

## Indicating mate preferences by mixing survey and process-generated data: the case of attitudes and behaviour in online mate search

Schmitz, Andreas; Skopek, Jan; Schulz, Florian; Klein, Doreen; Blossfeld, Hans-Peter

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Schmitz, A., Skopek, J., Schulz, F., Klein, D., & Blossfeld, H.-P. (2009). Indicating mate preferences by mixing survey and process-generated data: the case of attitudes and behaviour in online mate search. *Historical Social Research*, 34(1), 77-93. <https://doi.org/10.12759/hsr.34.2009.1.77-93>

### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

### Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>

# Indicating Mate Preferences by Mixing Survey and Process-generated Data. The Case of Attitudes and Behaviour in Online Mate Search

*Andreas Schmitz, Jan Skopek, Florian Schulz,  
Doreen Klein & Hans-Peter Blossfeld\**

**Abstract:** »Die Indikation partnerschaftlicher Präferenzen mithilfe der Mischung von Befragungs- und Prozessdaten. Das Beispiel von Einstellung und Verhalten bei der Online-Partnersuche«. Web-based process-generated data is produced by social agency of users and recorded by the respective provider without any originally scientific purpose. We support our idea of advantageous applications of process-generated data by outlining a research example that uses data generated by email contacting on an online dating website for the investigation of mate preferences. This approach follows the paradigm of indicating or 'revealing' preferences by observing choosing acts. Advantages and disadvantages of this approach in comparison to the traditional 'stated preference'-paradigm of survey research are discussed. Both approaches suffer different informational restrictions and induce different problems of valid inference. In conclusion we offer an outlook towards research strategies of an integration of the two quantitative paradigms.

**Keywords:** Process-generated Data, Online-Dating, Survey-Data, Mixed Methods, Triangulation.

## 1. Introduction

Social science research uses the internet as a medium for data acquisition particularly in form of email and web surveys. But in many internet applications additional process-produced data is generated by the (social) agency of its users and recorded by the respective provider; originally without any scientific purpose. In internet applications that have been labelled Social Web or Web 2.0 additional data is accumulated that provides information about interactions, relations and social networks. The detailed records of such process-produced

---

\* Address all communications to: Andreas Schmitz, Lehrstuhl für Soziologie I, Lichtenhaidstrasse 11, Postfach 1549, 96045 Bamberg, Germany;  
e-mail: andreas.schmitz@uni-bamberg.de.  
The project 'Processes of Mate Choice in Online-Dating' is conducted by the Otto-Friedrich-University of Bamberg under supervision of Prof. Dr. Hans-Peter Blossfeld, Chair of Sociology I, and funded by the German Research Foundation.

data, generated by the use of social online-applications, contain rich information about their users.

For instance, social networking, matchmaking and online dating services proliferated extensively in the last years and meanwhile exhibit a huge amount of users. Employing these sites and applications for their personal purposes the respective users are producing rich sets of data as by-products stored in databases hosted by the operating companies. This data contains types and sequences of different pages on the web site, changes in profile information, interactions and communications with other users as well as search acts. Since this kind of data is generated by the software-assisted social online activities without the intention of a scientific inquiry, we label it as *web-based process-generated data*.

In this paper we chose the indication of mate preferences as methodological problem in the field of human mate selection research for illustrating and discussing the capability of an empirical design working with web-based process-generated data. The question here is how to methodologically assess effective mate preferences which represent a latent phenomenon, i.e. are not directly measurable. We argue that web-based process-generated data and web-survey data display different informational dimensions and have specific limitations and strengths with regard to theory testing. Web-based process-produced data enables the researcher to comprehend actual mating activities and the resulting macro-structures of these activities (as for example an assortative mating structure). The actors' motives, their perception of themselves, of their potential partners and their social contexts on the other hand are not available when utilizing web-based process-produced data only. By conducting surveys additionally, the research design is amenable to collecting and displaying these subjective constructs necessary for an adequate discussion of the research question.

We discuss a complementary integration of web-based process-generated data and online survey data and provide an empirical example concerning contact behaviour. This methodological discussion can be located in the wider context of triangulation in social sciences research (Bryman 2006). While in most cases the triangulation of quantitative and qualitative data is discussed we will deal with the integration of different quantitative web data in an empirical example.

## 2. Preferences in Mate Search

To illustrate our discussion of using web-based process-generated data, we first outline a well known methodological problem from (mate) choice research. Our exemplary methodological application concerns the problem of how researchers can assess mate preferences of men and women, which structure matching and relationship formation in the dating and marriage market. In

common use a preference means to favour an alternative A over an alternative B. A mating preference in particular is referring to (i) favouritism of a specific trait over another (e.g. age might be more important than educational degree), (ii) favouritism of a specific value within a trait dimension over another value within the same dimension (e.g. high education over low education) and (iii) favouritism of one potential mate over the other, given all traits.

