suggests that awareness visualizations can be problematic when they are used as a resource for gauging contributions in general. We have demonstrated not only that this is true for fairness in particular, but also how the dynamics of group interactions influence the interpretation of meters, i.e. that people incorrectly attribute either good or bad behaviour to the meter scores. A prominent finding was that a meter's basic representation can arouse discontent through quantifying only a single type of behaviour and by failing to account for the

full range of contributions that participants can bring to the collective good. Meters were found to be of dubious reliability, with players stating that they can easily be manipulated or 'gamed', and offer little in terms of understanding the quality of an individual's work or how contextual details may have impacted performance. These findings may not be surprising given that meters are obviously limited, but are nonetheless valuable because they provide an understanding of the problems that arise when people try to make complete assessments of contribution based on limited information. This in turn casts doubt over the value of any attempt to distil the value of contributions into a unidimensional representation, either for in-the-moment judgements of fairness or for retrospective review of performance.

With this in mind, it is important to recognise that player judgements of fairness draw on—but are not uniquely determined by—transient awareness information, such as that provided by meters. In WoW, judgements are situated with respect to players' understanding of the game world, the contextual details of a task, prior experiences with collaborators, and knowledge of the demands placed on people by the particular work they are doing. While some negative consequences may arise through the limited scope of meters, we note that the extent to which these impact the actual work of a group depends on a team's ability to utilise the awareness information effectively. As described by our seventh participant, meters require their readers *"to be able to interpret what they're saying properly, and understand what they're saying properly. And as with all of these things, there are people that don't know how to do that"*.

Our study demonstrates that, while meters do not show everything relevant to a task, problems arise because people sometimes struggle to apply their contextual understanding in the face of the 'objective' data furnished by the meters. This may be because people either fail to recognise and understand a meter's limitations, or because they are aware of its limitations but do not account for them in practice. Failure to act in a way that is consonant with a meter's limitations means that, rather than delaying interrogations of undercontribution and its legitimacy, users may berate colleagues for underperformance in the here-and-now. This is a problem that is not unique to WoW—it is a danger for any context and may cause effects similar to those we have described here to arise if a meter-like representation were used elsewhere. Furthermore, our findings point towards a more general issue in that over-reliance on a representation like a meter may come to define the very way in which contributions are assessed. That is, visualizations like meters will inevitably undermine valuable contributions because they encourage judgements whose territory is scoped by the content of the visualization. People can 'repair' this problem by looking beyond the meters when assessing a person's worth, but problems arise because people sometimes fail to do this in practice, and may find it difficult because the other inputs people have given to the team are not recorded.

Lessons for Supporting Fairness Judgements

We believe these considerations point towards the following lessons for the design of visualizations that are intended to support fairness judgements in a collaborative system. Most prior work on visualizations has considered a single type of contribution in the context of relatively constrained tasks, e.g. contributions to group discussion [4, 11, 12, 27, 28], but our study focused on a context in which a variety of contributions are necessary for group success. Our findings demonstrate that, while visualizations can support some basic inferences about the legitimacy of contributions, it is misguided to assume that a unidimensional measure of performance (such as a meter) will guide full judgements of fairness where multidimensional teams are concerned. In situations where multiple types of contribution are relevant to a group's work, designers should avoid trying to support fairness judgements with visualizations that only account for a limited range of contributions.

A different strategy could be to provide multiple awareness mechanisms that show different types of contributions at the same time. In collaborative writing, for example, a visualization could show things like words-per-minute, time on task, and number of edits as separate measures. This might then provide better support for fairness judgements by giving a broader lens on an individual's net contribution to a team, perhaps allowing people to examine the way in which individuals have traded off different subtasks—low performance on one metric might be offset by high performance on another. Such a measure could be used both in real-time and retrospectively.

Alternatively, it could be possible to aggregate several contributions into something like a 'multidimensional meter' to give a better reflection of a person's inputs. However, this raises an additional challenge related to how a surface representation might combine subtasks that may not be equivalent. In something like software development, projects may involve writing, computer programming, and software design, as well as intangible contributions associated with project management. Comparison of these subcomponents in a meter may lack meaning because it is hard to gauge what each is worth relative to the next. One way to solve this might be to show a proportional rather than aggregate representation. This is something that can only be investigated with further work.

