Misrepresentation of Health Research in Exertion Games Literature

Joe Marshall
Mixed Reality Lab, School of Computer Science
University of Nottingham, UK
joe.marshall@nottingham.ac.uk

Conor Linehan
School of Applied Psychology
University College Cork,
Cork, Ireland
conor.linehan@ucc.ie

ABSTRACT
HCI often requires scholars to build upon research from fields outside their expertise, creating the risk that foundational work is misunderstood and misrepresented. The prevailing goal of “exergames” research towards ameliorating obesity appears to be built on just such a misunderstanding of health research. In this paper, we analyse all citations to a single influential study, which has been extensively cited to justify research on exergames. We categorise the 375 citations based on whether they represent the findings of that study accurately or inaccurately. Our findings suggest that 69% of exergames papers citing this study misrepresent the findings, demonstrating a systematic failure of scholarship in exergames research. We argue that exergaming research should cease focusing on games as treatment for obesity, and that HCI publications should demand more critical and scholarly engagement with research from outside HCI.

Author Keywords
Exertion; games; exertion games; health; obesity

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Exertion games, exercise games or exergames, are games in which the player must exercise in order to play the game [47]. Exertion games are often designed as motivational systems to encourage people to exercise, and thus to improve their health. One particular health benefit which is heavily promoted is the potential of exertion games to combat a growing “obesity epidemic” [60]. For example, exertion games are said to “control obesity” [4], be a method of “obesity prevention” [36] and to “prevent diseases related to childhood obesity” [34]. Sedentary video games are also often cited in the exergames literature as one of the causes of childhood obesity (see data below). However, there is surprisingly little evidence in the health literature that games either cause obesity or that exergames could function as a useful preventative measure. As exergames researchers ourselves, we believe it is worth reflecting on how and why these misconceptions have proliferated, since the focus on designing exergames primarily as health interventions for obesity constrains the types of design and evaluation carried out in exergames research.

While undertaking a review of evidence for exergaming, we noticed that one paper [65] from the discipline of child health research is heavily cited in exergames research in support of the link between video game play and childhood obesity. On close reading of that paper, we realized that the findings do not actually suggest a simple relationship, either causal or correlational, between game playing and obesity. In other words, this paper is being systematically misrepresented in exergames literature. We suggest that this mis-citation serves as an interesting case study of how research from outside a field can be misunderstood or misrepresented, and of the effect that this misrepresentation can have on research in that field. In the current paper, we set out to analyse how that one paper has been cited in exergames literature; whether citations have been accurate or inaccurate representations of the original research, whether there are differences in citation accuracy between exertion games and non-exertion games papers, and whether there are differences in accuracy between health and non-health papers.

The intention of the current paper is to critique the prevailing approach to the design of exertion games as obesity interventions, by demonstrating that there is little robust evidence to support that goal, and by highlighting problems with the interpretation of health research made by exertion games researchers. All datasets used for this analysis are included as supplementary materials. The contribution of the current paper is fourfold:

a) We present a case study that demonstrates that exergames literature commonly misrepresents the health research that it cites.
b) We critique the excessive and unjustified focus on obesity as an outcome in research on the design and evaluation of exergames. Following Marshall & Mueller’s argument in [41], we argue that this focus...
is likely to lead to poor game designs that people will not choose to play.

More generally, this paper demonstrates that in HCI research, and interdisciplinary research more generally, misunderstanding and misrepresentation can commonly occur when referring to literature outside the expertise of the authors, highlighting a problem that we must address as a field.

e) We make the argument that even when HCI work is based on aims which seem intuitively good for society (e.g. #CHI4GOOD, Games for Health), we should remain critical and consider whether such aims are soundly supported and realistic.

In the following sections of this paper, we first describe, and demonstrate the prevalence of, the argument that exertion games represent an effective treatment for obesity. We subsequently present a brief summary of the evidence, from both the medical and gaming literature, that suggests that games have not been demonstrated an effective treatment for obesity. Further to this, we show that exertion gaming research literature also argues that computer games cause obesity, an argument that also appears to lack evidence. In order to understand how these misconceptions persist in the exergames literature, we undertake an analysis of the citations made to a study on childhood obesity that is heavily cited in that literature. Our analysis demonstrates that exertion games research seriously misrepresents this prior research, with errors in 69% of citations, with several authors even stating that the cited study identified causal relationships, which as an observational study, it never could. We conclude by presenting a set of lessons both for exertion gaming, and for wider HCI as a field.

