# Effect of Conformity on Perceived Trustworthiness of News in Social Media

Senuri Wijenayake The University of Melbourne

Danula Hettiachchi The University of Melbourne

Simo Hosio University of Oulu

Vassilis Kostakos The University of Melbourne

Jorge Goncalves The University of Melbourne

Abstract—A catalyst for the spread of fake news is the existence of comments that users make in support of, or against, such articles. In this study we investigate whether critical and supportive comments can induce conformity in how readers perceive trustworthiness of news articles and respond to them. We find that individuals tend to conform to the majority's opinion of an article's trustworthiness (58%), especially when challenged by larger majorities who are critical of the article's credibility, or when less confident about their personal judgement. Moreover, we find that individuals who conform are more inclined to take action: to report articles they perceive as fake, and to comment on and share articles they perceive as real. We conclude with a discussion on the implications of our findings for mitigating the dispersion of fake news on social media.

■ INTRODUCTION Social media platforms are increasingly becoming the primary source of news and information for most people. In a recent survey, 68% of Americans reported to at least occasionally consume news through social media, with Facebook being the most commonly used platform [1]. People perceive social media to be a more convenient, cheaper and timely alternative to traditional news sources, with the additional opportunity to interact with others by commenting and sharing news articles [1].

However, the convenience, cost-efficiency and accessibility offered by online social media that

helped gain its fame, has also resulted in these platforms being exploited for the rapid dispersion of fake news *i.e.*, "news articles that are intentionally and verifiably false, and could mislead readers" [2]. Purveyors of fake news attempt to sway the public's opinion to accept biased or false information to support political propaganda, personal ideology or to gain financial incentives. While the Pew Research Center reports that most users tend to question the veracity of news they consume through social media [1], other reports (*e.g.*, [2], [3]) emphasising the influence of fake news on the 2016 U.S. presidential election sug-

1

gest that a majority of the general population is still largely susceptible to fake news.

As a result, determining what factors influence the perceived trustworthiness of news articles appearing on social media (*i.e.*, whether an article is fake or real) has become a critical research topic. While existing work highlight how comments appearing on news articles influence opinion formation in subsequent users [4], [5], their impact on how users perceive the trustworthiness of a news article is yet to be investigated. Furthermore, to fully comprehend the forces at play, it is vital to understand how socio-psychological concepts such as 'social conformity' may influence how people perceive trustworthiness and respond to news articles appearing on social media.

Therefore, this study intends to investigate whether and how the composition of user comments either supporting or criticising a news article posted on Facebook, may trigger conformity in how subsequent users assess its trustworthiness. Moreover, we aim to explore how users' perceived trustworthiness of a news article may influence how they respond to it in social media, which is also crucial to the dispersion of online news.

# **Related Work**

The influence of fake news stories that disseminate through social media has been undeniable since the 2016 U.S. presidential elections. Studies report that an average American encountered between 1 to 3 fake news stories on social media in the month leading to the election, which they often accepted as genuine information [2]. Others also report that a majority of such fake news stories were in favour of Donald Trump, which may have been a significant determinant of his victory in the election [3].

Previous work revealed that fake news spreads faster and deeper than genuine news articles in social media, due to the behaviour of human users and not social bots [6]. Therefore, existing literature exploring computational, expert-oriented and crowdsourcing approaches to determine the veracity of social media news articles [7] may not be sufficient to mitigate the dispersion of fake news. For instance, Facebook attempted to mitigate the dispersion of fake news by displaying disclaimers on certain news articles, to alert users that the article might be fake. Contrary to expectations, displaying disclaimers disputing an article's content was seen to further establish incorrect preconceptions of users [8].

Moreover, literature recognises the impact of user comments on opinions of news readers in online contexts. Studies suggest that user comments may have a higher impact on a reader's opinion than the article itself [4]. Furthermore, in cases where user comments disagree with an article's viewpoint, readers' opinions have been seen to align with that of user comments [5].