In describing meters, we identified two parameters relevant to fairness: support for social comparisons, and detail about contributions. Our study has drawn attention to additional distinctions that have not been identified in previous work on visualizations of group activity, but which are important for assessing fairness. First, we have seen how a visualization can be used for personal reflection versus public evaluation. Second, our findings draw attention to the use of meters to enable in-the-moment appraisal of individual and team performance, versus retrospective dissection of the work of the group as it unfolded. These two contrasts are orthogonal, and so represent four quadrants of a design space for thinking about visualizations that can support knowledge of contributions and, by extension, judgements of fairness.

Each of these quadrants raises its own challenges, but it might be possible to support them simultaneously. Since fairness can be evaluated in real-time and retrospectively, future mechanisms could attempt to support tripartite considerations of fairness by drawing on information collected before, during, and after an episode of collaborative work. This would help

to ensure that differential contributions are taken into account when assessing the performance of individuals. One way to do this might be to provide multiple perspectives on the same data, i.e. different sources of information about the same event. In WoW, for example, it is possible to look at a damage meter to assess contribution, but it might also be possible to look at a video replay of what players were doing on the raid in order to augment one's appreciation of a player's efforts.

At a minimum, designers could tailor their representations to match the ability of collaborators to attend to the information, as well as the opportunity that they each have to discuss it. Using WoW as the example to hand, it is difficult to discuss contributions during a raid. Although meters permit glanceability and lightweight social comparisons, discontent can arise because people make inferences about fairness based on information that was not designed as such. Thus much analysis of player behaviour occurs after the fact, whether that be through analysis of non-damage contributions or in deeper information processing through World of Logs. The discussion that occurs around this data is important because individuals acquire an opportunity to explain aspects of their behaviour that could not be measured by an awareness mechanism.

There is, however, a need to be realistic about the extent to which designers can in fact support fairness in collaborative systems, and whether they should at all. This is because there are situations in which it is unlikely that a visualization will ever be able to capture all of the meaningful inputs that people give to a group. We have seen that people can 'repair' deficiencies in a unidimensional visualization by invoking contextual knowledge to answer questions about fairness. Yet it is unsafe to rely on this as a compensatory phenomenon since people can be prone to snap judgements, and may be unaware of mitigating circumstances that affect a person's ability to contribute, meaning that they do not have this knowledge to apply when appraising the efforts of others.

What this suggests is that people could be reminded about the limitations of a visualization in advance, especially when it is not possible to represent everything an individual has contributed to a team. Based on the results of this study, designers might actually wish to discourage people from making inferences about fairness in a situation where it is only possible to provide a unidimensional representation. This suggestion may seem contrary to the aim of this paper, but is a legitimate consideration given the issues we have raised about meters. It may be better to provide visualizations only in situations in which contributions are tightly constrained, easily quantifiable, or remain closed to subjective interpretation.

Generalisation to Other Work Settings

There is an open question as to whether our findings would apply to other collaborative work situations. We do not believe that there is anything unique about WoW that would cause our findings to be applicable only to this context. Instead, we believe these issues may emerge in situations where similar enabling conditions are present. In WoW, collaboration is characterised by the following properties:

- Teams have a clear objective.
- Success requires a combination of different skills.
- The application of skill requires coordination in time and space.
- Team members must pay attention to the state of one another, as well as to themselves.
- The reward structure is clear.
- Reputation is formed over successive episodes of similar work and has a material impact on the inclusion of an individual in future team activities.
- Certain aspects of each person's work can be quantified.

These are things that are true of WoW, and so we have reason to believe that in other collaborative situations that share these characteristics, one might expect to find similar challenges to those we have identified.

There are, however, some particular features of WoW that shape the extent to which it is necessary to question another person's contribution, and thus also relate to the extent to which a visualization for judging fairness is even required. The first of these is that in WoW raids there is some benefit to doing as much as work possible—most of the time, the more damage players do, the faster the task is done. This is not true of situations in which the evaluation of work is more subjective or that do not involve a clear stopping point. Second, WoW raids are characterised by a need to interrogate work in-the-moment because there is an immediate risk of failure (i.e. a 'raid wipe'). Many other collaborative tasks do not involve such risks, but in those that do (perhaps something like mission-critical teamwork) having a visualization might prove helpful to ensure that contributions are made correctly.

