Predicting Mobile Advertising Response Using Consumer Co-Location Networks

Peter Pal Zubcsek and Zsolt Katona and Miklos Sarvary[1]

February 9, 2017

[1]Peter Pal Zubcsek is Senior Lecturer of Marketing at the Coller School of Management, Tel Aviv University, Israel. Zsolt Katona is Associate Professor of Marketing at the Haas School of Business, University of California at Berkeley, Berkeley, CA 94720. Miklos Sarvary is the Carson Family Professor of Business at Columbia Business School, New York, NY 10027. E-mail: peterz@tauex.tau.ac.il, zskatona@berkeley.edu, miklos.sarvary@columbia.edu. The authors wish to thank the following for their invaluable comments and suggestions: the editors and three anonymous reviewers at JM, Yakov Bart, Eyal Biyalogorsky, Shai Danziger, Jeff Inman, Barak Libai, Xueming Luo, Dominik Molitor, Venky Shankar, Martin Spann, Andrew Stephen, Christophe Van den Bulte, and seminar participants at Marketing Science 2013, Mapping Mobile @NYU Stern, MSI conference on Marketing in a Multi-Channel and Multi-Screen World, the 12th ZEW Conference on Economics of ICT, AMA Winter Educators 2015, ACR 2015, University of Florida, University of Pittsburgh, IDC Herzliya, and Tel Aviv University.

Predicting Mobile Advertising Response Using Consumer Co-Location Networks

Abstract

Building on results from economics and consumer behavior, the authors theorize that consumers' movement patterns are informative of their product preferences, and propose that marketers monetize this information using dynamic networks that capture co-location events (when consumers appear at the same place at approximately the same time). To support this theory, the authors study mobile advertising response in a panel of 217 subscribers. The dataset spans over three months during which participants were sent m-coupons from retailers in various product categories via a smartphone application. The data contain coupon conversions, demographic and psychographic information, and information both on the hourly GPS location of participants and on their social ties in the form of referrals. The authors find a significant positive relationship between co-located consumers' response to coupons in the same product category. In addition, they show that incorporating consumers' location information can increase the accuracy of predicting the most likely conversions by 19%. These findings have important practical implications for marketers engaging in the fast growing location-based mobile advertising industry.

Keywords: mobile commerce, mobile targeting, location-based advertising, price promotion, network analysis

2013 marked the first year when Americans spent more time consuming digital media (online and mobile) than TV (eMarketer 2013), and consumers' use of their mobile devices has continued to increase at a rapid pace, reaching almost three hours a day in 2015 (eMarketer 2015a). This shift of consumer attention to the mobile channel has also triggered changes to marketers' allocation of their advertising budgets: worldwide mobile ad spending is set to grow from $19.2 Billion in 2013 to over $100 Billion in 2016 (eMarketer 2015b), with mobile ads taking up the majority of all digital advertising spending globally for the first time.

The proliferation of smartphone and tablet devices has led to entirely new types of interactions between consumers and marketers: Mobile technology does not only enable consumers to access digital content on the go – it also allows marketers to collect information on consumers' location and target their advertising based on these data (Luo et al. 2014, Shankar et al. 2010). Tapping into this opportunity, a large number of apps in the mobile advertising ecosystem collect subscribers' location: a recent study found that the apps on a typical smartphone user's mobile device collectively report the subscriber's location up to several thousand times per week (Almuhimedi et al. 2015). After an app obtains consent to access a user's location, it periodically collects the user's geographic coordinates and transmits them to a server. From these data, marketers are able to construct consumers' dynamic location profiles that can in turn be used to enhance the customization and targeting of the marketers' offers.

Traditionally, marketers' knowledge of consumers' location was limited to their place of residence. The main uses of such static location data were to partition consumers' locations into contiguous geographic regions representing market segments, or to consider spatially correlated preferences based on the proximity of consumers (due to either location-specific shocks or the nature of observational word-of-mouth, e.g. Bell and Song (2007), Yang and Allenby (2003)). Early attempts to monetize mobile location data applied a similar logic, dynamically mapping consumers into segments characterized by a static partitioning of geographic locations. The best known such approach is geo-fencing – sending consumers promotional offers when they enter the vicinity of the retailer (Jagoe 2003, Schiller and Voisard 2004).

Targeting mobile coupons based on store distance has indeed been confirmed to be quite effective in certain contexts (Danaher et al. 2015, Molitor et al. 2015). However, Luo et al. (2014) showed that store distance effects are only prevalent for same-day offers, and that next-day coupons are in fact more effective in non-proximal targeting. This is good news for marketers since longer offer validity allows them to cast a wider web and target more prospects. However, it remains unclear whether collecting consumers' location data may improve the targeting of such longer-validity offerings beyond the extent that can be based on consumers' past behavior (Reinartz and Kumar 2003, Rossi, McCulloch, and Allenby 1996). On the other hand, it is known that collecting sensitive information such as location data may reduce prospects' willingness to opt in to receive advertising (Banerjee and Dholakia 2008, Barkhuus and Dey 2003).

Here, we present a new general method to improve the dynamic segmentation of consumers based on their past response to marketing activities and their location history. Our work builds on theoretical advancements in economics and consumer behavior suggesting that consumers' location choices may be indicative of their product preferences (Bettman, Luce, and Payne 1998, McFadden 2001). In particular, we propose that consumers who attend the same venues may exhibit commonalities in their taste. To test this theory, we construct a dynamically evolving network of co-location events – events when two or more participants are at the same place at the same time. Our work combines the technique employed by Bell and Song (2007) who created a static network of consumers through clustering them by their ZIP codes, and the dynamic methods proposed by Hui, Fader, and Bradlow (2009) to study consumers' movement patterns on a map.

Specifically, we propose that the likelihood of a consumer's undertaking of a particular activity (e.g., redeeming a coupon in a particular product category) is positively related not only to their past rate of engaging in the same activity but also to the corresponding past rate of each of their co-location network neighbors, i.e. those consumers whose mobility trace overlapped with that of the focal consumer. We note that our goal is not to separate the economic and behavioral explanations linking consumers' location behavior to their product choices. Rather, in the spirit of Leone and Schultz (1980) and Bass (1995), we document an important empirical pattern that is relevant for subsequent theory building.

We empirically test the above proposition on data from a pilot program of a mobile operator that, over several months, provided smartphone user participants (independently from their current or past locations) with digital coupons in four product categories, while collecting their location data up to a few dozen times each day. For every coupon offered in the program, we compute the co-location network corresponding to the day before the offer was launched. We then simultaneously estimate the effects of network position in the co-location network, referral network effects, and the impact of demo- and psychographics on coupon redemption in a fixed-effects logit model. Our key empirical findings include the following:

- We discover a significant positive link between the co-location of consumers and their response to coupon-based promotions, even after controlling for coupon characteristics, demographic and psychographic differences, referral network effects and unobserved individual heterogeneity. This suggests that consumers who frequent the same locations indeed have correlated preferences (even if they do not know each other). Constructing dynamic networks from co-location events appears to be an effective method to capture such preference similarities, and thereby enhance the segmentation and targeting efforts of marketers.
- Whereas consumer-level demographic or psychographic variables may also be used to uncover correlations between consumers' preferences, we find that variables derived from consumers' location history may be more effective predictors of their purchase behavior than traditional variables. In particular, we perform out-of-sample estimations to show an approximately 19%

increase of prediction accuracy at identifying the most valuable prospects. Given the high opportunity cost of inaccurately targeted ads – overloading consumers with ads might cause them to opt out from receiving *any* ads in the future – this improvement of targeting accuracy is of utmost importance to managers.

These are fundamental results confirming that location history can be effectively used to assess consumer preferences and improve the targeting of next-day mobile coupons. Therefore, our method may successfully complement current location-based advertising methods, which typically rely on the geo-fencing approach. More importantly, our work highlights the opportunity in capturing the dynamic interdependencies in prospects' location behavior. In today's world, where rich mobility data on consumers is abundant, marketers must look beyond methods that attempt to monetize consumer location data using only a static paradigm.

Location Data and Consumer Choice

For most of the 20th century, location-aware empirical methods were constrained by the scarce availability of data. In the early days of marketing, most firms could at best observe data on consumer activities (e.g., sales) aggregated at the store level. The introduction of individual loyalty cards and credit cards (Guadagni and Little 1983) made individual-level observation possible, but beyond loyal customers' residential address, most marketers could still only learn consumers' precise location upon their visit to a brick-and-mortar touch point (e.g., retail store) of the brand. Consequently, early applications of spatial models in marketing focused on grouping locations into geographical market segments (Anderson and De Palma 1988, Hofstede, Steenkamp, and Wedel 1999, Huff 1964, Zoltners and Sinha 1983), and the implications of the segment boundaries to the optimal pricing, promotion, and distribution strategies of the firm (Bronnenberg 2005, Bronnenberg and Mahajan 2001, Greenhut 1981). More recently, the impact of consumers' relative location to stores on online purchases has also been examined within the above paradigm. For example, Forman, Ghose, and Goldfarb (2009) demonstrated that when a store opens locally, people substitute away from online purchasing. Similarly, Ghose, Goldfarb, and Han (2012) found that Internet users accessing a microblog were more likely to browse for stores in their vicinity. [1]

Location Data in Mobile Marketing

The close to ubiquitous adoption of smartphones and tablets has dramatically improved marketers' ability to collect rich location data on consumers. To monetize the location data collected via smart devices, marketers initially employed methods reminiscent of traditional location-aware marketing models. In the most popular application of location data, geo-fencing, prospects are targeted with promotional offers when they enter the vicinity of the retailer (Reyck and Degraeve 2003, Schiller and Voisard 2004). In essence, this approach creates a dynamic segmentation of consumers based on one static partitioning of all physical locations in the market (where the surroundings of each store correspond to a specific partition). Confirming the results known from the literature on the redemption of non-digital coupons (Chiou-Wei and Inman 2008), many papers have independently demonstrated the increased promotion-sensitivity of mobile consumers closer to the physical location of the store (Luo et al. (2014), Fong, Fang, and Luo (2015), Danaher et al. (2015); see Andrews et al. (2016) for a comprehensive review). In the context of location-aware pull advertising, wherein consumers search for available offers in their vicinity, the results are similar (Ghose, Goldfarb, and Han 2012, Molitor et al. 2015).

These results are not without caveats. Combining field experiments (conducted at grocery retailers) with simulations, Hui et al. (2013) highlighted that once the marketer accounts for unplanned purchases, mobile advertising that requires shoppers to travel further away from their planned path may be more profitable than offers for an unplanned category near shoppers' planned path. Furthermore, both Luo et al. (2014) and Fong, Fang, and Luo (2015) showed that store-proximity effects may wear out by the day after the coupon is delivered. This may indicate that most consumers plan their shopping trip in advance (Dellaert et al. 1998, Popkowski Leszczyc, Sinha, and Sahgal 2004), or that consumers taking care of business unrelated to shopping find mobile advertising intrusive (Shankar et al. 2010). Making consumers opt-in to mobile promotions (Barwise and Strong 2002, Danaher et al. 2015) dampens the intrusiveness of mobile advertising but even so, it is still the case that a store engaging in geo-fencing is unable to send promotional messages to prospects whose typical movement patterns do not intersect the fence. Put differently, targeting based on distance should be more effective when consumers have already revealed their shopping intentions (Hui, Bradlow, and Fader 2009, Hui et al. 2013).

Dynamic Location Behavior and Consumer Preferences

The abovementioned models take location data as input to predict consumers' likelihood of choosing from a set of options (which may, for product purchase behavior, include the outside option) that is the same for the entire population of consumers. However, marketers are often also interested in understanding consumers' location choices (Bradlow et al. 2005).

The assumptions of standard economic theory state that each location choice is made to the end of maximizing some underlying utility function (McFadden 2001). Under this paradigm, consumers' choices of location *reveal* their (static) underlying preferences, which in turn may allow marketers to better predict their future choices. For instance, Bellovin et al. (2013) and de Montjoye et al. (2013) demonstrated how even the aggregation of noisy location data (collected from cell towers) could be used to predict individuals' demographic attributes and frequently-attended future locations.

On the other hand, the literature on constructive consumer choice (Gregory, Lichtenstein, and Slovic 1993) argues that, instead of well-articulated preferences, location choices may reveal situational factors such as the activated goals of the decision-maker (Bettman, Luce, and Payne 1998). Moreover, exposure to particular contextual factors present at the consumer's chosen locations (Kirchner et al. 2012) may influence the accessibility of their preferences, ultimately affecting their future choice behavior in non-locational domains (Berger and Fitzsimons 2008, Feldman and Lynch 1988). Examples for such context effects in mobile marketing include the effect of weather on advertising response (Li et al. 2015), the differential effect of being exposed to ads at home or at work (Reinaker et al. 2015), and the effect documenting higher ad response from passengers traveling on more (vs. less) crowded trains (Andrews et al. 2015).

Importantly, the approaches of both the economics and consumer behavior literatures acknowledge the possibility that consumers' dynamic location choices carry information about their (possibly dynamically changing) product preferences. When the locational context (e.g., points-of-interest) at the place of co-location is known to the marketer, it is indeed straightforward to link consumers' locations to their subsequent product choices. However, even in the absence of such contextual information, marketers may be able to monetize the information captured in consumers' location histories. To demonstrate this, below we present a novel method that uses networks to capture commonalities between consumers' preferences.

Using Networks to Predict Consumer Behavior

Besides the emergence of rich data on consumer locations, the technological advancements of the 21st century have also led to the emergence of large databases on the structure of consumer interactions, e.g. online friendship networks, telecommunication networks, or networks based on financial transactions (Sundararajan et al. 2013). Marketers had long recognized the powerful impact of word-of-mouth on consumer choice (Bass 1969, Brooks 1957, Herr, Kardes, and Kim 1991), and researchers soon showed how micro-level data can enhance models of product or service adoption (Katona, Zubcsek, and Sarvary 2011, Manchanda, Xie, and Youn 2008) and disadoption (Nitzan and Libai 2011).