In a recent study, Colliander [9] investigated the impact of comments posted by previous readers on a fake news article posted on Facebook, on the attitude and response of subsequent readers. The results indicate that exposure to comments critical of a fake news article adversely impacted participants' attitude on the article and lowered their likelihood to positively comment or share the article on Facebook. Alternatively, supportive comments were seen to favourably impact participants' attitude on the fake news article and increase their likelihood to comment positively and share the article. Furthermore, the above results were compared with an article consisting of supportive comments as well as a disclaimer alerting participants that the article might actually be fake, which did not lower participants' attitude or their likelihood to comment and share the post. The author concluded that comments from other users are more influential than a disclaimer from a social media platform, due to effects of 'social conformity', i.e., the human tendency to adjust personal opinions to agree with a group majority, seeking social approval (normative conformity) or presuming the majority to be 'correct' in uncertain situations (informational conformity) [10], [11], [12].

While Colliander's study established the influence of conformity to user comments on a Facebook news article, it only investigated conformity in the presence of unanimously critical or supportive comments, while in reality a news article could have a combination of supportive and critical comments (*e.g.*, a majority of supportive comments vs. a minority of critical comments and vice versa). This is especially crucial as literature strongly emphasises on the effect of majority – minority group compositions on conformity behaviour [10], [11], [12]. Furthermore, conforming to user comments could also depend on personal determinants of users such as their self-confidence [10], [11], [12], familiarity with the news article, gender, age and time spent on social media which are yet to be investigated.

Moreover, further work is required to investigate how users respond to articles they perceive as fake or real, across a wider variety of responses that could enhance or mitigate article dispersion (*i.e.*, their likelihood to 'react', report and factcheck the post in addition to commenting and sharing it), to fully understand the impact of conforming to others' opinions on the dispersion of news articles and whether it could effectively correct misinformation.

## Method

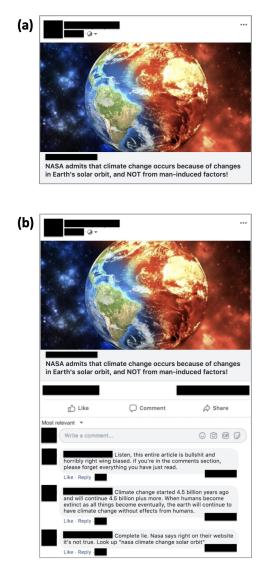
We deployed an online survey consisting of 28 Facebook posts including 14 fake and 14 real news articles. The use of Facebook posts for the study was inspired by previous work investigating conformity in social media [9]. All fake and real news articles were extracted from Snopes - a popular fact-checking website. The articles chosen were fact-checked by Snopes between Jan 2019 – Feb 2020 based on popular demand and were labelled as either 'false' (fake news) or 'true' (real news). We did not include news articles that may directly favour a specific political affiliation to avoid possible confounding effects.

We manipulated the total number of comments appearing on the Facebook post (2 - 4; a minimum of 2 to create a 'group' opinion anda maximum of 4 as previously seen sufficientto elicit conformity by Colliander [9]) and theirarrangement (in terms of the number of critical/supportive comments, and their presentationorder). This meant that each post tested a uniquecombination of critical and supportive comments, with either a majority of critical comments, amajority of supportive comments or with an equalnumber of critical and supportive comments (nomajority). The considered arrangements of comments were equally tested on both fake and realnews articles.

All supportive and critical comments included in each post were extracted from the relevant original news article. We defined a comment posted by a user that supports the authenticity of the article as a 'supportive' comment (*e.g.*, "*Scary scenario and quite plausible!*"), and any comment posted by a user questioning the authenticity of the article or directly criticising it as fake news or misinformation as a 'critical' comment (*e.g.*, "*People, please stop spreading fake news. At least some of you should be smarter than this*"). We note that in the original news articles, the "most relevant" user comments were displayed based on their popularity and engagement [13]. Hence, we chose the top critical and supportive comments from each post, to be utilised in the experiment.

The survey was deployed on Amazon Mechanical Turk with the participation of 52 US residents (26 women and 26 men) over the age of 18 years (M = 36, SD = 9.25). All participants were registered users of Facebook and had completed more than 1000 HITs with an approval rate above 95%, a commonly used pregualification criteria used in MTurk studies [14]. Furthermore, participants were provided a downloadable Plain Language Statement with the instructions and the requirements of the survey before accepting the task. Upon accepting the task, they were asked to provide their demographic information (age, selfdisclosed gender, level of education completed and the approximate number of hours spent on social media). Subsequent to submitting their information, participants could then initiate the survey.

