Sociodynamic discrete choice on networks in space: impacts of agent heterogeneity on emergent outcomes

Elenna R Dugundji
AMIDSt—Amsterdam Institute for Metropolitan and International Development Studies, Faculty of Social and Behavioral Sciences, Universiteit van Amsterdam, Nieuwe Prinsengracht 130, 1018 VZ Amsterdam, The Netherlands; e-mail: e.r.dugundji@uva.nl

László Gulyás
AITIA Inc.—Artificial Intelligence, Information Technology, Intelligent Agents, Czet Janos u. 48-50, 1039 Budapest, Hungary; e-mail: lgulyas@aitia.ai

Received 10 February 2006; in revised form 12 August 2008

Abstract. The reported research treats interactions between households and generated feedback dynamics in the adoption of various transportation mode alternatives. We consider a model in which an agent's choice is directly influenced by the percentages of the agent's neighbors and socioeconomic peers that make each choice. The model also accounts for shared unobserved attributes of the choice alternatives in the error structure. We address nonglobal interactions explicitly within different social and spatial network structures, combining advanced econometric estimation with computational techniques from multiagent-based simulation, and we present an empirical application of the model using pseudo-panel microdata collected by the Amsterdam Agency for Traffic, Transport and Infrastructure. Additional heterogeneity is introduced in the model through different mechanisms, such as individual-specific sociodemographic characteristics of the agents, individual-specific attributes of the choice alternatives, and the availability of alternatives. We conclude by highlighting limitations of our present study and recommendations for future work.

1 Introduction
A wide spectrum of policy measures have been put forward over the past decade to try to address the infamous rush-hour congestion in the ‘Randstad’, the western region of the Netherlands marked by the ring of cities Amsterdam – Utrecht – The Hague – Rotterdam. These measures include flexible work hours, congestion pricing, light rail, the facilitation of park-and-ride schemes, and road construction. The research reported here is a small part of a larger work aimed at understanding, measuring, and modeling the combined residential choice and transportation mode choice behavior of households residing in the greater Amsterdam region. The focus of the larger work is on the promotion and facilitation of multimodal transportation as a land-use transportation planning policy instrument for reducing road congestion (Joh, 2004; Krygsman, 2004; Maat et al, 2004; Timmermans et al, 2002). The contribution of this particular subproject is the treatment of social and spatial interactions between households and generated feedback dynamics in the adoption of various transportation mode alternatives.

Pioneered in the domain of travel demand by Ben-Akiva (1973), Domencich and McFadden (1975), and others, discrete-choice analysis has become an industry standard in land-use and transportation planning models. Some subsequent elegant and elaborate operational examples of the development of this methodology are due to Wegener (1996), Waddell (2002), Martinez and Aguila (2004), Hensher and Ton (2001), to cite just a few. Meanwhile, the field itself has flourished in the past thirty years, ultimately extending the basic random utility model to incorporate cognitive and behavioral processes, flexible error structures, and different types of data in so-called hybrid choice models (Ben-Akiva et al, 1999; 2002). However, as discrete-choice theory
is fundamentally grounded in individual choice, an outstanding challenge remains in the treatment of the interdependence of various decision makers' choices, be that via global or local interactions. The formulation of the nature of the interaction in turn raises the issues of networks, and of network evolution. When considering the domain of land use and transportation, social networks and spatial networks may be relevant (Dugundji et al, 2001).

Some examples of research issues we might like to address, in relation to interhousehold networks, include spatial coordination and social awareness or acceptance in the adoption of various transportation mode choices. If a certain critical mass of households is willing to choose public transit in a particular region or at a particular park-and-ride location, it can become economically viable to provide a high level of public transit service to that region or from that park-and-ride location. Being able to guarantee a high level of service might then in turn attract additional households. On the other hand, the lack of a sufficient transit ridership base can be a reason for a poor level of service, which in turn might discourage transit use by segments of the population that have other reasonable transportation mode alternatives at their disposal, which in turn could lead to further cutbacks in level of service, and so on. Such interhousehold feedback thus can have very important implications for the prediction of (system-wide) results over the course of time. If such feedback exists it can propel or hinder the adoption of a mode over time. In diverse literature this dynamically reinforcing behavior is referred to as a social multiplier, a cascade, a bandwagon effect, imitation, contagion, herd behavior, etc (Manski, 1995).

In the spirit of Aoki (1995), Brock and Durlauf (2001; 2002), and Blume and Durlauf (2002), we consider a model in which an agent's choice to adopt a discrete behavior or to buy a discrete product is influenced by the percentages of the agent's reference entities that make each choice. An important extension with respect to earlier work is that we develop results for a case in which there are shared unobserved attributes of the choice alternatives. We revisit a classic approach to statistical prediction in such a situation, given an observed sample of decision-making agents in a population—namely the nested logit model. Additionally, a key feature of our work is that we explicitly consider nonglobal interactions, in an example with a social and spatial network structure that we can visualize and analyze using geographic information systems tools and techniques.

We present an application of the model to transportation mode choice using pseudo-panel microdata collected in the greater Amsterdam region during the period 1992–97. Here we combine advanced econometric estimation (Dugundji and Walker, 2005) with computational techniques from the field of multiagent-based simulation. This paper extends our earlier work (Dugundji and Gulyás, 2003a; 2003b) by exploring effects of social geography and additional heterogeneity (beyond that induced via the nonglobal interactions on the sociospatial network) introduced in the model through different mechanisms, such as individual-specific sociodemographic characteristics of the agents, individual-specific attributes of the choice alternatives, and the availability of alternatives. Finally, we conclude by highlighting limitations of our present study in any extension for policy considerations on the adoption of innovation in transportation mode choice, and we give our suggested recommendations for future work.

2 Model
Since the early theoretical work by Aoki (1995), Brock and Durlauf (2001), and Blume and Durlauf (2002) on binary discrete-choice models, there have been a few extensions addressing both the complexity of the discrete-choice model kernel as well as the
complexity of the feedback effect and the utility specification. For example, Brock and Durlauf (2002) have extended their results on the behavior of binary logit models to multinomial logit models. Dugundji (2003; 2004) makes Brock and Durlauf’s multinomial results precise for trinary multinomial choice and extends the results for the case of a nested logit model with global interactions. Also, while the behavior over time derived in early work assumed that each decision maker was influenced by all other decision makers (so-called global interactions), we (Dugundji and Gulyás, 2003a; 2003b) have derived more general behavior for the case in which each decision maker is influenced by only a subset of decision makers (so-called nonglobal interactions).

Importantly, however, a key to the theoretical results is the assumption that the only explanatory variable in the model is the feedback effect. Though such a specification may be plausible if modeling a fad, it is much less intuitive for transportation mode choice, in which other explanatory variables could be assumed to be significant— including attributes of the alternatives, such as travel time, and characteristics of the decision-making agents, such as gender, age, and income. In a previous paper (Dugundji and Gulyás, 2005) we thus presented results using simulated data for a binary logit model with nonglobal interactions and other explanatory variables included in the utility for a parameter sweep of network density across a series of networks in the abstract class of random networks. In the current paper we present results for the behavior over time of a nested logit model with nonglobal interactions, using empirical data and an empirical treatment in which decision makers influence each other on the basis of socioeconomic group and spatial proximity of residential location. Table 1 highlights the contribution of the current work relative to the context of existing literature along the dimensions of the discrete-choice model kernel, the specification of the systematic utility, the interaction framework, and the approach to studying behavior over time.

An economic issue that arises in the empirical estimation of a feedback effect in discrete-choice models using multinomial logit or nested logit models, however, is that the error terms are assumed to be identically and independently distributed across decision makers. It is not obvious that this is in fact a valid assumption when we are specifically considering interdependence between decision makers’ choices. We might reason that, if there is a systematic dependence of each decision maker’s choice on an explanatory variable that captures the aggregate choices of other decision makers who are in some way related to that decision maker, as considered in the literature referenced above, then there might be an analogous dependence in the error structure. Stated alternatively, the same unobserved effects might be likely to influence the choice made by a given decision maker as well as the choices made by those in the decision maker’s reference group. In terms of transportation mode choice, for example, accessibility measures for residents in the same neighborhood could play such a role to the extent that they are unable to be captured directly through explanatory variables in the utility specification. In this case the use of transportation mode shares of neighbors living in the same zone as an explanatory variable will be correlated with the unobserved error of the given decision maker, which is a classic case of endogeneity. The results and coefficients of such a model are likely to be biased. To try to separate out effects, it is therefore first and foremost critically important to begin with a well-specified model which makes use of relevant available explanatory variables. In a companion paper to this one using exactly the same empirical dataset, Dugundji and Walker (2005) explore issues in the estimation of discrete-choice models with feedback effects, by specifically testing for correlation among agents in the error structure through the use of mixed generalized extreme-value family models. The reader interested in finer points of discrete-choice theory is referred to the companion paper.
In the current paper the focus is not estimation of the discrete-choice model, but rather the simulated evolution of choice behavior over time with positive feedback owing to network effects, using computational techniques from the field of multiagent-based simulation. To achieve this, the discrete-choice estimation results controlling overall mechanisms related to individual heterogeneous preferences are embedded in a computational version of the model, which was created using the Repast multiagent-based modeling platform (http://repast.sourceforge.net).