BACKGROUND

Recent years have seen growing interest from both commercial developers and researchers in developing digital games that require the undertaking of physical exercise as part of game play. Games such as Zombies, Run! [1] and Wii Fit [49] aim to combine game design and modern activity tracking technology (GPS, pedometry, MS Kinect) to create playful experiences around exercise. More recently, commentators have claimed that the popular game Pokemon Go successfully motivates players to exercise [32].

Many researchers have also developed games specifically intended as health interventions, to encourage reluctant populations to undertake exercise when they otherwise would not choose to do so. Promotion of physical exercise seems a worthwhile goal, and some projects have indeed demonstrated at least short term increases in the quantity and quality of exercise undertaken by game playing participants [38]. However, the relationship between game playing and measures of obesity is much more complex.

Obesity reduction as a specified goal of exergames research

It is important in the context of the current paper to first establish that exergames papers often mention obesity as a specific target outcome (either short- or long-term) of their work. In order to do so, we can provide some example papers for readers to consult. For example, exertion games are said to “control obesity” [4], to be a method of “obesity prevention” [36] and to “prevent diseases related to childhood obesity” [34]. However, in order to fully appreciate the prevalence of this argument in the literature, we felt it necessary to conduct a rough bibliometric analysis. The list of papers in this analysis are available in the supplementary materials provided with the paper.

We collected an initial dataset of 388 exertion games research papers by using IEEE Portal, ACM Library and Google Scholar searches for “exertion AND game”, “exergame”, “exercise AND game”. We then used Mendeley to search these for the word obesity. We found that 108 (27%) of these papers mentioned obesity, suggesting that over a quarter of all exertion games papers in this set suggest that they are in some way relevant to obesity. So, given the prevalence of this argument, we would expect to see strong evidence of games as useful obesity interventions.

Lack of evidence

Many short term studies have shown that playing exertion games involves a level of exertion above rest [66]. These studies often show that players of exergames do genuinely move while playing. For example, one recent study showed that if players in a lab setting played a specific fitness exertion game for 30 minutes (plus 15 minutes of warm up and cool down time on lab treadmills), they did indeed exercise for 30 minutes [5]. However, these studies offer no evidence as to whether exertion games can treat or prevent obesity, which is a long-term condition unlikely to be affected by once-off bouts of game playing.

Exertion games don’t increase overall activity long term

In longer term studies of exertion games, findings suggest that even when there is strong evidence that players played those games for significant amounts of time, there is not a consistent increase in all-day levels of exertion, or other health measures, in comparison to control [2,14]. This is thought to be in part due to compensation effects, where an increase in activity at one point in the day, or through one type of activity, is compensated by less activity over the remainder of the day. This is an effect seen also in relation to school physical activity [19]. Studies where players are freely allowed to choose when they play exergames also showed that players stopped playing very quickly. For example, one study showed an average of 3 minutes a day of game time after 6 weeks [54], a level unlikely to have impact upon health outcomes. This is in line with other studies of activity technology to aid weight loss which have yet to show effective long term results; one major study even showed that adding activity tracking devices to standard medical dietary and physical activity interventions had significantly less weight loss success compared to the standard intervention group; suggesting that poorly designed exercise technology has potential to actively harm weight loss outcomes [30].
Obesity is probably not caused by inactivity

Many studies have demonstrated correlations between inactivity and obesity [42]. However, there is strong evidence from longitudinal studies that the causal relationship may be the inverse to that generally assumed, that, in fact, inactivity is a symptom of obesity rather than a cause [19,42]. There is also evidence from comparative studies of different countries that overall activity levels do not significantly vary between developed and less developed countries, despite large variations in mean BMI between countries [17], suggesting that levels of physical activity are unlikely to be the main cause of obesity. Studies where physical activity has been objectively measured have also failed to show any prospective effects of physical activity on obesity [67]. Overall, public health research is clear that while physical activity has many health benefits, there is little evidence for exercise as a treatment or preventative measure for obesity [37]. Dietary intake has been found as such a strong predictor of obesity that changes in activity levels have little effect in comparison [39].

