Using Passively Collected Data to Investigate Social Travel

Thibaut Dubernet

Kay W. Axhausen

February 2019

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

1 Introduction

In developed countries, a continuous increase of the share of trips which are performed for leisure purposes could be observed in the last dozens of years (Schlich *et al.*, 2004; Axhausen, 2005). This represents a challenge for travel behavior modeling, as those trips are much more difficult to capture than commuting trips: they are performed more sporadically, and data about those trips is much more difficult to collect. Understanding better how destination choice for leisure trips is made is therefore essential to improve the accuracy of those forecasts. This increase in leisure travel has been anticipated early, and the social nature of such travel already hypothesized, for instance by Salomon (1985), who stated that *"one particular type of travel, that for* recreational *and* social *purpose, may increase when more leisure time is available"*. This forecast was later confirmed, for instance by Stauffacher *et al.* (2005), who analyzed the motives behind leisure activities, using the results of a Swiss 12 weeks leisure travel diary survey. They found social contact to be the most important, and that in addition respondents traveled with social contacts for more than 70% of leisure activities. This fact, among others, generated a growing interest in the social dimension of travel, and how travel decisions are influenced not only by the global state of the transportation system, but also by joint decisions and interactions with social contacts — a clear sign for this interest being the regular workshops organized on this theme (Dugundji *et al.*, 2008, 2011, 2012; Scott *et al.*, 2013; Goetzke *et al.*, 2015).

Interest in the relationship between mobility, social contacts and leisure behavior is not new (Stutz, 1973; Kemper, 1980), but enjoyed a renewed interest in recent years. Previous studies have been conducted with the idea that an important factor in leisure trip destination choice, or activity duration choice, is the ability to meet social contacts. Examples of empirical work include Carrasco and Habib (2009), Habib and Carrasco (2011) or Moore *et al.* (2013). All those studies show a significant influence of social contacts on the spatial and temporal distribution of activities. In addition, the influence of the social nature of human beings was shown to generate paradoxical effects. For instance, Harvey and Taylor (2000) show that persons working from home tend to travel *further* for leisure purpose, in order to fulfill their need for social contact, that they cannot fulfill at their workplace. A model ignoring such effects might thus substantially underestimate the traveled distances for such individuals.

Typically, co-participants in activities are classified in household and other contacts. Srinivasan and Bhat (2006) analyzed the American Time Use Survey to search for interaction patterns with household members and other contacts. They found that a significant proportion of activities of all types, be it during the week or the week end, are performed jointly. There are however systematic patterns that come out of the data: joint (out of home) activities during the week tend to be performed with non-household members, the opposite being true on the week-end. In addition, activities with household and family members tend to be longer than activities with friends. Kemperman *et al.* (2006) observed the same kind of effect between week-end and week day in the Netherlands.

Households are a typical unit of analysis in economics and transportation, and much work has been conducted on modeling the co-dependence of the mobility behavior of different household members. Those studies often use the classical random utility framework extended to group decision making. A classical way to cope with the possibly conflicting objectives of different members of the household is to specify a group level utility function. For instance, Zhang *et al.* (2005, 2007) develop a model where time for different activity types is allocated to household members, subject to time constraints (including equality of time participation in joint activities), using a group level utility function formulated as a multilinear combination of the individuals' utilities — that is, a linear combination of individual utilities and pair-wise product of individual utilities. Kato and Matsumoto (2009) use a linear combination of the utility functions of the household members as a group utility. The assumption behind this kind of models is the existence of "utility transfers": individuals accept to decrease their own utility if it allows to increase the utility of others by a certain fraction of their loss. Bradley and Vovsha (2005) focus on the "daily activity pattern" generation, with household "maintenance" tasks (*e.g.* shopping) allocation and possibility of joint activities. To do so, they assume a layered choice structure, choosing first a daily activity pattern for each member, and then assigning joint and maintenance activities. Gliebe and Koppelman (2005) also base their model on the daily activity pattern concept, choosing first a "joint outcome" (the sequence of individual and joint activities), and then an individual pattern for each household member. Those models rely on enumeration of the possible household level patterns. Gliebe and Koppelman (2002) also derived a constrained time allocation model, which predicts the time passed by two individuals in joint activities. Rather than postulating a group level utility function, the models of those authors specify a special distribution for the error terms of the individuals. In this setting, the error term of the individuals are correlated so that the probability of choosing a given joint output is the same for all individuals. Ho and Mulley (2013) also estimate models in which members of the household perform choices constrained by the choice of a household level travel pattern. Their data, as well as the parameters of the models, show high joint household activity participation on weekends, and a high dependence of joint travel on trip purpose and household mobility resources. Those results highlight the importance of representing joint household decisions, in particular when going beyond the "typical working day". Vovsha and Gupta (2013) formulate a time allocation model for multiple worker households, which considers a positive utility for members of the household to be home jointly, as it makes joint activities possible. The estimation results show a significant influence of this kind of synchronization mechanism. Most models listed in this paragraph are specific to given household structures; in particular, separate models need to be estimated for different household sizes.

