Spatial Subterfuge: An Experience Sampling Study To Predict Deceptive Location Disclosures

Shion Guha
Department of Information Science
Cornell University
Ithaca, NY 14853
sg648@cornell.edu

Stephen B Wicker
School of Electrical & Computer Engineering
Cornell University
Ithaca, NY 14853
sbw11@cornell.edu

ABSTRACT
Prior research shows that people often engage in deception when sharing location. Privacy concerns, social surveillance and impression management are the primary drivers of these types of behaviors. One methodological question that arises in this research context is the problem of reliable measurement to study predictors of deceptive location disclosure from usage data. In this note, we propose a simple experience sampling method (ESM) approach that is useful for studying this phenomenon. We describe our ESM deployment and report the results of a long term, quantitative study of 204 foursquare users over 1 year. Results indicate that physical distance, tie strength and order of visibility on the foursquare feed are significant predictors (with moderate to high effect sizes) of deceptive location disclosure. We connect these findings to the rich tradition of location disclosure behavior research in ubiquitous computing.

Author Keywords
foursquare;deception;ESM;impressions;visibility;privacy

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

INTRODUCTION AND MOTIVATION

Location sharing applications (LSAs) have proliferated [23] in the last decade. Many modern LSAs depend upon voluntary "check-ins" [6] as the primary mode of location disclosure. Previous work has outlined how impression management [6], privacy concerns [19] and social surveillance [11,16] play important roles in making the decision to disclose location voluntarily and that users often disclose false locations [10] as part of a holistic location sharing strategy. There is also a small but growing body of literature [14, 17] on deceptive location disclosure centered on understanding why and detecting how people would choose to engage in such behavior. Specifically, Zhang et al. [26] conducted a study comparing location traces from GPS/Wi-fi signals and foursquare check-ins and found some differences. They indicated that some of this disparity might be explained by deception arising from the gamification elements of foursquare. We argue that this is not just obvious but expected as part of a user's overall social activity [10,11,16]. In addition, there hasn't been significant progress on understanding factors (from automatically collected usage data) that affect such decision making. Morever, the methodological question around accurate and scalable collection of deceptive location disclosures remains to be investigated in greater depth.

Experience Sampling Method (ESM) has been a popular and often recommended choice [4, 13] in ubiquitous computing to measure phenomena traditionally intractable for appraisal. ESM has also gained traction in the rich tradition of location sharing user studies in ubiquitous computing research. Consolvo et al. [5] (n=16) studied the willingness to disclose location in a simulated LSA. They found that requestor identity, location granularity and the reason for requesting access to be significant factors predicting location disclosure. Khalil and Connelly [15] (n=20) analyzed how social relations and contextual clues can affect location sharing. Anthony et al. [1] (n=25) found that location-sharing decisions were often based on reasons beyond physical location. Finally, Patil et al. [18] (n=35) examined the impact of feedback and control on location sharing decisions through various LSA privacy settings.

Our study contributes to the above discussion by examining two research questions. First, can a simple ESM deployment be a useful way to capture deceptive location disclosure data? Second, what factors (from ESM and other automatically collected usage data) affect deceptive location sharing decisions? In the note that follows, we report the results of an ESM study on 204 foursquare users over a period of 12 months. Our main findings are:

First, using a straightforward ESM deployment within the foursquare application, we captured data regarding 312,467 check-ins over 12 months with 32,357 deceptive check-ins (9.7%). Second, the outcome of a binary, hierarchical logit model predicts **three** factors from foursquare usage data

UbiComp '15, September 7-11, 2015, Osaka, Japan.

ACM 978-1-4503-3574-4/15/09...$15.00.
http://dx.doi.org/10.1145/2750858.2804281

that significantly influence deceptive location disclosures on foursquare - physical distance between a user and a particular friend (OR=1.667, p<0.01), tie strength between a user and a specific friend (OR=0.556, p<0.05) and the order of visibility in the foursquare feed (OR=1.315, p<0.01). We also discuss how these findings connect and contribute to existing theoretical and empirical work on privacy, surveillance and impression management.