Exertion gaming papers argue that computer games cause obesity

A major element of the argument used to promote exertion games for obesity is that computer games “have led children to adopt a sedentary behavior, causing an increase in obesity.” [16] This causal relationship, from time spent computer gaming, to sedentary behavior, and onwards to obesity, is questionable. For instance, the argument relies on an assumed causality from sedentary behavior to obesity, which is not currently supported in the literature [18,44], and also assumes that people don’t compensate for computer game inactivity during the rest of the day. According to the ‘activity stat’ hypothesis, increasing activity at one point in the day may lead children to be less active at other points and vice versa, an effect that has been observed in large scale studies [19].

Goldfield et al. [25] present a review of the current research into exertion gaming as a tool for obesity management and treatment; they conclude that whilst exertion games clearly motivate people to do physical activity at the time of playing, there is limited evidence for real life benefits and conclude that “In the meantime, physical activity in the natural environment with associated benefits of fresh air, vitamin D, connection with nature, and meaningful social interactions should be promoted over exergaming.” [25]. Use of exertion games in schools has also been criticized by public health researchers for encouraging excessive focus on unhealthy body image [53], and by sociologists as part of a wider problem of using ‘crisis discourses’ such as the ‘obesity epidemic’ as a way for commercial interests to exploit public school funds [60].

While the causal link between game play and obesity is not supported by evidence from health research, a version of this argument is presented many times in published exertion game literature, complete with supporting citations from public health research. Through reviewing exertion games papers for a separate study, we noticed that many articles cited a single study as evidence for link between game playing and obesity, entitled; “Linking obesity and activity level with children's television and video game use” by Vandewater, Shim and Caplovitz [65]. This article is cited to support claims such as “too much time spent playing computer and video games is said to lead to physical inactivity and obesity” [43] or that “playing the video games with these devices has become a factor in reduced activity levels and childhood obesity risk” [55] and “Video games are considered the main reason for physical inactivity” [20].

According to citations, which we refer to as VSC, demonstrates a link between computer game play, sedentary behavior and increased weight. However, on reading VSC, it does not demonstrate this causal link at all, because:

- VSC is a single point of time cross-sectional study, not an RCT or prospective study, so it only demonstrates correlations, and clearly states that it cannot suggest causality.
- VSC did not find correlation between videogame time and overall sedentary behavior. It did find strong correlation between obesity and sedentary behavior, although as described above the direction of causality is unclear.
- Whilst VSC demonstrates a significant correlation between obesity and game play, the correlation is curvilinear, indicating that “children with higher weight status played moderate amounts of electronic games, while children with lower weight status played either very little or a lot of electronic games.” [65]

VSC concludes that: “data available to date do not support the notion that turning off the television or unplugging the video game console amounts to a ‘magic bullet’ which will reduce the prevalence of childhood obesity.” [65]

Challenges in interdisciplinary literature reviews

Much of the impetus behind exergame research appears to derive from their claimed potential to encourage exercise and reduce obesity. However, as demonstrated in the brief literature review above, there is little to no evidence of them doing so to this point, nor realistic hope of them doing so effectively in the near future, due to the relative importance of diet in predicting obesity, and the possibility that inactivity is in fact a symptom of obesity rather than a cause.

The continued propagation in the literature of unsupported assumptions raises questions over; i) how well exergame researchers have read, understood and reported the literature investigating the link between activity and obesity, and ii) the robustness of practices and processes through which the reviewing of health literature is undertaken in HCI. Building productively upon research findings from outside your own field is an acknowledged challenge of inter-disciplinary research in HCI [59], and one which we suggest may be significantly impairing the quality of exergames research.
STUDY
In this section, we investigate the types of errors made in referring to VSC, with the intention to understand better the nature and scale of misunderstanding of this single example of a health study used to motivate HCI work.

Method
We set out to perform a structured analysis of citations to VSC. To do this we first used Google Scholar to obtain a list of all 572 articles citing the Vandewater, Shim and Caplovitz [65] paper (VSC) (on 08 April 2016).

We then filtered these by removing:
- 68 non-English articles
- 37 articles to which we had no access
- 1 self-citation by Vandewater
- 34 references in bibliography but no citation in text
- 6 duplicate entries
- 51 “citation only” references with no article.

This left us with a dataset of 375 articles referencing VSC. For each article, we extracted text surrounding the citation of VSC and marked articles as exergaming (n=97) or not (n=278). Exergaming included anything considering games where players must exercise to play, such as: medical interventions using commercial exergames [63], energy expenditure during exergames [52], technical characteristics of exergames [34], and design of outdoor digital games [16].