Household level decision processes have also been modeled with approaches which significantly differ from the classical random utility framework. Golob and McNally (1997) propose a structural equation model, which predicts time allocation and trip chaining based on the sociodemographics of a household. Golob (2000) also used a structural equation model to model the dependence of time allocations of the two heads (man and woman) of a household.

Another class of approaches, more oriented toward multiagent simulation than analysis, is the use of optimization algorithms to generate households plans. They handle the household scheduling problem by transforming it into a deterministic utility maximization problem. Contrary to the previously presented approaches, those alternatives do not lead to the estimation of a model against data. The first of those approaches was introduced by Recker (1995). By extending increasingly the formulation of the Pick-Up and Delivery Problem With Time Windows, a well studied combinatorial optimization problem, he formulates the problem of optimizing the activity sequence of members of a household as a mathematical programming problem. However, due to the complexity of the problem, the full problem cannot be solved exactly by standard operations research algorithms, and the activity durations are not part of the optimized dimensions. Chow and Recker (2012) designed an inverse optimization method to calibrate the parameters of this model using measured data. Also, the formulation from Recker (1995) was later extended by Gan and Recker (2008) to introduce the effects of within day rescheduling due to unexpected events. Another attempt to generate plans for households uses a genetic algorithm, building on a previous genetic algorithm for individual plan generation (Charypar and Nagel, 2005; Meister *et al.*, 2005). This algorithm optimizes sequence, duration and activity choice for a household, rewarding the fact that several members of the household perform the same activity simultaneously, in the way also used by Vovsha and Gupta (2013). Finally, Liao *et al.* (2013) formulate the problem of creating schedules for two persons traveling together as finding the shortest path in a “supernetwork”, and solve this problem using exact shortest path algorithms. They however note that their model is specific to the two person problem, and that extension to larger numbers of agents may prove to be computationally expensive. All those approaches remained experimental, and were not integrated into multiagent simulation tools.

Another class of methods aiming at multiagent simulations consists of rule based systems, which use heuristic rules to construct household plans. Miller *et al.* (2005) develop such a model for household mode choice. The main difference with an individual mode choice model is the consideration of household level vehicle allocation. In their model, individuals first choose modes individually. If a conflict occurs, the allocation that maximizes the household level utility is chosen. The members which were not allocated a vehicle will fall back on their second best choice, and/or examine shared rides options. Arentze and Timmermans (2009) develop a rule base model which relies on a simulated bargaining process within the household. Though such models can easily represent complex decision processes, their calibration and validation is cumbersome.

Another field of empirical research studies the spatial characteristics of social networks. For instance, Carrasco *et al.* (2008) studied the relationship between individual's socioeconomic characteristics and the spatial distribution of their social contacts. This kind of empirical work allows to specify and estimate models able to generate synthetic social networks, given sociodemographic attributes and home location. Another kind of data collection is the one of Kowald (2013), that uses the technique of *snowball sampling*, where random individuals are asked to list social contacts, that are in turn contacted and asked the same set of questions. Based on this data, Arentze *et al.* (2012) estimated a model capable of synthetizing social networks with realistic geographical and topological properties. This kind of model is essential if one wants to include social network interactions in microsimulation model.