THE PRESENT STUDY

Research Context

We chose foursquare [7] as our sandbox for probing our research questions for two main reasons. First, foursquare is the largest and most popular LSA with over 55 million users and 6 billion check-ins [7] as of March 2015. We note here that in July 2014, the check-in mechanics were split off into another application named Swarm [22] while foursquare continued focusing on localized venue recommendations. This had no significant bearing on our study since all our users continued using Swarm and the APIs for ESM deployment and data collection remained constant for the study duration. Second, foursquare has no common privacy management features such as access control lists or location obfuscation settings. The only strategy is the decision to (not) check-in and is therefore, ideal for our research questions. A very exhaustive resource describing foursquare system features can be found freely online [9] for anyone wishing to (re)acquaint themselves with this LSA.

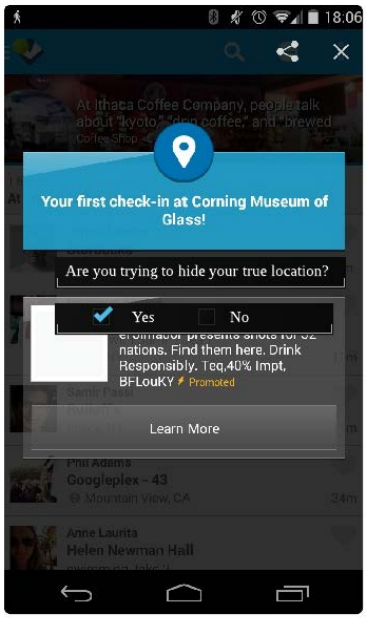

Figure 1. ESM prompt after every foursquare check-in

Experience Sampling Method

We used the foursquare API [8] to inject our custom application within the regular foursquare application. This manifested itself in the form of a prompt after every check-in as illustrated in Figure 1. Users were asked to report if they considered a specific act of location disclosure to be deceptive in nature (yes/no). We catalogued every choice as well as foursquare usage data summarized in Table 1.

Data Collection and Participants

We used a mixture of snowball and convenience sampling to recruit our participants. Our primary strategies were to reach out to our personal contacts as well as to propagate our study advertisement on popular social media (Facebook, Twitter etc.), our university campus (northeastern United States) noticeboards and internal mailing lists. This was a three-step process. First, participants clicked on a link that brought them to an online informed consent form that also outlined the objectives of the study. If they completed and submitted this form, they were redirected to an online survey asking them to answer brief demographical questions (name, email, gender, age, income and race), which we adapted from the US Census categories. Finally, they were again directed to a website to approve access to our custom application through the foursquare API and hence, by extension, their foursquare usage data.

Participants were not compensated for their effort but were asked to voluntarily contribute to research. This was a point of ancillary interest to us namely – would it be possible to deploy a long term ESM study without significant incentives or compensation? Our university's Institutional Review Board approved this study design.

We collected data about 204 foursquare users (99 male, 105 female) for a period of 12 months (January 2014 - December 2014). 75 of these users (~37%) were prior foursquare friends. We periodically prompted our participants via our custom application to rate their friendship with particular foursquare friends (on a scale of 1-10; 1 being the lowest and vice versa). This gave us an overall sense of tie strength about specific friends. Table 1 describes our participants in terms of relevant foursquare usage and demographics for the period of our study.

We did not find any significant variation in foursquare usage among our participants during our study when compared against their previous foursquare usage. In addition, we did not find any significant variations in self-reported deceptive check-ins among our participants. Neither could we discern any particular demographic factor that significantly differentiated more deceptive users from less deceptive users.

Variables	Min.	Max.	Mean	Std. Dev.
Total Check-ins	414	4815	1366	505.74
Avg. Check-ins / Day	1.13	13.19	3.74	1.24
Deceptive Check-ins	42	507	141	92.57
Deceptive Check-ins/Day	0.11	1.38	0.39	0.23
Avg. 'Likes' on check-ins	186	4985	1884	788.59
Avg. Tie Strength	1	10	7.2	1.78
No. of Friends	3	462	87	31.65
Age	20	68	32.4	4.73
Income(USD)	32000	180000	76500	24300.52

Table 1. Descriptive Statistics of Participants

Variables and Measurements

Here, we outline our different variables of interest in our regression model and describe their measurement.

Choice - This is our dependent variable and takes a binary value representing whether the user considered a particular check-in to be deceptive (1) or not deceptive (0) in nature.