Much of the empirical social network literature has been focusing on empirically separating social influence from the inherent similarity (or homophily, McPherson, Smith-Lovin, and Cook (2001)) of connected individuals in the network (Durlauf and Ioannides 2010, Iyengar, Van den Bulte, and Valente 2011), which is an often difficult task due to the reflection problem (Aral, Muchnik, and Sundararajan 2009, Manski 1993). However, in the typical scenario when the marketer cannot control the (mostly static) structure of the consumer network, the agnostic approach of naïvely quantifying the relationship between network characteristics and consumer choice (Hill, Benton, and Van den Bulte 2013, Reingen et al. 1984) can provide a major contribution to the firm's bottom line. In a recent example, Goel and Goldstein (2014) took this route and showed how incorporating the structure of communication links from a very large instant messaging network can improve the accuracy of predictions for a variety of consumer choice behaviors including response to advertising or retail purchases.

In the same spirit, we propose a novel method to monetize location data. In particular, we propose to use network theory for characterizing commonalities between consumer location patterns, and use the so-derived network characteristics to improve the prediction accuracy of traditional consumer choice models. We present our proposed approach next.

Predicting Consumer Choice Using A Network of Co-Locations

Traditionally, marketers used to assess consumer choice likelihood using demographic variables and data on consumers' past purchase behavior (Rossi, McCulloch, and Allenby 1996). However, the literatures linking location behavior to consumer preferences suggest that consumers' dynamic location choices may also be indicative of their product preferences. [2] In particular, to the extent that consumers' location choices are indicative of their preferences in a product category, the choices of individuals who often attend the same venues may exhibit similarities. To quantify these similarities, we take an approach reminiscent of Goel and Goldstein (2014). Specifically, we propose to construct a dynamic network to capture the correlations of consumer preferences reflected by co-location events – when consumers are at the same place at approximately the same time. Furthermore, we propose that the likelihood of a consumer's engaging in a particular activity (e.g., redeeming a coupon in a particular product category) is then positively related not only to their past rate of engaging in the same activity but also to the corresponding past rate of each of their co-location network neighbors, i.e. those consumers whose mobility trace recently overlapped with that of the focal consumer.

Thus, our proposed method is a form of naïve collaborative filtering (Herlocker, Konstan, and Riedl 2002) based on co-location patterns.[3] It presents a novel means of dynamically mapping geographic locations to consumer segments, wherein the role of specific locations may exhibit variation both across time and, via others' location choices, across consumers. In this regard, we improve over Andrews et al. (2015) and Hui, Bradlow, and Fader (2009) by considering multiple location observations to determine co-location between consumers. Table 1 contrasts our approach to existing work that considers dynamic location information to predict consumer choice.

Importantly, our method does not aim to disentangle the economic and behavioral explanations linking consumers' location behavior to their product choices. To identify specific drivers of consumer choice, one would have to run causal experiments (Cooke and Zubcsek 2017, Lurie et al. 2016). Rather, in the spirit of Leone and Schultz (1980) and Bass (1995), we document an important empirical pattern that is relevant for later theory building. We proceed with an empirical test of our approach on a location-enhanced dataset of mobile coupon response.

[– Insert Table 1 around here –]

Application: Improving Mobile Coupon Targeting Via Subscriber Co-Location Networks

Smartphone app-based advertising platforms provide a natural context to empirically test our modeling approach. Relying on the GPS sensors in participants' mobile devices, these platforms are capable of collecting more precise consumer location data than cell tower-based methods, also in areas that are not restricted to the vicinity of participating retailers' store locations. In addition, the reach of the smartphone channel has grown past 2 Billion consumers (eMarketer 2016). Mobile devices are becoming consumers' personal companions (Shankar and Balasubramanian 2009) – a trend witnessed by the increasing amount of time consumers spend on their mobile devices (eMarketer 2015a). Marketers' reallocation of advertising budgets from the online to the mobile channel (eMarketer 2015b) highlights the importance of further studying this application area.

In-app display ads constitute a rapidly emerging format of mobile advertising (Bart, Stephen, and Sarvary 2014, Grewal et al. 2016, Shankar et al. 2010). Whereas the display advertising category is often thought of as just banner and poster ads on the mobile web, it increasingly also includes other formats, such as video ads, sponsored stories in social media newsfeeds and rich media advertising. Recent research has demonstrated the effectiveness of these marketing vehicles in a variety of contexts (Burns and Lutz 2006, Katz 2014, Rosenkrans 2009).[4]

The most advanced in-app display advertising techniques attempt to leverage the targeting opportunities arising from the high level of portability and personalness of mobile devices (Barwise and Strong 2002, Ghose and Han 2011, Shankar and Balasubramanian 2009). Inter alia, data on online consumer behavior can be complemented with rich data on offline consumer behavior (e.g., time and location data), and offline purchase incentives can be delivered in an online way (e.g. mobile coupons or *m-coupons*).

Our paper falls into the rapidly emerging topic in mobile advertising that closely builds on the mobile nature of the advertising medium, and studies the impact of subscriber location on mobile coupon redemption (Danaher et al. 2015, Luo et al. 2014, Molitor et al. 2015). Importantly, seminal work on mobile coupons (Banerjee et al. 2011, Danaher et al. 2015, Dickinger and Kleijnen 2008, Fong, Fang, and Luo 2015, Luo et al. 2014, Reichhart, Pescher, and Spann 2013) has presented findings and coupon redemption rates consistent with the literature that analyzed traditional coupon formats (Bawa and Shoemaker 1987, Inman and McAlister 1994, Lichtenstein, Netemeyer, and Burton 1990). This suggests that, pending data availability, better understanding the link between consumer location choices and offer redemption behavior in the context of mobile coupons may also carry managerial relevance for marketers using traditional coupon formats.

Data

Our data came from a mobile operator in a Pacific country. The operator ran a pilot program in two major metropolitan areas to study consumer response to in-app mobile advertising from January to early April of 2012. Subscribers invited to the pilot program – either by the operator or by one of the program members – had to respond to a short demographic survey and were asked to install a new app on their smartphone. During the program, the operator distributed digital coupons (*offers*) from participating retailers (to every participant currently in the program). In practice, for each offer, the mobile app received the coupon information from the server and displayed a notification on participants' smartphones. Upon clicking through the notification, the app asked a lead-in question indicating the category of the offer like "Are you interested in an offer for groceries?" At this point, participants could either discard the offer or continue to watch more details including the participating brand(s) and – if applicable – product(s),[5] the discount value of the coupon and the "when-you-spend" amount.[6] After viewing these details, participants could hit accept anytime during the validity period of the offer and receive the in-store discount after making the required purchase and showing the accepted coupon. Unfortunately, the exact time of redemption was not recorded. We only have data on whether and how a participant responded to any given offer over the duration of its validity. Our panel contains a total of 15,353 observations on 96 offers sent to 217 participants. The panel is incomplete, however, because some participants signed up after certain offers had already expired (registration was open throughout the period of the pilot).

Participants signing up for the pilot also agreed to have their GPS location information regularly transmitted to the mobile operator by the app. However, location data could only be captured this way when both the location services were enabled on participants' smartphones and the device was able to detect the necessary satellite signals. Whereas the app was set up to submit location information about four times every hour, the actual rate of transmission was lower. To correct for this dispersion and balance the data, in the dataset that the operator provided, the coordinates of all observations within the same hour are averaged, resulting in up to one observation per hour. For each hour during which there was no successful transmission, there is no location observation in the dataset. We have an average of 11.29 hourly location observations per day per participant in our panel. However, for about 40% of the observations, there is no location information on the participant.

The dataset does not contain information about potential Points-of-Interest (POI) in the geographic region where the study was conducted. Obvious POIs would include the stores that participated in the coupon program, yet unfortunately, the operator collecting the data lacked this information for the overwhelming majority of advertisers. Moreover, in some cases, the coupons were valid for a whole network of stores (including several dozen store locations in town), while other campaigns were specific to a particular store, making store-location data difficult to work with. We are also missing information about subjects' home address, which was not collected for confidentiality reasons. To gauge the impact of these data issues, we conducted a variety of robustness analyses. For instance, in one of our validity tests, we attempted to infer participants' home location from the GPS data to make sure that co-location did not simply reflect co-habitation (e.g., membership in the same family).

Methods

We modeled consumers' offer redemption behavior using a logistic regression. To control for unobserved heterogeneity, we included offer fixed effects and participant random effects in our main models. Estimating our models with offer fixed effects was a natural choice for two reasons. First, the individual offers in the program were qualitatively very different, and henceforth the baseline response rate also exhibited a large variation. Second, the cardinality of offers is much lower than that of participants in our panel. Concerning participant random effects, we note that some participants did not redeem any offer during the period studied, so we could only include fixed effects at the cost of throwing away all the observations for these participants.

We formally specified our model as

$$\log \frac{\Pr(Y_{ij}=1)}{1-\Pr(Y_{ij}=1)} = \beta_0 + \beta_1 \mathbf{W_i} + \beta_2 \mathbf{X_{ij}} + \beta_3 \mathbf{Z_{ij}} + \xi_i + \eta_j, \tag{1}$$

where Y_{ij} is an indicator of consumer i's response to offer j. If consumer i redeemed offer j, then $Y_{ij} = 1$, and 0 otherwise. $\mathbf{W_i}$ is a set of covariates measuring consumers' individual characteristics, $\mathbf{X_{ij}}$ are variables that are both consumer- and offer-specific (e.g. the length of time an offer was available for a particular consumer). $\xi_i \sim N(0, \rho^2)$ represent participant random effects and η_j represent offer-specific fixed effects. Finally, $\mathbf{Z_{ij}}$ are variables that capture the network effects between pairs of consumers for a given offer. Central to our interest is the network based on consumers' co-location. In addition, to control for the otherwise unobserved similarity of consumers, we also constructed a network connecting consumers based on whether one has referred the other to the program. Table 2 describes the variables used in our main regressions. We detail these below.

[– Insert Table 2 around here –]

Consumer-level and Offer-level Covariates ($\mathbf{W_i}$ and $\mathbf{X_{ij}}$)

The data contain numerous demographic and psychographic variables. Upon signing up, participants had to answer 24 profile questions. Dropping all profile variables that had more than two possible answers and did not translate to an ordinal scale left us with 9 profile variables (see Table 2). In addition, for each consumer-offer pair, we included three other variables. First, we included the number of days the offer could be used by the participant (Offer_Length). (This only varied for participants who joined the program while the particular offer was available.) Second, we added the number of days the participant had been in the program prior to the first day that the offer was made available (Days_Since_Joined). Finally, we added Category_Redeem_Rate, the rate at which the participants had been responding to prior offers within the category of the offer – consumer packaged goods, food & beverages, retail, and recreation – considered. Table 3 reports the summary statistics for these variables per offer category. Whereas it is straightforward to also calculate overall response rates, due to the high (.41) correlation between the overall and within-category response rates, we decided to include only the category-specific variable in our model.

[– Insert Table 3 around here –]

Network Variables ($\mathbf{Z_{ij}}$)

To study how similarities between individuals' location behaviors may correspond to similarities in coupon redemption, while controlling for the similarity of socially connected individuals, we introduced variables derived from two participant networks, co-location and referral. In both cases, our network-based variables were constructed in two steps. We first constructed a dynamic network between consumers, then we defined a neighborhood effect for each offer based on the network structure a day before the launch of the offer and neighboring consumers' response to prior offers in the same category.

We defined both networks as time-varying undirected simple graphs over participants as nodes. In the co-location network (denoted by the matrix $\mathbf{C_j}$), two participants were connected to each other if they had been at the "same location" according to at least one of the hourly GPS observations during the last day preceding the launch of offer j. The GPS coordinates of the hourly location observations (where available) gave us the latitude and longitude of each participant. For computational reasons, we used this information to determine co-location the following way. We defined a rectangular grid of 0.002° spacing and determined which grid cell each observation belonged to. (The grid spacing was chosen to be approximately equal to the lowest reported GPS signal accuracy of 250m.) Two participants were then considered to be co-located when their observation for the same hour fell into the same grid cell (cf. Figure 1). The co-location matrix was thus formally defined as

$$C_j(i,h) = \begin{cases} 1 & \text{if consumers } i \text{ and } h \text{ were } \textit{simultaneously} \text{ recorded in the same GPS grid cell at least once during the day preceding the launch of offer } j\text{, and} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

[– Insert Figure 1 around here –]

A positive effect of co-location would indicate the similarity of the co-located consumers' preferences. Importantly, however, Pan, Aharony, and Pentland (2011) point out that such an effect may be driven by the presence of social relationships between most co-located actors. In our main model, we attempted to control for such relationships both via participant random effects, and by means of adding information about referrals between participants to the model.

To control for referrals, we considered the referral network, denoted by $\mathbf{R_j}$. This network was also defined as a time-varying simple network, wherein two participants were related if one of them had invited the other to the program. (Albeit referral itself assumes asymmetric roles, we defined an undirected referral network to capture not just word-of-mouth influence, but any correlation between the advertising response behavior of two linked participants.) Formally,

$$R_j(i,h) = \begin{cases} 1 & \text{if both consumers } i \text{ and } h \text{ joined the pilot before the launch of offer } j \text{ and one referred the other to the program, and} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Note that the variation in the referral network was monotonic in time, i.e. the network kept growing since for each referral relationship in our data, we included the network link starting from the day when the second of the two participants joined the pilot.

The final step was to construct the $\mathbf{Z_{ij}}$ neighborhood effect variables for each offer in each network. In line with Goel and Goldstein (2014) and Provost, Martens, and Murray (2015), we expected that a participant was more likely to engage in a certain behavior if they were connected to someone who had previously engaged in that behavior. To capture this phenomenon, in both networks, for each participant and offer, we took the sum of the within-category response rates of all network neighbors prior to the launch of the offer, and included zeroes where this was not applicable. Formally, for offer j, in the $\mathbf{C_j}$ graph,

$$Z_{ij} \ni \text{Colocation\_Category\_Redeem\_Rate}_{i,j} = \begin{cases} \sum\limits_{h|C_j(i,h)=1} \text{Category\_Redeem\_Rate}_{h,j} & \text{if participant } i \text{ had at least one neighbor in the co-location network corresponding to the launch of offer } j\text{, and} \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

For the reference network, we defined the variable Referral_Categ_Redeem_Rate$_{i,j}$ in the exact same way using the $\mathbf{R_j}$ network instead of $\mathbf{C_j}$.