We begin the model specification by reviewing some basic notation and assumptions in choice modeling in section 2.1. We deliberately choose the nested logit model as a starting point for the current paper in order to extend incrementally the existing literature, rather than jumping immediately to a more complex discrete-choice model. In section 2.2 we present a strategy for introducing social and spatial network interdependencies in choice models as a generic feedback effect. In section 3 we review the data available to us, we make a hypothesis regarding a network of commonality

<table>
<thead>
<tr>
<th>Discrete choice kernel</th>
<th>Systematic utility</th>
<th>Behavior over time</th>
<th>Interaction framework</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 1. Contribution of the current work in the contest of existing literature.

<table>
<thead>
<tr>
<th>Discrete choice kernel</th>
<th>Systematic utility</th>
<th>Behavior over time</th>
<th>Interaction framework</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Binary logit feedback effect only**
  - equilibrium solution only
  - transition dynamics
  - Brock and Durlauf (2001)
  - Dugundji and Gulyás (2003a; 2003b)

- **Multinomial logit feedback effect only**
  - equilibrium solution only
  - transition dynamics
  - Brock and Durlauf (2002)
  - Dugundji (2003; 2004)

- **Nested logit feedback effect only**
  - transition dynamics
  - Dugundji (2003; 2004)

- **Cross-nested logit not addressed**

- **Empirically defined**
  - transition dynamics

In the current paper the focus is not estimation of the discrete-choice model, but rather the simulated evolution of choice behavior over time with positive feedback owing to network effects, using computational techniques from the field of multiagent-based simulation. To achieve this, the discrete-choice estimation results controlling overall mechanisms related to individual heterogeneous preferences are embedded in a computational version of the model, which was created using the Repast multiagent-based modeling platform (http://repast.sourceforge.net).

We begin the model specification by reviewing some basic notation and assumptions in choice modeling in section 2.1. We deliberately choose the nested logit model as a starting point for the current paper in order to extend incrementally the existing literature, rather than jumping immediately to a more complex discrete-choice model. In section 2.2 we present a strategy for introducing social and spatial network interdependencies in choice models as a generic feedback effect. In section 3 we review the data available to us, we make a hypothesis regarding a network of commonality

---

By the designation ‘empirically defined’ is understood a fully specified systematic utility containing the feedback effect as well as attributes of alternatives and/or characteristics of decision-making agents. Model specifications without the feedback effect are not relevant for the purposes of this particular work.

Nonidentifiable: the feedback effect is perfectly correlated with an alternative specific constant (see Dugundji and Gulyás, 2005).
between different agents in different places on the basis of common attributes in terms of social grouping between places, and we estimate coefficients for the models presented in section 2. In section 4 we consider the evolution of choice behavior over time. Although it is important to begin with a well-specified model before adding subsequent elaborations, the techniques presented here could be applied to any choice model. In this paper we will find that the simulated evolution of choice behavior over time with positive feedback owing to network effects is critically sensitive to the discrete choice estimation results defining the individual heterogeneous preferences. Heterogeneity matters! The emergent outcomes are, remarkably, dramatically different: (1) with the inclusion of heterogeneity in the systematic utility (in the extension of the model from one including the feedback effect only to one including empirically defined systematic utility); and also (2) with the inclusion of shared unobserved heterogeneity (in the extension of the model from a multinomial logit to a nested logit). This finding in turn opens a wide spectrum of research along various dimensions for more detailed understanding of the impact of agent heterogeneity on emergent outcomes for any policy-related considerations.

2.1 Nested logit: shared unobserved heterogeneity
Discrete-choice theory allows prediction on the basis of computed individual choice probabilities for heterogeneous agents' evaluation of alternatives. In accordance with the notation and convention in Ben-Akiva and Lerman (1985), the so-called nested logit model is specified as follows. Assume a sample of $N$ decision-making entities indexed $n = 1, \ldots, N$ each faced with a single choice among mutually exclusive elemental alternatives $i$ in the choice subset $C_n$ of some universal choice set $C$. Suppose that the choice set $C_n$ faced by decision-making entity $n$ is partitioned into $M$ mutually exclusive and collectively exhaustive `nests' $C_{mn}$ of elemental alternatives indexed $j = 1, \ldots, J_m$, which are assumed to be correlated.

$$
C = \{C_1, C_2, \ldots, C_M\},
C_m \cap C_{m'} = \emptyset, \quad m \neq m',
\bigcup_{m=1}^{M} C_m = C.
$$

(1)

In general the composite choice set $C_n$ will vary in size and content across agents: not all elemental alternatives $i$ in the universal choice set may be available to all agents. The overall correlation structure of alternatives is, however, assumed to be the same across agents, aside from availability. See figure 1.

![Figure 1](image-url)

**Figure 1.** Depiction of the two-level nested logit model, partitioned into $M$ mutually exclusive and collectively exhaustive ‘nests’ of elemental alternatives indexed $j = 1, \ldots, J_m$ which are assumed to be correlated. The nested logit model will reduce to the multinomial logit model if the scale parameter $\mu_m$ for the lower nests is equal to the scale parameter $\mu$ for the upper nest.
Let $U_{in} = V_{in} + e_{in}$ be the utility that a given decision-making entity $n$ is presumed to associate with a particular elemental alternative $i$ in its choice set $C_n$, where $V_{in}$ is the deterministic (to the modeler) or so-called ‘systematic’ utility and $e_{in}$ is an error term. The error term represents unobserved heterogeneity. Such unobserved heterogeneity may arise from unobserved attributes of the choice alternatives, unobserved characteristics of the decision-making entities, or simply measurement errors in observed attributes and/or characteristics. Also, in the case where instrumental variables are used as a proxy for variables which are not observable, the error term is relevant for capturing unobserved heterogeneity.

Without loss of generality, the systematic utility is assumed to be defined by a linear-in-parameters function $V_{in} = h_i + V(S_{in}, z_{in})$ of observable characteristics $S_{in}$ of the decision-making entity and observable attributes $z_{in}$ of the choice alternative for a given decision-making entity. The term $h_i$ is an alternative specific constant included to explicitly account for any underlying bias for one alternative over another alternative. In other words, $h_i$ reflects the mean of $e_{jn} - e_{in}$ — that is, the difference in the utility of alternative $i$ from that of $j$ when all else is equal.

Similarly, let $U_{mn} = V_{mn} + e_{mn}$ be the composite utility that the given decision-making entity $n$ is presumed to associate with a particular choice subset $C_{mn}$. This assumption thus allows for the possibility of observed and unobserved heterogeneity which is shared among the elemental alternatives within a nest. In section 3 we will study an empirical application of trinary mode choice in which the elemental alternatives are public transit, bicycle, and automobile. A shared attribute not observed in the particular subset of raw data made available to the authors might be, for example, crow’s flight distance between residential location and work or business location. We might hypothesize, namely, that the distance above a certain threshold might give a shared preference for public transit and automobile modes over bicycle.

Let $A_{in}$ be an alternative availability indicator variable defined as 1 if elemental alternative $i$ is available to decision-making entity $n$ and defined as 0 otherwise. As derived in chapter 5.2 of Ben-Akiva and Lerman (1985), under the assumption of Gumbel-distributed disturbances $e_{in}$ with a scale parameter $\mu_{in}$, the probability $P(i|C_{mn})$ that agent $n$ chooses alternative $i$ within nest $C_{mn}$, conditional on having chosen that nest, has a convenient closed-form expression, given by:

$$P(i|C_{mn}) = \frac{A_{in} \exp(\mu_{in} V_{in})}{\sum_{j \in C_{mn}} A_{jn} \exp(\mu_{in} V_{jn})}. \quad (2)$$

As derived in chapter 10.3 of Ben-Akiva and Lerman (1985), under the assumption of Gumbel-distributed disturbances $e_{mn}$ with a scale parameter $\mu$, the probability $P(C_{mn}|C_n)$ that agent $n$ chooses the particular choice nest $C_{mn}$ among the set of $M$ nests also has a convenient closed-form expression, given by:

$$P(C_{mn}|C_n) = \frac{\exp(\mu V_{mn})}{\sum_{m' \in M} \exp(\mu V_{m'n})} \exp \left\{ \mu \left[ \tilde{V}_{mn} + (1/\mu) \ln \sum_{j \in C_{mn}} A_{jn} \exp(\mu_{in} V_{jn}) \right] \right\}$$

$$= \frac{\exp \left\{ \mu \left[ \tilde{V}_{mn} + (1/\mu) \ln \sum_{j \in C_{mn}} A_{jn} \exp(\mu_{in} V_{jn}) \right] \right\}}{\sum_{m' \in M} \exp \left\{ \mu \left[ \tilde{V}_{m'n} + (1/\mu_{in}) \ln \sum_{j \in C_{m'n}} A_{jn} \exp(\mu_{in} V_{jn}) \right] \right\}}. \quad (3)$$
The joint probability \( P(i|C_n) \) that the decision-making entity \( n \) chooses elemental alternative \( i \) within the nest \( C_{mn} \) among all possible alternatives in its choice set \( C_n \) is then given by:

\[
P(i|C_n) = P(i|C_{mn})P(C_{mn}|C_n) \tag{4}
\]