Applying a rational choice perspective (Coleman 1990), the formation of romantic relationships can be considered analytically as the result of a dynamic interplay between opportunity structures, restricting or facilitating chances of meeting certain people on the ‘marriage market’ on the one hand, and mechanisms guiding individual choice for specific partners with given attributes on the other hand (Kalmijn 1998: 395). The latter point has been usually dealt with the concept of ‘mate preferences’, assuming that choices of men and women for certain partners are guided by their internalized ‘tastes’ for specific resources like socioeconomic status, physical appearance, situational aspects, personality traits. Therefore, regardless of whether those preferences are shaped by genetic dispositions, specific paths of socialization, or restrictions from social structural conditions (Eagly et al. 1999: 408), or whether those preferences are stable over actors and time, it is commonly accepted that an investigation of micro-level concepts is indispensable for explaining marriage or relationship outcomes.

### 3. Possible Data Sources

There are several empirical approaches to measure or indicate preferences in mating research and social science in general, differing in their potential to provide information required by theory and each one representing different methodological problems.

Empirical Studies devoting the revelation of individual mate preferences are for instance conducted by means of content-analysis of personal ads, ex post reconstruction using formed couples, traditional self-report methods as surveys and the observation of choosing acts (e.g. Logan et al 2008: 559; Fisman et al. 2006; Feingold 1991: 981; South 1991: 928; Buss et al. 1986: 559; Bolig et al. 1984: 587).

Blossfeld (2009) criticises the validity of interpretations based on data about prevailing marriages, as they necessarily are post hoc and do not differentiate between the causal effect of preferences and structural aspects like the combined effects of assortative marriage, assortative union dissolution and assortative remarriage. Since comparing couple constellations from ex-post seems to be the most inappropriate approach for measuring mate preferences, we will neglect this in the further discussion. Although occasionally employed by past research, we won’t concern methods of content analysing personal ads printed in newspapers in our paper (c.f. Feingold 1991 or Bolig et al. 1984).

In the following we will focus on limitations of web surveys as traditional approaches indicating ‘stated preferences’ by asking respondents for desired traits<sup>1</sup>. Here after we discuss two modern empirical approaches indicating ‘revealed preferences’: ‘the speed-dating design’ and the ‘online-dating design’.

### 3.1. Online Surveys as a Tool for Indicating Stated Mate Preferences

A common method in mate search research is to explain mate search and mate choice by subjective preferences, mostly surveyed using standardized questionnaires. This approach can be characterized by its concept of preference indication. The ‘*stated preferences*’ are here measured or indicated by actor’s choice intentions or choice rankings and ratings within sets of hypothetical decision situations. In the context of ‘virtual’ mate choice online surveys are an obvious tool of measuring mate preference as stated preferences.

Subjective opinions and attitudes about partnership and love can be operationalized derived from mate search theory and surveyed in a flexible way, as changes in the design of the particular questionnaire can be realized without high costs or time investment. Specific preferences as for example the hair colour, the intended form of relationship (e.g. friendship vs. marriage) or the age of a possible partner can be surveyed and used to explain individual mating behaviour.

However, in the past years, the potentials and limitations of online surveys have been discussed extensively (see for example Couper et al. 2006: 217; Welker et al. 2007 or Baur et al. 2008). The most obvious problem is the potential of generalization of the collected data. This problem corresponds with the difficulties of a random sampling process, which is the basis for inference statistical generalizations. The sampling process requires a definition of the total population, which is often difficult concerning samples of web surveys. In the context of online mate search a population is not self-explanatory nameable, as inference in the truest sense is restricted to a) online-daters of b) the analyzed platform. But even with a clearly defined population, problems regarding the data quality arise, due to systematic non-response-selection-bias. This leads to a considerable challenge of analysis and statistical inference. For example the amount and impact of certain mate preferences might be over- or underrated if the respondents systematically differ from the non-respondents. In the present case this problem is given, as online daters differ from online users and online users differ from offline-populations.

But even if the sample process was conducted properly and in a controlled way, there are some methodological weaknesses of survey data in general. First, they picture intentions and not actions, thereby inducing a systematic bias

---

<sup>1</sup> Some points of critique apply to survey research in general.

in direction of the respondents conscious (strategic) and unconscious deviations from their 'real' behaviour. In the context of mate preference research, unconscious preferences and strategic answers to the questionnaires become a considerable problem for inference, as we expect them to appear frequently and systematically. The concept of preference is essentially resting on the assumption of conscious perception and decision. Thus, applied on human mating behaviour, which is not necessarily completely conscious, we might fail to assess the realistic importance of specific traits.