UserVenue - This is a matrix with four columns. First, *user_vencat* represents the category (nominal) of the venue that the user is currently checked-in. Second, *avg_rating* is a continuous variable which represents the current rating (out of 10) provided by foursquare based on a proprietary algorithm which is an indicator of the overall popularity of that venue. Third, *tot_likes* is a count variable which represents the total number of "likes" that users have given a particular venue. Finally, *venue_checkins* is a count variable representing the total number of unique check-ins to that venue at the time of check-in by that particular user.

FriendMatrix - This is a matrix of a user's friends (*friend_id*) (nominal variable), their last check-in venue (*friend_venue*) (nominal variable), the timestamp of their last check-in *(friend_time),* the physical distance (in kilometers) from these venues to the user's venue *(friend_distance),* venue category (i.e. American Restaurant or Academic Building) (*friend_vencat*) (nominal variable) for each check-in. In addition, if a friend has disclosed their gender *(gender)* on their foursquare profile, we were able to capture this information as well.

FriendOrder - This is a matrix of the order (descending) in which a users' friends appeared on her check-in feed for each check-in consisting of two columns - *friend_id* (nominal variable) and *order* – an ordinal variable representing the descending order rank of each friend in a user's foursquare feed which they view whenever they choose to check-in to a location.

FriendTie - This is a continuous variable representing the self-reported tie strength between a user and a particular friend on a scale of 1-10. This gives us a broad idea of how strongly participants feel connected to specific friends.

Analysis

We wrote a script in R to fit a binary hierarchical logistic regression model adapted from the model proposed by Wong & Mason [25] to test our dependent variable *(Choice)* against our predictors. We fit different models with stepwise regression using all our predictor variables. The final model selection was done on the basis of a combination of Likelihood Ratio, AIC and BIC scores. Results of the final model are presented in Table 2.

Parameters	Variable Type	Std. β	Odds Ratio
intercept	*Continuous*	0.16	1.191
tot_likes	*Count*	0.07	0.271
friend_distance	*Continuous*	0.24	1.667**
order	*Ordinal*	-0.21	0.553*
gender	*Nominal*	0.09	0.162
FriendTie	*Continuous*	0.23	1.315**
AIC=2459	**BIC**=2386	*p <0.05 **p<0.01	

Table 2. Results of Selected Binary Logit Model

Our model indicates that there are three main factors predicting deceptive check-ins on foursquare. First, we see that the physical distance between venues that the user and a particular friend is checked-in at that moment acts as a significant predictor (OR=1.667,p<0.01) of deceptive check-ins. Second, the order in which friends appear on the foursquare check-in feed is also an important predictor (OR=0.556, p<0.05). Since our *order* variable was constructed in a descending order (farthest users first), we interpret results as being that visibility of *closest* users is an important predictor of deceptive location disclosure. Finally, the tie strength between that user and her visible (on the foursquare feed) friends is also a strong predictor (OR=1.315, p<0.01) of deceptive location sharing. We discuss these in greater detail in the next section.

RESULTS AND DISCUSSION

The findings in Table 2 contribute to the intersection of three bodies of literature. Briefly, user behavior in LSAs in the broader research community has been theorized from privacy [19], surveillance studies [11] and impression management [10]. Vision and visibility [11] are two oft-cited concepts in these literatures. The (potential) visibility of a user's location disclosure often leads to privacy concerns [10] about who might be able to see these disclosures. Conversely, in a friend network, social surveillance is driven by platform visibility and usage [11].

Moreover, visibility of a location disclosure is, at many times, used to cultivate favorable impressions [3, 10] to one's friend network by selectively disclosing location. Our results contribute to this conversation by implying that deceptive location disclosures take advantage of their potential visibility and constitute part of an overall strategy to manage these multiple interests and norms. We suggest that the decision to deceptively share location is one of several strategies undertaken by a user in conjunction with how they perceive their actual visibility on the social network. This is part of bag of strategies (others include non-use, limited use, obfuscation etc.) to manage different types of boundary negotiation processes like privacy and surveillance concerns and impression formation and management.