It is the allowance for the possibility of *shared unobserved heterogeneity* at the nest level which is the signature of the nested logit model. Note, however, that, without loss of generality, the shared observable attributes at the nest level may be defined at the elemental alternative level. This is what is done in practice in typical model estimation software packages, such as the freely available optimization toolkit Biogeme developed by Bierlaire (http://roso.epfl.ch/biogeme). Similarly, the explicit definition of availability at the nest level is superfluous. Also note that the nested logit model will reduce to the multinomial logit model if the scale parameter \( \mu_m \) for the lower nests is equal to the scale parameter \( \mu \) for the upper nest.

\[
P(i|C_n) = \frac{A_{in} \exp(\mu_{in})}{\sum_{\forall j \in C_{mn}} A_{jn} \exp(\mu_{jn})} \exp \left[ \mu \tilde{V}_{mn} + (\mu/\mu_m) \ln \sum_{\forall j \in C_{mn}} A_{jn} \exp(\mu_{jn}) \right],
\]

where \( \mu_m = \mu \).

\[
P(i|C_n) = \frac{A_{in} \exp(\mu_{in})}{\sum_{\forall j \in C_{mn}} A_{jn} \exp(\mu_{jn})} \exp(\mu \tilde{V}_{mn}) \exp \left[ \ln \sum_{\forall j \in C_{mn}} A_{jn} \exp(\mu_{jn}) \right],
\]

2.2 Interaction mechanism framework: induced heterogeneity

We introduce a feedback effect among agents by allowing the systematic utility \( V_{in} \) to be a first-order function, linear in parameter \( \beta \), of the proportion \( x_{in} \) of a given decision maker’s reference entities, who have chosen elemental alternative \( i \).

\[
V_{in} = \beta f(x_{in}) = \beta x_{in} + h_i + V(S_n, z_{in}) \tag{6}
\]

The variable \( x_{in} \) is termed a ‘field variable’. As motivated by Aoki, “Knowledge of a field variable relieves agents (at least partially) of the need for detailed information on interaction patterns. Any macroeconomic variable that serves this decentralizing function is called a field variable” (page 151). The seminal work of Aoki (1995), Brock and Durlauf (2001; 2002) and Blume and Durlauf (2002) assumes either a global (fully connected) network, a uniform network, or a mean-field effect averaging out individual variations. That is, the proportions evolve in time, but all agents perceive the same proportions \( x_i \) at a given moment in time, namely the mode shares taken across the entire sample. Our model extends the seminal work in that we consider nonglobal interactions.
In our case, even when the only variable in the systematic utility is the field variable, heterogeneity is *induced*, since different agents perceive different proportions at the same moment in time, depending on the choices of their particular reference entities. In other words, the subscript \( n \) on the variable \( x_{in} \) is significant in our case, whereas in the seminal work mentioned above the subscript is superfluous. In fact, in our case, even agents who happen to be within the same reference group \( g \) perceive (slightly) different proportions of reference agents \( n' \) choosing each elemental alternative \( i \), because an agent’s own choice is not included in these proportions. Let \( y_{in} \) be a choice indicator variable defined as 1 if decision-making entity \( n \) chooses elemental alternative \( i \) and defined as 0 otherwise. If \( N_g \) is the number of agents in a reference group \( g \), then in our case:

\[
  x_{in} = \frac{1}{N_g} \sum_{n' \in g, n' \neq n} y_{in'}.
\] (7)

Provided suitable data are available, the approach could be further generalized—for example, to allow different weights on the ties with different reference entities.

In the conceptualization of the interaction framework a distinction is hypothesized between social versus spatial interactions and between identifiable versus aggregate interactions (Dugundji and Gulyás, 2003a; 2003b). We speak of interaction between ‘identifiable’ decision makers when the links in the network are well known and explicitly defined on an individual decision maker by decision maker basis. We speak of interaction between ‘aggregate’ decision makers when interdependence is assumed to take place only at an aggregate level, with links being defined, for example, more generally on the basis of decision-maker characteristics. We speak of ‘spatial’ network interactions when the interdependence represents a confluence of decision makers in geographic terms. For example, decision makers may be linked on the basis of spatial proximity of residential location, work location, or some other geographical point of reference, such as school, childcare, shopping, healthcare, leisure/recreation, or other relevant activity location. We speak of ‘social’ network interactions when decision makers are linked on the basis of social circles. The decision makers need not be proximally situated in geographical terms and the interaction is not necessarily centered at a particular geographic point of reference; interaction may take place at a distance, so to speak.

The framework for the mechanisms of interaction is proposed as follows:
- Interactions among individuals within households: for example, joint residential location choice in a dual-income earner household (Timmermans et al, 1992); coordinating activity schedules and travel patterns within a household.
- Interactions between identifiable households proximally situated in a spatial network: for example, nuisance from neighbors and, conversely, satisfaction with neighbors are very strong factors in the inclination to move house, both for under 55 years and for over 55 years age groups in the Netherlands (Hooimeijer and van Ham, 2000); coordinating carpooling with neighbors or coworkers.
- Interactions between identifiable households associated in a social network, not necessarily proximally or tangentially situated in a spatial network: for example, attraction to a particular municipality in the choice of residential location because friends or family live there; awareness about the availability of certain alternatives in the choice-set generation process through information transmission in the social network—via friends, family, neighbors, and/or coworkers—whether that be the suitability of a particular neighborhood in residential location choice, the suitability of using a park-and-ride transferium for a commute, or the existence of a carpool facility.
- Interactions between a household and the aggregate actions of other households proximally situated in a spatial network: for example, high volatility or, conversely, stagnancy of turnover in housing stock in a particular neighborhood affecting the general desirability of a neighborhood or the possibility to move there; social pressure to own a car because other neighbors or other coworkers, on average, do, regardless of whether there is any direct social contact with these persons; improved feasibility for a higher level of public transit service associated with a higher volume of public transit ridership in a particular region.

- Interactions between a household and the aggregate actions of other households associated in a social network, not necessarily proximally or tangentially situated in a spatial network: for example, preference for a particular type of housing situation (as opposed to preference for a specific municipality); social acceptance of cycling or public transit because friends, family, neighbors, and/or coworkers also cycle or use public transit.

- Interactions between a household and the aggregate actions of other households in a (sub)population, not necessarily associated in a social network nor proximally or tangentially situated in a spatial network: because of a general trend or societal bandwagon effect.

Furthermore, an important distinction can be understood in this particular problem domain among (social and/or spatial) network interactions which impact on choices, such as transport mode choice, which do not necessarily endogenously affect the household's reference position in a network (e.g., whether a household chooses carpool versus transit in a multimodal trip, or chooses a unimodal trip, will not spatially affect the fact of who the household's neighbors or coworkers are), as opposed to network interactions impacting ‘sorting’ type choices, such as residential location choice, which obviously endogenously impact the household's reference position in a spatial network and potentially also within a social network (e.g., in moving to a new neighborhood, a household by definition acquires new neighbors).

In summary, illustrative examples of such interactions along the above described dimensions for the *exogenous* network case in the given problem domain—that is, transportation mode choice—are provided in table 2. The nested logit model is seen as a promising direction for coupling the exogenous network case (transportation mode choice) and the endogenous network case (residential choice).

The research reported here explores interactions between a decision maker and the aggregate actions of other decision makers proximally situated in a spatial network, and interactions between a decision maker and the aggregate actions of other decision makers associated in a socioeconomic network (mechanisms 2 and 4 in table 2). Technically, however, interactions between identifiable decision makers (mechanisms 1

### Table 2. Interaction mechanism framework—some illustrative examples.

<table>
<thead>
<tr>
<th>Interaction between ...</th>
<th>Identifiable decision makers</th>
<th>Aggregate interdependence a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial network b</td>
<td>1 coordinating carpooling with neighbors</td>
<td>2 feasibility of high level of public transit service</td>
</tr>
<tr>
<td>Social network c</td>
<td>3 awareness about mode choice alternatives</td>
<td>4 social acceptance of cycling or transit</td>
</tr>
</tbody>
</table>

a Global interactions between a decision maker and the aggregate actions of other decision makers in the entire sample population (general societal bandwagon effects) may be addressed as the special limiting case of a fully connected network.

b Intrahousehold interactions can be seen as a particular special case of spatial interactions.

c Not necessarily proximally situated in a spatial network; interaction may take place at a distance.
Typically, survey data for interaction between identifiable decision makers would include explicit information on the relevant networks for each decision maker for the decision of interest. The members of the networks might then in turn be surveyed. In travel-demand data collection a typical practice is to sample households from the population and then survey all members of that household above a certain age. As of the time of embarking on the current research, we were unaware of any travel-demand datasets that would take, for example, a snowball sampling approach collecting explicit information on inter-household networks of decision makers. As suggested in table 2, some examples of research questions we might like to answer relating to interhousehold networks include spatial coordination or feasibility and social awareness or acceptance in the take-up of various transportation mode choices. In the absence of survey data on interaction between identifiable decision makers at the interhousehold level, we turn instead to consider aggregate interactions between decision makers and use a priori beliefs about the social and/or spatial dimension of interactions to formulate the connectivity of the network.