We manually classified papers into seven categories by considering text surrounding citation to VSC. C outright representing VSC, U: uncertain, and 5 separate categories describing different types of misrepresentation: M1: implying a simple positive linear correlation between gaming and obesity, M2: explicitly describing such correlation, M3: describing causal link, M4: citing it in support of something irrelevant, M5: misrepresentation in relation to TV. Table 1 shows examples of each category.

As we are interested specifically in arguments made relating to games, for the analysis presented below we exclude papers that made miscellaneous errors (M4) or errors relating to television (M5), leaving 342 papers (exergaming = 83, non-exergaming = 259).

Dataset Availability
Whilst the results we present here are extreme, it is clear that there is an element of subjectivity in the analysis of the citations. In order to avoid skewing the results we deliberately made liberal use of the uncertain category, even where the paper was clearly making an argument about obesity and gaming, but it was unclear how the citation to VSC supported that. We invite readers to consult our raw data and analysis variables in supplementary material files - Vandewater-r raw.xlsx: List of VSC citing papers including citation quote and classification, ignoring tv.sav: SPSS variables used for analysis, with M4 & M5 removed, exergame-papers-full.bib, exergame-papers-mention-obesity.bib: Exertion games paper lists used on page 3

<table>
<thead>
<tr>
<th>Category</th>
<th>Description and examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>C: Correct Citation</td>
<td>“children with higher weight status played moderate amounts of video games, but children with lower weight status either played very little or a lot of videogames.” [10]</td>
</tr>
<tr>
<td>U: Uncertain</td>
<td>“obese children are more sedentary” [57]</td>
</tr>
<tr>
<td>M1: Implying simple positive correlation</td>
<td>“increased television viewing and video game use is associated with overweight in children” [27]</td>
</tr>
<tr>
<td>M2: Directly claims purely positively describing a correlation</td>
<td>“children who had greater average game-time minutes also had higher BMIs than the children with lower average game-time minutes.” [48]</td>
</tr>
<tr>
<td>M3: Attributing causality</td>
<td>“these advances in technology have led children to adopt a sedentary behavior, causing an increase in obesity” [16]</td>
</tr>
<tr>
<td>M4: Irrelevant / misleading quote</td>
<td>“Videogames, more so than television, are the culprit for the negative physical and social outcomes” [61]</td>
</tr>
<tr>
<td>M5: VSC found no TV &amp; obesity correlations:</td>
<td>&quot;higher levels of TV viewing have been associated with higher body mass&quot; [29]</td>
</tr>
</tbody>
</table>

Table 1. Example categorizations of citation errors
H2: Exertion gaming makes worse mistakes in citing VSC than non-exertion gaming <Supported by our data>
As well as making more mistakes than non-exertion gaming, we can see from the results that exertion gaming makes more of the worse types of mistakes. We verified this with a Mann-Whitney U-Test using the ordering Correct, Uncertain, M1: implies linear correlation, M2: claims linear correlation, M3: attributes causality. Error severity for exergames (mean rank = 149.34) was statistically significantly higher than for non-exergames (mean rank = 240.65), n = 342, U = 5009, z = -8.091, p < .001. In particular, 26% of exertion games papers suggested that VSC describes a causal link between games and obesity, in comparison to 2% in other articles.

Is this finding purely due to disciplinary effects?
The study of exertion games is highly interdisciplinary. As well as within HCI, exertion games are studied in a range of areas such as sports science, public health, child development (where they are often called “active video games or AVGs). Many, but not all of these disciplines are very far removed from the child health research area that VSC is published in, which could potentially lead to misunderstanding of VSC. One potential confounder of our initial results could be that errors in citation are purely caused by the fact that exertion games work is often published in non-health venues by researchers who are unlikely to be health researchers and hence may not understand the complex statistical presentation of VSC, whereas non-exergaming citations may be mainly by health researchers who would understand this work.