This integration of social networks in multiagent simulation frameworks has already been attempted by other authors. Due to their disaggregated description of the world, such models are particularly well suited to the representation of complex social topologies. Han *et al.* (2011) present experiments of using social networks to guide activity location choice set formation in the FEATHERS multiagent simulation framework. Using a simple scenario with 6 agents forming a *clique*, they consider the influence of various processes like information exchange and adaptation to the behavior of social contacts to increase the probability of an encounter. They do not, however, represent *joint decisions*, such as the scheduling of a joint activity. The same kind of processes have been investigated by Hackney (2009), using more complex network topologies, within the MATSim framework. Ronald *et al.* (2012) and Ma *et al.* (2011, 2012) present agent based systems which do integrate joint decision making mechanisms, based on rule based simulations of a bargaining processes. Frei and Axhausen (2011) demonstrate a simple joint planning model, where (a) social contacts decide to perform a joint activity if it improves the utility of all co-participants (b) location of a joint activity is chosen to maximise a group utility. They are not yet integrated into any operational mobility simulation platform.

This paper is part of a project that aims at including joint mobility behavior, in particular in the leisure context, into an multi-agent mobility simulation. A method to represent this type of social behavior in the simulation is already implemented (Dubernet, 2017). This paper focuses on the estimation of models of joint travel, with the aim of including them into the simulation. To do so, it relies on a very detailed dataset that allows the exploration of joint travel at an unprecedented level of details.

2 Data

The main difficulty for studying travel to joint activities is the difficulty to find data about travel of all co-participants. Arentze (2014) solved this by designing a SP experiment where respondents are not only presented with their situation, but also with the one of their co-participants.

Moving from stated to revealed data is not easy, as it requires having a high probability to observe co-participants jointly. The data from Stopczynski *et al.* (2014) offers this opportunity. In this study, first-year student were offered a modern smartphone for free use for one year, under the condition that they accepted various forms of data to be collected and analyzed. This data included geographical location (GPS), Bluetooth and WiFi scans, SMS, call and Facebook activity. About one third of the students accepted. This dataset is particularly well suited to the analysis of joint travel for several reasons:

- the target group (freshmen) is likely to be a relatively isolated social group in the city, with significant connectivity between students and few with non-student residents.

- the high sampling rate means that for each student, the probability to have at least one of its social contacts in the dataset is quite high.

Of course, the dataset has its limitations. In particular, it is likely that a substantial number of interactions occur with individuals that are not part of the dataset. This means that one cannot identify "solitary activities" with certainty, and can only estimate a lower bound for party size. Another limitation is that the sample is obviously not representative of the population as a whole. It can however help better understand the phenomena behind joint travel decisions.

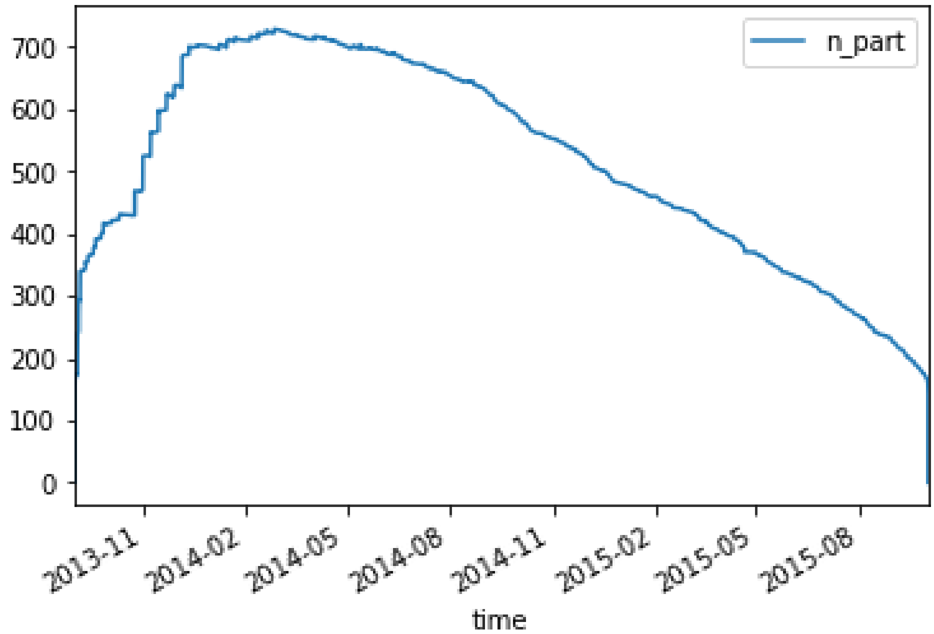
This paper uses this data to analyze joint travel patterns. The relevant subset of the data consists of:

- Mobility data. This data was processed from the raw GPS logs, by first separating traces between movements and stops.