Secondly, preferences measured as statements do not necessarily conform to future actions, as they are biased by intentional and unintentional variations of 'true' values and by the actors' cognitive incapacity to reflect the structural boundary conditions of their agency. The problem here is that the actors might fail to evaluate their preference order, i.e. the relevance of specific cues for contact initiation and contact maintenance. Research on mate choice applying the stated preference perspective therefore has to conduct great investments in terms of measuring and predicting realistic preferences.

Furthermore survey data do feature hypothetical decision elements and do not contain the factual actor's choice sets, which is a necessary requirement for a theory-adequate application of preference-based research as well as a theoretical condition for transferring hypothetical situations of surveys to empirical situations of actions. If the respondents are hypothetically asked to choose between different mating cues, the resulting individual and aggregated pattern will barely reflect the reality.

With regard to mating behaviour, survey data cannot focus processes of interactions from their starting point and in their sequential development and it does usually not contain information about potential and rejected potential mates. Aspects which are of great importance from a methodological point of view, as posterior inferences based on results of (inter)action in the case of realized mate preferences, are not in accordance with the idea of causality.

Finally, the question emerges to what extent surveys are adequate methodical instruments to acquire data about unintended consequences of human activities (Baur 2008) for example about a specific (partner) *market*. The well-known phenomenon of assortative mating processes (Blossfeld 2009) describing the macro-phenomenon of actors mating each other hierarchically according to their market value or their resources, for example, can hardly be modelled by using subjective statements of preferences.

### 3.2. Revealing Mate Preferences

Observing choices, choice sets and interactions directly offers an alternative approach of conceptualising actors' preferences. This '*revealed preferences*'

approach enables the researcher to utilize realized actions instead of intentions as indicators of preferences by observing the action in particular<sup>2</sup>.

There are several methodological advantages of a revealed preferences approach. Revealed preferences are independent from selective participation of the respondents and independent from selective response patterns of a particular respondent as well. Secondly, conscious and unconscious deviations from 'true' preferences are less problematic, leading to more valid indicators of preference. Furthermore, revealing preferences in form of observing action, theoretically (and sometimes empirically) implies observing situations of individual choice, interaction and eventually complete 'markets' as for example assortative patterns of mating.

Within this 'observational paradigm' of revealed mate preferences some studies use experimental conditions, for example observing interactions within 'speed dating' designs (Finkel et al. 2007). One advantage of such a design is the possibility to observe all occurring dyadic interactions, outperforming traditional individualistic approaches. Those analyzed dyads include matches and non-matches and can therefore be used to analyze emerging relations from the very beginning in contrast to traditional approaches starting their interpretation from marital matches.

A new approach of studying issues of mate choice processes is constituted by the analysis of interactions within online mate markets like dating sites on the internet (Fiore et al. 2005). Instead of observing matches as it is done within a speed dating design, contacts are digitally recorded. Comparing an online dating design with the speed dating design the generalisations of findings gained do not suffer the restriction of small cases and the systematic biases of the experimental situation itself in the former case. From a theoretical point of view, it is an additional advantage within the online dating design that the researcher is able to reconstruct the first moments of partner choice processes as well as subsequent episodes and long-term partner formation. This is advantageous as past research emphasizes that the "process underlying ... [dating and mate selection] ... is one of sequentially filtering-out, gradually narrowing the field of eligibles from a pool of several to a specific few and eventually to one individual" (Edwards 1969: 521). Finally, it is possible within the online dating design to account for the "features of the environment (...) (that) might cue a different set of preferences" (Kurzban et al. 2007: 15) and possibly leading to a distortion of empirical results. Using process-generated interaction data of an online dating site thereby enables the researcher to reconstruct early processes of partner choice of men and women without spurious effects of an experimental design.

---

<sup>2</sup> From a methodological perspective this approach is by no means without controversy. In particular the danger of a theoretical tautology and problems resulting from observational fallacies regarding true preferences are discussed by Sen (1986).

The structural patterns of contact relationships we observe in online dating as well as in real-life settings represent individual choosing behaviour. However, we state that choosing in online dating is to a less extent socially biased than opportunity settings of everyday life since there exist basically less systematic institutional or spatial boundaries that constrain meeting chances of people. Major dating sites typically cover a vastly differentiated population – from younger to older, lower or higher educated, from single to divorced people. Every person of this population can be ‘reached’ by the choosing actor in the same amount of time and site clicks. *Choice sets are less restricted within the online-dating design.*

Due to this relative absence of structured interaction a choice act as such can be observed and analyzed more clearly. The selective moment of subjective choices thereby enables the researcher to methodologically utilize individual behaviour in explanatory approaches. So while being able to control for the structural probability of contact relationships between people belonging to certain educational groups, we can assess the structural relevance and importance of choice related aspects of mate preferences. Selective choosing for example with regard to educational can be analyzed given an observable opportunity structure. *Choice sets can be observed directly within the online dating design.*