We predicted that co-location relationships would correspond to correlated coupon redemption rates. The network variables constructed above provided a simple way to capture this main effect and also to control for the effect of referrals. Importantly, however, the co-location variable contained positive values for less than 20% of the program participants per offer. Furthermore, we needed to make sure that positive values of *Colocation_Category_Redeem_Rate* were not simply more prevalent for participants whose location was observed more often. Therefore, in our regressions we controlled for *GPS activity*, the number of nonzero location observations available to us during the period wherein co-location was studied (i.e. the day preceding the launch of the offer in question).

Results

The results of our estimations are presented in Table 4. Model 0 was our baseline model that estimated redemption likelihood in terms of only the simple offer-related variables. Model 1a added traditional predictor variables (demographics and psychographics). In contrast, Model 1b augmented Model 0 by incorporating the location variables. Models 2a and 2b extended Models 1a and 1b, respectively, by controlling for referrals. Model 3, the *full model*, included all demographic, psychographic, location, and referral variables on top of those included in Model 0. (Thus, Models 0-1a-2a-3 and 0-1b-2b-3 were nested in each other.)

[– Insert Table 4 around here –]

The effects on the category- and participant-specific variables show strong face validity and are generally consistent across all models. For instance, the data confirm the intuition that the longer an offer was available to a participant, the more likely that participant was to use it. Additionally, there appears to be weak evidence for that (after controlling for offer-specific mean response rates) the tenure of participants in the program increased their response rate. We speculate that this may be due to many less-adventurous participants joining the program toward its end – in case it took them some time to get familiar with the system, their number could lead to such an effect.

Our next set of observations focus on the effect of central interest: that of co-location. Models 1b, 2b, and 3 all estimated the co-location network effect, varying the degree of control via demographic variables and the referral network. Confirming our predictions, we found that co-location had a significant effect over and above those of other variables. Specifically, in all three models, Colocation_Category_Redeem_Rate was significant at the $p = .01$ level (Table 4). This evidences the positive relationship between the coupon redemption likelihood of participants who were co-located on the day prior to the launch of the offer.

Turning to the control variables, we first note that we found referral effects to be quite strong. This result, consistent with the results of Goel and Goldstein (2014), underscores the importance of controlling for participant heterogeneity via the referral network. Concerning demographics, the results provide consistent support for the pattern that female participants, and participants with more education had a higher coupon redemption rate. The results in Table 4 thus indicate that there were some younger (and consequently less educated) male participants whose engagement at redeeming offers was lower than those for the average user. A closer investigation revealed that these participants were generally less engaged with the program. In the Web Appendix, we report additional tests wherein we controlled for this respondent heterogeneity in multiple ways. We found our main results to be robust to alternative specifications.

Demo- and Psychographics versus Co-Location

The above findings demonstrate how predicting coupon redemption behavior becomes more accurate as co-location is incorporated in the model. These results are reminiscent of those in Goel and Goldstein (2014). Albeit only for age and gender, Goel and Goldstein also demonstrated (see their figures 6b and 7) that (social) network variables in the absence of demographic predictors may achieve better results than using only demographic information in the estimations. It is therefore natural to investigate the same issue in our data, and compare the performance of the location variables to that of demographics.

It is paramount to highlight that such questions carry tremendous importance to mobile marketers. For example, in many emerging markets, the vast majority of mobile subscriptions are prepaid (Castells 2007). In a prepaid customer relationship, the operator often does not possess reliable demographic information about the customer. However, location data may be captured either at the cell tower or at the GPS level (Bellovin et al. 2013), and CRM databases can store prior customer response to marketing offers.

In-Sample Model Fit

To compare the predictive power of demo- and psychographic variables to that of location variables, we first studied the fit of Models 0-3. All six models included the offer-specific variables as it is fair to assume that these would naturally be available to marketers. Further, in the models with co-location (Models 1b, 2b, and 3), we included the GPS Activity variable to ensure that "participant activity" did not contribute to any network effects identified by the estimation. Generally, adding the demo- and psychographic variables to a model improved its fit less than adding the location variables: LR(Model 1a, Model 0)=$\chi^2(9) = 16.76$, $p = .053$; LR(Model 3, Model 2b)=$\chi^2(9) = 14.69$, $p = .100$; LR(Model 1b, Model 0)=$\chi^2(2) = 24.50$, $p < .001$; LR(Model 3, Model 2a)=$\chi^2(2) = 18.89$, $p < .001$. Comparing the log-likelihoods of Models 1a and 2a to that of Models 1b and 2b, respectively (Table 4), we found that despite using fewer predictors, the model incorporating location variables achieved a higher model fit. In accordance with this, on both the AIC and BIC tests, (Table 4), Model 2b performed best.

Out-of-Sample Estimations

Next, we also performed out-of-sample estimations. First, we split the data in time such that about 75% of the observations be used as training data, and tested the accuracy of model predictions on the remaining 25% of observations. The ROC values were .7177 for Model 0, .7285 for Model 1a, .7453 for Model 1b, .7346 for Model 2a, .7443 for Model 2b, and .7469 for Model 3. None of the pairwise differences between the ROC values were significant at $p = .05$, which is not surprising given that, accounting for the offer fixed effects, even the baseline model included over 100 variables, to which we added a total of nine demographic and psychographic, and up to three location and referral variables.

It is important to note, however, that marketers should care more about the sensitivity than the specificity of the model. The reason for this is that in most mobile advertising platforms, consumers may limit their daily exposure to a small number of ads, thereby vastly increasing the opportunity cost of inaccurate targeting. To assess the specificity of the model at such low rates of targeting, we next devised and carried out a round robin test. Specifically, we randomly divided our observations into ten approximately equal sets and performed ten tests, in each of which nine of the sets were pooled to form the training data and the tenth (denoted by T in this paragraph) was used to evaluate the predictions of the models calibrated on the training data. In each test, we ranked the predicted probabilities and defined our prediction list as the list containing the m nodes with the highest predicted adoption probabilities (for various values of m). To test predictive power, we calculated the proportion of offer redemptions in the prediction list as a function of the fraction of observations included in our sample. (We expressed m as a fraction of $|T|$ to balance the slight variation of the number of observations across the ten evaluation datasets.) Finally, we averaged the ten success rate functions corresponding to different choices of the training set.

We carried out the round-robin prediction test for the six models from our main analysis. In line with our above argument, we focused on the success rate of the models when no more than 10% of the participants (based on their high estimated redemption probability) were predicted to redeem the offer. The average lift of Model 1b over Model 1a was 18.82% and 8.81% in the range of predicting up to 5% and 10% of redemptions, respectively. Controlling for referrals (Model 2b vs. Model 2a), the average improvement of prediction success was 19.09% and 9.57%, respectively. These differences evidence that location variables may be better predictors of the most likely coupon redemptions than demo- and psychographic variables. Since calculating the co-location variables is a fairly straightforward task even for firms with moderately sophisticated computational infrastructure, we believe our method can be transformed into an effective and practical tool to improve targeting accuracy.

Naturally, it is likely that a smoother definition of the networks used in the estimations could slightly alter the range in which the location-enhanced models outperform their "location-free" counterparts. For instance, including variables derived from a social network based on communication interactions (voice calls, SMS, etc.) would likely increase the discriminating power of our model over most participants, not only those few who were involved in referrals. Unfortunately, we do not possess such additional data. On the other hand, our location data are rich enough to allow for some flexibility in our approach. In the next sections, we explore various relaxations of the definition of co-location to study the trade-off between the amount of relationships in the so-derived network and the ability of these relationships to uniquely reflect common underlying shocks impacting the related consumers' behavior.

Eliminating Store Distance Effects

Mobile subscribers may be co-located for a wide variety of reasons. They may dine or shop at the same location, pass through the same public transportation hub, etc. Based on the rational and constructive consumer choice literatures, we theorized that such co-occurrences may reflect similar preferences. Importantly, however, our network definitions will also deem participants related if they live in the same household or work in the same office (building or district). Whereas such instances should clearly correspond to similar preferences, this also admits the possibility that our results are driven by the main effect identified in Molitor et al. (2015): If two people live in the same household (or next to each other), then any store may be at the same distance from them. (Recall that, as we do not know the exact time that offers were delivered at, we neither possess the location of participants at the time of coupon delivery.)

Including the referral network variables in our main models did account for such social effects. However, given that our referral network sample may only contain the strongest social relationships between participants, we performed further tests to study the impact of multiple participants in or near the same household or at the same workplace, respectively. In particular, for various cut-off thresholds $0.2 \leq F_i \leq 0.5$, we defined two participants as "co-habiting" if they were co-located at the same grid cell between 1am and 5am on at least an F_i fraction of the days that they spent in the program. (We note that this is consistent with the method of Reinaker et al. (2015), which corresponds to $F_i \geq 0.33$ in our approach.) For example, in the case of $F_i = 0.5$, this search returned two connected pairs of subscribers with no overlap between the pairs. Removing all four participants from our data decreased the average degree in the co-location network by over 15%, indicating that other co-location relationships had a greater offer-to-offer variability.

Similarly to "co-habitation," for any threshold F_i, we defined two participants to be in a "co-working" relationship based on the co-location observations in the morning between 9am and 12pm and in the afternoon between 2pm and 5pm. As the two analyses yielded similar results, herein we focus on the results of the co-habitation analysis. (The results for the co-working analysis are presented in Table WA5 in the Web Appendix.) The results of our estimations on the reduced samples of participants are presented in Table 5. We found that the relationship between the location network variables and coupon redemption remained significant even after removing the potentially co-habiting individuals from the sample. We conclude that our results cannot be merely driven by store-distance effects.

[– Insert Table 5 around here –]

Hot Spots versus "Cooler" Locations

Could busy urban hubs drive the co-location effect? In principle, it is possible that a few popular locations, such as shopping malls visited by many users, reveal something context-specific about the preferences of consumers who appear at those locations. Clearly, the presence or absence of such effects would have very different implications for marketers planning to target consumers with mobile coupons. To this end, we examined whether popular places or less-frequently visited locations reveal more about consumer preferences.

Apart from the obvious managerial take-aways, this problem is also interesting from a theoretical perspective. In the related prior literature, Hui, Bradlow, and Fader (2009) found that consumers in (currently) busier areas of the supermarket are, in general, less prone to make a purchase than in less busy areas. On the other hand, Andrews et al. (2015) showed that SMS coupons (which could be converted by sending the appropriate response code in a reply message) had a higher take-up rate in more-crowded subway trains than in less-crowded ones.

In both of these papers, the nature of the physical environment formed a small natural upper bound on the number of jointly co-located consumers. Relaxing the strict simultaneity constraint and focusing on the aggregate visitation count of locations over 10 days, Provost, Martens, and Murray (2015) found mixed support for the inverse relationship between location popularity and the strength of the co-location effect on publisher choice. We conducted a similar test to determine whether frequently-attended hot spots or "cooler" (less-visited) locations explained more of consumers' coupon redemption behavior. Just like Provost, Martens, and Murray (2015), we proxied the popularity of a given location (here: grid cell) by counting the total number of visitations during the entire observed period. As an alternative definition, we counted only the aggregate number of distinct users at any location (counting even frequently-visiting consumers only once). Figure 2 displays a heat map of one metropolitan area, defining the popularity of any grid cell as the total number of GPS observations in that cell during the period studied. (The heat map for the alternative definition of location popularity is displayed in Figure WA1 in the Web Appendix.)

[– Insert Figure 2 around here –]

For both definitions of location popularity, we conducted the following test. We split the locations into two groups and derived two separate co-location networks based on co-location events that occurred at hot spots and at less-visited locations. To obtain the maximum power, we created the split so that our 15,353 coupon redemption observations were represented in the two resulting networks as evenly as possible. For the heat map in Figure 2, the 97 locations, each with more than 595 location observations in our sample became the hot spots. For the heat map shown in Figure WA1, cells that were visited by at least 10 participants became the hot spots. For both definitions of location popularity, we then created the $\text{Coloc\_High-Freq\_Category\_Redeem\_Rate}_{i,j}$ and $\text{Coloc\_Low-Freq\_Category\_Redeem\_Rate}_{i,j}$ variables just as in Equation 4, and estimated Model 3 replacing the co-location variable by the above two variables.

For brevity, we report the detailed results of our analysis in Table WA7 in the Web Appendix. Concerning the variables of interest, we found that the co-location effect was statistically significant only in the network representing co-location events at infrequently-visited locations ($\beta = 4.912$, $z = 3.20$ for the split based on total observation count, and $\beta = 3.498$, $z = 3.11$ for the split based on the count of distinct users) but not for the network capturing frequently-visited locations ($\beta = 1.916$, $z = 1.67$ and $\beta = 2.990$, $z = 1.53$, respectively). We conclude that our effects are not driven by co-location at hot spots such as business districts, transportation hubs, etc. Instead, the increased predictive power of Models 2 and 3 in our main analysis stems from capturing the co-location events that occurred at less visited places on the map.

Alternative Definitions of Co-Location

This section explores the robustness of our results to alternative definitions of co-location. To this end, we first relax the definition of co-location in geographical space and time, respectively. We then look at aggregating more (days worth of) co-location events into the co-location network. Finally, we prune the co-location network to only retain links between participants who met at more than one location during the program. The summary statistics of the so-created networks – we present these along with the characteristics of the networks used in our main models – are reported in Table 6, and the results of our alternative models in Table 7. We detail our methods and findings below.

[– Insert Table 6 around here –]

[– Insert Table 7 around here –]

"Almost Co-location:" Space

In our main model, we defined co-location as being in the same cell on a grid of 0.002° spacing. Due to the accuracy of our GPS data (captured by smartphones), further precision was not possible. However, one may question if even such a strict definition is required to study participants' preferences. To examine this issue, we redefined our co-location network so that it admitted relationships between any two participants who were at the same time in cells that could be blocks of the same cell on a grid of $0.004°$ spacing (see Figure 3). We then estimated Model 3 using the so-derived co-location variables.