Specifically, our findings point towards an intuition that these are staccato-like decisions based upon real time factors. Guha and Birnholtz [10] introduced the concept of *check-in transience* - arguing that the last check-in matters more amidst all these different norms. Two of our predictive factors (physical distance and order of visibility) are indeed, very real time indicators. Users make a quick compromise to be disingenuous between their interests and concerns vis-à-vis who they can see on their foursquare feed quickly. Physical distance is also quite curious. We manually examined (at random) many of the venues where deceptive check-ins were reported and found two main geographical settings that might explain some of this phenomenon. In densely populated areas like Manhattan/San Francisco or college towns like Ann Arbor/Ithaca where foursquare is extensively used and where about 82% (167) of our participants are from, users are more likely to have friends living, working or commuting nearby. One of the original design goals of foursquare was to promote spontaneous social coordination of nearby friends and indeed; subsequent research [10] has confirmed this usage. Therefore, in practical daily foursquare use in these areas, you are more likely to see friends who are within a reasonable traveling distance. However, this doesn't mean that a user would *always* want a serendipitous encounter but might, for impression management reasons be loath to broadcast her reluctance. Deceptive location disclosure, especially to a venue farther away, then acts as a useful stratagem to navigate such tricky social waters. Reynolds et al. [20] theorized that such acts of deception are not just expected but also play a relational role due to the ubiquity of such communication platforms. Therefore, these acts of deceptive location sharing are not completely unexpected as personal boundary navigation mechanisms propagated to the spatio-digital realm.

Of course, we don't mean to imply that foursquare check-ins are always rife with deception given that the total proportion of deceptive check-ins in our study is approximately 9.7% of the total number of check-ins. However, even if we assume that 1 in every 10 foursquare check-ins involve some sort of deception, then this is still a sizable number, in line with other work [21] on social media deception.

We also found tie strength to be a significant predictor of deceptive location disclosure. Contrary to the other two factors described above, tie strength is usually not a momentary factor but accrues over socio-temporal interactions. Previous studies [6, 10, 11] have uncovered that the foursquare friend network generally tends to comprise stronger ties than other popular social media platforms e.g. Facebook, Google+. In addition, Wiese et al. [24] reported that tie strength remains the best predictor of location sharing. We theorize, along these lines that tie strength is also an important predictor of *deceptive* location sharing. Users maintain different levels of online and offline engagement with their audience. Given a quasi-public space such as foursquare and the competing, differential concerns of users, it is not unreasonable to assume that tie strength matters. For instance, a user might report high tie strength with her father but wouldn't necessarily want to disclose her actual location (or be deceptive about it) [10] all the time. She might have concerns that he might see her check-in to a venue that she wouldn't want him to know she was currently in. On the other hand, she might also report high tie strength with other friends with whom she might not want to engage in such illusion but may use different media (e.g. SMS, email etc.) in conjunction with LSA usage to confirm social coordination.

We would also like to comment on a methodological implication from this work. Our ESM deployment demonstrates that it is possible to run simple, long-term ESM studies. We followed Consolvo et al [4] as a guideline. However, follow-up interviews with participants are also necessary and important to capture their feelings about their participation in such studies. Moreover, this is a long-term study of deception and seems to be significantly longer than other major studies of deception [2, 12, 17, 20, 21] in social media. While each of these studies are important in their contribution to the area and many of them used ESM deployments as their primary lens of inquiry, they either used custom built applications or were limited in terms of sample size, statistical power and deployment time. We make the case that it is possible to use ESM in a low-cost manner given two important design considerations - low effort and simplicity for users and seamless background integration with existing user applications.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their constructive and detailed reviews. This project was supported by National Science Foundation Grant No.1016203.