### 3.3. Properties of Web-Based Process-Generated Data

Once registered on a dating site, the actors can present themselves by creating an *user profile* – a self-presentation in the sense of a personal advertisement – by filling out online questionnaires asking for socio-economic and socio-demographic information, personal information like hobbies, likes and dislikes. Additionally, on typical sites users can enhance their self-presentation by uploading pictures. Compared to traditional personal ads in newspapers, usually being restricted regarding the number of characters and mostly dedicated to a single issue, online profiles on dating sites basically offer a bulky way of presenting the own ‘identity’ and being accessible for others twenty-four hours a day. This profile data make up the first type of our data. Albeit this data is recorded in the business process of the provider it corresponds somehow to survey data as the users are asked for their statements rather than to be observed.

The second type of data is the relational information of what is happening between users. Typically users are exploring the profile database of the site, are checking profiles of other users and, if being attracted to the opposite somehow, will try to get in contact by writing email messages using a messaging function common for most dating sites. In a way, this is the most interesting component of our dataset, as this kind of relational data combined with the user profiles allows us a detailed temporal reconstruction of the social process of



contact formation and interaction between mate searchers on an observational level. For instance, we are able to retrace, which profiles of other users given users looked at and which profiles he or she chose for contacting via email afterwards.

In our case, the raw data comes in anonymized dump files from the operative SQL database of the maintaining company. This is a relational data format, which first has to be prepared (for the specific research question) and then converted to flat file formats readable for most statistical packages. Albeit the ‘traffic’ in a given time period varies over different sites with different popularity, according to our experience, the researcher has to deal with an unusual huge amount of data (like hundreds of thousand of user profiles, million visits and messages). These circumstances may demand some advanced technical skills from the researcher.

Since this data represent ‘real’ dating behaviour recorded within a setting being not necessarily arranged for scientific intentions, there are particular aspects to be considered by the researcher striving for utilizing this data. It is important to note, that, as long as the most significant sites are hosted by companies, and not universities, academic and commercial interests and goals have always to be balanced out. This means that in practice the scientific optimum of data collection in terms of measurement from a researcher’s point of view not necessarily has to coincide with the commercial optimum in terms of site usability and user experience from a company’s point of view. Thus, another crucial task within cooperation is to achieve a balance of high level data quality for later data analysis (like measurement quality of profile questionnaires) on the one hand and a high value for site users that make the site alive on the other hand. The extraction of the process-generated data from the operative web databases of the provider is a further issue that has to be treated carefully. As those databases contain some personal information of the users (like email addresses or payment information) the anonymity of the person behind a user has to be ensured by a suitable retrieving process. The provider has to strip all information from the profile and interactional data that points to the personal identity of the users before handing out the database dumps to the researcher. Finally, the providing company has to archive an additional amount of possibly ‘out-dated’ information (events like clicks on profiles, messages and profile changes) over at least the period of observation to permit a complete reconstruction of events in the later analysis.

#### 4. Example: Homo- and Heterophily in Choosing Contact Partners

For giving an illustrative example of using online process-generated data, we will pick up the empirical research program regarding mate preferences of women and men looking for romantic partners already outlined in the introduc-

tion. Coming from a social structural research interest, we intend to shed light on the question of how men and women choose partners according to their educational level.

We investigate the social structural ‘driving forces’ that are working already in the early phases of the mate choice process. In order to capture the ‘step-wise’ filtering-process (Murstein 1976) from the very beginning, where people originally decide about contacting certain other people for engaging in dating relationships, we analyze data about contacting behaviour from the German dating website mentioned above.

Table 1: Educational Homo- and Heterophily in Making First-Contacts by Gender of Site Users

	Mean fraction of total contacts by a user			
	‘Man contacts Women’		‘Woman contacts Man’	
	Observed	Expected	Observed	Expected
‘Hyperphily’ $E_I < E_R$	29.2	(31.6)	40.7	(38.4)
‘Homophily’ $E_I = E_R$	35.0	(30.4)	40.9	(30.2)
‘Hypophily’ $E_I > E_R$	35.8	(38.1)	18.5	(31.4)
Total	100.0	(100.0)	100.0	(100.0)

*Legend:* *Hyperphily* is the case, if the initiator (I) of a first-contact dyad displays a lower educational level (E) in his/her user profile compared to the receiver (R). *Homophily* means both display the same educational level, while *hypophily* describes those cases where the initiator contacts somebody below his own educational level. Since the number of different contacts per user is not restricted and there is quite a variation in the number of contacts sent by a user, we weighted all contact events by the dividing by the number of total contacts per user. That gives us finally an interpretation along users and not events. Source: Skopek et al. (2009).