[– Insert Figure 3 around here –]

Intuitively, making the definition of co-location more inclusive should induce two effects. On the one hand, the density of the co-location network should increase. On the other hand, each relationship in the co-location network should be a weaker reflector of the similarities between the preferences of the two linked participants. We observed both of these effects in our data. In particular, Table 6 shows that the average degree in the co-location network almost precisely quadrupled from the average degree derived from the basic definition of co-location. On the other hand, we found that whereas Colocation_Category_Redeem_Rate was significant at the $p = .05$ level ($\beta = 1.411$, $z = 2.47$), its coefficient was substantially lower than the corresponding equivalent reported in Table 4.

"Almost Co-location:" Time

In our main model, we defined co-location as being in the same cell on the grid during the same hour. It is natural to ask how the model performs with the time constraint relaxed. To that end, we created an alternative co-location network where two participants were connected if they were at the same location (i.e., inside the same grid cell) during any two-hour window during the day preceding the launch of the offer. We then estimated Model 3 using the so-derived co-location variables.

We expected the same effects as in the previous section, and again, we observed both trends in the data. The average degree in the co-location network increased by about 70%. The effect of co-location remained significant at the $p = .05$ level ($\beta = 2.342$, $z = 2.86$) but both its coefficient and the significance level decreased relative to that reported in Table 4.

Varying Window Length

In our main model, two participants are neighbors in the co-location network if they were co-located at least once during the last day preceding the launch of the offer. Table 7 reports the results for the models wherein the co-location network is defined based on such co-location events for $w = 2$, 3, and 4 days before the launch of the offer. The results are similar to those discussed above. As w increased, more distant events were incorporated in the network, leading to a higher degree (cf. Table 6). However, these distant events were less revealing about participants' current preferences, ultimately weakening the effect of the $\text{Colocation\_Category\_Redeem\_Rate}_{i,j}$ variable.

Requiring Multiple Locations of Co-Occurrence

Given the positive link between co-location and coupon redemption behaviors, it is interesting to consider whether multiple co-location shortly preceding the launch of an offer indicates more commonalities between the corresponding consumers' preferences. Such events, however, were very rarely observed in our data and hence we are unable to answer this question. To focus on participants who co-occurred at different locations, we therefore defined a co-location network that factored in participants' location behavior during the entire program. Specifically, we pruned the co-location network (defined in Equation 2) such that we only kept connections between co-located participants who, over the entire period observed, co-occurred at a minimum of two different locations. (E.g., we did remove connections between individuals who co-located very often, but did so always at the same location.)

The results (presented in Table 7) are nearly identical to our baseline model. The directions of the coefficients are the same and with the exception of some demographic variables, the significance levels also did not change. These results lend additional robustness to our main findings. In particular, since we do not observe the physical location(s) associated with each offer, one could argue that our co-location effect just captures the fact that different consumers regularly return to the same location near a given store, and this is driving our results.[7] Defining co-location based on multiple distinct locations alleviates this concern.

Additional Tests of Validity

Besides the "co-working" analysis and the additional results concerning hubs versus less-frequented locations, in the Web Appendix, we also detail additional validity tests we performed. Specifically, we estimated our model (1) using alternative means to control for participant heterogeneity, (2) with category-specific slopes to capture the impact of past coupon redemption behavior, and (3) using data only on those offers that were viewed by participants (to the end of testing the impact of their coupon selection behavior). We note that, for the variables of interest, the results of these tests are consistent with those reported herein.

Discussion and Concluding Remarks

As digital media consumption and the time consumers spend on mobile devices are steadily on the rise, marketers are reallocating their budgets to reach consumers increasingly via their smartphones. However, marketers' use of dynamic location data is mostly limited to geo-fencing – targeting consumers when they are in the vicinity of the retailer (Jagoe 2003). Recent academic research (Fong, Fang, and Luo 2015, Luo et al. 2014) pointed out that such methods ignore many valuable prospects outside the "geo-fence," leaving money on the table. The core objective of this paper was therefore to explore a novel use of dynamic consumer location data, with applications to segmenting consumers and targeting mobile advertisements. Based on results from economics (McFadden 2001, Rossi, McCulloch, and Allenby 1996) and consumer behavior (Bettman, Luce, and Payne 1998, Feldman and Lynch 1988), we theorized that consumers' movement patterns are informative of their product preferences, and proposed to exploit this relationship by means of constructing a network of co-locations – events when two or more consumers appeared at the same place at approximately the same time.

Applying the proposed new theory, we studied the behavior of 217 smartphone users who participated in a pilot advertising program run by a mobile operator in the Pacific region. Over three months, participants received about 100 mobile coupons in four different categories. Albeit the offers were broadcasted to participants in the program irrespective of their location, the data also contained hourly GPS records for the participants, allowing us to reconstruct their mobility patterns. We created a dynamically evolving co-location network, wherein any two participants were connected if they had appeared at the same place at about the same time during the period studied. We then modeled the joint effects of network position in the so-derived co-location network, offer characteristics, demographics, psychographics, and referral network position on advertising response.

Our results show that location history can be a relatively effective predictor of promotion response behavior. Specifically, we found that the past coupon response rate of a consumer in a given category did positively correlate with the coupon response likelihood of a consumer that they "co-located" with, and that this effect on mobile coupon response was present *over and above* the effects of both traditional variables such as demographic and psychographic data, and those of referral network position. Moreover, our validity tests demonstrated that the effect of co-location was not driven by "co-habitation" or "co-working" (individuals residing or working in the same area), supplying further evidence that consumers' location patterns provide information about their preferences beyond what is revealed by their social relationships. In addition, comparing co-locations at frequently-visited "hot spots" to all other co-location events indicated that trips to unusual destinations contain more information on consumers' preferences than do trips to busy locations visited by many people. These results lend further credibility to the idea that consumers' location choices can be indicative of their product preferences.

Finally, we estimated the value of information captured by the location variables in scenarios with high costs of false positives. An example for such a setting is an advertising platform that gets commission from all ads that lead to sales while being (e.g., legally) constrained by user-defined limitations such as delivering only up to 3 ads per day to any participating consumer's smartphone. In such environments, the primary goal of marketers is to identify the prospects who are the most likely to respond to each ad, respectively. Using out-of-sample estimations, we demonstrated the practical value of our methods – we found that, even after controlling for participant and offer heterogeneity, incorporating location data in the models increased the sensitivity of selecting the top 5% of prospects by as much as 19%.

Limitations and Avenues for Future Research

Our results suggest that managers need to respond to the new data challenges emerging in the mobile marketing ecosystem by embracing new methods in the area of location-based marketing. To illustrate the potential of such methodological developments, it is important to remember that our analysis was based on location information only representing abstract coordinates. We did not have information about the POIs in the geographic region covered by consumers' movements.[8] Notwithstanding these restrictions, just by studying the co-location network of consumers, we were able to demonstrate significant effects while controlling for individual and offer characteristics and even for social relationships via the referral network between participants. One can only speculate that the presence of POI information, let alone purchase history data could greatly enhance the effect of location measures.

Specifically, combining our data with the exact time of coupon redemptions and detailed information on the network of participating stores could allow us to separately test co-location effects on same-day conversions, and on all other coupon redemptions. In line with Luo et al. (2014), we anticipate store-distance effects to be prevalent for same-day conversions, potentially weakening the co-location effect. Furthermore, obtaining the home and work locations of participants could provide further insights into the specific circumstances and specific type of offers that led consumers to plan trips further off their usual path of travel, a key question in shopper marketing (Shankar et al. 2011). Similarly, observing the list of items purchased by an individual upon redeeming a coupon could improve our model by providing context to the purchase occasion. Understanding customers shopping goals could in turn allow marketers to “contextualize” the various points on consumers’ geographic path to purchase using the techniques of Hui et al. (2013), leading to a deeper understanding of the cognitive processes leading to coupon conversions (Cooke and Zubcsek 2017). Finally, obtaining data on participants’ voice and short messaging communications could allow us to differentially account for strong and weak social ties in our models. Admittedly, most referrals in our data likely indicate strong ties but, as Pentland (2014) points out, weak ties may be more central to the spreading of behaviors (such as redeeming a specific offer).

It is clear from the above discussion that there may be a number of variables that may correspond to boundary conditions for the co-location effect documented in this paper. Notwithstanding these limitations, in the spirit of Leone and Schultz (1980) and Bass (1995), our work aims to take an important step towards establishing an empirical generalization linking consumers’ preferences to their dynamic movement patterns. In addition, we note that the proposed methodology is robust and relatively easy to implement in a practical setting.[9] By matching consumers’ location in space and time, a simple algorithm can create the co-location network variables, which can then be used in standard statistical models. It is also straightforward to generate a co-location measure based on our methods that can be used in scoring models. These models can compare the effect of co-location with other covariates in the particular context and investigate further the interactions with other factors. Although our work does not uncover the details of the process through which co-location is related to consumer behavior, we demonstrate that this is not necessary for identifying and targeting the best prospects. In sum, we believe that our proposed methodology can be particularly useful for marketing practice and that it is an important step towards understanding the role of location dynamics in mobile marketing.

References

Almuhimedi, Hazim, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal (2015). Your location has been shared 5,398 times!: A field study on mobile app privacy nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, pp. 787–796.

Anderson, Simon P. and Andre De Palma (1988). Spatial price discrimination with heterogeneous products. *The Review of Economic Studies*, 55 (4), 573–592.

Andrews, Michelle, Jody Goehring, Sam K. Hui, Joseph Pancras, and Lance Thornswood (2016). Mobile promotions: A framework and research priorities. *Journal of Interactive Marketing*, 34, 15–24.

—, Xueming Luo, Zheng Fang, and Anindya Ghose (2015). Mobile ad effectiveness: Hyper-contextual targeting with crowdedness. *Marketing Science*, 35 (2), 218–223.

Aral, Sinan, Lev Muchnik, and Arun Sundararajan (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106 (51), 21544–21549.

Banerjee, Syagnik and Ruby Roy Dholakia (2008). Mobile advertising: Does location based advertising work? *International Journal of Mobile Marketing*, 3 (2), 68–75.

—, Amit Poddar, Scott Yancey, and Danielle McDowell (2011). Measuring intangible effects of m-coupon campaigns on non-redeemers. *Journal of Research in Interactive Marketing*, 5 (4), 258–275.

Barkhuus, Louise and Anind K. Dey (2003). Location-based services for mobile telephony: a study of users' privacy concerns. In *INTERACT*, Citeseer, vol. 3, pp. 702–712.

Bart, Yakov, Andrew T. Stephen, and Miklos Sarvary (2014). Which products are best suited to mobile advertising: A field experiment of mobile display advertising. *Journal of Marketing Research*, 51 (3), 270–285.

Barwise, Patrick and Colin Strong (2002). Permission-based mobile advertising. *Journal of Interactive Marketing*, 16 (1), 14–24.

Bass, Frank M. (1969). A new product growth model for consumer durables. *Management Science*, 15, 215–227.

— (1995). Empirical generalizations and marketing science: A personal view. *Marketing Science*, 14 (3_supplement), G6–G19.

Bawa, Kapil and Robert W. Shoemaker (1987). The effects of a direct mail coupon on brand choice behavior. *Journal of Marketing Research*, 24 (4), 370–376.

Bell, David R. and Sangyoung Song (2007). Neighborhood effects and trial on the internet: Evidence from online grocery retailing. *Quantitative Marketing and Economics*, 5 (4), 361–400.

Bellovin, Steven M., Renee M. Hutchins, Tony Jebara, and Sebastian Zimmeck (2013). When enough is enough: Location tracking, mosaic theory, and machine learning. *New York University Journal of Law & Liberty*, 8, 556–629.

Berger, Jonah and Gráinne Fitzsimons (2008). Dogs on the street, Pumas on your feet: How cues in the environment influence product evaluation and choice. *Journal of Marketing Research*, 45 (1), 1–14.

Bettman, James R., Mary F. Luce, and John W. Payne (1998). Constructive consumer choice processes. *Journal of Consumer Research*, 25 (3), 187–217.

Bradlow, Eric T., Bart Bronnenberg, Gary J. Russell, Neeraj Arora, David R. Bell, Sri Devi Duvvuri, Frankel Ter Hofstede, Catarina Sismeiro, Raphael Thomadsen, and Sha Yang (2005). Spatial models in marketing. *Marketing Letters*, 16 (3-4), 267–278.

Bronnenberg, Bart J. (2005). Spatial models in marketing research and practice. *Applied Stochastic Models in Business and Industry*, 21 (4-5), 335–344.

—, Sanjay K. Dhar, and Jean-Pierre Dubé (2007). Consumer packaged goods in the United States: National brands, local branding. *Journal of Marketing Research*, 44 (1), 4–13.

— and Vijay Mahajan (2001). Unobserved retailer behavior in multimarket data: Joint spatial dependence in market shares and promotion variables. *Marketing Science*, 20 (3), 284–299.

Brooks, Robert C. (1957). "Word-of-mouth" advertising in selling new products. *The Journal of Marketing*, 22 (2), 154–161.

Burns, Kelli S. and Richard J. Lutz (2006). The function of format: Consumer responses to six on-line advertising formats. *Journal of Advertising*, 35 (1), 53–63.

Castells, Manuel (2007). *Mobile communication and society*. Mit Press.

Chiou-Wei, Song-Zan and J. Jeffrey Inman (2008). Do shoppers like electronic coupons?: A panel data analysis. *Journal of Retailing*, 84 (3), 297–307.

Cooke, Alan D. and Peter P. Zubcsek (2017). The connected consumer: Connected devices and the evolution of customer intelligence. *Journal of the Association for Consumer Research*, forthcoming.