In the case study to be discussed, we have rich socioeconomic data for each respondent as well as the geographic location of each respondent’s residence. This allows

**Figure 2.** Commuter mode count for decision makers grouped by residential district. Sampled residential locations are given in terms of the centroid of a traffic analysis zone, indicated in the figure with small gray circles. There are nine sampled residential districts each with a distinctively characteristic population in part reflective of the period of construction of the district. Adhering to the original numbering of districts in the raw dataset, these are: (1) Amsterdam Center, (2) Amsterdam West, (3) Amsterdam South, (4) Amsterdam East, (6) Amsterdam North, (8) Amsterdam Far West, (9) Amstelveen, (10) Amsterdam Southeast, (16) Far Amstelveen. The number of respondents in the sample per district is indicated in the inset. The nonsampled regions of the Municipality of Amsterdam represent the industrial harbor district to the west, the commercial office park district to the southeast, and the primarily agricultural district to the north.
us to define aggregate interactions by grouping agents into geographic neighborhoods or into socioeconomic groups, in which the influence is assumed to be more likely. Figure 2 illustrates transportation mode shares for decision makers in the sample grouped by residential district.

In the simplest case these groups are assumed to be mutually exclusive and collectively exhaustive and each agent $n$ belongs to one and only one group $g$. The agent is influenced by the average choice behavior of his or her group, and the influence of other groups is assumed to be negligible. At a global level the picture is a fragmented or disconnected network of clustered groups. If we are interested in equilibrium behavior, the consequences of such an assumption are important: there is no transmission of influence across groups, and the global picture is a weighted average behavior of the separate clusters. Thus, we consider the case with overlapping groups, with agents, for example, connected by social group as well as by residential district. This leads to a giant cluster for the empirical example under consideration, with the important implication that influence can spread throughout the entire population. Such a network is abstractly visualized in figure 3 using the freely available software packet Pajek, developed by Batagelj and Mrvar (http://vlado.fmf.uni-lj.si/pub/networks/pajek). Nodes represent fully connected residential districts, as in figure 1; the darkness and width of lines gives an indication of the number of links between districts induced by socioeconomic group. In section 3.1 we will review the socioeconomic and geographic data available to us. In section 3.2 we will make a hypothesis regarding a network of commonality between different agents in different residential districts on the basis of common attributes in terms of social grouping between districts. In section 3.3 we will estimate coefficients for the model with field variables for residential district and socioeconomic group. In section 3.4 we will estimate a simple nested logit model with global (fully connected) network and with the only explanatory variable in the systematic utility being the field variable.

Figure 3. Abstract visualization of network interdependence defined by residential district plus socioeconomic group, using the freely available software packet Pajek developed by Batagelj and Mrvar (http://vlado.fmf.uni-lj.si/pub/networks/pajek). Nodes represent fully connected residential districts; the numbering of the districts indicated in parentheses corresponds to the same district numbering as in figure 2. The darkness and width of the lines give an indication of the number of links between districts induced by socioeconomic group; this is in turn a reflection both of the number of respondents in the sample living in a given district indicated in figure 2, and of the similarity of the districts in terms of residential group stratification.
3 Empirical application

The data used in this paper originate from travel questionnaires administered by the Municipality of Amsterdam Agency for Infrastructure, Traffic and Transport (dIVV) during the period 1992–97 in Amsterdam and a neighboring suburb to the south of the city, Amstelveen. In the absence of survey data on interaction between identifiable decision makers at interhousehold level, we consider aggregate interactions and use a priori beliefs to group agents into geographic neighborhoods and socioeconomic groups in which the influence is assumed to be more likely. Nonglobal field variables are defined on the basis of the overlapping groups and model coefficients are estimated.

3.1 Data

The dataset made available by the Municipality of Amsterdam is a subset of the full modal split database, containing direct home–work trips and direct work–home trips, where the purpose of the trip at the nonhome location is classified as either ‘work’ or ‘business’. Geographical location is given in terms of the centroid of a traffic analysis zone. The data received include records of trips for which respondents have indicated one of the following transportation mode choices:

- pt—public transit, including external transit system modes (intercity train, intercity bus) as well as internal system modes (tram, metro, local train, local bus);
- bi—bicycle mode (travel by moped or motorcycle is also included in this category);
- au—automobile driver or automobile passenger.

The final sample used in the case study contains 2913 decision-making agents. Raw variables available for use in the model are described in table 3.

### Table 3. Description of the raw variables in the transportation mode choice dataset.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type of variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Available indicator for public transit mode</td>
<td>$A_{ni}, i = pt$</td>
<td>1 if public transit alternative is available for agent $n$, 0 otherwise</td>
</tr>
<tr>
<td>Car ownership</td>
<td>$S_n$</td>
<td>1 if agent $n$ owns an automobile, 0 otherwise</td>
</tr>
<tr>
<td>Gender</td>
<td>$S_n$</td>
<td>1 if agent $n$ is female, 0 if male</td>
</tr>
<tr>
<td>Income</td>
<td>$S_n$</td>
<td>income range of agent $n$ coded using governmental classification</td>
</tr>
<tr>
<td>Age</td>
<td>$S_n$</td>
<td>age range of agent $n$: 12–17 years; 18–29 years; 30–44 years; 45–59 years; 60 years and older</td>
</tr>
<tr>
<td>Education</td>
<td>$S_n$</td>
<td>education level achieved by agent $n$: elementary education; lower vocational education; high school education; post high school education; other</td>
</tr>
<tr>
<td>Residential location</td>
<td>$S_n$</td>
<td>residential location of agent $n$ given by centroid of a traffic analysis zone</td>
</tr>
<tr>
<td>In-vehicle time for public transit mode</td>
<td>$z_{ni}, i = pt$</td>
<td>in-vehicle travel time in minutes by public transit</td>
</tr>
<tr>
<td>Out-of-vehicle time for public transit mode</td>
<td>$z_{ni}, i = pt$</td>
<td>out-of-vehicle travel time by public transit in minutes (access, egress, waiting, transferring, etc)</td>
</tr>
<tr>
<td>Travel time for bicycle mode</td>
<td>$z_{ni}, i = bi$</td>
<td>travel time in minutes by bicycle</td>
</tr>
<tr>
<td>Travel time for automobile mode</td>
<td>$z_{ni}, i = au$</td>
<td>travel time in minutes by automobile</td>
</tr>
<tr>
<td>Parking time for automobile mode</td>
<td>$z_{ni}, i = au$</td>
<td>time in minutes to park automobile</td>
</tr>
</tbody>
</table>

Note. pt, public transit; bi, bicycle; au, automobile.
3.2 Field variables for residential district and socioeconomic group

Next we turn to the specification of the aggregate interdependence. We begin with the broad classification by residential district, as shown earlier in figure 2. There are 9 districts represented in the sample, ranging in size from 223 sampled respondents to 461 sampled respondents. The mean size is 323 respondents with standard deviation 74, skewness 0.32, and kurtosis 0.19. Next, using the three variables age, income, and education, 13 socioeconomic groups are defined (see table 4). The socioeconomic groups range in size from 99 sampled respondents to 385 sample respondents. The mean size is 224 respondents with standard deviation 111, skewness 0.33, and kurtosis –1.8.

In both figure 2 and table 4 we find distinct clusters of agents for which the commuter mode share per district and per socioeconomic group deviates notably from the overall modal split in the sample as a whole. For example, in district 1 (Amsterdam Center) the mode share by bicycle is highest, whereas at the periphery of the metropolitan region, in district 8 (Amsterdam Far West) and in district 16 (South Amstelveen), the mode share by car is the highest. This is intuitive with respect to the urban density in general, and concentrations of residential locations and work locations in particular. What is not immediately obvious from the geographical distributions of mode shares per district in the thematic map in figure 2 is how much of the mode share is due to attributes of the transportation alternatives and characteristics of the decision makers, and how much of this may be due to social influence by neighbors and residential self-selection. For the socioeconomic group of commuters with above-average incomes in table 4, it is similarly intuitive that the mode share by car is higher than those groups with the same age category and same education level with lower incomes. Again, what is not immediately obvious simply from studying the cross-tabulation in table 4 is how much of the variation in mode shares is due to having more spending power for travel by car and how much is due to social influence—to travel by car when income is a certain level, and to status pressures.

Table 4. Commuter mode share and sample count by socioeconomic group.