REFERENCES

1. Anthony, D., Henderson, T., & Kotz, D. 2007. Privacy in location-aware computing environments. *IEEE Pervasive Computing*, 6(4). pp. 64-72.
2. Birnholtz, J., Guillory, J., Hancock, J., & Bazarova, N. 2010. On my way: Deceptive texting and interpersonal awareness narratives. In *Proc. CSCW'10*. pp. 1-4.
3. Boesen, J., Rode, J. A., & Mancini, C. 2010. The domestic panopticon: location tracking in families. In *Proc. Ubicomp'10*.pp. 65-74.
4. Consolvo, S., & Walker, M.2003. Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing,* 2(2), pp. 24-31.
5. Consolvo, S., Smith, I. E., Matthews, T., LaMarca, A., Tabert, J., & Powledge, P.2005. Location disclosure to social relations: why, when, & what people want to share. In *Proc. CHI'05*.pp. 81-90.
6. Cramer, H., Rost, M., & Holmquist, L. E.2011. Performing a check-in: emerging practices, norms and 'conflicts' in location-sharing using foursquare. In *Proc. MobileHCI'11*. pp. 57-66.
7. foursquare. URL: https://foursquare.com/about/
8. foursquare API. URL: https://developer.foursquare.com/
9. foursquare platform feature descriptions. URL:https://support.foursquare.com/hc/en-us
10. Guha, S., & Birnholtz, J.2013. Can you see me now?: location, visibility and the management of impressions on foursquare. In *Proc. MobileHCI'13*. pp. 183-192.
11. Guha, S., & Wicker, S. B. 2015. Do Birds of a Feather Watch Each Other? Homophily and Social Surveillance in Location Based Social Networks. In *Proc. CSCW'15*. pp. 1010-1020.
12. Hancock, J., Birnholtz, J., Bazarova, N., Guillory, J., Perlin, J., & Amos, B. 2009. Butler lies: awareness, deception and design. In *Proc. CHI'09*. pp. 517-526.
13. Hsieh, G., Li, I., Dey, A., Forlizzi, J., & Hudson, S. E. 2008. Using visualizations to increase compliance in experience sampling. In *Proc. Ubicomp*. pp. 164-167.
14. Iachello, G., Smith, I., Consolvo, S., Abowd, G. D., Hughes, J., Howard, J., Potter,F., Scott,J., Sohn,T., Hightower,J., & LaMarca, A. 2005. Control, deception, and communication: Evaluating the deployment of a location-enhanced messaging service. In *UbiComp 2005*. pp. 213-231.
15. Khalil, A., & Connelly, K. 2006. Context-aware telephony: privacy preferences and sharing patterns. In *Proc. CSCW'06*. pp. 469-478.
16. Page, X., Knijnenburg, B. P., & Kobsa, A. 2013. Fyi: communication style preferences underlie differences in location-sharing adoption and usage. *In Proc. Ubicomp'13*. pp. 153-162.
17. Page, X., Knijnenburg, B. P., & Kobsa, A. 2013. What a tangled web we weave: lying backfires in location-sharing social media. In *Proc. CSCW'13*. pp. 273-284.
18. Patil, S., Schlegel, R., Kapadia, A., & Lee, A. J. 2014. Reflection or action?: how feedback and control affect location sharing decisions. In *Proc. CHI'14*. pp. 101-110.
19. Pontes, T., Vasconcelos, M., Almeida, J., Kumaraguru, P., & Almeida, V.2012. We know where you live: privacy characterization of foursquare behavior. In *Proc. Ubicomp'12*. pp. 898-905.
20. Reynolds, L., Smith, M. E., Birnholtz, J. P., & Hancock, J. T. 2013.Butler lies from both sides: actions and perceptions of unavailability management in texting. In *Proc. CSCW'13*. pp. 769-778.
21. Smith, M. E., Hancock, J. T., Reynolds, L., & Birnholtz, J.2014. Everyday deception or a few prolific liars? The prevalence of lies in text messaging. *Computers in Human Behavior*, 41, pp. 220-227.
22. Swarm. https://www.swarmapp.com/
23. Wicker, S. B. 2013. *Cellular Convergence and the Death of Privacy*. Oxford University Press.
24. Wiese, J., Kelley, P. G., Cranor, L. F., Dabbish, L., Hong, J. I., & Zimmerman, J. 2011.Are you close with me? are you nearby?: investigating social groups, closeness, and willingness to share. In *Proc. Ubicomp'11*.pp. 197-206.
25. Wong, G. Y., & Mason, W. M. 1985. The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, 80(391), pp. 513-524.
26. Zhang,Z.,Zhou,L.,Zhao,X.,Wang,G.,Su,Y.,Metzger,M, Zheng,H., & Zhao,Y.B. 2013. On the validity of geosocial mobility traces. *Proc. HotNets.* Article 11, 7 pages.