Finally, WoW raids involve a wide range of tasks, but some collaborative situations are characterised by a single type of contribution, e.g. providing ideas to a brainstorm. The task one seeks to support is likely to define the extent to which a visualization is necessary and helpful, and whether negative side effects might occur. One of the key parameters we identified in meters is that they provide a certain level of detail about contributions. However, since meters were found to encourage evaluation apprehension, some situations might be best served by occluding detail about contributions. One example is the case of brainstorming, where people are supposed to contribute ideas without any fear of how those ideas will be perceived. Evaluation apprehension is known to be problematic in this context [10], so a visualization here could simply show the number of ideas a person has contributed, without displaying what those ideas are.

Limitations and Future Work

While we characterise groups in WoW as work teams, we recognise that people in WoW enter a special environment that is separated from their real life. In real world work groups, consequences of group interaction can bleed over into other aspects of life, whereas in WoW this is not the case. That said, reputation is important in WoW and so there is still a need to be mindful of how a player's wider online identity can be affected by the consequences of game group activity.

One issue that we were unable to investigate in detail concerns the variable importance of damage meters between different types of raid group. In temporary, one-off raids (known as 'pick up groups' or PUGs) players may be more prone to rely on metered information than in guild settings where trust among teammates is likely to have been established well in advance. Although our informants had participated in both types of raid group, most of their interview comments were made without respect to particular raid settings, making it difficult for us to tease apart the two scenarios. Future studies might seek to investigate how the dynamics of more or less casual work impact the use of meters for assessment of contributions.

Second, it is not clear whether some of the side effects that we identified would also arise in other collaborative scenarios. While we do not doubt that findings from WoW are useful when considering other collaborative work scenarios, it may be the case that there are qualities of gaming environments that encourage certain behaviours over others—the issue of promoting competition may be especially relevant to contexts where play is prevalent. This is an issue that can only be explored through further work.

Finally, our study had relatively few participants, and although we are confident that our data was sufficient to support a coherent set of claims about meters, future work should look to verify our findings. Since the time at which our data was collected, an additional expansion has been released for WoW. This change to the landscape of the gaming environment may have introduced additional issues for raid groups and the use of meters that go beyond those captured in our analysis.

CONCLUSION

The purpose of this study was to explore how it is that a collaborative system might support judgements of fairness. Using meters as a case in point, we saw that a visualization can enable basic inferences about contributions and fairness, but the incomplete nature of the visualization can lead to other problems that impact harmonious collaboration. Independent of our interest in fairness, this study supports earlier suppositions from the literature on group mirrors. Leshed et al. [29] proposed that a visualization might encourage competition when it allows someone to be identified as a 'top contributor', and that there is potential for visualizations to be manipulated in pursuit of desired outcomes. Our study confirms the presence of these behaviours in the use of meters. Furthermore, this study contributes to the literature on visualization of teamwork by providing an extended picture of how an awareness mechanism is used to infer contributions in a non-laboratory setting. We have drawn attention to the consequences that arise when a limited representation is used in the context of teams and tasks where multiple contributions are required. This work confirms Taylor's intuitions about the way in which players interpret meters [37] while providing a more complete description of the issues that cloud the use of these artefacts.

Our investigation of meters indicates that, while it may be possible to design to permit superficial judgements about contributions, supporting fairness is clearly a complex and multi-dimensional design problem. Any designer seeking to enable considerations about fairness will need to be mindful of the potential consequences that may arise as a result of providing awareness information about the work of others. However, it remains to be seen whether the sorts of issues identified in our study might arise when contribution awareness is implemented in other collaborative systems. What is required are further explorations of the design space and its four quadrants, separating visualizations for in-the-moment appraisals from those intended for review after-the-fact, as well as those for personal reflection from those available to public scrutiny.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers and associate chairs for their extremely helpful comments on this work.