<table>
<thead>
<tr>
<th>Socioeconomic groupa</th>
<th>Public transit</th>
<th>Bicycle</th>
<th>Automobile</th>
<th>Sample count</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 – 29, LO/LBO</td>
<td>0.31</td>
<td>0.24</td>
<td>0.45</td>
<td>112</td>
</tr>
<tr>
<td>12 – 29, MO/other</td>
<td>0.32</td>
<td>0.27</td>
<td>0.41</td>
<td>385</td>
</tr>
<tr>
<td>12 – 29, HO</td>
<td>0.34</td>
<td>0.30</td>
<td>0.36</td>
<td>329</td>
</tr>
<tr>
<td>30 – 44, LO/LBO</td>
<td>0.24</td>
<td>0.20</td>
<td>0.55</td>
<td>177</td>
</tr>
<tr>
<td>30 – 44, MO/other, 0 – zkf</td>
<td>0.29</td>
<td>0.26</td>
<td>0.45</td>
<td>353</td>
</tr>
<tr>
<td>30 – 44, HO, 0 – zkf</td>
<td>0.21</td>
<td>0.41</td>
<td>0.37</td>
<td>361</td>
</tr>
<tr>
<td>30 – 44, MO/other, zkf+</td>
<td>0.17</td>
<td>0.11</td>
<td>0.72</td>
<td>115</td>
</tr>
<tr>
<td>30 – 44, HO, zkf+</td>
<td>0.14</td>
<td>0.22</td>
<td>0.63</td>
<td>338</td>
</tr>
<tr>
<td>≥ 45, LO/LBO</td>
<td>0.16</td>
<td>0.23</td>
<td>0.61</td>
<td>175</td>
</tr>
<tr>
<td>≥ 45, MO/other, 0 – zkf</td>
<td>0.21</td>
<td>0.27</td>
<td>0.52</td>
<td>175</td>
</tr>
<tr>
<td>≥ 45, HO, 0 – zkf</td>
<td>0.20</td>
<td>0.35</td>
<td>0.46</td>
<td>101</td>
</tr>
<tr>
<td>≥ 45, MO/other, zkf+</td>
<td>0.15</td>
<td>0.15</td>
<td>0.70</td>
<td>99</td>
</tr>
<tr>
<td>≥ 45, HO, zkf+</td>
<td>0.17</td>
<td>0.24</td>
<td>0.59</td>
<td>193</td>
</tr>
<tr>
<td>Total</td>
<td>690</td>
<td>779</td>
<td>1444</td>
<td>2913</td>
</tr>
</tbody>
</table>

*a Defined on the basis of age category in years, education level, and income category. Education is coded: elementary education and lower vocational education (LO/LBO); high school education and other (MO/other); post high school education (HO). Income category is based on the Dutch governmental classification: zkf+ indicates above-average incomes, 0 – zkf is all other incomes. Where income level is not explicitly specified, respondents from all incomes falling in the given age and education group are included.
To understand these questions we need to delve deeper. One way of doing this is by estimating various discrete-choice models with different specifications (Dugundji and Walker, 2005).

In the absence of detailed data on the exact interaction framework between identifiable decision makers at interhousehold level, we turn instead to consider aggregate interactions between decision makers. We hypothesize a network of commonality between different agents in different residential districts, based on common attributes in terms of social grouping between districts. This way we take into account, in a general way, possible socioeconomic influences and possible feedback owing to local neighborhood effects. Three field variables are defined, as in table 5.

### Table 5. Definition of field variables for transportation mode choice model.

<table>
<thead>
<tr>
<th>Variable name&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Type of variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{sdptnsl}$ $x_n$, $i = pt$</td>
<td>share of agent’s fellow district residents and socioeconomic peers in the sample choosing to commute by public transit</td>
<td></td>
</tr>
<tr>
<td>$d_{sdbinsl}$ $x_n$, $i = bi$</td>
<td>share of agent’s fellow district residents and socioeconomic peers in the sample choosing to commute by bicycle</td>
<td></td>
</tr>
<tr>
<td>$d_{sdaunsl}$ $x_n$, $i = au$</td>
<td>share of agent’s fellow district residents and socioeconomic peers in the sample choosing to commute by automobile</td>
<td></td>
</tr>
</tbody>
</table>

Note. pt, public transit; bi, bicycle; au, automobile; nsl denotes ‘no self-loops’.

<sup>a</sup>The designation ‘dsd’ (district sociodemographics) refers to the hypothesized network of commonality between different agents in different residential districts, on the basis of common attributes in terms of social grouping between districts. This is to be contrasted with the designation ‘full’ later in table 8 in the definition of a global field variable for the treatment with a fully connected or uniform network. In Dugundji and Walker (2005), various other treatments are additionally hypothesized and estimated.

The designation ‘nsl’ in table 5 refers to ‘no self-loops’. That is, the mode choice of a given decision-making agent $n$ is not included when considering the average behavior of the reference group perceived by that agent. In equation (7) this is given by explicitly taking the sum over all $n'$ in reference group $g$, for $n' \neq n$. We do this to avoid endogeneity. In practice, however, if the reference group is very large, the influence of the agent’s own choice in the group average will be negligible. The effect of self-loops versus no self-loops is studied in detail in Dugundji and Gulyás (2005) across a range of network densities.

### 3.3 Specification of utility functions and estimation of model coefficients

Various piecewise linear specifications of all travel-time-related variables, as well as age, were tested against linear, quadratic, and logarithmic forms of these variables. Similarly, various generic forms of the categorical variables were tested against alternative specific forms. Considering various a priori hypotheses of behavior in the greater Amsterdam region, and after statistical comparison of the alternative nonlinear specifications of variables against the linear versions thereof, using loglikelihood ratio tests and nonnested tests (Ben-Akiva and Lerman, 1985), the definitions in table 6 of observable characteristics $S_n$ of the decision-making agents, observable attributes $z_{in}$ of the choice alternatives for a given decision-making agent, and alternative specific constants $h_i$ are ultimately used in a baseline multinomial logit model, in addition to the field variables $x_{in}$ defined in table 5.
In addition, an availability indicator variable for the bicycle mode is defined, using the raw variable for travel time for the bicycle mode, as: 1 if travel time by bicycle for decision-making entity \( n \) is less than seventy-five minutes, and 0 otherwise.

Using the notation of uppercase letters to denote estimated coefficients and the notation of lowercase letters to denote variables, the linear-in-parameters systematic utilities for the baseline model are thus specified as follows for internal/external system public transit (pt), bicycle/motorcycle/moped (bi), and automobile driver/passenger (au) modes (see table 6 for definitions):

\[
\begin{align*}
V_{\text{PT}} &= (\text{DSDNSL} \times \text{dspdptsl}) + (\text{ASC}_\text{PT} \times \text{aspt}) + (\text{GEND}_\text{PT} \times \text{gender}) \\
&\quad + (\text{AGE4559}_\text{PT} \times \text{age4559}) + (\text{LNAGE}_\text{PT} \times \text{lnage}) \\
&\quad + (\text{IVTSQ}_\text{PT} \times \text{ivtsqpt}) + (\text{OVT}_\text{PT} \times \text{ovtpt}) , \\
V_{\text{BI}} &= (\text{DSDNSL} \times \text{dsdbinsl}) + (\text{AOWMIN}_\text{BI} \times \text{aowmin}) + (\text{TT}_\text{BI} \times \text{ttbi}) , \\
V_{\text{AU}} &= (\text{DSDNSL} \times \text{dsdaunsl}) + (\text{ASC}_\text{AU} \times \text{ascau}) + (\text{CAROWN}_\text{AU} \times \text{carown}) \\
&\quad + (\text{GEND}_\text{AU} \times \text{gender}) + (\text{LNTT}_\text{AU} \times \text{lnttau}) \\
&\quad + (\text{PARKSQ}_\text{AU} \times \text{parksqua}) .
\end{align*}
\]

### Table 6. Definition of constants, characteristics and attributes in mode choice model.

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Type of variable</th>
<th>Definition a</th>
</tr>
</thead>
<tbody>
<tr>
<td>aspt</td>
<td>( h_{i, i = \text{pt}} )</td>
<td>1 if alternative ( i ) is the public transit mode, 0 otherwise</td>
</tr>
<tr>
<td>ascau</td>
<td>( h_{i, i = \text{au}} )</td>
<td>1 if alternative ( i ) is the automobile mode, 0 otherwise</td>
</tr>
<tr>
<td>carown</td>
<td>( S_n )</td>
<td>1 if agent ( n ) owns an automobile, 0 otherwise, defined in the systematic utility for the automobile mode</td>
</tr>
<tr>
<td>gender</td>
<td>( S_n )</td>
<td>1 if agent ( n ) is female, 0 if male, defined alternative specifically for public transit mode and automobile mode</td>
</tr>
<tr>
<td>aowmin</td>
<td>( S_n )</td>
<td>1 if income category AOW social minimum, 0 otherwise, defined in the systematic utility for the bicycle mode</td>
</tr>
<tr>
<td>lnage</td>
<td>( S_n )</td>
<td>natural logarithm of age in years as given by midpoint of the age category, defined in the systematic utility for public transit</td>
</tr>
<tr>
<td>age4559</td>
<td>( S_n )</td>
<td>age 45 to 59 defined piecewise continuously in the systematic utility for public transit: ( \max[0, \min(\text{age} - 45, 15)] )</td>
</tr>
<tr>
<td>ivtsqpt</td>
<td>( z_{i, i = \text{pt}} )</td>
<td>in-vehicle travel time in minutes by public transit, squared, defined in the systematic utility for public transit</td>
</tr>
<tr>
<td>ovtp</td>
<td>( z_{i, i = \text{pt}} )</td>
<td>out-of-vehicle travel time by public transit in minutes, defined in the systematic utility for public transit</td>
</tr>
<tr>
<td>ttbi</td>
<td>( z_{i, i = \text{bi}} )</td>
<td>travel time in minutes by bicycle, defined for bicycle mode</td>
</tr>
<tr>
<td>lnttau</td>
<td>( z_{i, i = \text{au}} )</td>
<td>natural logarithm of travel time in minutes by automobile, defined in the systematic utility for automobile</td>
</tr>
<tr>
<td>parksqua</td>
<td>( z_{i, i = \text{au}} )</td>
<td>time in minutes to park automobile, squared, defined in the systematic utility for automobile</td>
</tr>
</tbody>
</table>

Note. pt, public transit; bi, bicycle; au, automobile. AOW refers to the income level of the basic state pension for people aged 65 years and over.

a It is possible to identify maximum \( J-1 \) alternative specific constants \( h_i \) and maximum \( J-1 \) alternative specific variables for each characteristic \( S_n \) of the decision-making agents. For our case of a trinary model choice model, this yields a maximum of 2 in both cases. The reason why defining three alternative specific constants in our trinary choice case would be nonidentifiable is because all that matters is the difference between the alternative specific constants \( h_j - h_i \) for elemental alternatives \( i, j \) in \( J \). Similarly, defining three alternative specific variables for a given characteristic \( S_n \) would be nonidentifiable, because all that matters is the effect of the alternative specific variables on the relative utility between the elemental alternatives.

