Sibling similarity in family formation

Paper prepared for the Annual Meeting of the Population Association of America, 2014

Anette Fasang & Marcel Raab

Humboldt University Berlin & Berlin Social Science Center (WZB)

Jani Erola & Aleksi Karhula

University of Turku, Finland

Abstract

Sibling studies have been widely used to analyze the impact of family background on socioeconomic and, to a lesser extent, demographic outcomes. We contribute to this literature with a novel research design that combines sibling comparisons and sequence analysis to analyze longitudinal family formation trajectories of siblings and unrelated persons. This allows us to scrutinize in a more rigorous way, whether there is sibling similarity in family formation trajectories and if siblings' shared background characteristics, such as parental education and early childhood family structure can account for similarity in family formation. We use Finnish register data from 1987 until 2007 to construct complete longitudinal family formation trajectories in young adulthood for siblings and unrelated dyads (N=14,259 dyads). Findings show that siblings' family formation is moderately but significantly more similar than for unrelated dyads, also after controlling for crucial parental background characteristics. Shared parental background characteristics add surprisingly little to account for sibling similarity in family formation. Instead of shared parental background, gender and the respondents' own education are more decisive forces in the stratification of family formation. Yet family internal dynamics seem to reinforce this stratification, such that siblings have a higher probability to experience similar family formation patterns. Particularly patterns that go along with economic disadvantage are concentrated within families. This is in line with a growing body of research highlighting the importance of family structure in the reproduction of social inequality.

INTRODUCTION

Family of origin effects on family behavior are at the center of a controversial debate in family demography. On the one hand, growing evidence on profound social change in family formation with the proliferation of cohabitation, lone parenthood and diverse family forms questions, whether the family of origin still matters for peoples' family behavior (Elzinga & Liefbroer, 2007; Fussell & Furstenberg, 2005; Popenoe, 1993; Shanahan, 2000). On the other hand, a number of studies compellingly demonstrate continuing importance of the family of origin for peoples' family behavior (e.g. Liefbroer & Elzinga, 2012). The broader relevance of the topic is highlighted by another line of research emphasizing the importance of family structure for the reproduction of social inequality and the perpetuation of family disruption across generations (Carlson and England 2011, McLanahan & Percheski, 2008).

Sibling studies have been the method of choice to study family of origin effects - everything siblings share - on socioeconomic outcomes such as education and occupational status (e.g. Conley, 2008). To a lesser extent they have also been employed to examine family of origin effects on demographic outcomes such as family behavior (e.g. Lyngstad & Prskawetz, 2010). To date, sibling studies on family formation largely focus on isolated fertility transitions (Kuziemko, 2006; Lyngstad & Prskawetz, 2010). Yet, these fertility transitions are embedded in family formation trajectories that typically lead through cohabitation, marriage, then having one or more children, and possibly separation and re-partnering in between.

This paper combines the sibling approach with sequence analysis to scrutinize the impact of shared parental background characteristics on sibling similarity in holistic family formation trajectories. We address three research questions: (1) are family formation trajectories of siblings more similar than among comparable unrelated persons, (2) can shared

parental background characteristics account for sibling similarity in family formation, and (3) in which way is siblings' family formation more similar?

Drawing on Finnish register data our empirical analysis employs (a) conditional assignment to build an analysis sample of sibling dyads and comparable unrelated dyads, (b) sequence analysis to measure similarity in family formation within these dyads, (c) dyadic regression analysis to assess the impact of shared parental background characteristics, and (d) cluster analysis to examine in which way siblings' family formation is similar. To our knowledge this is the first study that combines a sibling design with sequence analysis to study family of origin effects on family formation. This analytical strategy aims to contribute to the literature in three regards.

First, to our knowledge this is the first study to establish the simple descriptive fact that siblings are more similar to one another also in their holistic family formation trajectories and not only in isolated fertility transitions. This is possible through the combination of a sibling design with sequence analysis. We thereby recognize the importance of linked lives within families (Elder 2003) as well as the interdependence of multiple family formation events across the life course and the diversity of family formation processes (Wu & Li, 2005, p. 112).

Second, we compare sibling dyads with a control group of unrelated dyads that are identical with regard to parental background characteristics. Siblings naturally share parental background characteristics, such as parental education and early childhood family structure that unrelated persons to not share. This simple compositional effect alone might generate sibling similarity in family formation. Our sibling design enables us to scrutinize to what extent these shared background characteristics account for sibling similarity in family formation trajectories.