For analyzing homo- and heterophilous mate preferences regarding the educational level we open up a randomly chosen observational window (first half of six months period from January 2007 to June 2007) of mutual contacting events on the site. Within this timeframe, the profiles have been reconstructed that were initially contacted by a sample of 12.608 ‘active’ users from age 14 to 79 (58.9% male, on average 36 years old) resulting in 116.138 ‘first-time’ message events. Only contacts between women and men were selected, although some same sex contacts occurred in the observational period. The sub-

stantial variable in this analysis is the educational level which the users are exposing in their profiles via a standardized selector<sup>3</sup>.

Table 1 shows the mean structure of contacting behaviour regarding educational constellations between initiator and receiver of first-contacts. The table is separated vertically by initiator's gender and horizontally by the educational constellations we labelled 'homophily', (sender and receiver have same education), 'hyperphily' (senders contacted receivers with a higher educational level) and 'hypophily' (senders contacted receivers below their own educational level).

Moreover, looking on real events generated by the process of online mate search, we can distinguish between observed and expected rates of homo- and heterophily. The expected rate would be observed on average if the process of selecting contact partners would follow a random choice mechanism with respect to the educational variable<sup>4</sup>. If choice would be random in this sense, the distribution of educational levels in men's and women's profiles on the site would predict the amount of observed contacting homophily and heterophily. Therefore, looking at the example of educational homophily, a high observed value compared to the expected value would be an indicator for the high importance of educational similarity for the choosing actor.

For a detailed analysis and discussion of results see Skopek et al. (2009), but for the moment, let us emphasize that for both male and female initiators, educational similarity to the receiver – the case of homophily – indeed is an important aspect. This holds especially for choosing women, where homophily has the lowest expectation but the highest observed fraction. Quite surprising is the extreme low fraction of women 'contacting downwards' in educational terms (on average female users distribute 18.5% of their first-contacts to men displaying a lower educational level) compared to the expectation (31.4% could be expected taking structural conditions into account). This result partly supports hypotheses derived from traditional gender role conceptions, as found in the literature (cf. Blossfeld et al. 2003). Finally, male users much more than female users seem to choose close to the expectation, suggesting a choice process for men that is to a less extent oriented toward educational status of potential mates. Literature assigning a primacy of physical appearance at the expense of socioeconomic status resources for the male mate selection preferences is supporting this result (e.g. Li et al. 2002: 947).

---

<sup>3</sup> The educational levels used in this analysis comprehend four categories that are related to the German educational system namely 'Hauptschule', 'Realschule/Lehre', 'Abitur' and 'Hochschule'.

<sup>4</sup> Technically spoken Skopek et al. (2009) employ a model of random choice based on an expectation matrix assuming independence of initiator's and receiver's educational level for the contacting process. This matrix of the expected contact distribution is then compared with the actual observed distribution of contacts. For details on constructing random choice models, see Verbrugge (1977: 576).

Of course, this reduction to education-related mechanisms is a quite narrow view on the ample motives that guide the very subjective choosing behaviour of men and women in partner search. Other aspects like physical appearance, emotional appeal, the style of presenting or just a feeling of trustworthiness of the other might be reported by the respective individuals as even primary reasons for making a contact. However, in our simple example we were able to unfold individual mechanisms at work that lead to structural consequences in such a way that there are objective barriers along social dimensions present in the dating process.

A sole recourse on stated preference methodology in terms of surveying subjective preference would in contrast lack the possibility of quantifying the objective impact on the emerging matching structures. Some actors might state preferences for humorous or sympathetic mates compared to education-related traits, but it is not possible to compare their relative impact on individual agency and collective consequences. The methodological advantage of our empirical analysis therefore consists in showing the observed preference impact as a *significant* one.

## 5. Limitations of Web-Based Process-Generated Data for Mate Research

There is a severe methodological challenge inherent in data situations as described. Both data sources, process-generated data *and* survey data, can not be used to draw inference on the population of the online daters in general or even on the (German) offline population just by implication. All collected information refers to the specific sub-population of the particular platform. Given a multiple differentiated market of online dating, some statistical efforts have to be done to correct for systematic processes of self-selection into one particular social web site (Schmitz et al. 2008). But even more fundamental problems arise if the area of generalization is related to offline (mating) behaviour.

For Germany, a recent academic study with a representative sample shows that in 2007 about 12 percent of the internet users (corresponding to about five million persons in the German population) already used online dating sites (Schulz et al. 2008). The composition of those users significantly differs from the German population and the internet users by means of socio-demographic characteristics. The analysis showed that users of online dating are by trend low educated males and high educated women. In addition it seems to be probable that the users of online-dating do also differ with regard to unobserved characteristics as for example in their ideas about a relationship and behaviour in a relationship. In addition, there is some amount of reservations against online-dating so that persons with traditional romantic partnership conceptions might be under-represented in the data. Beyond this overall limitation of gener-

alizability, there are specific restraints of web-based process generated data of this kind.