In addition, an availability indicator variable for the bicycle mode is defined, using the raw variable for travel time for the bicycle mode, as: 1 if travel time by bicycle for decision-making entity \( n \) is less than seventy-five minutes, and 0 otherwise.

Using the notation of uppercase letters to denote estimated coefficients and the notation of lowercase letters to denote variables, the linear-in-parameters systematic utilities for the baseline model are thus specified as follows for internal/external system public transit (pt), bicycle/motorcycle/moped (bi), and automobile driver/passenger (au) modes (see table 6 for definitions):
After the estimation of the baseline multinomial logit model, the estimation of three successive nested logit models—first with public transit nested with bicycle, then with public transit nested with car, and finally with bicycle nested with car—shows the first nesting structure to be the most significant in terms of log-likelihood ratio test and in terms of a $t$-test on the nest coefficient (see figure 4). The third structure was not indicated. The nested logit model thus adds one additional parameter to the multinomial specification, namely the scale parameter $\mu$ for the transit–bicycle nest. Table 7 provides estimation results for the baseline multinomial logit and final nested logit model.

![Figure 4. Depiction of the nesting structure of a trinary discrete-choice model with unobserved heterogeneity that is shared between internal/external system public transit and bicycle/motorcycle/moped mode alternatives. It is not possible to identify the scale parameter both of the upper level nest and of the lower level nest(s), since what matters is only the ratio; in practice, the upper level nest is typically fixed to 1.](image)

| Table 7. Estimation results with sociogeographic network interdependence. |
|---------------------------|---------------------------|---------------------------|
| Variable description | Multinomial logit | Nested logit |
| | coefficient estimate | $t$-statistic | coefficient estimate | $t$-statistic |
| Share of each agent’s fellow residents and sociodemographic peers in the sample choosing each mode, defined generically | 1.91 | 4.54 | 1.93 | 5.59 |
| Alternative specific constant, defined for transit | 0.15 | 0.18 | 0.20 | 0.50 |
| Alternative specific constant, defined for car | 0.32 | 0.65 | -1.11 | -2.14 |
| Car ownership, defined for car | 2.54 | 24.68 | 2.53 | 24.84 |
| Gender, defined for transit | 0.56 | 4.64 | 0.24 | 3.12 |
| Gender, defined for car | 0.45 | 3.70 | 0.28 | 2.46 |
| Low income, defined for bicycle | -0.48 | -2.92 | -0.17 | -1.87 |
| Natural logarithm of age, defined for transit | -0.72 | -3.10 | -0.30 | -2.12 |
| Age 45 to 59 defined piecewise continuously, for transit | $4.09 \times 10^{-2}$ | 2.13 | $1.94 \times 10^{-2}$ | 1.80 |
| In-vehicle time, squared, defined for transit | $-3.95 \times 10^{-4}$ | -4.42 | $-2.90 \times 10^{-4}$ | -3.68 |
| Out-of-vehicle time, defined for transit | $-2.52 \times 10^{-2}$ | -2.87 | $-1.91 \times 10^{-2}$ | -3.26 |
| Travel time, for bicycle | $-8.10 \times 10^{-2}$ | -14.96 | $-3.75 \times 10^{-2}$ | -4.38 |
| Natural logarithm of travel time, for car | $-1.40 \times 10^{-2}$ | -7.11 | -0.50 | -1.97 |
| Parking time, squared, for car | $-1.17 \times 10^{-2}$ | -7.51 | $-1.36 \times 10^{-2}$ | -8.35 |
| Scale parameter, for transit-bicycle nest | – | – | 2.51 | 2.48 |

Note. Summary statistics—initial log-likelihood $= -2977$; final log-likelihood multinomial logit model $= -2063$; likelihood ratio test multinomial logit model $= 1829$; final log-likelihood nested logit model $= -2055$; likelihood ratio test nested logit model $= 1844$.

All $t$-statistics are against 0, except for the scale parameter, which is against 1.
Most estimated coefficients in table 7 for the multinomial logit versus the nested logit model are within two standard errors of each other. A noteworthy exception is the scale parameter estimated for the nested logit model (four standard errors difference). We will see in section 4 that the inclusion of shared unobserved heterogeneity in this way has a critical impact on the simulated evolution of choice behavior over time.

3.4 Reference model with fully connected network and global field variable only

For the purpose of comparison in the extension from existing literature and ‘docking’ of the multiagent-based simulation to be developed in section 4, we also estimate a simple nested logit model with a fully connected network and with the only explanatory variable in the systematic utility being the field variable. Since we do not yet include any sociodemographic information about the agents, since we do not yet include agent-specific attributes of the home-to-work or work-to-home trips, since we do not yet take availability of transportation mode alternatives for specific agents into consideration, since the network is assumed here to be fully connected, and since the model includes self-loops—that is, each agent counts its own choice in evaluating the choices made by reference agents—the descriptions of the systematic utilities of the agents are perfectly homogeneous. This is a case for which the steady-state solutions of the sociodynamic system can be solved analytically, as derived in Dugundji (2003; 2004). Three global field variables are defined, as in table 8.

Table 8. Definition of global field variables for reference mode choice model.

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Type of variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>fullptsl</td>
<td>$x_i, i = pt$</td>
<td>mode share of all agents in the sample choosing to commute by internal and/or external system public transit</td>
</tr>
<tr>
<td>fullbisl</td>
<td>$x_i, i = bi$</td>
<td>mode share of all agents in the sample choosing to commute by bicycle, motorcycle, or moped</td>
</tr>
<tr>
<td>fullausl</td>
<td>$x_i, i = au$</td>
<td>mode share of all agents in the sample choosing to commute by automobile, either as a driver or as a passenger</td>
</tr>
</tbody>
</table>

The designation ‘sl’ refers to ‘self-loops’. That is, the mode choice of a given decision-making agent $n$ is indeed included here when considering the average behavior of the reference group perceived by that agent, for the special case of a fully connected network, where the reference group is the entire sample. As a consequence, there is no variation of the field variable among agents with a fully connected network. That is why it is not possible to estimate an alternative specific constant in this case, as mentioned earlier in table 1. The field variable will be perfectly correlated with the alternative specific constant in the case of a global (fully connected) or uniform network.

The reason why we include self-loops in this case is because if we did not include self-loops the field variables would be a perfect predictor of choice when considering a fully connected network. Endogeneity is assumed not to be relevant here, since the influence of the agent’s own choice in the average over the entire sample will be negligible.

Using the notation of uppercase letters to denote estimated coefficients and the notation of lowercase letters to denote variables, the linear-in-parameters systematic utilities for the simple nested logit model are specified as follows for internal/external system public transit (pt), bicycle/motorcycle/moped (bi), and automobile driver/passenger (au) modes:

$$V_{PT} = \text{FULLSL} \times \text{fullptsl},$$  
$$V_{BI} = \text{FULLSL} \times \text{fullbisl},$$  
$$V_{AU} = \text{FULLSL} \times \text{fullausl}.$$  
(9)
The estimation of three successive nested logit models—first with public transit nested with bicycle, then with public transit nested with automobile, and finally with bicycle nested with automobile—indicates the second nesting structure to be valid, namely, a transit–automobile nest (see figure 5). Table 9 provides the estimation results for this simple nested logit model.

Figure 5. Depiction of the nesting structure with unobserved heterogeneity that is shared between public transit and automobile driver/passenger model alternatives. Such a shared attribute not observed in the particular subset of raw data made available for analysis might be, for example, crow's flight distance (above a certain threshold) between residential location and work or business location. The inclusion of travel time attributes in the systematic utilities, however, may be a reasonable proxy, and thus a possible reason why the model estimated in section 3.3 indicates a different nesting structure. In the case of a well-specified model, the nesting structure is able to capture any remaining shared unobserved heterogeneity at a more refined level.

Table 9. Estimation results for simple nested logit model with fully connected network.

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Coefficient estimate</th>
<th>Standard error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share of agents in the entire sampling choosing each mode, with self loops, defined generically</td>
<td>2.76</td>
<td>0.16</td>
<td>1.78 (against 0)</td>
</tr>
<tr>
<td>Scale parameter for transit–automobile nest</td>
<td>1.03</td>
<td>0.05</td>
<td>0.67 (against 1)</td>
</tr>
</tbody>
</table>

Note. Summary statistics—null log-likelihood = −3200; final log-likelihood = −3035; likelihood ratio test = 331.