Danaher, Peter, Michael Smith, Kulan Ranasinghe, and Tracey S. Danaher (2015). Where, when and how long: Factors that influence the effectiveness of mobile phone coupons. *Journal of Marketing Research*, 52 (5), 710–725.

de Montjoye, Yves-Alexandre, César A Hidalgo, Michel Verleysen, and Vincent D. Blondel (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, 3, 1376.

Dekimpe, Marnik G, Philip M Parker, and Miklos Sarvary (2000). Global diffusion of technological innovations: A coupled-hazard approach. *Journal of Marketing Research*, 37 (1), 47–59.

Dellaert, Benedict G. C., Theo A. Arentze, Michel Bierlaire, Alyos W. J. Borgers, and Harry J. P. Timmermans (1998). Investigating consumers' tendency to combine multiple shopping purposes and destinations. *Journal of Marketing Research*, 35 (2), 177–188.

Dickinger, Astrid and Mirella Kleijnen (2008). Coupons going wireless: determinants of consumer intentions to redeem mobile coupons. *Journal of Interactive Marketing*, 22 (3), 23–39.

Dreze, Xavier and François-Xavier Hussherr (2003). Internet advertising: Is anybody watching? *Journal of Interactive Marketing*, 17 (4), 8–23.

Durlauf, Steven N. and Yannis M. Ioannides (2010). Social interactions. *Annu. Rev. Econ.*, 2 (1), 451–478.

eMarketer (2013). Digital set to surpass TV in time spent with US media (August 1). http://www.emarketer.com/Article/Digital-Set-Surpass-TV-Time-Spent-with-US-Media/1010096.

— (2015a). Growth of time spent on mobile devices slows (October 7). http://www.emarketer.com/Article/Growth-of-Time-Spent-on-Mobile-Devices-Slows/1013072.

— (2015b). Mobile ad spend to top $100 billion worldwide in 2016, 51% of digital market (April 2). http://www.emarketer.com/Article/Mobile-Ad-Spend-Top-100-Billion-Worldwide-2016-51-of-Digital-Market/1012299.

— (2016). Slowing growth ahead for worldwide internet audience (June 7). http://www.emarketer.com/Article/Slowing-Growth-Ahead-Worldwide-Internet-Audience/1014045.

Feldman, Jack M. and John G. Lynch (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology*, 73 (3), 421–435.

Fong, Nathan M., Zheng Fang, and Xueming Luo (2015). Geo-conquesting: Competitive locational targeting of mobile promotions. *Journal of Marketing Research*, 52 (5), 726–735.

Forman, Chris, Anindya Ghose, and Avi Goldfarb (2009). Competition between local and electronic markets: How the benefit of buying online depends on where you live. *Management Science*, 55 (1), 47–57.

Ghose, Anindya, Avi Goldfarb, and Sang Pil Han (2012). How is the mobile internet different? Search costs and local activities. *Information Systems Research*.

— and Sang Pil Han (2011). An empirical analysis of user content generation and usage behavior on the mobile internet. *Management Science*, 57 (9), 1671–1691.

Goel, Sharad and Daniel G. Goldstein (2014). Predicting individual behavior with social networks. *Marketing Science*, 33 (1), 82–93.

Greenhut, Melvin L. (1981). Spatial pricing in the United States, West Germany and Japan. *Economica*, 48 (189), 79–86.

Gregory, Robin, Sarah Lichtenstein, and Paul Slovic (1993). Valuing environmental resources: A constructive approach. *Journal of Risk and Uncertainty*, 7 (2), 177–197.

Grewal, Dhruv, Yakov Bart, Martin Spann, and Peter Pal Zubcsek (2016). Mobile advertising: A framework and research agenda. *Journal of Interactive Marketing*, 34, 3–14.

Guadagni, Peter M. and John D. C. Little (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science*, 2 (3), 203–238.

Herlocker, Jon, Joseph A Konstan, and John Riedl (2002). An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. *Information Retrieval*, 5 (4), 287–310.

Herr, Paul M, Frank R Kardes, and John Kim (1991). Effects of word-of-mouth and product-attribute information on persuasion: An accessibility-diagnosticity perspective. *Journal of Consumer Research*, pp. 454–462.

Hill, Shawndra, Adrian Benton, and Christophe Van den Bulte (2013). A comparison of product network and social network based recommendation engines for twitter users, working Paper, University of Pennsylvania.

Hofstede, Frenkel Ter, Jan-Benedict E. M. Steenkamp, and Michel Wedel (1999). International market segmentation based on consumer-product relations. *Journal of Marketing Research*, 36 (1), 1–17.

Huff, David L. (1964). Defining and estimating a trading area. *The Journal of Marketing*, 28 (3), 34–38.

Hui, Sam K., Eric T. Bradlow, and Peter S. Fader (2009). Testing behavioral hypotheses using an integrated model of grocery store shopping path and purchase behavior. *Journal of Consumer Research*, 36 (3), 478–493.

—, Peter S. Fader, and Eric T. Bradlow (2009). Path data in marketing: An integrative framework and prospectus for model building. *Marketing Science*, 28 (2), 320–335.

—, J. Jeffrey Inman, Yanliu Huang, and Jacob Suher (2013). The effect of in-store travel distance on unplanned spending: Applications to mobile promotion strategies. *Journal of Marketing*, 77 (2), 1–16.

Inman, J. Jeffrey and Leigh McAlister (1994). Do coupon expiration dates affect consumer behavior? *Journal of Marketing Research*, 31, 423–428.

Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente (2011). Opinion leadership and social contagion in new product diffusion. *Marketing Science*, 30 (2), 195–212.

Jagoe, Andrew (2003). *Mobile Location Services: The Definitive Guide*. Upper Saddle River, NJ: Prentice Hall.

Jank, Wolfgang and P. K. Kannan (2005). Understanding geographical markets of online firms using spatial models of customer choice. *Marketing Science*, 24 (4), 623–634.

Katona, Zsolt, Peter Pal Zubcsek, and Miklos Sarvary (2011). Network effects and personal influences: The diffusion of an online social network. *Journal of Marketing Research*, 48 (3), 425–443.

Katz, Helen (2014). *The media handbook: A complete guide to advertising media selection, planning, research, and buying*. Routledge.

Kirchner, Thomas R., Jennifer Cantrell, Andrew Anesetti-Rothermel, Jennifer Pearson, Sarah Cha, Jennifer Kreslake, Ollie Ganz, Michael Tacelosky, David Abrams, and Donna Vallone (2012). Individual mobility patterns and real-time geo-spatial exposure to point-of-sale tobacco marketing. In *Proceedings of the conference on Wireless Health*, ACM, p. 8.

Leone, Robert P. and Randall L. Schultz (1980). A study of marketing generalizations. *Journal of Marketing*, 44 (1), 10–18.

Li, Chenxi, Andy Reinaker, Zhang Cheng, and Xueming Luo (2015). Weather and mobile purchases: 10-million-user field study, available at SSRN 2585064.

Lichtenstein, Donald R., Richard G. Netemeyer, and Scot Burton (1990). Distinguishing coupon proneness from value consciousness: an acquisition-transaction utility theory perspective. *The Journal of Marketing*, 54 (3), 54–67.

Luo, Xueming, Michelle Andrews, Zheng Fang, and Chee Wei Phang (2014). Mobile targeting. *Management Science*, 60 (7), 1738–1756.

Lurie, Nicholas H., Jonah A. Berger, Zoey Chen, Beibei Li, Hongju Liu, Charlotte Mason, David M. Muir, Grant M. Packard, Joseph Pancras, Ann Schlosser, Baohong Sun, and Rajkumar Venkatesan (2016). Everywhere and at all times: Mobility, consumer decision making, and choice, available at SSRN 2843891.

Manchanda, Puneet, Jean-Pierre Dubé, Khim Yong Goh, and Pradeep K. Chintagunta (2006). The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, 43 (1), 98–108.

—, Ying Xie, and Nara Youn (2008). The role of targeted communication and contagion in product adoption. *Marketing Science*, 27 (6), 961–976.

Manski, Charles F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60 (3), 531–542.

McFadden, Daniel (2001). Economic choices. *The American Economic Review*, 91 (3), 351–378.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415–444.

Mittal, Vikas, Wagner A. Kamakura, and Rahul Govind (2004). Geographic patterns in customer service and satisfaction: An empirical investigation. *Journal of Marketing*, 68 (3), 48–62.

Molitor, Dominik, Philipp Reichart, Martin Spann, and Anindya Ghose (2015). Measuring the effectiveness of location-based pull advertising: A randomized field experiment, working paper, Ludwig-Maximilians Universität, Munich, Germany.

Nitzan, Irit and Barak Libai (2011). Social effects on customer retention. *Journal of Marketing*, 75 (6), 24–38.

Pan, Wei, Nadav Aharony, and Alex Pentland (2011). Composite social network for predicting mobile apps installation. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, AAAI, pp. 821–827.

Pentland, Alex (2014). *Social physics: How good ideas spread-the lessons from a new science*. Penguin.

Popkowski Leszczyc, Peter T. L., Ashish Sinha, and Anna Sahgal (2004). The effect of multi-purpose shopping on pricing and location strategy for grocery stores. *Journal of Retailing*, 80 (2), 85–99.

Provost, Foster, David Martens, and Alan Murray (2015). Finding similar mobile consumers with a privacy-friendly geosocial design. *Information Systems Research*, 26 (2), 243–265.

Reichhart, Philipp, Christian Pescher, and Martin Spann (2013). A comparison of the effectiveness of e-mail coupons and mobile text message coupons for digital products. *Electronic Markets*, 23 (3), 217–225.

Reinaker, Andy, Michelle Andrews, Xueming Luo, and Zheng Fang (2015). Higher roi for mobile advertisements? Tailoring ad copy to customer locational contexts, working paper, Temple University.

Reinartz, Werner J. and V. Kumar (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, 67 (1), 77–99.

Reingen, Peter H., Brian L. Foster, Jacqueline Johnson Brown, and Stephen B. Seidman (1984). Brand congruence in interpersonal relations: A social network analysis. *Journal of Consumer Research*, 11 (3), 771.

Reyck, Bert de and Zeger Degraeve (2003). Broadcast scheduling for mobile advertising. *Operations Research*, 51 (4), 509–517.

Rosenkrans, Ginger (2009). The creativeness and effectiveness of online interactive rich media advertising. *Journal of Interactive Advertising*, 9 (2), 18–31.

— and Keli Myers (2012). Mobile advertising effectiveness. *International Journal of Mobile Marketing*, 7 (3), 5–24.

Rossi, Peter E., Robert E. McCulloch, and Greg M. Allenby (1996). The value of purchase history data in target marketing. *Marketing Science*, 15 (4), 321–340.

Schiller, Jochen and Agnès Voisard (2004). *Location-based services*. Elsevier.

Shankar, Venkatesh and Sridhar Balasubramanian (2009). Mobile marketing: a synthesis and prognosis. *Journal of Interactive Marketing*, 23 (2), 118–129.

—, J. Jeffrey Inman, Murali Mantrala, Eileen Kelley, and Ross Rizley (2011). Innovations in shopper marketing: Current insights and future research issues. *Journal of Retailing*, 87, S29–S42.

—, Alladi Venkatesh, Charles Hofacker, and Prasad Naik (2010). Mobile marketing in the retailing environment: current insights and future research avenues. *Journal of Interactive Marketing*, 24 (2), 111–120.

Sundararajan, Arun, Foster Provost, Gal Oestreicher-Singer, and Sinan Aral (2013). Information in digital, economic and social networks. *Information Systems Research*, 24 (4), 883–905.

Yang, Sha and Greg M. Allenby (2003). Modeling interdependent consumer preferences. *Journal of Marketing Research*, 40 (3), 282–294.

Zhang, Jie, Michel Wedel, and Rik Pieters (2009). Sales effects of attention to feature advertisements: A bayesian mediation analysis. *Journal of Marketing Research*, 46 (5), 669–681.

Zoltners, Andris A. and Prabhakant Sinha (1983). Sales territory alignment: A review and model. *Management Science*, 29 (11), 1237–1256.

Footnotes

[1]The abovementioned models constrain the direct interdependency between consumers' actions mostly to within their geographic regions. A common way this problem has been addressed in the literature is by assuming that the geodesic distance of decision-making individuals moderates the likelihood of them engaging in the same behavior, e.g. due to network externalities, geographically-varying common shocks, or simply a higher chance of observational word-of-mouth (Bell and Song 2007, Bronnenberg, Dhar, and Dubé 2007, Dekimpe, Parker, and Sarvary 2000, Jank and Kannan 2005, Mittal, Kamakura, and Govind 2004, Yang and Allenby 2003).

[2]Demonstrating a positive link between consumers' web browsing and location behaviors over a span of ten days, Provost, Martens, and Murray (2015) provided initial evidence for this claim.

[3]Specifically, we define the dynamic location-based similarity in a dichotomous fashion, and we use equal weights to aggregate network neighbors' past behavior. We thank two anonymous reviewers for highlighting this connection.

[4]Mobile banner ads have also been shown to bear many similarities to online banner ads (Bart, Stephen, and Sarvary 2014, Rosenkrans and Myers 2012). For instance, there is a general consensus that digital banner ads primarily impact consumer behavior via memory and recall rather than leading to instant click-throughs directing traffic to the advertiser's website (Dreze and Hussherr 2003, Manchanda et al. 2006, Zhang, Wedel, and Pieters 2009).

[5]The applicability of the offers varied from a single product of a given brand to *any* purchase at *any* coffee shop in town.

[6]According to data from the mobile operator, participants saw 85.66% of the offers. Not-seen offers mostly correspond to either initial technical difficulties, or to participants losing interest in the program altogether. Within observations on participants who had not redeemed any coupons yet, the rate of seen offers varied between 76.9-78.9% in each category. For observations on participants who had redeemed at least one coupon, the "seen" rates were between 97.6-97.8% for each category. Together, these results indicate a limited impact of coupon category selection. Nonetheless, for completeness, we also estimated the model excluding the observations corresponding to not-seen coupons. The results of interest, reported in Table WA1 in the Web Appendix, are essentially the same as those for the full set of observations.

[7]We would like to thank an anonymous reviewer for pointing out this possibility.