The survey was designed to capture the change in participants' personal opinion of an article's trustworthiness after being exposed to others' comments. At first, each post was displayed without user comments, *i.e.*, only the cover image and the headline used in the original article were visible as shown in Figure 1 (a). Participants were instructed to imagine the post appearing on their news feed as shared by one of their distant friends on Facebook. They were then asked to rate their familiarity of the article (on a scale of 0 - 100, 0 = 100 familiarity and 100 = high familiarity), their opinion on the article's trustworthiness (on a scale of 0 - 100, 0 = criticaland 100 = supportive) and their confidence on the provided trustworthiness rating (on a scale of 0 - 100, 0 = 100 confidence and 100 = 100confidence). Participants were then prompted to indicate whether and how they would respond to the post by reacting, commenting, fact-checking,



**Figure 1.** An example of a fake news article used in the survey, both with and without user comments.

sharing or reporting it (each on a scale of 0 - 100, 0 = extremely unlikely and 100 = extremely likely), based on their initial assessment of the article's trustworthiness.

Upon submitting their initial answers, we displayed the complete post with user comments as shown in **Figure 1** (b). Participants were instructed to read the comments carefully and determine whether the displayed user comments are mostly supportive or critical of the post's trustworthiness. In response, participants could select one option among "supportive", "critical" or "equally distributed among the two". This question was used as a Gold Standard question to determine whether participants have read through the comments with adequate attentiveness before moving on to the next step.

After displaying the user comments and answering the Gold Standard question, participants were again requested to provide their opinion on the article's trustworthiness and their confidence on the new trustworthiness rating. Subsequently, they were also prompted to provide new ratings to reflect their likelihood to react, comment, factcheck, share or report the post after reading user comments. This approach allowed us to capture the effect of previous user comments on participants' personal opinion on the article's trustworthiness and how they subsequently chose to respond to the post. This process was repeated for each post in the survey.

The experimental design was approved by the Ethics Committee of our university. The experiment lasted for approximately 30 - 45 minutes per participant. Participants who answered the survey in full, with at least 80% of correct answers for the Gold Standard questions received a payment of \$10 (USD) for participation.

# Results

All 52 participants answered the 28 survey items which resulted in a total of 1456 responses, equally distributed among fake and real news articles. Of these, in 1040 responses participants faced a clear majority, either supportive or critical of the article's trustworthiness. In the remaining 416 responses, the previous user comments had an equal number of supportive and critical comments.

For the purpose of this study we define conformity as a binary variable, *i.e.*, increasing trustworthiness rating after seeing a majority of supportive comments and reducing the trustworthiness rating after seeing a majority of critical comments were considered as conformity. We observed conformity in 604 out of the 1040 responses where there was a clear majority (supportive or critical), leading to a 58% conformity rate. All participants conformed at least once during the survey, with an average of 11.62 conformity responses (*SD* = 3.44) per participant. We then investigated the impact of the following variables on the conformity behaviour of our participants, to understand factors that may have influenced their behaviour.

- Majority opinion: Supportive or critical majority.
- **Majority size**: Size of the majority (range: 2 4).
- **Minority size**: Size of the minority (values: 1 or 0).
- **Group size difference**: Difference between the majority group size and the participant's group size (range : 0 4).
- Num. of comments: Total number of previous user comments (range : 2 4).
- Num. of critical comments: Num. of comments critical of the article's trustworthiness (range : 0 4).
- Num. of supportive comments: Num. of comments supportive of the article's trustworthiness (range : 0 4).
- **Familiarity**: Participant's familiarity of the article (range: 0 100).
- Initial confidence: Participant's confidence in the initial trustworthiness rating prior to revealing user comments (range: 0 100).
- Gender: Participant's self-disclosed gender.
- Age: Participant's age (range: 20 59).
- Social media usage: Hours spent on social media per week by the participant (range: 1 30).
- User ID: A unique identifier assigned to each participant during the survey.

We used the R package *lme4* to perform a generalised linear mixed-effects model (GLMM) analysis of the relationship between the aforementioned variables and participant conformity. A GLMM allows us to identify the effect of a set of predictor variables on an outcome variable (conformity) while following an arbitrary (*i.e.*, possibly non-normal) distribution. We specified participant (User ID) as a random effect to account for individual differences in our model.