REFERENCES

1. Khaled Bachour, Frederic Kaplan, and Pierre Dillenbourg. 2010. An interactive table for supporting participation balance in face-to-face collaborative learning. *IEEE Transactions on Learning Technologies* 3, 3 (2010), 203–213.
2. Jeffrey Bardzell, Jeffrey Nichols, Tyler Pace, and Shaowen Bardzell. 2012. Come meet me at Ulduar: progression raiding in world of warcraft. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 603–612. `http://doi.acm.org/10.1145/2145204.2145296`
3. Shaowen Bardzell, Jeffrey Bardzell, Tyler Pace, and Kayce Reed. 2008. Blissfully productive: grouping and cooperation in world of warcraft instance runs. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*. ACM, New York, NY, USA, 357–360. `http://doi.acm.org/10.1145/1460563.1460621`
4. Tony Bergstrom and Karrie Karahalios. 2007. Seeing more: visualizing audio cues. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part II*. Springer-Verlag, 29–42.
5. Jeremy Birnholtz and Steven Ibara. 2012. Tracking changes in collaborative writing: edits, visibility and group maintenance. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 809–818. `http://doi.acm.org/10.1145/2145204.2145325`
6. Charles M Brooks and Janice L Ammons. 2003. Free riding in group projects and the effects of timing, frequency, and specificity of criteria in peer assessments. *Journal of Education for Business* 78 (2003), 268–272.
7. Barry Brown and Marek Bell. 2004. CSCW at play: 'There' as a collaborative virtual environment. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)*. ACM, New York, NY, USA, 350–359. `http://doi.acm.org/10.1145/1031607.1031666`
8. Mark Chen. 2008. Communication, Coordination, and Camaraderie in World of Warcraft. *Games and Culture* 4, 1 (2008), 47–73.

9. Morten Deutsch. 1975. Equity, equality and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues* 31 (1975), 137–149.
10. Michael Diehl and Wolfgang Stroebe. 1987. Productivity loss in brainstorming groups: toward the solution of a riddle. *Journal of Personality and Social Psychology* 53, 3 (1987), 497–509.
11. Joan Morris DiMicco, Katherine J. Hollenbach, Anna Pandolfo, and Walter Bender. 2007. The impact of increased awareness while face-to-face. *Human–Computer Interaction* 22, 1–2 (2007), 47–96.
12. Joan Morris DiMicco, Anna Pandolfo, and Walter Bender. 2004. Influencing group participation with a shared display. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)*. ACM, New York, NY, USA, 614–623. **http://doi.acm.org/10.1145/1031607.1031713**
13. Paul Dourish and Victoria Bellotti. 1992. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM Conference on Computer Supported Cooperative Work (CSCW '92)*. ACM, New York, NY, USA, 107–114. **http://doi.acm.org/10.1145/143457.143468**
14. Nicolas Ducheneaut, Nicholas Yee, Eric Nickell, and Robert J. Moore. 2006. "Alone together?": exploring the social dynamics of massively multiplayer online games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 407–416. **http://doi.acm.org/10.1145/1124772.1124834**
15. Thomas Erickson and Wendy A. Kellogg. 2000. Social translucence: an approach to designing systems that support social processes. *ACM Trans. Comput.-Hum. Interact.* 7, 1 (March 2000), 59–83. **http://doi.acm.org/10.1145/344949.345004**
16. Leon Festinger. 1954. A theory of social comparison processes. *Human Relations* 7, 2 (1954), 117–140.
17. Robert Folger and Russell Cropanzano. 2001. Fairness theory: Justice as accountability. In *Advances in Organizational Justice*, J. S. Greenberg and R. Cropanzano (Eds.). Stanford University Press, Stanford, California, 1–55.
18. Jolene Galegher and Robert Kraut. 1990. CMC for intellectual teamwork: a field experiment in group writing. In *Proceedings of the 1990 ACM Conference on Computer Supported Cooperative Work (CSCW '90)*. ACM, New York, NY, USA, 65–78. **http://doi.acm.org/10.1145/99332.99343**
19. Carl Gutwin and Saul Greenberg. 2002. A descriptive framework of workspace awareness for real-time groupware. *Journal of CSCW* 11, 3 (2002), 411–446.
20. Carl Gutwin, Saul Greenberg, and Mark Roseman. 1996. Workspace awareness support with radar views. In *Conference Companion on Human Factors in Computing Systems (CHI '96)*. ACM, New York, NY, USA, 210–211. **http://doi.acm.org/10.1145/257089.257286**
21. Carl Gutwin, Reagan Penner, and Kevin Schneider. 2004. Group awareness in distributed software development. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)*. ACM, New York, NY, USA, 72–81. **http://doi.acm.org/10.1145/1031607.1031621**
22. Carl Gutwin, Mark Roseman, and Saul Greenberg. 1996. A usability study of awareness widgets in a shared workspace groupware system. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work (CSCW '96)*. ACM, New York, NY, USA, 258–267. **http://doi.acm.org/10.1145/240080.240298**
23. Guido Hertel, Henk Aarts, and Marcel Zeelenberg. 2002. What do you think is 'fair'? Effects of ingroup norms and outcome control on fairness judgments. *European Journal of Social Psychology* 32 (2002), 327–341.
24. Guido Hertel, Carsten Deter, and Udo Konradt. 2003. Motivation gains in computer-supported groups. *Journal of Applied Social Psychology* 33, 10 (2003), 2080–2105.
25. Scott E. Hudson and Ian Smith. 1996. Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work (CSCW '96)*. ACM, New York, NY, USA, 248–257. **http://doi.acm.org/10.1145/240080.240295**
26. Chyng-Yang Jang. 2007. *Managing Fairness: Reward Distribution in a Self-organized Online Game Player Community*. Springer.
27. Jeroen Janssen, Gijsbert Erkens, Gellof Kanselaar, and Jos Jaspers. 2007. Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers & Education* 49, 4 (2007), 1037–1065.
28. Gilly Leshed, Jeffrey T. Hancock, Dan Cosley, Poppy L. McLeod, and Geri Gay. 2007. Feedback for guiding reflection on teamwork practices. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work (GROUP '07)*. ACM, New York, NY, USA, 217–220. **http://doi.acm.org/10.1145/1316624.1316655**
29. Gilly Leshed, Diego Perez, Jeffrey T. Hancock, Dan Cosley, Jeremy Birnholtz, Soyoung Lee, Poppy L. McLeod, and Geri Gay. 2009. Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 537–546. **http://doi.acm.org/10.1145/1518701.1518784**
30. Gerald S. Leventhal. 1980. What should be done with equity theory? In *Social exchange*, K. J. Gergen, M. S. Greenberg, and R. W. Willis (Eds.). Springer, 27–55.
31. Robert J. Moore, Nicolas Ducheneaut, and Eric Nickell. 2007a. Doing virtually nothing: Awareness and accountability in massively multiplayer online worlds. *Journal of CSCW* 16 (2007), 265–305.