Compared to offline process-generated data web based generated data is less affected by reactivity and institutional selectivity as well as by observational bias brought on by persons in charge due to subjective variations or different particular times, but validity is still an relevant issue. Procedural reactivity, a systematic behavioural deviation induced by the observation process, might occur as the users of a dating platform have to be informed about the recording process. On the other hand we expect this procedural reactivity to be low and hardly affecting the choosing behaviour of the actors. But then, it is not sure that a recorded activity represents the activity of one specific user. It is possible that several users share one account, that one user has several accounts and that fake profiles are conducted.

More fundamentally, we have to consider that by applying the revealed preferences paradigm to empirical phenomena, the users are restricted in their choice behaviour in respect of the technical organisation of the web-page. Maybe a user would reveal another choice behaviour if he or she would have the possibility to act in a different way. The impact of the preference towards a (non)-smoking partner for example can not be addressed if the possibility of acting concerning this particular cue is not designed by the provider. Consequently the absence of possibilities to methodically imply influences of new choice characteristics represents a serious limitation of the observational paradigm, as we might ignore relevant aspects of the mating behaviour. The resulting characterisation of this mating behaviour may lead to difficulties of validity.

Furthermore we have to take into account that an actor in the course of his life experience already adjusted his choosing behaviour by reflecting his own mate value in the calculation. In this sense it might be the case that the actor's contacting behaviour would reveal rather strategies resulting from trade-offs than underlying preferences for certain traits. For example an actor might have a preference for physically attractive or educated persons on the one side, but on the other side might restrain from contacting very attractive persons since he expects to get rejected from those attractive persons. The observation of specific contact patterns therefore complicates the valid interpretation of findings as preference-guided behaviour.

Thus, the *methodological* advantage of a relative absence of time-, space- and co-presence impact turns to a *theoretical* disadvantage. Online performed agency is not just a virtual equivalent of offline agency. A straight *statistical* transfer grounded on the (implicit) assumption of the equivalence of online and offline situations is by no means valid. From our point of view, it is wise to use the synthesised insights of a particular web environment not with regard to *statistical inference* but in form of *substantiated hypotheses* for subsequent offline research purposes.

## 6. Towards a Complementary Data Integration

Process-produced data enable the researcher to comprehend actual activities and the corresponding structures of these activities. But as discussed above there are several theoretical limitations inherent to this kind of observational information. Even if process-generated data recorded for instance by an online dating service can describe actual structures, a complementary usage of the respondent's cognition is still required in order to avoid a narrowing structuralistic perspective.

„Thus, observation necessarily requires an insight in and an appropriate interpretation of the subjective meaning and the social relevance of a specific activity or a behavioural sequence. Without such an understanding, the observation would be blind and irrelevant for social science research.” (Mayntz et al. 1971: 87, cited from Friedrichs 1980, S. 270).

Mating behaviour as an example of social agency in general can be analytically conceived as a syndrome of previous experience, current calculations and intentions, dynamic adaptation of preferences and inert dispositions.

In contrast to survey data process-generated (web) data lack valid information contents about those intentions, attitudes as well as temporal and local circumstances. It appears to be advisable then to capture the personal (mating) history of the actors as well as their specific (mating) preferences and general dispositions by means of retrospective and panel surveys.

By combining the continuously archived process-generated data of the platform and different questionnaires sources, the possibilities of data analysis can be decisively improved. Similar to the data situation of offline conducted process-generated data and survey data, a combination or “triangulation” of both data sources appears promising, since each user can be described on the basis of more comprehensive information.

“By combining process-generated data with survey data, which have been collected specifically for the research question, a maximum of significance and validity of the results can be reached” (Hartmann et al. 2007: 10).

In the case of applying the stated preferences approach to the users contact behaviour, we can not be sure whether our interpretation holds for particular individuals or even in general. In the case of preference based behaviour in online dating, we therefore have to consider the different boundary conditions of the different (types of) individuals to understand their actions. Comparing revealed preferences based on the electronic records with stated preferences measured via interviews or questionnaires might give valuable hints where strategic considerations of the actors might arise and where the interpretation of the researcher has to be modified. Divorced women, for instance, will probably see their position in a different way than young single woman, but might not necessarily differ from each other with regard to their objective position within the interaction patterns. Different intentions (superficial or long-term partner-

ship), inventory of knowledge (what are the rules of interaction) and strategies (dispersive vs. purposive), to give but a few examples, are surely essential for a comprehensive interpretation of mating behaviour on the dating platform. Those aspects of preference guided behaviour can not be recorded within process-generated data, but have to be complementary surveyed with web-surveys.