To explore whether such disciplinary effects explain away the observed difference in error rates between exertion games and non-exertion games research, we added a discipline variable based on the publication venue of each article. We categorized the data into two sets: health disciplines closely linked to VSC (e.g. medical, health, child development), where authors could be expected to understand medical research results, 217 papers, and unrelated areas (e.g. computer science, media studies), 125 papers. These results are presented in Figure 2 and Table 3. Numerical breakdown of health vs non health articles. We used this breakdown to explore the hypothesis that:

H3: Poor citation in exergames literature is caused purely by people from non-health related disciplines being unable to understand health literature <Not supported>
If we first consider health versus non-health publications as a whole, it appears that the hypothesis can be supported – that health publications make less bad mistakes than publications from other disciplines; (Mann-Whitney U-Test, mean ranks=non-health: 196.11, health:157.32), n=342, U=10486.5, p<0.001. However, as seen in Figure 2 the overall disciplinary difference is explained by differences specifically in exertion game research. Of the two groups of exergaming papers, error severity was significantly lower for those published in health research (mean rank=32.06) than other disciplines (mean rank=49.62), n=83, U=488, p=0.001. Conversely, looking

---

**Figure 1. Accuracy of citation in exertion gaming (n=83) and non-exertion gaming (n=259) papers**

**Table 1. Numerical breakdown of citation errors by field**

<table>
<thead>
<tr>
<th>Category</th>
<th>Other</th>
<th>Exertion Gaming</th>
</tr>
</thead>
<tbody>
<tr>
<td>C: Correct</td>
<td>174 (67%)</td>
<td>17 (21%)</td>
</tr>
<tr>
<td>U: Uncertain</td>
<td>17 (7%)</td>
<td>8 (10%)</td>
</tr>
<tr>
<td>M1: Implies positive linear correlation</td>
<td>32 (12%)</td>
<td>21 (25%)</td>
</tr>
<tr>
<td>M2: Directly claims positive linear correlation</td>
<td>30 (12%)</td>
<td>15 (18%)</td>
</tr>
<tr>
<td>M3: Attributes Causality</td>
<td>6 (2%)</td>
<td>22 (26%)</td>
</tr>
<tr>
<td>Totals</td>
<td>259</td>
<td>83</td>
</tr>
</tbody>
</table>

Note: After the title of each hypothesis, we mark whether we found it to be supported by the data or not. This aims to ensure that anyone skim-reading the paper does not assume H3 is supported, and incorrectly cite our work.

**H1: Exertion gaming researchers are more likely to make mistakes in citing VSC than non-exertion gaming researchers <Supported by our data>**
This hypothesis explores whether citation behavior is different in papers on exergames than other papers. The hypothesis is strongly supported by our data; exertion gaming papers are far more likely to make mistakes than non exertion gaming work. Discarding uncertain results (i.e. only categories C vs M1, M2 & M3), we performed a chi-square test for association, comparing exertion games (23% correct) and non-exertion games (72% correct), which was statistically highly significant, $\chi^2(1) = 57.952$, p<0.001.

---

**Figure 2. Distribution of publication venue**

---

**Table 2. Classification of disciplines (n=342)**

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health</td>
<td>275</td>
</tr>
<tr>
<td>Other</td>
<td>67</td>
</tr>
</tbody>
</table>

---

**Table 3. Error rates by publication venue**

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Non-Health</th>
<th>Health</th>
<th>Mann-Whitney U-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Rank</td>
<td>Mean Rank</td>
<td>U</td>
</tr>
<tr>
<td>Health</td>
<td>196.11</td>
<td>157.32</td>
<td>10486.5, p&lt;0.001</td>
</tr>
<tr>
<td>Other</td>
<td>49.62</td>
<td>32.06</td>
<td>32.06</td>
</tr>
</tbody>
</table>
DISCUSSION
False claims regarding a causal link between game playing and obesity have propagated in the literature on exertion games. In order to better understand the pervasiveness of this misrepresentation, as well as the scholarly practices that support its propagation, we carried out a study examining all citations made to an influential paper from child health research cited to support such claims. This showed that exergames work makes mistakes in representing VSC more frequently, and not purely for reasons of cross-disciplinary misunderstanding. The findings of our study suggest that exertion gaming research misrepresents the results reported in VSC far more than is the case in papers with a different focus, despite blinded peer review of most work.

Due to the findings of the current study, it seems an appropriate point in time to consider why this poor scholarly behavior may be occurring in the field of exertion games. From working closely in this field ourselves, we propose four factors which may lead to the state of affairs where exertion gaming misrepresents health research to a worse extent than other types of research. We believe it is likely to be due to a combination of some or all of these factors.