- "Context" data, identifying the type of amenity the respondent performs, based on rule-based home identification, knowledge of the study plan and open street map data.

Although the measurements took place in a period of about two years, the number of participants varied significantly with time, as illustrated in Fig. 1. As a high number of participants is essentials for the analyses conducted here, the focus is put on the months from February to August 2014, which are the months during which the number of active respondents peaks.

Figure 1: Number of Participants with Time



3 Results

3.1 Detection of Communities

Social networks are characterized by communities, or groups of people who often meet together. Identifying those communities is no easy task: groups will form and dissolve over time, and all members of the community do not necessarily need to be present at each and every meeting. In this context, a community is a group of nodes that are often strongly connected to each other. Aslak *et al.* (2018) proposed using a flow formulation on a multi-layer network, where each layer is a time step, to identify recurring communities. This approach was used here, relying on the Infomap software package (www.mapequation.org).

The underlying time-dependent social network was generated from the observed co-presence as follows:

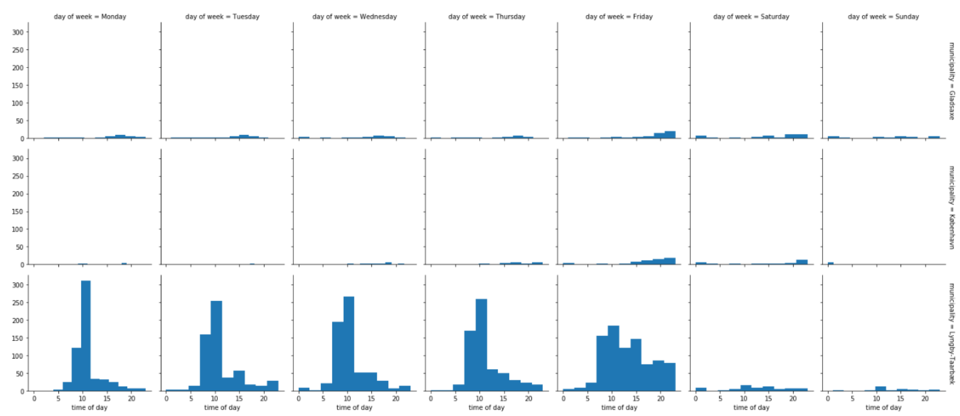
- only stops at locations where food or drinks are known to be served are kept
- stops on the location of classes where the participant is registered are dropped (to avoid artifacts from classrooms located in the vicinity of cafeterias)
- remaining stops that are located less than 20 meters from each other are considered as "copresent".

3.2 Behavior of Communities

The identified joint activities follow a clear spatio-temporal pattern, where most daytime week activities happen on Campus, and night and week-end activities happen with higher probability in the city of Copenhagen itself. This is visible in Fig. 2, which shows the number of activities per time of day, per day of the week, in three municipalities.

The DTU Campus is particular in that it is located at about 15km from the city of Copenhagen itself, with all amenities needed for student life on Campus (including accommodation and dining facilities). As a result, most trips observed in the dataset are very short and do not allow to study the impact of the spatial distribution of opportunities on joint location choice behavior.

Figure 2: Spatio Temporal Pattern



Although the community detection algorithm identifies recurring communities, assigning exactly one community to each meeting results in most communities being "active" only once.

4 Outlook

This study takes advantage of an unusual dataset, where a large fraction of students in a large European university were tracked over a long period. This dataset unfortunately does not allow to look at the mobility patterns of several social contacts participating in a joint activity in great details, due to the concentration of student activities on Campus.

Although the community detection algorithm did not allow to identify actually recurring communities, a model of activity participation is in preparation, where social ties are considered in isolation, without consideration for pre-identified communities.

5 References

Arentze, T. A. (2014) Individuals' social preferences in joint-activity choice: The role of fairness and asymetric evaluation of costs and rewards, paper presented at the *93rd Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2014.

Arentze, T. A., M. Kowald and K. W. Axhausen (2012) A method to model population-wide social networks for large scale activity-travel micro-simulation, paper presented at the *91st Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2012.