Note that, in the case of the simple nested logit model with the only explanatory variable in the systematic utility being the global field variable, the nesting structure in figure 5 is differently indicated than in the earlier case in figure 4 with the full empirical model. The reason for this is that, without other explanatory variables in the definition of the systematic utilities, the nesting structure is forced to try to capture grossly the effect of all of these omitted variables as shared unobserved heterogeneity. In the case of a well-specified model, the nesting structure is able to capture any remaining shared unobserved heterogeneity at a more refined level. The fine-tuned shared unobserved heterogeneity, when observed characteristics and observed attributes are properly accounted for, may have a different structure than that of the gross shared unobserved heterogeneity for which there is little in the model.

4 Simulated evolution of choice behavior over time
Using the Repast multiagent-based modeling platform (http://repast.sourceforge.net), we created computational versions of the models presented in section 2. The discrete-choice estimation results from section 3 controlling overall mechanisms related to the individual preferences were embedded in the multiagent-based computational model.
In this way we were able to study the simulated evolution of choice behavior over time with positive feedback owing to network effects.

Related methodological applications which explore agent-based modeling approaches to social influence in general on networks are found in Hedstrom (2004), Rolfe (2004), Rahmandad and Sterman (2004), Stauffer (2001), Flache and Hegselmann (2001), Deffuant et al (2002), Urbig (2003), and Stauffer et al (2004), among others in a large stream of interdisciplinary research on opinion dynamics addressed both by social scientists and by physical scientists alike. This paper notably differs from these other works in its focus on issues that arise in the empirical application of multiagent-based models when discrete-choice estimation results control the mechanism related to individual preferences.

In section 4.1 we investigate the simulated temporal dynamics for the simple nested logit model estimated in section 3.4 with the global field variable only. In section 4.2 we investigate the simulated temporal dynamics for the baseline empirically defined multinomial logit model estimated in section 3.3. In section 4.3 we investigate the simulated temporal dynamics for the empirically defined nested logit model estimated in section 3.3.

4.1 Temporal dynamics for nested logit model with global field variable only
Using the Repast multiagent-based modeling platform, we investigated the simulated temporal dynamics for the simple nested logit model estimated in section 3.4 with the fully connected network and with the only explanatory variable in the model being the field variable. Example time series results are shown in figure 6.

Figure 6. Example time series with different random seeds for the simple nested logit model, with only a generic global field variable in the systematic utilities for each mode. The black series represents the mode share of agents choosing to commute by automobile either as a driver or passenger. The light gray series represents the mode share of agents choosing to commute by bicycle/moped/motorcycle. The medium gray series represents the mode share of agents choosing to commute by internal and/or external system public transit. Each run is allowed to iterate for 600,000 time steps. This is, on average, 200 revisions of choices with asynchronous decision making for the sample size of roughly 3000 agents.
The temporal dynamics show a moderately slow transition to either of two steady states. One of the emergent equilibrium solutions has a mode share for bicycle of 0.70 and mode shares for public transit and automobile each of 0.15. The other emergent equilibrium solution has a mode share of approximately 0.70 for automobile and mode shares for public transit and bicycle each of approximately 0.15.

Note that, because of the symmetry of the system in which public transit and automobile are nested together and the agents are otherwise homogeneous, at any mode share value for which there is an equilibrium solution for public transit, there should theoretically exist a dual equilibrium solution with an analogous mode share value for automobile, and vice versa. This implies that there should be a third equilibrium solution with a mode share of approximately 0.70 for public transit and mode shares for automobile and bicycle each of approximately 0.15. However, given that the initial starting conditions for the sample are almost 50% for automobile commuters, and less than 25% for public transit users, we find, in the computational simulations, that this third stable solution never has a chance to emerge.

When the global field variable is the only component of the systematic utility, there are simplifications that make the steady-state behavior analytically solvable. In earlier work Dugundji (2003; 2004) has indeed derived five equilibrium solutions of the simple nested logit dynamic system with the particular estimated parameter values from table 9. The stable steady-state solutions 1 and 2 in table 10 are confirmed in figure 6. Solution 3 is the theoretically expected solution which we do not find in practice in the computer simulations, as mentioned in the previous paragraph. For random seeds 5 and 6 in figure 6 we do see the temporary appearance of the saddle node solution 4 at the beginning of the time series, although, as expected, the solution does not persist. Saddle node solution 5 never has an opportunity to emerge because of the values of the initial mode shares, and would not be expected to persist anyway.

It is also instructive to note that, since the scale parameter for the transit–automobile nest in table 9 is close to unity, the emergent mode shares in the first equilibrium solution of 70% for bicycle commuters and 15% each for public transit and automobile commuters are analogous to the emergent equilibrium solution with a mode share of approximately 70% for automobile commuters and approximately 15% each for public transit and bicycle commuters. In numerical terms the emergent mode shares are close to a perfectly symmetrical case of a trinary multinomial logit model. The significant impact, however, of the scale parameter for the public transit–automobile nest being 1.03 instead of exactly 1 is that the first equilibrium solution with 70% for bicycle commuters has a chance to emerge. In a pure multinomial case, although the equilibrium solution exists theoretically, the solution would not emerge in practice, since the initial starting conditions for the sample are almost 50% for automobile commuters, and less than 25% for bicycle commuters. Without the nesting to account for shared unobserved heterogeneity between public transit and automobile, in a perfectly homogeneous multinomial case with initial conditions of almost 50% automobile commuters,

<table>
<thead>
<tr>
<th>Solution</th>
<th>Stability</th>
<th>Mode share</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>bicycle</td>
<td>transit</td>
<td>automobile</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(most) stable</td>
<td>0.700</td>
<td>0.150</td>
<td>0.150</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>stable</td>
<td>0.158</td>
<td>0.143</td>
<td>0.698</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>stable</td>
<td>0.158</td>
<td>0.698</td>
<td>0.143</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>saddle node</td>
<td>0.267</td>
<td>0.237</td>
<td>0.496</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>saddle node</td>
<td>0.267</td>
<td>0.496</td>
<td>0.237</td>
<td></td>
</tr>
</tbody>
</table>

Table 10. Analytical benchmark equilibrium solutions for the simple nested logit model.
the positive feedback effect of the automobile mode in a global (fully connected) or uniform network will, over time, always ‘win’ in practice over the other two modes with initial conditions of mode share of about 25%.

Although this simple model is too stylized to be applied seriously for policy purposes in transportation mode choice without any explanatory variables in the model except the hypothesized field variable, the exercise of estimating the simple model, investigating the emergent behavior over time and corroborating the computational results with analytical results is good practice in that it gives us confidence that the computational model is behaving as we expect. The exercise is also useful as a baseline in understanding corner solutions in parameter space before proceeding to more complex empirical models.

4.2 Transition dynamics for a multinomial logit model with an empirically defined systematic utility and an empirically defined field variable for aggregate interaction

We next consider a more empirically interesting case with heterogeneous agents. The model is the empirically defined trinary multinomial logit transportation mode choice model, with choice alternatives public transit (pt), bicycle/motorcycle/moped (bi), and automobile driver/passenger (au), and the coefficients estimated in section 3.3. Using the Repast multiagent-based modeling platform, we investigate the simulated temporal dynamics over time owing to the hypothesized aggregate feedback effect among agents defined by residential district and socioeconomic group. Example time series results for different random seeds are shown in figure 7 for the multinomial logit case.

![Figure 7. Example time series with different random seeds for the trinary multinomial logit model with empirically defined systematic utilities for each mode and an empirically defined field variable for aggregate interaction. The black series represents the mode share of agents choosing to commute by automobile either as a driver or passenger. The light gray series represents the mode share of agents choosing to commute by bicycle/moped/motorcycle. The medium gray series represents the mode share of agents choosing to commute by internal and/or external system public transit. Each run is allowed to iterate for 600 000 time steps. This is, on average, 200 revisions of choices with asynchronous decision making for the sample size of roughly 3000 agents.](image-url)
The temporal dynamics in figure 7 show the relatively rapid emergence of only one single equilibrium solution with a mode share of approximately 60% for automobile commuters, approximately 25% for public transit, and approximately 15% for bicycle commuters. The general picture in figure 7 is thus notably different than that in figure 6 along four different dimensions: the magnitude of the ‘winning’ mode share; the emergence of only one steady-state solution; the symmetry breaking of the ‘nonwinning’ mode shares; and the speed of the transition to the steady state.

We have already noted in section 4.1 that the simulated behavior over time of the simple nested logit model estimated in table 9 is close to multinomial, since the scale parameter on the nesting structure is only 1.03. Because the value of the coefficient on the field variable is 1.91 for the multinomial logit model in table 7, as compared to the value of 2.76 in the simple nested logit model in table 9—that is, since the coefficient is lesser in magnitude in table 7—we might expect that the emergent effect on mode share of the ‘winning’ mode is lesser in magnitude, which it indeed is.

Also, as noted in section 4.1, and since we now have a true multinomial logit model in section 4.2, it is understandable that the automobile mode always ‘wins’, given the existing initial conditions of almost 50% automobile commuters; the other modes never have a chance to ‘win’ over time over the automobile mode given the starting conditions—especially since the coefficient on the feedback is even less strong here than in section 4.1.

We see in section 4.2 that the symmetry of the mode share of ‘nonwinning’ modes in the emergent equilibrium solution is broken. Instead of the mode share of the ‘nonwinning’ modes in the steady-state solution each being (approximately) 15%, as in figure 6, we see that, in figure 7, the mode share of public transit commuters is consistently higher, at a level of approximately 25%, than the mode share of bicycle commuters at approximately 15%. This symmetry breaking is due to the heterogeneity of the agents in the model estimated in table 7 with fully specified, empirically defined systematic utilities, as compared with the homogeneous agent case in table 9 with only a generic global field variable in the systematic utilities.