1) Reading errors
Papers such as VSC with large numbers of statistical results are easy to misunderstand. It is possible some researchers simply skim-read such papers and did not read the details of the main conclusions, or that they read only the abstract (although the abstract does clearly state that the correlation observed between games and obesity is curvilinear). Non-exergaming, non-health authors show no evidence of misunderstanding VSC, so this cannot be a full explanation, but unless we believe all researchers making errors cite papers such as VSC with large numbers of statistical results, some degree of failure to read cited literature thoroughly must be part of the problem.

2) Hasty Literature Reviews
Preparing concise yet comprehensive reviews of previous work in an interdisciplinary area such as exertion gaming presents challenges. While researchers are likely to have a good idea of the exertion gaming literature itself, given that much exertion gaming work is done by researchers with computer science or HCI backgrounds, they are unlikely to have a broad knowledge of health literature. To achieve scope and breadth necessary, researchers may take short cuts. As has probably occurred in some work studied here, instead of carefully reading source material, authors validate their statements through citing the same studies as are cited in influential papers in their field. Literature review sections also tend towards briefly describing rather than analyzing and criticizing previous work. Because of this, misunderstandings can propagate within sub-disciplines.

3) Reviewer Errors
In interdisciplinary fields, “scientific ‘peers’ can no longer be reliably identified, because there is no longer a stable taxonomy of codified disciplines from which ‘peers’ can be drawn” [50]. Reviewers in HCI will see research cited from

![Figure 2. Health vs non health breakdown](image)

**Table 2. Numerical breakdown of health vs non health articles**

<table>
<thead>
<tr>
<th></th>
<th>Non-health</th>
<th>Health</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Other</td>
<td>Exertion Gaming</td>
</tr>
<tr>
<td>C: Correct</td>
<td>49 (62%)</td>
<td>6 (13%)</td>
</tr>
<tr>
<td>U: Uncertain</td>
<td>6 (8%)</td>
<td>5 (10%)</td>
</tr>
<tr>
<td>M1: Implies positive linear correlation</td>
<td>13 (17%)</td>
<td>9 (19%)</td>
</tr>
<tr>
<td>M2: Directly claims positive linear correlation</td>
<td>8 (10%)</td>
<td>6 (13%)</td>
</tr>
<tr>
<td>M3: Attributes Causality</td>
<td>2 (3%)</td>
<td>21 (45%)</td>
</tr>
<tr>
<td>Totals</td>
<td>78</td>
<td>47</td>
</tr>
</tbody>
</table>

*These findings suggest that H3 can be rejected – we have no evidence that health researchers simply do a better job of representing health research. Research on exergames is uniquely bad in consistently misrepresenting VSC.*
fields unfamiliar to them; in these fields they are a) unlikely to have read all citations, b) unlikely to have time to read all cited papers, and c) if they do, are unlikely to read in sufficient depth to catch subtle errors in citation. We believe this may explain why 21 exertion gaming papers passed peer review in non-health fields despite suggesting causality is detectable from a cross-sectional study. As an example of this potential, the first author recently reviewed a paper for a major ACM conference which cited VSC to support a claim that gaming causes obesity, and was the only one of four reviewers to even consider the erroneous claim.

4) Cherry Picking to Justify Pre-existing Agenda

We suggest that, in some of the studies we reviewed, the idea of using exergaming to fight obesity was taken prior to any critical review of the literature. It is possible that, rather than beginning projects by openly surveying the health literature to understand problems that must be urgently addressed, researchers are forming research agendas based on their non-expert understanding of social issues, and later seeking out literature that they believe supports this understanding. For example, one exergame article begins “the western world faces an escalating obesity problem, mainly due to lack of activity” [45]. This is a highly contentious statement, presumably based on a lay understanding of obesity and exercise and is clearly not the result of an exhaustive reading of health literature. The VSC paper was cited to provide support for an already existing argument (that inactivity caused by games is the cause of obesity). Having decided on such justifications, authors then presumably set out to find research supporting their justifications.

From years of experience reviewing interdisciplinary game research, we suggest “literature reviews” are actually often constructed post-hoc to support findings. This observation is supported by the negative result of our third hypothesis, which suggests that exertion gaming’s misrepresentation of VSC is not purely due to misunderstanding of work from different disciplines. We suggest exergaming work may have a particular agenda to present evidence that “games cause obesity” and “exercise can cure obesity” which is not necessarily present in other fields.