[8]The type of POIs can be quite varied depending on the context of the marketing campaign or study. In most cases, the store locations make a natural candidate. However, more general POIs may also prove to be useful. For instance, in the context of a political campaign, visiting historical sites could be a good indicator of political affiliation.

[9]Note that the general methodology used in our estimations could just as easily be replicated on other data from which consumer networks can be generated. The point is that such networks may be based on the recording of a variety consumer behaviors. The advantage of such 'agnostic' approaches is that they need not rely on strong behavioral assumptions. Instead, they may benefit from the well-documented pattern of homophily that generally characterizes behavior-based networks.

Tables

Table 1: Comparison of Consumer Choice Models Using Dynamic Location Data

Paper	Factors *moderating* the impact of location on consumer choice[a]				Geography Covered	Time Period Covered	Location Observations per Participant
	Time	**Idiosyncratic**	**Structure of Co-location**	**Location History**			
Andrews et al. (2015)[b]	Yes	Other consumers' location	Clusters (passengers on same train)	No	Subway in an Asian city	< 30 days	1
Danaher et al. (2015)	No	No	N/A	No	Shopping mall (ca. 400 stores)	654 days	16.87
Fong, Fang, and Luo (2015)	Yes	No	N/A	No	4.5km strip connecting two cinemas	7 days	1
Ghose, Goldfarb, and Han (2012)	No	No	N/A	No	South Korea	98 days	34.22
Hui, Bradlow, and Fader (2009)	No	Other consumers' location	Clusters (shoppers in same zone)	Via latent variables	Supermarket (96 zones)	21 days	99.80
Li et al. (2015)	Yes	No	N/A	No	China	32 days	1
Luo et al. (2014)	Yes	No	N/A	No	2km circle around a cinema	3 days	1
Molitor et al. (2015)	No	No	N/A	No	Large Western-European country	98 days	3.26[c]
Reinaker et al. (2015)	No	Home / work	Clusters (co-workers)	Home location	Hospitals in an Asian country	30 days	N/A (1 used)
This work	Yes	Other consumers' location	Network	24-96 hours	Two major cities in a Pacific country	84 days	1,202.94

[a] A few other moderators have been considered in specific papers. Ghose, Goldfarb, and Han (2012) considered different location effects for PCs and mobile devices, while Molitor et al. (2015) allowed the effects to reflect which order the coupons were presented in, and whether the distance of the store was displayed on the m-coupon.
[b] In Andrews et al. (2015), six-car subway trains were considered as the location of a consumer.
[c] On average, the redemption behavior of 27.43 offers belong to a single location observation (session).

Table 2: Description of Variables

Offers		**Mean**	**(St. Dev)**
$\text{Redeemed}_{i,j}$	The dependent variable. Equals to 1 if participant i redeemed offer j and is 0 otherwise.	.049	(.22)
$\text{Offer\_Length}_{i,j}$	The number of days for which offer j was available to participant i. (If i joined after offer j was launched, this number may be less than the duration of offer j.)	3.841	(.90)
$\text{Days\_Since\_Joined}_{i,j}$	The number of days that participant i had been in the program *before* offer j became available.	28.87	(19.52)
Category-$\text{\_Redeem\_Rate}_{i,j}$	The fraction of offers that participant i redeemed from those that were in the same category as offer j and that had expired before offer j became available.	.026	(.07)
Location			
$\text{GPS Activity}_{i,j}$	The number of hours with location observations from participant i during the day before offer j became available to i.	11.31	(9.87)
Colocation_Category-$\text{\_Redeem\_Rate}_{i,j}$	The sum of “Category_Redeem_Rate” for the neighbors of participant i in the co-location network defined by “same time, same place” events on the day before offer j became available to i; or 0 if the above did not apply.	.0065	(.04)
Individual Characteristics (Profile Variables)			
Age_i	Age of participant i (in years).	34.95	(9.75)
Gender_i	Gender of participant i (1=Male, 2=Female).	1.40	(.49)
Education_i	Highest degree completed by participant i (1=Primary school, 6=PhD or above).	3.35	(1.00)
Experience_i	Professional experience of participant i (1=Has not started career, 4=Established in career).	3.09	(1.09)
Income_i	Income bracket of participant i (1=lowest, 4=highest).	2.91	(1.34)
Hedonic_i	Participant i's tendency to consume hedonic products (1=lowest, 3=highest).	2.25	(.60)
Well-read_i	Participant i's tendency to read books (1=lowest, 4=highest).	2.36	(.86)
Hard-working_i	Bracket of part. i's ideal workweek length (1=lowest, 4=highest).	1.93	(1.15)
Spender_i	Participant i's tendency to expend money (1=lowest, 5=highest).	2.12	(.76)
Referral (Control)			
Referral_Category-$\text{\_Redeem\_Rate}_{i,j}$	The sum of “Category_Redeem_Rate” for the neighbors of participant i in the referral network on the day before offer j became available to i; or 0 if the above did not apply.	.0064	(.03)

Table 3: Descriptive Statistics of Offers by Category

Offer Category	**Observations**	**Offer Length Mean (St.Dev.)**	**After-Coupon Minimum Price[a] Mean (St.Dev.)**	**Overall Redemption Rate Mean (St.Dev.)**
CPG	3534	3.89 (.38)	12.78 (10.44)	.06 (.24)
Food & Bev	4296	3.95 (.83)	9.57 (4.98)	.03 (.18)
Retail	5136	4.07 (.83)	20.34 (13.24)	.04 (.19)
Recreation	2387	3.85 (.47)	3.25 (6.10)	.08 (.28)

[a] Equals to the when-you-buy sum decreased by the value of the coupon.

Table 4: Parameter Estimates, Main Models (z-Scores in Parentheses)

Variable	Model 0		Model 1a		Model 1b		Model 2a		Model 2b		Model 3	
Offer Length	.366**	(3.14)	.371**	(3.18)	.291*	(2.50)	.353**	(3.02)	.280*	(2.39)	.286*	(2.44)
Days Since Joined	.015	(1.68)	.012	(1.42)	.016	(1.76)	.016	(1.78)	.018*	(2.06)	.016	(1.79)
Category_Redeem_Rate	1.054	(1.71)	1.014	(1.65)	.845	(1.36)	.861	(1.39)	.717	(1.15)	.693	(1.11)
GPS Activity					.016*	(2.48)			.016*	(2.46)	.016*	(2.45)
Colocation_Categ_Redeem_Rate					3.858***	(3.99)			3.225**	(3.26)	3.267**	(3.30)
Age			.040	(1.91)			.035	(1.72)			.036	(1.75)
Gender			.785*	(2.30)			.748*	(2.21)			.716*	(2.13)
Education			.359*	(2.13)			.351*	(2.10)			.349*	(2.10)
Experience			-.154	(-.74)			-.134	(-.65)			-.133	(-.65)
Income			-.103	(-.67)			-.090	(-.59)			-.104	(-.69)
Hedonic			-.171	(-.62)			-.182	(-.67)			-.163	(-.61)
Well-read			-.224	(-1.23)			-.206	(-1.14)			-.214	(-1.20)
Hard-working			-.175	(-1.27)			-.144	(-1.05)			-.158	(-1.16)
Spender			.242	(1.08)			.175	(.78)			.178	(.81)
Referral_Categ_Redeem_Rate							5.971***	(3.65)	5.380**	(3.16)	4.769**	(2.81)
Observations	15,353		15,353		15,353		15,353		15,353		15,353	
Participants	217		217		217		217		217		217	
Offers	96		96		96		96		96		96	
Log Likelihood	-2,082.15		-2,073.77		-2,069.90		-2,067.09		-2,065.00		-2,057.65	
AIC	4,362.30		4,363.54		4,341.80		4,352.19		4,333.99		4,337.30	
BIC	5,118.57		5,188.56		5,113.34		5,184.84		5,113.18		5,194.88	

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. The models include offer fixed effects and participant random effects.

Table 5: Parameter Estimates, "Co-Habiting" Individuals Removed

Variable	Maximum frequency of nights when co-located: 50%		40%		30%		20%	
Offer Length	.284*	(2.44)	.273*	(2.35)	.269*	(2.31)	.263*	(2.26)
Days Since Joined	.015	(1.77)	.015	(1.73)	.016	(1.76)	.015	(1.60)
Category_Redeem_Rate	.691	(1.09)	.647	(1.01)	.688	(1.07)	.643	(1.00)
GPS Activity	.016*	(2.42)	.016*	(2.42)	.016*	(2.40)	.014*	(2.03)
Colocation_Category_Redeem_Rate	3.258**	(3.25)	3.165**	(3.14)	3.253**	(3.13)	2.904**	(2.67)
Age	.036	(1.72)	.038	(1.80)	.037	(1.74)	.034	(1.53)
Gender	.744*	(2.21)	.762*	(2.23)	.723*	(2.06)	.757*	(2.13)
Education	.318	(1.88)	.319	(1.87)	.333	(1.92)	.311	(1.76)
Experience	-.134	(-.65)	-.144	(-.69)	-.145	(-.69)	-.163	(-.77)
Income	-.101	(-.66)	-.117	(-.75)	-.123	(-.77)	-.092	(-.57)
Hedonic	.-217	(-.79)	-.216	(-.78)	-.210	(-.74)	-.123	(-.42)
Well-read	-.204	(-1.14)	-.202	(-1.11)	-.215	(-1.17)	-.173	(-.93)
Hard-working	-.154	(-1.13)	-.156	(-1.13)	-.146	(-1.03)	-.114	(-.79)
Spender	.177	(.79)	.192	(.85)	.205	(.90)	.086	(.34)
Referral_Category_Redeem_Rate	4.951**	(2.78)	4.924**	(2.76)	4.761**	(2.65)	4.386*	(2.21)
Observations	15,042		14,890		14,668		14,116	
Participants	213		211		208		200	
Offers	96		96		96		96	
Log Likelihood	-2,021.02		-1,988.18		-1,954.24		-1,859.99	

* $p < .05$, ** $p < .01$, *** $p < .001$. The models include offer fixed effects and participant random effects.

Table 6: Descriptive Statistics of Networks Considered

Network	Degree: Mean (st.dev)	Degree: Mean number of non-isolate nodes[a]	Degree: Mean (st.dev.) if non-isolate	Clustering: Mean number of nodes defined for[a]	Clustering: Mean (st.dev) if defined
Co-location	.164 (.49)	20.33	1.292 (.64)	4.38	.401 (.45)
Co-location, overlapping cells of size .004°	.655 (1.55)	41.19	2.543 (2.14)	22.46	.481 (.37)
Co-location, overlapping two-hour intervals	.280 (.74)	28.16	1.589 (1.00)	9.97	.358 (.41)
Co-location, two-day lookback	.263 (.66)	29.23	1.439 (.84)	8.35	.314 (.41)
Co-location, three-day lookback	.344 (.80)	35.25	1.564 (.99)	11.80	.279 (.38)
Co-location, four-day lookback	.420 (.92)	40.29	1.666 (1.12)	14.77	.272 (.37)
Co-location, at multiple locations (in entire sample)	.090 (.33)	12.46	1.152 (.44)	1.573	.192 (.40)
Referral	.096 (.39)	12.24	1.249 (.76)	2.66	0 (0)

[a] Mean per offer (across the 96 offers considered).

Table 7: Parameter Estimates, Alternative Definitions of Co-location

					Considering Co-locations Preceding Offer Launch By							
Variable	**0.004° cells**		**Two-Hour Window**		**2 Days**		**3 Days**		**4 Days**		**Multiple Locations**	
Offer Length	.285*	(2.44)	.286*	(2.45)	.314**	(2.65)	.323**	(2.72)	.323**	(2.72)	.243*	(2.20)
Days Since Joined	.016	(1.80)	.016	(1.79)	.015	(1.77)	.015	(1.75)	.015	(1.74)	.020*	(2.27)
Category_Redeem_Rate	.762	(1.23)	.734	(1.18)	.706	(1.13)	.695	(1.11)	.698	(1.12)	.730	(1.17)
GPS Activity	.016*	(2.40)	.016*	(2.42)	.0089	(1.23)	.0073	(.95)	.0078	(.97)	.018**	(2.75)
Colocation_Category_Redeem_Rate	1.411*	(2.47)	2.342*	(2.86)	2.364**	(2.75)	1.579*	(2.02)	1.321	(1.87)	3.202**	(2.92)
Age	.037	(1.79)	.036	(1.77)	.036	(1.76)	.036	(1.77)	.036	(1.78)	.018	(.93)
Gender	.723*	(2.15)	.716*	(2.13)	.726*	(2.16)	.732*	(2.18)	.730*	(2.17)	.405	(1.27)
Education	.336*	(2.02)	.342*	(2.06)	.354*	(2.13)	.351*	(2.11)	.350*	(2.11)	.223	(1.37)
Experience	-.130	(-.63)	-.134	(-.65)	-.136	(-.66)	-.138	(-.67)	-.139	(-.68)	-.136	(-.66)
Income	-.100	(-.67)	-.103	(-.68)	-.100	(-.67)	-.097	(-.64)	-.096	(-.64)	-.124	(-.82)
Hedonic	.-156	(-.57)	-.162	(-.60)	-.165	(-.61)	-.166	(-.61)	-.162	(-.60)	-.356	(-1.34)
Well-read	-.218	(-1.22)	-.216	(-1.21)	-.208	(-1.16)	-.207	(-1.16)	-.208	(-1.16)	-.325	(-1.83)
Hard-working	-.165	(-1.21)	-.159	(-1.17)	-.151	(-1.11)	-.151	(-1.11)	-.152	(-1.12)	-.244	(-1.81)
Spender	.180	(.81)	.180	(.82)	.180	(.81)	.180	(.81)	.180	(.81)	.0088	(.04)
Referral_Category_Redeem_Rate	5.152**	(3.08)	5.092**	(3.03)	4.855**	(2.85)	5.106**	(3.01)	5.180**	(3.06)	5.531**	(3.32)
Observations	15,353		15,353		15,353		15,353		15,353		15,353	
Participants	217		217		217		217		217		217	
Offers	96		96		96		96		96		96	
Log Likelihood	-2,060.06		-2,059.09		-2,062.26		-2,064.46		-2,064.72		-2,063.04	

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. The models include offer fixed effects and participant random effects.