A complementary data integration of those questionnaire data and web-based process-generated data allows (amongst others) for a data triangulation of the concept of preference. In our empirical example we already applied a simple combination of web-based process information (contact behaviour) and survey information (profile specifications). From a substantial point of view it is also promising for instance to (1) compare stated and revealed preferences to investigate for (conscious or unconscious) separations of perception and action by comparing answer patterns retrieved from surveys with contact patterns retrieved from the web generated process data. Secondly the relative impact of stated dispositions on (2) individual and (3) structural observed consequences can be empirically analyzed. Another resulting possibility is to (4) observe the influence of objective choice sets as they exist in form of possible contacts on individual perceptions, which can be inquired with surveys. Finally further research can be done on (5) the dynamic interplay of perception and agency within time, as for example by assessing the influence of (inquired) stated preferences on future observed behaviour recorded in form of web-based process-generated data.

## 7. Conclusion

We choose the example of the stated and revealed preference approach to the measurement of mate preferences, as those conceptions naturally correspond to process-generated data and survey data respectively. The latent phenomenon of preferences in social sciences can be located between subjective intentions and actual conduct of choice acts. To empirically assess mate preferences there are several methodological possibilities.

Usually in social sciences the development and test of a theory postulates several informational elements to be measured. The discussion showed that (web-) survey data and web-based process-generated data display different informational dimensions and have specific limitations and strengths in different fields. (Web-)Survey data allows for measuring attitudes, opinions and value orientations, but it is limited by means of predicting actual choosing behaviour, mating interaction and emergent results of agency. Web-based process-generated data in contrast enables the researcher to reveal individual mating (inter-) action as well as emergent results of this (inter-) action, but this data has no information about internalized aspects, as for example self-perceived mating opportunities.

The purpose of our empirical example was to demonstrate, how data retrieved from social web applications can be triangularly employed in empirical research on substantial questions of social sciences. Due to the relational structure of our data it was possible to objectify subjective information by observing the reaction of actors towards traits given in the profiles. This simple example of triangulating behavioural and survey information allowed for picturing emergent educational homogeneity of dating-site users

Of course, the field of application is not narrowed to marriage market and mate choice issues, but right here, the described data offer a valuable empirical access to social processes that can hardly be measured by other empirical settings. Though, empirical analyses are not restricted to ‘measuring’ preferences in one-sided choices of single actors (as done in our small example).

Survey data will not become obsolete for social science research by applying the observational paradigm, but the methodological and theoretical challenges of (web) questionnaire research can be tackled with additional observational information. Therefore, approaches utilizing online process-generated data and combining it with survey data promise to be a valuable contribution to the methodical repertoire of mate choice research. . In doing so, the continuing investigation of congruence and divergence of the different measurement approaches will allow for more valid interpretations in case of congruence and for informative evidence in case of divergence.

Furthermore, it seems that chances for the triangulation of survey and process-generated data emerge especially in a social web context. Online-activities leave digital behavioural traces and – compared to offline data collection – these digital traces can be collected easily. For future research in the field of social internet activities it would make sense to inquire about already existing electronic process-produced data and the possible use for social science research and to fathom the possibilities of a complementary survey. As more and more aspects of everyday life become related to the internet, it can be expected that the possibilities of the complementary use of electronic process-generated data and traditional survey data will become a standard tool of empirical social web research.

## References

- Baur, Nina (2008): Mixing Process-Generated Data in Market Sociology, submitted.
- Baur, Nina und Michael J. Florian (2008): Stichprobenprobleme bei Online-Umfragen. In: Jakob, Nikolaus/Schoen, Harald/Zerback, Thomas (Eds.) (2008): Sozialforschung im Internet: Methodologie und Praxis der Online-Befragung. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Becker, Gary S. (1976): The Economic Approach to Human Behaviour. Chicago: University of Chicago Press.