**IMPLICATIONS FOR HCI AND EXERTION GAMES**

In this paper, we have shown that large numbers of peer reviewed exertion games articles make fundamental mistakes in their representation of health research. We believe that this has serious implications for exertion games and also for the wider project of HCI. We present these implications in the form of 3 lessons:

**Exertion gaming should focus on areas where it can realistically be useful**

In previous work [41], we argue that an excessive focus on obesity is likely to lead to poor game designs. The lack of focus in many exergames projects on designing interesting experiences means that those games also likely to be ineffective, since people will not choose to play uninteresting games. We argue that if exertion gaming researchers wish to create games with health benefits, they should be realistic in considering what health benefits are possible, and rigorous in evaluating whether the game interventions being proposed have any practical likelihood to be effective. Neither of those traits are prevalent in the majority of research studied in the current paper. Exercise has many health benefits unrelated to obesity, such as increasing strength and stamina, reducing chances of disease and positively affecting mental health and cognitive function in both adults and children [8,31], so there are many reasons to encourage it, but an excessive focus on obesity and calorie burn risks missing the opportunity to encourage useful exercise behavior.

Further to that, we must do a better job of considering the long term outcomes of this work [25]. It is not enough to create a game for the general public and say that it will have health benefits without considering likelihood of low ongoing use. Arguably, games designed for and evaluated in therapeutic (e.g. [11]) or school settings (e.g. [38]), where there is an element of coercion or real incentives to play games over the long term, may provide more realistic potential for impact. As an example of a study which presents realistically achievable effects, Gao & Mandryk [21] describe acute, short term cognitive benefits which occur directly after exertion. These effects are well supported by health literature, and because they are short term in nature, can be demonstrated in the short-term studies in the paper.

Beyond a health focus, as game researchers and sports people, we also strongly believe that hard physical exercise is in itself a positive and exciting thing which can be used to create games which are fun and interesting, and that exertion game researchers should not discount the design of highly physical exercise games purely for entertainment’s sake, i.e. treating them as interesting ‘body games’ [40] or ‘sports’ [46], rather than as a device for producing health benefits.

**HCI must promote a culture of respect for other fields**

Exertion games are just one example of an area where HCI intersects with other research fields. HCI is inherently interdisciplinary, requiring an understanding both of technology and people, but also of the areas in which the technology is applied [12]. HCI has previously been criticized for shallow interdisciplinarity, where authors do not fully understand work from other fields, leading to “violence being done to the origins of imported approaches and concepts.” [59]. For example there have been heated debates about (mis)use of methods from ethnography [15] and design research [24,56] within HCI, and about whether HCI research needs to become more ‘scientific’ by focusing on replicating existing results [28].

What we have seen in our exertion games case study appears to be an underlying lack of respect for the health research that is being cited. Rather than survey the wider literature prior to practical work in order to inform the nature of the research to be done, we believe motivational sections of articles may be treated as somewhere to cite prior work in order to justify
research that has already been done or theoretical positions that have already been taken.

As HCI reviewers a key implication is that we need to pay more attention to the first two pages of papers where people motivate and situate their work within wider research, rather than focusing critical attention primarily on experimental results, methodology and discussion. We need to take a more critical approach to why people are doing work, how it is motivated and whether that motivation is realistic.

Further to this, as chairs and editors seeking reviewers, this work highlights the fact that when work aims to be highly grounded in application areas such as health promotion, it is vital that the reviewer panel includes reviewers who are able to evaluate claims relating to that application area. For the field of health, we note that the SIGCHI conference on Human Factors has this year assembled a health subcommittee which will hopefully address some of our concerns. This could help avoid situations such as has occurred in exertion games, where mistakes spread within a sub-area of HCI as people read each other’s work without studying source material in depth and as Reeves describes “without specialist attention, weak strains are sustained/incubated within HCI;” [59].

We need to be critical about “HCI for Good” to avoid accidentally causing harm

Linehan et al’s concept of “Games Against Health” [35] argues that the Games for Health movement goes against the ideas of user-centric design, in that large communities of users clearly enjoy behaviors such as playing sedentary video games and eating junk food, and that in aiming to discourage such behaviors, designers are deliberately going against what users actually want. Even if we do not accept their view that trying to encourage healthy behaviors is patronizing “cultural imperialism”, we need to approach claims that HCI work is aimed at societal good in a critical manner.