Figures

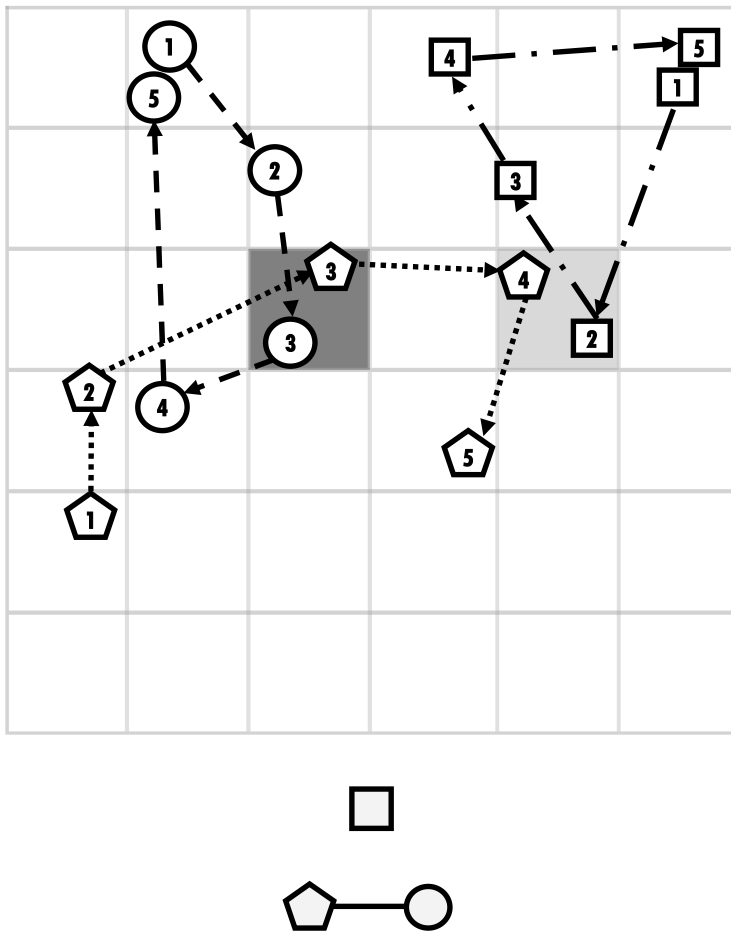

Figure 1: Top: Illustrating our definition of co-location on a grid. Bottom (below grid): The co-location network derived from this information. The three shapes and the numbers inside represent three consumers and the same five consecutive time periods, respectively. The dark gray grid cell in row 3, column 3 witnessed a co-location (between the circle and pentagon shapes) in time period 3. This relationship is indicated by the solid line between these symbols in the network shown below the grid. However, the light gray grid cell in row 3, column 5 does not correspond to such an event as the pentagon and the square were in the same cell at two different times. Thus, the network derived from the above movement patterns of consumers would only have one relationship, as shown.

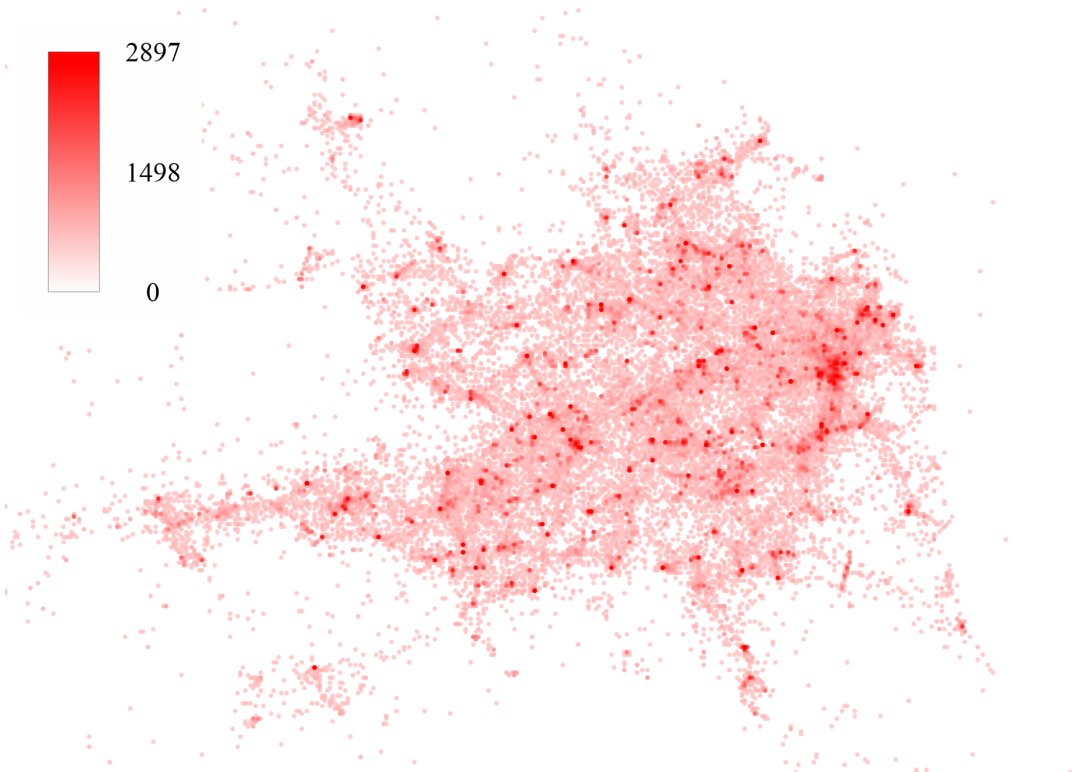

Figure 2: Heat map of all GPS observations in one of the metropolitan areas covered by the program.

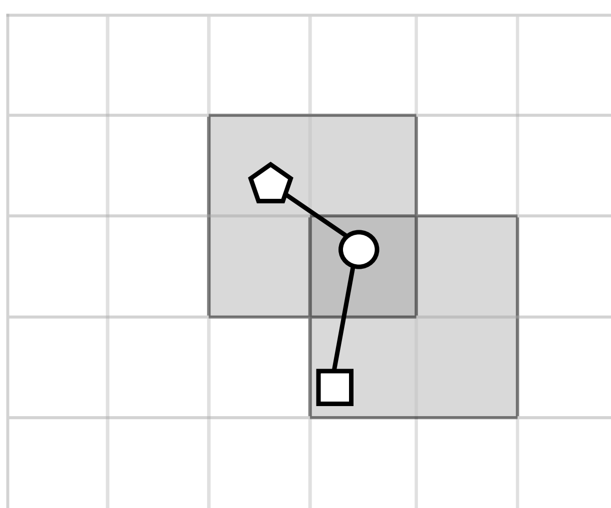

Figure 3: Relaxing the main definition of co-location (the different shapes on the grid representing the location of three consumers at a given hour). Whereas no two consumers are in the same cell (resulting in no co-location event considered in our main estimations), under the relaxed definition, the circle is co-located with both the pentagon and the square. (However, there is no 2×2-cell grid rectangle that contains the latter two symbols, so the pentagon and the square are not co-located with each other.)

Web Appendix

Predicting Mobile Advertising Response Using Consumer Co-Location Networks

Estimating Category-Specific Slopes for Coupon Redemption

In the main text, we restricted the effect of the historic category-specific redemption rates to have the same coefficient across categories. However, we also estimated Model 3 using separate slopes, i.e. by splitting each of the Category_Redeem_Rate, Referral_Category_Redeem_Rate, and Colocation_Category_Redeem_Rate variables into four variables for the four product categories (creating a total of 12 variables out of 3). Comparing model fit, the AIC (4,339.90 vs. 4,344.74) and BIC (5,194.88 vs. 5,261.42) values gave preference to the simpler model. Similarly, the model with 12 redemption rate variables had a log-likelihood of -2,052.37. Comparing this to the value for Model 3 reported in Table 4, we get that LR(Category-specific slopes, Uniform slopes)=$\chi^2(9) = 10.56, p > 0.3$. We therefore decided to report only the results of the models using common slopes for all offer categories.

Testing the Impact of Coupon Selection

In our main models, we did not take into account whether participants saw the coupons sent to their device. As we explained in Footnote 6 of the paper, participants did not appear to have selected coupons for viewing based on the category of the coupon. For completeness, we nevertheless estimated Model 3 using only those observations that corresponded to offers seen by the participants. The results, reported in Table WA1, show that the effect of co-location is virtually the same (within 1%) as in Table 4 of the main manuscript. Furthermore, we can also observe that, for each of the variables in the model, the sign of the coefficients is the same and the 95% confidence intervals overlap between the two estimations. We therefore conclude that coupon selection (due to either participant preferences or technical issues) did not drive the effect of co-location.

Table WA1: Parameter estimates, considering only opened ("seen") offers

Variable	**Model 3**	
Offer Length	.276*	(2.48)
Days Since Joined	.022**	(2.70)
Category_Redeem_Rate	.622	(1.00)
GPS Activity	.0074	(1.14)
Colocation_Category_Redeem_Rate	3.290**	(3.34)
Age	.0088	(.49)
Gender	.310	(1.06)
Education	.209	(1.41)
Experience	-.081	(-.43)
Income	-.125	(-.89)
Hedonic	-.309	(-1.26)
Well-read	-.296	(-1.81)
Hard-working	-.229	(-1.83)
Spender	.044	(.22)
Referral_Category_Redeem_Rate	4.809**	(2.88)
Observations	13,152	
Participants	215	
Offers	96	
Log Likelihood	-2,032.08	

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$
The model includes offer fixed effects and participant random effects.

Additional Controls for Participant Heterogeneity

Quiz Engagement

During the period studied, participants on average received less than two offers per day. To increase their engagement with the mobile app, participants were also invited to respond to quizzes. Quizzes were brief surveys with questions pertaining to a broad range of topics including personality traits, interests, lifestyle, work, and also product preferences in certain product categories. Responding to quizzes was entirely optional and participants did not receive any reward for their answers.

Whereas most quizzes were not admissible to our main models due to violating the basic principles of questionnaire design, participants' level of engagement with quizzes throughout the program is a good indicator of their level of general engagement with the program. In this section, we discuss two different ways in which this idea allows quiz responses to be used as robustness

Table WA2: Parameter estimates, Controlling for Quiz Activity

Variable	**Model 3 + Quiz Activity**		**Frequent Quiz Responders Only**	
Offer Length	.295*	(2.52)	.327*	(2.39)
Days Since Joined	.015*	(2.01)	.023*	(2.33)
Category_Redeem_Rate	.775	(1.26)	.779	(.89)
GPS Activity	.012	(1.85)	.0053	(.71)
Colocation_Category_Redeem_Rate	3.183**	(3.24)	4.203**	(3.19)
Age	.012	(.69)	-.037	(-1.60)
Gender	.334	(1.19)	-.744*	(-1.83)
Education	.263	(1.89)	.067	(.39)
Experience	-.070	(-.40)	.222	(.94)
Income	.0077	(.06)	-.022	(-.15)
Hedonic	-.0085	(-.04)	-.122	(-.33)
Well-read	-.184	(-1.21)	.036	(.17)
Hard-working	-.073	(-.63)	.124	(.72)
Spender	.058	(.30)	.270	(1.12)
Quiz Activity (No. Answered)	.010***	(7.52)		
Referral_Category_Redeem_Rate	4.200*	(2.58)	4.576*	(2.35)
Observations	15,353		7,952	
Participants	217		113	
Offers	96		96	
Log Likelihood	-2,024.31		-1,365.95	

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. The models include offer fixed effects and participant random effects.

tests for our main model. First, we extended Model 3 by adding the variable that captured the total number of quizzes that a participant responded to throughout the program. Second, we selected all – twenty – quizzes which translated to an ordinal scale and had an at least 75% completion rate in our data, and added the responses to these quizzes as covariates to Model 3. Requiring a response to all twenty included quizzes removed less engaged participants from the sample, addressing some potential biases that may have been caused by participants' self-selection into the program. (For instance, while we had only 40.1% female participants in the full sample, there were 47.8% females in the restricted sample.)

The results of our estimations are reported in Table WA2. The coefficients are consistent with those reported in 4 of the main manuscript, except for those of demo- and psychographic variables. This indicates that these models are good validity tests of the co-location effect. In that regard, our efforts provided further support for our theory: The coefficient of the co-location variable is indeed significant in both models reported in Table WA2.

Participant (and Offer) Fixed Effects

In our main regressions, we reported the estimates of models with offer fixed and participant random effects. The reasoning behind this choice is that many participants did not respond to any offer during the period studied. Some of these participants were otherwise active, e.g. they supplied location information and responded to some quizzes. Some even accepted offers inside the app, but ultimately did not redeem the accepted coupons. From the marketer's perspective, several explanations for the lack of redemption may be indistinguishable. For instance, while it is possible that many of these participants simply did not like the offers enough, it may also be that there was some problem with observing their coupon redemptions (at least at certain retailers). To estimate fixed effect models, these participants need to be removed from the sample. However, if the marketer is trying to estimate the impact of independent variables on predictions, then in case these participants are still good prospects to redeem future offers, removing them from the sample would bias all estimates upward. Therefore, we decided to estimate our main models using participant random effects.

Our results on the demographic and psychographic variables nevertheless indicate the presence of participant heterogeneity not captured by random effects. To this end, we re-estimated our model with both offer (η_j) and participant (ξ_i) fixed effects. (To avoid multicollinearity, we therefore dropped the demo- and psychographic variables from this analysis.) The results are reported in Table WA3.