All statistically significant predictors included in the final model (following model selection through incremental addition of variables based on their predictive power) are shown in **Table 1**. We performed a likelihood ratio test with the null model and found that our model is statistically significant ( $\chi^2 = 427.95$ , p < 0.001) and explains 33.2% of the variance in accuracy (R =0.58,  $R^2 = 0.33$ ). To ensure the validity of the model, we then checked for the existence of

# Table 1. Effect of statistically significant predictors on participant conformity.

Predictor	Coefficient	P-value
Group size difference : Majority opinion (critical)	1.00	< 0.001
Group size difference	0.63	< 0.001
Initial confidence	-0.01	< 0.001

between the predictor and conformity behaviour. Absolute value of the coefficient determines the effect size.

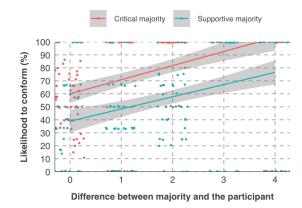
multicollinearity. Our predictors report variance influence factors less than 1.10, well below the often-used threshold of 5 to detect multicollinearity [15].

We observed statistically significant main effects from group size difference (difference between the majority group size and the participant's group size) and the self-reported initial confidence level of participant. Moreover, the group size difference also interacted with the majority's opinion (either supportive or critical of the article) to display the highest effect on conformity. Next, we present a more detailed look of the significant features.

Group size difference, Majority opinion and Initial confidence

Our results reveal that participants were more inclined to conform to the majority as the size difference between the majority and themselves increased (despite the influence of other variables), signifying a main (positive) effect from group size difference on conformity. Furthermore, the impact of the group size difference on conformity heightened when participants were challenged by critical majorities than supportive majorities. As shown in **Figure 2**, while both lines display an upward trend, the likelihood of participants conforming to critical majorities is consistently higher than their likelihood of conforming to supportive majorities.

Furthermore, participants' confidence on their initial trustworthiness rating of an article (prior to seeing user comments) displayed a statistically significant negative effect on their conformity behaviour. When participants were less confident about their initial trustworthiness rating they were more likely to be influenced by the majority's opinion. While the initial confidence level of participants ranged between 0 - 100 in both non-conforming and conforming responses, mean initial confidence values were at 72.09 (SD =



**Figure 2.** Participants' likelihood to conform to supportive and critical majorities as the group size difference increases from 0 - 4.

28.77) and 61.84 (SD = 25.64) respectively.

We did not note any significant effect from participants' reported familiarity of the post, gender, age or social media usage on their conformity behaviour during this analysis. Moreover, the total number of comments appearing in the post, the number of critical/supportive comments, or their presentation order had no notable influence on participant conformity.

### Participants' responses to news articles

Upon establishing the presence of conformity in how people perceive trustworthiness of online news, we then investigated whether and how participants' conformity behaviour impact their response to news articles (*i.e.*, their likelihood to react, comment, share, fact-check or report the Facebook post). We ran paired t-tests on the likelihood ratings reported by participants for each of the aforementioned response types, before and after viewing others' comments. This analysis was conducted across conforming and non-conforming responses separately, when participants were facing either supportive and critical majorities. Our results are summarised in **Table 2**.

We observe statistically significant mean differences among the before and after response likelihood ratings in conformity responses. Participants who conformed to a majority supportive of an article's trustworthiness were significantly more inclined to fact-check, share, comment on and react on the Facebook post (respectively), after viewing user comments. Alternatively, par-

Table 2. Mean difference between the likelihood ratings provided before and after seeing user comments for each response type.

Response Type	Conforming Responses		Non-conforming Responses	
	Supportive Majority	Critical Majority	Supportive Majority	Critical Majority
React	5.06 (p <0.001)	-11.20 (p <0.001)	-3.74 (p=0.002)	-1.61 (p=0.232)
Comment	5.28 (p < 0.001)	-5.75 (p <0.001)	-0.83 (p=0.480)	-0.37 (p=0.787)
Share	6.60 (p <0.001)	-8.81 (p < 0.001)	-2.02 (p=0.026)	-0.72 (p=0.270)
Fact-check	7.32 (p <0.001)	-6.81 (p <0.001)	-2.62 (p=0.081)	-7.91 (p <0.001)
Report	-2.32 (p=0.029)	14.77 (p <0.001)	3.40(p=0.008)	11.78 (p < 0.001)

(likelihood increased). Negative mean differences indicate that the initial ratings are higher in value that the subsequent ratings (likelihood reduced).

ticipants who conformed to a majority critical of an article's trustworthiness reported higher inclination to report the post, while also lowering their inclination to react, share, fact-check and comment on the post.