32. Robert J. Moore, E. Cabell Hankinson Gathman, Nicolas Ducheneaut, and Eric Nickell. 2007b. Coordinating joint activity in avatar-mediated interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 21–30. **http://doi.acm.org/10.1145/1240624.1240628**
33. Bonnie Nardi and Justin Harris. 2006. Strangers and friends: collaborative play in world of warcraft. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW '06)*. ACM, New York, NY, USA, 149–158. **http://doi.acm.org/10.1145/1180875.1180898**
34. Lindsay Reynolds, Jeremy Birnholtz, and Allen Lee. 2012. The Effect of Communication Channel and Visual Awareness Display on Coordination in Online Tasks. In *Proceedings of the 2012 iConference (iConference '12)*. ACM Press, New York, NY, USA, 120–128. **http://doi.acm.org/10.1145/2132176.2132192**
35. Anselm Strauss and Juliet M. Corbin. 1998. *Basics of Qualitative Research*. Sage.
36. Sara Streng, Karsten Stegmann, Heinrich Hussmann, and Frank Fischer. 2009. Metaphor or Diagram?: Comparing Different Representations for Group Mirrors. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7 (OZCHI '09)*. ACM, New York, NY, USA, 249–256. **http://doi.acm.org/10.1145/1738826.1738866**
37. T. L. Taylor. 2006. Does WoW change everything? How a PvP Server, Multinational Player Base, and Surveillance Mod Scene Caused Me Pause. *Games and Culture* 1, 4 (2006), 318–337.
38. Tom R Tyler. 2006. Psychological perspectives on legitimacy and legitimation. *Annual Review of Psychology* 57 (2006), 375–400.
39. Fernanda B. Viégas, Martin Wattenberg, and Kushal Dave. 2004. Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM, New York, NY, USA, 575–582. **http://doi.acm.org/10.1145/985692.985765**
40. Dakuo Wang, Judith S. Olson, Jingwen Zhang, Trung Nguyen, and Gary M. Olson. 2015. DocuViz: Visualizing Collaborative Writing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1865–1874. **http://doi.acm.org/10.1145/2702123.2702517**
41. Dmitri Williams, Nicolas Ducheneaut, Li Xiong, Yuanyuan Zhang, Nick Yee, and Eric Nickell. 2006. From tree house to barracks: The social life of guilds in world of warcraft. *Games and Culture* 1 (2006), 338–361.
42. Nick Yee. 2006. The labor of fun: How video games blur the boundaries of work and play. *Games and Culture* 1 (2006), 68–71.