It is easily seen from Table WA3 that the coefficient estimates of location and referral effects are essentially the same as those in Table 7. We conclude that controlling for participant heterogeneity via participant random effects is sufficient. However, there is one important difference between the results in Tables 7 and WA3. In particular, in the smaller sample restricted to participants with nonzero response rates, we find that the tenure of participants in the program decreased their response rate. (Capturing unobserved similarities via the network variables

Table WA3: Parameter estimates, Offer and Participant Fixed Effects

Variable	Model 0		Model 2	
Offer Length	.322**	(2.83)	.285*	(2.29)
Days Since Joined	-.052***	(-3.58)	-.016	(-.63)
Category_Redeem_Rate	.096	(.16)	-.208	(-.33)
GPS Activity			.013	(1.86)
Colocation_Category_Redeem_Rate			3.260**	(3.25)
Referral_Category_Redeem_Rate			4.348*	(2.33)
Observations	9,014		9,014	
Participants	115		115	
Offers	96		96	
Log Likelihood	-1,797.06		-1,784.79	
AIC	4,020.11		4,001.58	
BIC	5,533.80		5,536.59	

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

weakened this effect, however.) On the one hand, this is consistent with a wear-out effect often seen in similar field experiments. On the other hand, it indicates that the weak positive effect of tenure in our main model may in part be driven by "more hesitant" participants joining the program toward the end of the period we study.

Modeling Individual Heterogeneity via Latent Classes

As we noted above, the mass of consumers who did not engage with the program make it impossible to address consumer heterogeneity via fixed effects on the entire panel. It is also important, however, to verify that the co-location effect presented in Table 4 is robust to relaxing the assumption of Normally distributed participant random effects. To do so, we estimated latent class models of the following form:

$$\log \frac{\Pr(Y_{ij} = 1)}{1 - \Pr(Y_{ij} = 1)} = \beta_0 + \beta_1 \mathbf{W_i} + \beta_2 \mathbf{X_{ij}} + \beta_3 \mathbf{Z_{ij}} + \zeta_i + \eta_j, \tag{5}$$

where $\zeta_i \in \zeta_1, \zeta_2, \ldots, \zeta_K$ for $K \in 2, 3, \ldots$ representing the number of latent classes, and for π_k denoting the probability of class k, $\sum_{k=1}^{K} \pi_k \cdot \zeta_k = 0$.

The estimation of the latent class models only converged for $K = 2, 3$, and 4. (Given the prevalence of non-responders in our data, this could be expected.) Of these three values of K, the model with 4 classes had the highest log-likelihood. Table WA4 reports the detailed results for this model.

First, we note that the coefficients of all regressors that varied across offers, were consistent with our results reported in Table 4. (On the other hand, the impact of demographics changed substantially with the structure of individual random effects.) In particular, the effect of co-location is remarkably close to that estimated using a Normal distribution of individual effects. Second, comparing the fit of the models with $K = 4$ latent classes, we see that, again, the log-likelihood of Model 3 was the highest, while on the AIC and BIC criteria Model 2b prevailed. Finally, we also ran the likelihood-ratio tests on the nested models to find that adding location variables improved model fit more than adding demographic predictors did: LR(Model 1a, Model 0)$= \chi^2(9) = 21.81$, $p = .010$; LR(Model 3, Model 2b) $= \chi^2(9) = 13.54$, $p = .140$; LR(Model 1b, Model 0)$= \chi^2(2) = 24.96$, $p < .001$; LR(Model 3, Model 2a)$= \chi^2(2) = 16.34$, $p < .001$. Overall, we conclude that the effect of co-location is robust to the latent-class specification of participant heterogeneity.

Table WA4: Parameter estimates, individual heterogeneity captured via latent classes

Variable	**Model 0**		**Model 1a**		**Model 1b**		**Model 2a**		**Model 2b**		**Model 3**	
Offer Length	.355**	(3.08)	.356**	(3.02)	.288*	(2.49)	.338**	(2.88)	.281*	(2.42)	.282*	(2.42)
Days Since Joined	.012	(1.66)	-.0052	(-1.05)	.013*	(2.23)	.00083	(.14)	.014*	(2.42)	.0085	(1.54)
Category_Redeem_Rate	1.363**	(2.68)	1.173*	(2.24)	.789	(1.47)	.983	(1.89)	.686	(1.22)	.648	(1.24)
GPS Activity					.011	(1.89)			.011	(1.93)	.011*	(1.98)
Colocation_Category_Redeem_Rate					4.163***	(4.85)			3.978***	(4.54)	3.371***	(3.73)
Age			.047***	(5.98)			.042***	(5.27)			.059***	(7.05)
Gender			.095	(.78)			.014	(.12)			.119	(.90)
Education			-.088	(-1.57)			-.146*	(-2.45)			-.136*	(-2.24)
Experience			.0014	(.02)			.00098	(.01)			-.248**	(-3.41)
Income			.172**	(2.84)			.166**	(2.76)			.211**	(3.25)
Hedonic			.342**	(3.30)			.313**	(2.77)			.244*	(2.12)
Well-read			-.348**	(-4.80)			-.373***	(-3.99)			-.314***	(-4.01)
Hard-working			-.169**	(-3.07)			-.164**	(-3.07)			-.169**	(-2.78)
Spender			.010	(.11)			-.079	(-.80)			-.081	(-.86)
Referral_Category_Redeem_Rate							5.817***	(4.53)	6.030***	(4.59)	4.577***	(3.60)
Observations	15,353		15,353		15,353		15,353		15,353		15,353	
Participants	217		217		217		217		217		217	
Offers	96		96		96		96		96		96	
Log Likelihood	-2,071.15		-2,060.25		-2,058.67		-2,054.22		-2,052.82		-2,046.05	
AIC	4,340.30		4,336.49		4,319.35		4,326.45		4,309.65		4,314.10	
BIC	5,096.57		5,161.51		5,090.89		5,159.10		5,088.83		5,162.04	

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. All models include offer fixed effects.

Store Distance Effects: Excluding "Co-Working" Participants

In the main manuscript, we reported the results of the models on the sample that was reduced by removing "co-habiting individuals." In the same vein, we also attempted to identify "co-working" relationships based on the co-location observations in the morning between 9am and 12pm and in the afternoon between 2pm and 5pm. Specifically, for any threshold $F_i \in \{.2, .3, .4, .5\}$, we defined two participants to be in a "co-working" relationship if they were co-located at the same grid cell during the above intervals on an at least F_i fraction of the workdays they spent in the program. This approach removed slightly more participants from the data than the removal of "co-habiting" participants. However, in the "co-working" analysis, we found no difference between the thresholds .2 and .3.

The results of our estimations are reported in Table WA5. Similarly to the results in Table 5, the effect of co-location was robust to our changes. Interestingly, however, the effect of referral weakened as F_i decreased. This indicates that some referral relationships may indeed correspond to people who worked in the same neighborhood. We note, however, that the significance of the co-location variable indicates that this possibility could not account for the effect of co-location observed in the data.

Table WA5: Parameter Estimates, "Co-Working" Individuals Removed

Variable	Maximum frequency of days when co-located during working hours 50%		40%		30%[a]	
Offer Length	.269*	(2.29)	.250*	(2.12)	.251	(1.96)
Days Since Joined	.014	(1.61)	.014	(1.56)	.013	(1.37)
Category_Redeem_Rate	.850	(1.32)	.810	(1.26)	.720	(1.11)
GPS Activity	.017*	(2.50)	.018**	(2.61)	.017*	(2.41)
Colocation_Category_Redeem_Rate	3.437**	(3.31)	3.104**	(2.91)	3.084**	(2.83)
Age	.040	(1.89)	.038	(1.75)	.046*	(1.99)
Gender	.803*	(2.33)	.769*	(2.14)	.692	(1.83)
Education	.326	(1.86)	.307	(1.70)	.333	(1.79)
Experience	-.199	(-.95)	-.193	(-.89)	-.245	(-1.09)
Income	-.106	(-.68)	-.127	(-.79)	-.139	(-.83)
Hedonic	.-274	(-.98)	-.275	(-.96)	-.247	(-.82)
Well-read	-.175	(-.96)	-.164	(-.87)	-.119	(-.60)
Hard-working	-.127	(-.92)	-.147	(-1.02)	-.147	(-.99)
Spender	.263	(1.16)	.212	(.85)	.262	(.99)
Referral_Category_Redeem_Rate	4.543*	(2.48)	4.098*	(2.00)	2.275	(1.01)
Observations	14,820		14,535		14,051	
Participants	210		206		200	
Offers	96		96		96	
Log Likelihood	-1,916.65		-1,852.01		-1,753.93	

[a] The results for thresholds .2 and .3 are identical.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. The models include offer fixed effects and participant random effects.

Hubs versus Less-Frequented Locations: Additional Analyses

Figure 2 in the main manuscript displayed a heat map of locations based on the total number of observations in them. Therein, we also discussed an alternative definition of location popularity based on the number of different subscribers who appeared in a given cell. Figure WA1 displays the corresponding heat map for the same area.

In Table WA6, we present the network statistics for both definitions of location popularity. Note that, since any participant may be co-located (possibly with different alters) at both types of locations on a particular day, the two degree metrics within a network may total to a higher sum than the degree reported in Table 6 in the main manuscript. Finally, for completeness, we also report the detailed results of our regressions corresponding to the networks of popular and infrequently-visited locations in Table WA7.

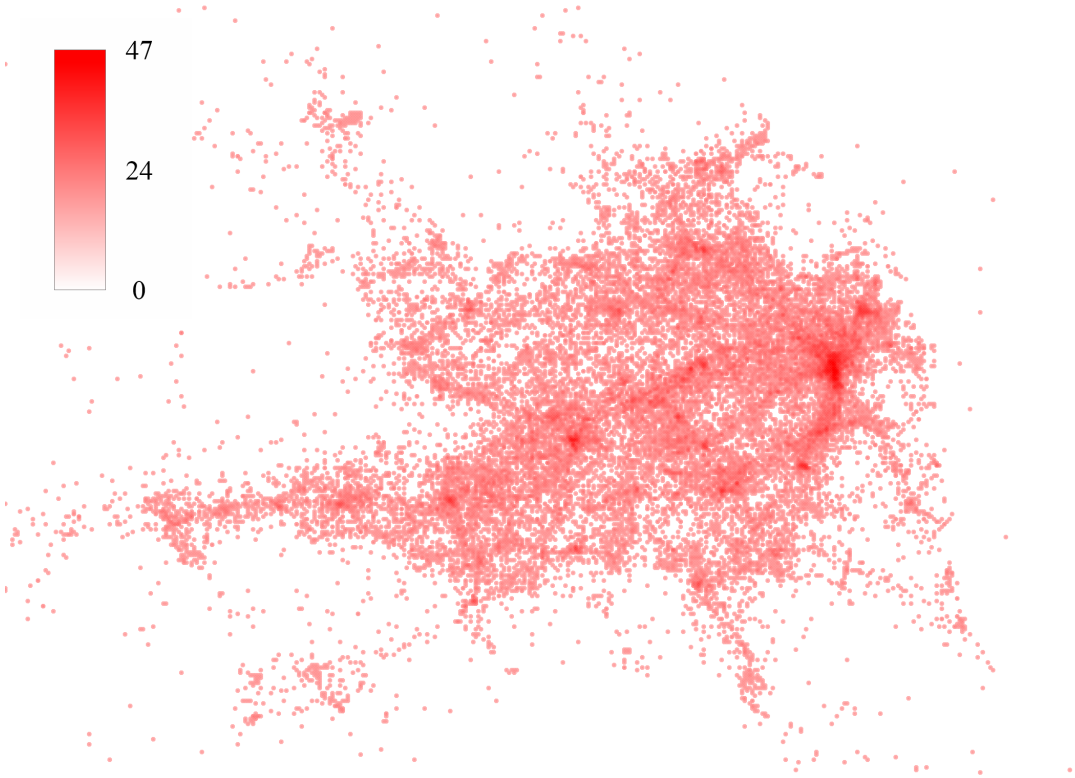

Figure WA1: Heat map of GPS grid cells by the number of subs that visited them during the period studied (in one of the metropolitan areas covered by the program).

Table WA6: Descriptive Statistics of Networks Based on Location Popularity

	Degree			Clustering	
Location Popularity Definition	Mean (st.dev)	Mean number of non-isolate nodes[a]	Mean (st.dev.) if non-isolate	Mean number of nodes defined for[a]	Mean (st.dev) if defined
Total Observations					
– High Frequency	.086 (.34)	11.29	1.21 (.52)	1.98	.688 (.43)
– Low Frequency	.086 (.34)	11.48	1.20 (.50)	1.86	.241 (.40)
Distinct Participants					
– High Frequency	.099 (.41)	10.98	1.441 (.73)	3.65	.449 (.46)
– Low Frequency	.068 (.26)	10.47	1.041 (.22)	.39	.333 (.45)

[a] Mean per offer (across the 96 offers considered).

Table WA7: Parameter estimates, Separate Co-location Networks for High- and Low-Frequency Locations

Variable	**GPS Cells Split by Observations**		**GPS Cells Split by Participants**	
Offer Length	.285*	(2.43)	.285*	(2.44)
Days Since Joined	.016	(1.80)	.016	(1.79)
Category_Redeem_Rate	.764	(1.23)	.688	(1.10)
GPS Activity	.016*	(2.42)	.016*	(2.44)
Coloc_High-Freq_Category_Redeem_Rate	1.916	(1.67)	2.990	(1.53)
Coloc_Low-Freq_Category_Redeem_Rate	4.912**	(3.20)	3.498**	(3.11)
Age	.037	(1.80)	.036	(1.75)
Gender	.724*	(2.16)	.716*	(2.13)
Education	.343*	(2.07)	.350*	(2.11)
Experience	-.129	(-.63)	-.134	(-.65)
Income	-.102	(-.68)	-.104	(-.69)
Hedonic	-.152	(-.56)	-.165	(-.61)
Well-read	-.211	(-1.18)	-.214	(-1.19)
Hard-working	-.160	(-1.18)	-.157	(-1.16)
Spender	.186	(.85)	.178	(.81)
Referral_Category_Redeem_Rate	4.865**	(2.87)	4.698**	(2.76)
Observations	15,353		15,353	
Participants	217		217	
Offers	96		96	
Log Likelihood	-2,056.71		-2,057.14	

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. The models include offer fixed effects and participant random effects.