This goes beyond simply preventing poor research, or presenting research that will not do harm. False claims to do good have the potential to form part of a wider agenda that is actively harmful. For example, as we describe above, effective strategies to prevent or combat obesity will necessarily prioritize dietary interventions above exercise based interventions. Exertion gaming research however assumes a simplistic model of “energy balance”, summarized by one exertion gaming author as “One of the key reasons for the increasing obesity epidemic is positive energy balance, that is, the condition where one’s energy intake exceeds one’s energy expenditure” [6]. This model suggests that people can lose weight via exercise alone, or that obesity can be blamed on lack of activity alone (e.g. “the western world faces an escalating obesity problem, mainly due to lack of activity” [45]). The wider health literature does not support this simple energy balance model, or that lack of activity is the main cause of obesity.

Despite being largely discredited within academic obesity work, this simplistic concept of energy balance continues to be widely promoted in the general media, and as we see from exertion gaming research is believed by many non-experts. One particular and well documented reason for this heavy promotion of energy balance is the existence of commercial food and drink interests who wish to minimize concerns about the unhealthiness of their products and avoid regulations such as ‘sugar taxes’. This promotion even reaches into academia, for example Coca Cola was widely criticized in 2015 for secretly funding an “independent” academic group called the “Global Energy Balance Network” which promoted the idea that inactivity, not unhealthy dietary practices was the key to fighting obesity (the Network was disbanded in December 2015 and participating universities returned funding after leaked emails demonstrated that it was explicitly a “political campaign” on behalf of the soft drink industry [51]).

In effect, exertion game research, in its desire to “do good”, has in fact been repeating and promoting an unscientific message that is heavily supported by providers of unhealthy food aiming to obscure more scientifically founded but less well funded public health messages relating to diet. We wonder whether if HCI had existed in the 1930s, we would see researchers developing persuasive systems to encourage smoking, based on successful campaigns by tobacco companies to engender a public belief that smoking was a healthy habit, something that doctors recommended. [23].

We suggest that we should be wary of attempts to “do good” in HCI which are not founded on deep critical understandings of underlying research and wider societal issues. For example, many research projects exist which suggest the use of wireless cloud devices, internet of things nodes or smartphones [33] to drive behavior change in energy use. However, unless such projects make serious consideration of their use of energy, both that involved in running the wireless cloud infrastructure that supports it [13] and the embedded energy involved in building smartphones and networks [58], they may be counter-productive [9].

As an example of good critical practice in HCI, the field of HCI4D has engaged strongly with “post-colonial” design, arguing that many assumptions in mainstream HCI reflect the nature and norms of western society [7] and do not necessarily apply in other cultures. However, as the example of exertion games and obesity demonstrates, it is important to be aware that HCI work may in fact also reflect incorrect assumptions about our own society.

CONCLUSIONS

In this article we present an analysis of the use of health claims and citation of health research to justify exertion games work which aims to combat obesity. In our case study, we consider a comprehensive dataset of citations to one study and show that 58% of the exertion games papers sampled (69% of our dataset) misrepresent the results of that study. That such a large number of errors can both have been made
by authors, and have got past extensive peer review panels such as those used by CHI is a sign of systematic failure of scholarship in the field.

We note that this kind of error in citation is not unique to HCI; whilst we were unable to find any comparable analysis of a single cited article, sampling based studies in other fields have identified citation errors in roughly 20% of citations [26,64]. If we make a reasonable assumption that such errors are not uniformly distributed amongst source articles, it is highly likely that similar examples exist in other areas of HCI and beyond. In future work, we aim to consider other areas where HCI makes extensive use of research from disciplines that are not core to HCI, and consider whether such problems are endemic in those areas also.

Finally, we should note that, despite these criticisms, as researchers in the field we are excited about the growth of exertion and full body gaming. We believe firstly that focused, evidence based takes on health such as in work on short-term cognitive benefits of exergaming [21] may offer practical health applications. Such work applies wider considerations of why physical activity is positive for participants than focusing on obesity or other unrealistic goals. As game enthusiasts, we hope that this work will encourage the exertion gaming community to focus more broadly on the wider positive nature and effects of physical play and to create new and exciting exertion gaming experiences which are not bound by the need to count calories or measure obesity related results.

ACKNOWLEDGMENTS
Joe Marshall’s work was funded in part by Leverhulme Trust (ECF/2012-677) & EPSRC (EP/M000877/1).

REFERENCES


35. Conor Linehan, Sabine Harrer, Ben Kirman, Shaun