Finally, we note that the convergence to a steady-state solution is relatively rapid in figure 7, as compared with figure 6. Since the coefficient on the field variable is less strong here than in section 4.1, the rapidity of the convergence would not seem to be purely due to the magnitude of the coefficient on the feedback. We hypothesize two distinct mechanisms for the rapidity of the convergence phenomenon. As with the symmetry breaking of the ‘nonwinning’ mode shares phenomenon, one hypothesis could again be the fact of the observed heterogeneity of the agents in the fully specified, empirically defined systematic utilities, as compared with the homogeneous agent case. Another hypothesis could be simply the efficiency of information flow in the smaller reference group defined by residential district and socioeconomic group versus the time necessary for information flow in the entire sample. Further research is necessary to separate these two effects. Dugundji (2007) compares the behavior over time of information flow with various hypothesized empirically defined aggregate interaction effects.

4.3 Transition dynamics for a nested logit model with an empirically defined systematic utility and an empirically defined field variable for aggregate interaction

Finally, using the Repast agent-based modeling platform, we created a computational version of our empirically defined nested logit model with heterogeneous agents. As in section 4.2, we investigated the simulated temporal dynamics over time owing to the hypothesized aggregate feedback effect among agents defined by residential district
and socioeconomic group. Example time series results for different random seeds are shown in figure 8 for the nested logit case.

Comparing the time series for the multinomial and the nested logit cases, we obtain dramatically different results yet again for the steady-state solutions of the system. The temporal dynamics in figure 8 show the rapid emergence of a perhaps surprising single equilibrium solution with a mode share of approximately 93% for public transit commuters, approximately 5% for automobile, and approximately 2% for bicycle commuters.

Since, in practice, the rapid time evolution of behavior from an initial case in which almost 50% of commuters in the sample choose automobile, to a steady-state solution in which 93% of commuters choose public transit, is unexpected and counterintuitive, further research is necessary in order to understand how this result arises. The result is particularly significant when we realize that most estimated coefficients in table 7 for the multinomial logit versus the nested logit model are within two standard errors of each other. A noteworthy exception of course is the scale parameter estimated for the nested logit model (four standard errors difference).

In any case, the effect of considering shared unobserved heterogeneity through the introduction of the scale parameter—that is, the effect of common unobserved attributes of the choice alternatives in the error structure—clearly cannot be ignored in an empirical application of a discrete-choice model with network-dynamic interactive feedback.

Figure 8. Example time series with different random seeds for the trinary nested logit model with empirically defined systematic utilities for each mode and an empirically defined field variable for aggregate interaction. The black series represents the mode share of agents choosing to commute by automobile either as a driver or passenger. The light gray series represents the mode share of agents choosing to commute by bicycle/moped/motorcycle. The medium gray series represents the mode share of agents choosing to commute by internal and/or external system public transit. Each run is allowed to iterate for 600,000 time steps. This is, on average, 200 revisions of choices with asynchronous decision making for the sample size of roughly 3000 agents.
5 Conclusions

We have extended previous work on discrete choice with social interactions in important ways. First, we present a framework for conceptualizing the interdependence of decision makers’ choices, making a distinction between social and spatial network interdependencies and between identifiable and aggregate agent interdependencies. In our empirical application we consider a model in which an agent’s choice is directly influenced by the percentages of the agent’s neighbors and socioeconomic peers making each choice; given the availability of appropriate data, our approach is, in principle, directly extendable to the identifiable agent case. We introduce additional heterogeneity in the model through different mechanisms, such as individual-specific sociodemographic characteristics of the agents, individual-specific attributes of the choice alternatives, and the availability of alternatives. Finally we introduce unobserved heterogeneity by accounting for common unobserved attributes of the choice alternatives in the error structure. We observe that these extensions generate dramatically different temporal dynamics, and thus cannot be ignored in any true empirical application.

A careful specification of both observed and unobserved heterogeneity matters critically for emergent temporal outcomes when there is sociodynamic feedback in the model, even when the feedback takes the simple form of an aggregate field variable. Heterogeneity has an impact on the magnitude of the mode shares, on the speed of the transition to the steady state, and very fundamentally on the number of possible observable steady-state solutions.

In order to separate out effects, more research is needed to explore systematically different model configuration treatments. In particular, it would be interesting to understand better the importance of observed heterogeneity (the effect of availability of alternatives, the effect of various explanatory sociodemographic agent characteristics, the effect of various agent-specific attributes of alternatives) relative to the importance of shared unobserved heterogeneity. Also, it is important to understand the detailed effect of induced heterogeneity in different network structures, including the speed of information flow across them.

Given the availability of suitable temporal data, another interesting extension to this research would be to estimate a time lag coefficient, to represent the importance of an agent’s choice on the existing choice. We might intuitively expect that there may be a resistance to change. The introduction of such a time lag variable might result in a more intuitively expected steady-state solution than the perhaps surprising result we discovered in section 4.3.

Also very important for any policy application, particularly for transportation mode choice, would be the introduction not only of positive feedback but also of negative feedback into the model in order to account for congestion effects in addition to agglomeration effects. The absence of negative feedback to account for congestion effects in the model over time may also possibly be one of the reasons for the perhaps surprising result we discovered in section 4.3. If too many commuters travel by automobile, roads may become congested, and traffic jams may result. The increased travel time owing to slower-moving or still-standing traffic over some route segments may discourage commuters from choosing the automobile mode. The result, in practice, is that there is some maximum automobile mode share that roads can handle and which commuters will tolerate. This will vary according to the road system of a given metropolitan region and the norms and values of a given population. On the other hand, if too many commuters travel by public transit, seats may not be available on certain route segments, which in turn may discourage certain strata of the population from traveling by public transit. Some strata of the population may be sensitive to crowded...
public situations in general, even regardless of whether there is a seat available or not. Furthermore, for example, with the train system there is a maximum number of trains that can ride over a given shared rail system out of safety considerations. It may not be possible in practice to transport the entire commuting population over a given rail system, particularly if freight transport by train runs over the same rail system as passenger trains. The result in practice is that, just as with the road system, there is some maximum public transit mode share that the public transit system can handle and which commuters will tolerate. This too will vary according to the public transit system of a given metropolitan region and the norms and values of a given population.

Note, however, that norms and values of a given population may change over time. These can be represented as the global interactions between a decision maker and the aggregate actions of other decision makers in the entire population, and may be addressed as the special limiting case of a fully connected or uniform network. Temporal data would be required in order to separate out a global field variable from an alternative specific constant.

Acknowledgements. The authors would like to gratefully acknowledge discussion with Frank le Clercq, Loek Kapoen, Joan Walker, Moshe Ben-Akiva, Mehran Kardar, Xiao-Gang Wen, Masanao Aoki, Robert Axtell, Scott Page, John Miller, George Kampis, József Vánca and András Márikus at various stages in the development of the research, as well as the very helpful comments of anonymous reviewers who significantly helped to improve the readability. The work forms Project II of the AMADEUS program directed by Harry Timmermans and coordinated by Theo Arentze, and the financial and institutional support of the Amsterdam Institute for Metropolitan and International Development Studies (AMIDSt). Special thanks are due to Guus Brohm and Yvon Weening of the Municipality of Amsterdam Agency for Infrastructure, Traffic and Transport (dIVV) for assistance with data in the empirical application, and to Len de Klerk, Gert van der Meer, Tina Zettl, Gimene Spaans, Ernst Berkhout, Michel Hageman, Barbara Lawa, Karin Retél-de Groot, Wim de Lange, Loes van Dort, the Facilitheek, and the NVD meldkamer, for institutional support in running simulations on computers in student halls during nights and weekends at the Department of Geography, Planning and International Development Studies in the Faculty of Social and Behavioral Science at the University of Amsterdam (UvA-FMG). Additional simulations were performed at SARA Computing and Networking Services, Science-park Amsterdam (http://www.sara.nl), with special thanks to Willem Vermin, Bert van Corler, the HPC team, Jan Hartman, and Marcel Heemskerk for institutional support. The authors claim full responsibility for any errors.

References
Ben-Akiva M, 1973 *Structure of Passenger Travel Demand Models* PhD dissertation, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA
Blume L, Durlauf S, 2002, “Equilibrium concepts for social interaction models”, SSRI working papers, Social systems Research Institute, University of Wisconsin, Madison, WI

Domencich T, McFadden D, 1975 Urban Travel Demand (North-Holland, Amsterdam)


Dugundji E R, Walker J L, 2005, “Discrete choice with social and spatial network interdependencies: an empirical example using mixed GEV models with field and ‘panel’ effects” Transportation Research Record number 1921, 70 – 78


Krygsman S, 2004 Activity and Travel Choice(s) in Multimodal Public Transport Systems PhD dissertation, Department of Geography, University of Utrecht, Utrecht


Urbig D, 2003, “Attitude dynamics with limited verbalisation capabilities” *Journal of Artificial Societies and Social Simulation* 6(1), http://jasss.soc.surrey.ac.uk/6/1/2.html


Conditions of use. This article may be downloaded from the E&P website for personal research by members of subscribing organisations. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.