Arentze, T. A. and H. J. P. Timmermans (2009) A need-based model of multi-day, multi-person activity generation, *Transportation Research Part B: Methodological*, **43** (2) 251–265.

Aslak, U., M. Rosvall and S. Lehmann (2018) Constrained information flows in temporal networks reveal intermittent communities, *Physical Review E*, **97** (6) 062312.

Axhausen, K. W. (2005) Social networks and travel: Some hypotheses, in K. P. Donaghy, S. Poppelreuter and G. Rudinger (eds.) *Social Dimensions of Sustainable Transport: Transatlantic Perspectives*, Ashgate, Aldershot.

Bradley, M. A. and P. Vovsha (2005) A model for joint choice of daily activity pattern types of household members, *Transportation*, **32** (5) 545–571.

Carrasco, J. A. and K. M. N. Habib (2009) Understanding the social embeddedness of activity-travel participation: The case of frequency and duration of social activities, paper presented at the *12th International Conference on Travel Behaviour Research (IATBR)*, Jaipur, December 2009.

Carrasco, J. A., E. J. Miller and B. Wellman (2008) How far and with whom do people socialize? empirical evidence about the distance between social network members, *Transportation Research Record*, **2076**, 114–122.

Charypar, D. and K. Nagel (2005) Generating complete all-day activity plans with genetic algorithms, *Transportation*, **32** (4) 369–397.

Chow, J. Y. and W. W. Recker (2012) Inverse optimization with endogenous arrival time constraints to calibrate the household activity pattern problem, *Transportation Research Part B: Methodological*, **46** (3) 463 – 479.

Dubernet, T. (2017) Explicitly correlating agent's daily plans in a multiagent transport simulation: Towards the consideration of social relationships, Ph.D. Thesis, ETH Zurich, Zurich.

Dugundji, E. R., A. Páez and T. A. Arentze (2008) Social networks, choices, mobility, and travel, *Environment and Planning B*, **35** (6) 956–960.

Dugundji, E. R., A. Páez and T. A. Arentze (2012) Urban mobility and social-spatial contact — introduction, *Environment and Planning A*, **44** (5) 1011–1015.

Dugundji, E. R., A. Páez, T. A. Arentze, J. L. Walker, J. A. Carrasco, F. Marchal and H. Nakanishi (2011) Transportation and social interactions, *Transportation*, **45** (4) 239–247.

Frei, A. and K. W. Axhausen (2011) Collective location choice model, *Working Paper*, **686**, IVT, ETH Zurich, Zurich.

Gan, L. P. and W. W. Recker (2008) A mathematical programming formulation of the household activity rescheduling problem, *Transportation Research Part B: Methodological*, **42** (6) 571–606.

Gliebe, J. P. and F. S. Koppelman (2002) A model of joint activity participation, *Transportation*, **29** (1) 49–72.

Gliebe, J. P. and F. S. Koppelman (2005) Modeling household activity-travel interactions as parallel constrained choices, *Transportation*, **32** (5) 449–471.

Goetzke, F., R. Gerike, A. Páez and E. R. Dugundji (2015) Social interactions in transportation: Analyzing groups and spatial networks, *Transportation*, **42** (5) 723–731.

Golob, T. F. (2000) A simultaneous model of household activity participation and trip chain generation, *Transportation Research Part B: Methodological*, **34** (5) 355–376.

Golob, T. F. and M. G. McNally (1997) A model of activity participation and travel interactions between household heads, *Transportation Research Part B: Methodological*, **31** (3) 177–194.

Habib, K. M. N. and J. A. Carrasco (2011) Investigating the role of social networks in start time and duration of activities: Trivariate simultaneous econometric model, *Transportation Research Record*, **2230**, 1–8.

Hackney, J. K. (2009) Integration of social networks in a large-scale travel behavior microsimulation, Ph.D. Thesis, ETH Zurich, Zurich.

Han, Q., T. A. Arentze, H. J. P. Timmermans, D. Janssens and G. Wets (2011) The effects of social networks on choice set dynamics: Results of numerical simulations using an agent-based approach, *Transportation Research Part A*, **45** (4) 310–322.

Harvey, A. S. and M. E. Taylor (2000) Activity settings and travel behaviour: A social contact perspective, *Transportation*, **27** (1) 53–73.