- Blossfeld, Hans-Peter (2009): Assortative Marriage in Comparative Perspective. In: *Annual Review of Sociology*, 35, Forthcoming.
- Blossfeld, Hans-Peter/Timm, Andreas (2003): *Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*. Dordrecht: Kluwer Academic Publisher.
- Bolig, Rosemary/Stein, Peter J./Mckenry, Patrick C. (1984): The Self-Advertisement Approach to Dating: Male-Female Differences. In: *Family Relations*, 33 (4), 587–592.
- Bovelet, Daniel (2006): The Use of Research Funding Databases for Research Assessment Information Systems. In: *Conference Criss 2006*.
- Bosnjak, Michael (2003): Web-basierte Fragebogenuntersuchungen. Methodische Möglichkeiten, aktuelle Themen und Erweiterungen. In: *Informationszentrum Sozialwissenschaften (IZ)/Arbeitskreis Sozialwissenschaftlicher Institute e.V. (ASI) (Ed.), Online-Erhebungen*. Bonn: IZ, 109-133.
- Bryman, Alan (2006): Integrating quantitative and qualitative research: how is it done? In: *Qualitative Research*, 6 (1), 97-113.
- Buss, David M./Barnes, Michael (1986): Preferences in Human Mate Selection. In: *Journal of Personality and Social Psychology*, 50 (3), 559–570.
- Cameron, Samuel/Collins, Alan (2000): *Playing the Love Market*. London: Free Association Books.
- Coleman, James S. (1990): *Foundation of Social Theory*. Cambridge, MA: The Belknap Press of Harvard 85 University Press.
- Couper, Mick/Coutts, Elisabeth (2006): Online-Befragung. Probleme und Chancen verschiedener Arten von Online-Erhebungen. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 44, (2004), 217-243.
- Eagly, Alice H./Wood, Wendy (1999): The Origins of Sex Differences in Human Behaviour: Evolved Dispositions Versus Social Roles. *American Psychologist*, 54, 408-423.
- Edwards, John N. (1969): Familial Behaviour as Social Exchange. *Journal of Marriage and the Family*, 31, 518-526.
- Feingold, Alan (1990): Gender Differences in Effects of Physical Attractiveness on Romantic Attraction. A Comparison Across Five Research Paradigms. In: *Journal of Personality and Social Psychology*, 59 (5), 981–993.
- Finkel, Eli J./Eastwick, Paul W./Matthews, Jacob (2007): Speed dating as an invaluable tool for studying romantic attraction: A methodological primer. In: *Personal Relationships*, 14, 149-166.
- Fiore, Andrew T./Donath, Judith S. (2005): Homophily in Online Dating: When Do You Like Someone Like Yourself. *Computer-Human Interaction*, 2005.
- Fisman, Raymond/Iyengar, Sheena S./Kamenica, Emir/Simonson, Itamar (2006): Gender Differences in Mate Selection. Evidence from a Speed Dating Experiment. In: *Quarterly Journal of Economics*, 75 (1), January (2008).
- Friedrichs, Jürgen (1980): *Methoden empirischer Sozialforschung*. Opladen: Westdeutscher Verlag, 270.
- Hartmann, Josef/Krug, Gerhard (2007): Verknüpfung von Befragungs- und Prozessdaten-Selektivität durch fehlende Zustimmung der Befragten? IAB Discussion Paper No.13.

- Hitsch, Günter J./Hortacsu, Ali/Ariely, Dan (2006): What Makes You Click? – Mate Preferences and Matching Outcomes in Online Dating. MIT Sloan Research Paper No. 4603-06.
- Kalmijn, Matthijs (1998): Intermarriage and Homogamy. Causes, Patterns, Trends. In: *Annual Review of Sociology*, 24, 395–421.
- Kurzban, Robert/Weede, Jason (2007): Do advertised preferences predict the behaviour of speed daters? In: *Personal Relationships*, Vol. 14.
- Li, Norman P./Bailey, Michael J./Kenrick, Douglas T./Linsenmeier, Joan A. W. (2002): The Necessities and Luxuries of Mate Preferences. Testing the Tradeoffs. In: *Journal of Personality and Social Psychology*, 82 (6), 947–955.
- Logan, John Allen/Hoff, Peter/Newton, Michael (2008): Two-Sided Estimation of Mate Preferences for Similarities in Age, Education, and Religion. In: *Journal of the American Statistical Association*, 103 (482), 559–569.
- Murstein, Bernhard I. (1976): *Who Will Marry Whom? Theories and Research in Marital Choice*. New York: Springer Publications.
- Sen, Amartya (1986): Behaviour and the Concept of Preference. In: Elster, Jon (Ed.): *Rational Choice*. Oxford: Basil Blackwell, 60–81.
- Schmitz, Andreas/Messingschlager, Martin/Blossfeld, Hans-Peter/Rässler, Susanne (2009): Multiple Imputation mit web-generierten Prozessdaten zur Behandlung fehlender Werte von Web-Befragungen (Working paper).
- Schulz, Florian/Skopek, Jan/Klein, Doreen/Schmitz, Andreas (2008): Wer nutzt Online-Dating in Deutschland? In: *Zeitschrift für Familienforschung*, 3, (2008).
- Skopek, Jan/Schulz, Florian/Blossfeld, Hans-Peter (2009): Partnersuche im Internet. Bildungsspezifische Mechanismen bei der Wahl von Kontaktpartnern. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, (forthcoming).
- South, Scott J. (1991): Sociodemographic Differentials in Mate Selection Preferences. In: *Journal of Marriage and the Family*, 53 (4), 928–940.
- Verbrugge, Lois M. (1977): The Structure of Adult Friendship Choices. In: *Social Forces*, 56 (2), 576–597.
- Welker, Martin/Wenzel, Olaf (2007): *Online-Forschung 2007. Grundlagen und Fallstudien*, Köln: Halem.