Our results do not indicate statistically significant mean differences between the before and after response likelihood ratings in non-conforming responses against supportive majorities. However, when challenged by critical majorities participants reported higher inclination to report the post, and lower inclination to fact-check the post despite their non-conforming behaviour. The likelihood ratings reported for sharing, commenting and reacting on the post did not significantly change when faced with critical majorities in nonconforming responses.

### Discussion

As human behaviour contributes more towards the dispersion of fake news than bots [6], its mitigation requires a thorough understanding of how people derive conclusions on a news article's trustworthiness. This study investigated how a combination of critical and supportive comments posted by others on a Facebook news article could influence subsequent readers' perception of the article's trustworthiness as well as their response to it.

Our findings confirm that readers frequently adjust their personal opinion on a news article's trustworthiness to agree with the opinion of a majority of previous readers, demonstrating conformity behaviour as seen in previous work [9]. As our study utilised combinations of both supportive and critical comments, we emphasise that a unanimous majority (critical or supportive) was not essential to trigger conformity, in contrast to previous observations [9]. Moreover, we note that participants were more likely to adopt the majority's opinion on an article's credibility as the number of comments reflecting the majority's sentiment (or the majority's size) increased. This is inline with observations from previous studies on online conformity [10], [11], [12]. More interestingly, the influence of the majority's size on participant conformity was higher when the majority was critical of an article's trustworthiness, than when the majority was supportive.

In addition, participants disregarded the majority's opinion when confident of their initial judgements, but were eager to adopt the majority's opinion when unsure of their initial judgements. Literature explains this behaviour as 'informational' conformity, where individuals conform to the majority presuming it to be 'correct' in uncertain situations (which is usually the case in online settings [10], [11], [12]). However, contrary to previous perceptions [9], an individual's familiarity with a news article or the time they spend on social media had no impact on their conformity behaviour.

Furthermore, readers who conformed to a critical majority were more inclined to take action against the dispersion of the news article (by reporting it) and were less inclined to contribute towards its further dispersion (by reacting, sharing and commenting on it), than readers who did not conform to the majority. Similarly, readers who conformed to a majority supportive of an article's trustworthiness were significantly more likely to share, comment and react on the news article to enable its further dispersion, than those who did not conform to a supportive majority.

Therefore, our observations not only confirm results of prior studies where the critical or supportive nature of user comments have been seen to influence how readers' perceive news articles as fake or real [9], but also provide insights on how conforming to others' opinions influence readers to align their responses to reflect the majority's (supportive or critical) opinion of an article. Hence, platforms should consider how user comments appearing underneath news articles on social media could be utilised to mitigate the dispersion of fake news, and encourage the dispersion of real news articles. Mitigating the dispersion of fake news on social media

Social media platforms (*e.g.*, Facebook) have been exploring how platform design could provide more context and facts on news articles to its readers, to assist them determine an article's credibility [8]. Our findings suggest that readers are receptive to others' comments and tend to mimic the majority's opinion on an article's trustworthiness derived through others' comments. This implies that displaying a filtered set of comments could be an effective approach to signal trustworthiness and credibility of news articles to readers and expose them to different perspectives on the topic, in comparison to displaying suggestions for related articles (as currently seen on Facebook).

Moreover, our results also have important implications on how platforms can effectively inform readers about fake news articles. Previous research note how disclaimers alerting readers of fake news had counterproductive results as they further entrenched personal beliefs of individuals [8]. In comparison, we recommend displaying a filtered set of critical user comments (unanimous or otherwise) underneath confirmed fake news articles, which is likely to encourage readers to adopt a critical opinion of its credibility, despite their personal perceptions due to conformity influences.

Conforming to the majority's opinion on an article's credibility also encouraged readers to align their responses with the majority's sentiment. Thus, by displaying a majority of critical user comments for potential fake news articles, platforms can mitigate their further dispersion. This is especially crucial as literature attributes the rapid dispersion of fake news to human behaviour [6]. Alternatively, for confirmed real news articles, displaying a majority of supportive comments could encourage further dispersion.