Ho, C. and C. Mulley (2013) Tour-based mode choice of joint household travel patterns on weekend and weekday, *Transportation*, **40** (4) 789–811.

Kato, H. and M. Matsumoto (2009) Intra-household interaction in a nuclear family: A utility-maximizing approach, *Transportation Research Part B: Methodological*, **43** (2) 191–203.

Kemper, F. J. (1980) Social contacts of an urban-population within an activity-space framework, *Geographische Zeitschrift*, **68** (3) 199–222.

Kemperman, A., T. A. Arentze and H. J. P. Timmermans (2006) Social commitments and activity-travel scheduling decisions, *Transportation Research Record*, **1977**, 242–249.

Kowald, M. (2013) Focusing on leisure travel: the link between spatial mobility, leisure acquaintances and social interactions, Ph.D. Thesis, IVT, ETH Zurich.

Liao, L., T. A. Arentze and H. J. P. Timmermans (2013) Multi-state supernetwork framework for the two-person joint travel problem, *Transportation*, **40** (4) 813–826.

Ma, H., T. A. Arentze and H. J. P. Timmermans (2012) Incorporating selfishness and altruism int dynamic joint activity-travel scheduling, paper presented at the *13th International Conference on Travel Behaviour Research (IATBR)*, Toronto, July 2012.

Ma, H., N. Ronald, T. A. Arentze and H. J. P. Timmermans (2011) New credit mechanism for semicooperative agent-mediated joint activity-travel scheduling, *Transportation Research Record*, **2230**, 104–110.

Meister, K., M. Frick and K. W. Axhausen (2005) A GA-based household scheduler, *Transportation*, **32** (5) 473–494.

Miller, E. J., M. J. Roorda and J. A. Carrasco (2005) A tour-based model of travel mode choice, *Transportation*, **32** (4) 399–422.

Moore, J., J. A. Carrasco and A. Tudela (2013) Exploring the links between personal networks, time use, and the spatial distribution of social contacts, *Transportation*, **40** (4) 773–788.

Recker, W. W. (1995) The household activity pattern problem: General formulation and solution, *Transportation Research Part B: Methodological*, **29** (1) 61–77.

Ronald, N., T. A. Arentze and H. J. P. Timmermans (2012) Modelling social interactions between individuals for joint activity scheduling, *Transportation Research Part B: Methodological*, **46** (2) 276–290.

Salomon, I. (1985) Telecommunications and travel: Substitution or modified mobility?, *Journal of Transport Economics and Policy*, **19** (3) 219–235.

Schlich, R., S. Schönfelder, S. Hanson and K. W. Axhausen (2004) Structures of leisure travel: Temporal and spatial variability, *Transport Reviews*, **24** (2) 219–237.

Scott, D. M., E. R. Dugundji and A. Páez (2013) The social dimension of activity, travel and location choice behavior, *Journal of Transport Geography*, **31**, 212–215.

Srinivasan, S. and C. R. Bhat (2006) Companionship for leisure activities: an empirical analysis using the american time use survey, paper presented at the *Innovations in Travel Demand Modeling (ITM'06)*, 129 – 136, Austin, May 2006.

Stauffacher, M., R. Schlich, K. W. Axhausen and R. W. Scholz (2005) The diversity of travel behavior: Motives and social interactions in leisure time activities, *Working Paper*, **328**, IVT, ETH Zurich, Zurich.

Stopczynski, A., V. Sekara, P. Sapiezynski, A. Cuttone, M. M. Madsen, J. E. Larsen and S. Lehmann (2014) Measuring large-scale social networks with high resolution, *PLoS ONE*, **9** (4) e95978.

Stutz, F. P. (1973) Intra-urban social visiting and leisure behavior, *Journal of Leisure Research*, **5** (1) 6.

Vovsha, P. and S. Gupta (2013) A model for work activity schedules with synchronization for multiple-worker households, *Transportation*, **40** (4) 827–845.

Zhang, J., H. J. P. Timmermans and A. W. J. Borgers (2005) A model of household task allocation and time use, *Transportation Research Part B: Methodological*, **39** (1) 81–95.

Zhang, J., H. J. P. Timmermans and A. W. J. Borgers (2007) Utility-maximizing model of household time use for independent, shared, and allocated activities incorporating group decision mechanisms, *Transportation Research Record*, **1807**, 1–8.