In conclusion, comments posted by readers on social media news articles could have untapped potential to assist platforms mitigate the dispersion of fake news. Our work is an initial step towards understanding how supportive and critical user comments trigger conformity in how subsequent users perceive and respond to news articles on social media. Further work could compare our results to other approaches currently used by social media platforms (such as related articles and disclaimers) to determine their comparative effectiveness. Moreover, while we considered the number and the presentation order of critical and supportive comments appearing on posts, we did not consider the popularity of each comment (*i.e.*, the number of reactions and replies each comment received) which could also determine its influence. We encourage future work to explore these avenues to further expand our understanding on how user comments could be utilised to mitigate human-induced dispersion of fake news in social media.

# REFERENCES

- E. Shearer and K. E. Matsa, "News use across social media platforms 2018," 2018.
- H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of economic perspectives*, vol. 31, no. 2, pp. 211–236, 2017.
- H. J. Parkinson, "Click and elect: how fake news helped donald trump win a real election," *The Guardian*, vol. 14, 2016.
- H. S. Yang, "The effects of the opinion and quality of user postings on internet news readers' attitude toward the news issue," *Korean Journal of Journalism & Communication Studies*, vol. 52, no. 2, pp. 254–281, 2008.
- E.-J. Lee and Y. J. Jang, "What do others' reactions to news on internet portal sites tell us? effects of presentation format and readers' need for cognition on reality perception," *Communication research*, vol. 37, no. 6, pp. 825–846, 2010.
- S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.
- K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, 2017.
- T. Lyons, "Replacing disputed flags with related articles," 2017.
- J. Colliander, ""this is fake news": Investigating the role of conformity to other users' views when commenting on and spreading disinformation in social media," *Computers in Human Behavior*, vol. 97, pp. 202–215, 2019.
- S. Wijenayake, N. van Berkel, V. Kostakos, and J. Goncalves, "Impact of contextual and personal determinants on online social conformity," *Computers in Human Behavior*, vol. 108, pp. 106–302, 2020.
- 11. S. Wijenayake, N. van Berkel, V. Kostakos, and J. Goncalves, "Measuring the effects of gender on

online social conformity," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 145:1–145:24, 2019.

- S. Wijenayake, N. van Berkel, V. Kostakos, and J. Goncalves, "Quantifying the effect of social presence on online social conformity," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW, pp. 55:1–55:22, 2020.
- Facebook, "What does most relevant mean on a facebook page post?," 2020.
- E. Peer, J. Vosgerau, and A. Acquisti, "Reputation as a sufficient condition for data quality on amazon mechanical turk," *Behavior research methods*, vol. 46, no. 4, pp. 1023–1031, 2014.
- J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, and R. Tatham, *Multivariate Data Analysis*. New Jersey, NJ, USA: Pearson, 2010.

Senuri Wijenayake is a PhD candidate at the University of Melbourne, Australia. Her research interests include social computing and human-computer interaction. Senuri completed her B.Sc. in Information Technology at the University of Moratuwa, Sri Lanka. Contact her at swijenayake@student.unimelb.edu.au.

**Danula Hettiachchi** is a PhD candidate at the University of Melbourne and holds a B.Sc. in Computer Science and Engineering from the University of Moratuwa, Sri Lanka. His primary research interests are crowdsourcing and social computing. Contact him at danula.hettiachchi@unimelb.edu.au

**Simo Hosio** is an Associate Professor at the University of Oulu, Finland. His research interests include crowdsourcing, digital health, and ubiquitous computing. Simo received a PhD in Computer Science from the University of Oulu. Contact him at simo.hosio@oulu.fi

**Vassilis Kostakos** is a Professor of at the University of Melbourne, Australia. His research interests include ubiquitous computing, human-computer interaction, and social computing. Vassilis received a PhD in Computer Science from the University of Bath. Contact him at vassilis.kostakos@unimelb.edu.au.

Jorge Goncalves is a Senior Lecturer at the University of Melbourne, Australia. His research interests include crowdsourcing, online communities, and ubiquitous computing. Jorge received a PhD in Computer Science from the University of Oulu. Contact him at jorge.goncalves@unimelb.edu.au.