Third, by employing sequence analysis and cluster analysis we can examine qualitative patterns of family formation and the nature of similarity of siblings' family

formation. Beyond directly testing the impact of parental background characteristics on sibling similarity, we thereby examine in which way siblings' family formation is more similar, i.e. which substantive family formation patterns are more likely to 'run in the family'.

BACKGROUND AND THEORY

Family of origin effects on family behavior fall into two broad categories: shared parental background effects and mutual sibling influence. Sibling similarity in family formation can result from either or both of them.

Research on *parental background effects* focuses on intergenerational transmission – the degree to which parents and their children are similar in the occurrence and timing of focal family events: fertility (Barber, 2001; Murphy & Knudsen, 2002; Murphy, 1999), marriage (Feng, Giarrusso, Bengtson, & Frye, 1999; Jennings, Axinn, & Ghimire, 2012; van Poppel, Monden, & Mandemakers, 2008), or divorce (Amato & DeBoer, 2001; Amato, 1996; Diekmann & Engelhardt, 1999; Wolfinger, 2000). Most of these studies find relatively modest but significant similarity between parents' and their children's family behavior. Two recent studies compare holistic family formation trajectories of parents and their children directly in a dyadic approach (Fasang & Raab, 2013; Liefbroer & Elzinga, 2012). Liefbroer and Elzinga (2012) conclude that in spite of profound social change in family formation over the past decades, there is persistent continuity in parents' and their children's family formation. Fasang & Raab (2013) show that in addition to strong intergenerational transmission, i.e. similarity, there are other systematic patterns between parental and filial family formation: a pattern of delayed transmission, and a pattern of intergenerational contrast, where children show very different family behavior from their parents.

Research on *sibling similarity* in family behavior has concentrated on particular aspects of fertility. Findings include that older siblings, particularly brothers, affect younger siblings' sexual initiation (Haurin & Mott, 1990; Widmer, 1997), sisters influence one another

in the transition to first birth but not notably in second parity transitions (Lyngstad & Prskawetz, 2010), the number of siblings affects own family size preferences (Axinn, Clarkberg, & Thornton, 1994) as well as own fertility (Murphy & Knudsen, 2002). Further, girls' teenage pregnancy and premarital birth increase their sister's risk of experiencing these events as well (East & Jacobson, 2001; East, 1998; Powers & Hsueh, 1997; Powers, 2001). In these studies it is generally unclear whether the sibling effect extends to the entire family formation process or only specific transitions and what the relative importance of shared parental background characteristics and mutual sibling influence is in generating the observed effects.

We add to previous research by examining whether there is sibling similarity in holistic family formation trajectories, acknowledging that different events in the family formation process such as cohabiting, marrying and having a child, do not happen independently from one another. In addition, we scrutinize the role of shared background characteristics for generating sibling similarity in holistic family formation trajectories. We thereby offer a detailed examination of one of the two broad categories of family of origin effects - parental background and mutual sibling influence - that potentially generate sibling similarity in family formation. In particular, we examine parental background in terms of parental marital history as an indicator of children's early childhood family structure and parental education as an indicator for socio-economic background. While these two indicators can certainly not capture the full "package" of shared parental background effects, they are two central factors that have been shown to correlate considerably with other observable und unobservable parental background characteristics, such as different parenting styles (Chan & Koo, 2010; Lareau, 2003). Subsequently, we discuss empirical findings and theoretical considerations on these two parental background characteristics. We first give an account of the mechanisms at work more broadly and then focus on their potential role in generating sibling similarity in family formation.

Parental marital history and early childhood family structure

Research on the intergenerational transmission of family behavior persuasively shows that the structure of the family of origin is associated with family formation in adulthood. Children from disrupted families have a higher risk of divorce than children from intact families (Amato, 1996; Diekmann & Engelhardt, 1999). Moreover, children of divorce have lower overall rates of marriage (Erola, Härkönen, & Dronkers, 2012; South, 2001), although they tend to be at higher risk for teenage marriage (Wolfinger, 2003). In addition, children who experience several marital transitions or alternative living arrangements during childhood have a higher "risk of forming a first cohabitational union" (Teachman, 2003). More generally, instable living arrangements during childhood promote instable family trajectories in adulthood (Carlson & England, 2011).

Family structure is closely linked to economic inequalities. Children from disrupted families often grow up in single-parent households and have an increased risk of experiencing economic hardship (Amato, 1996; McLanahan & Percheski, 2008), which negatively affects children's status attainment (McLanahan & Sandefur, 1994; Sigle-Rushton & McLanahan, 2004). Low socioeconomic status, in turn increases marital instability (Amato, 1996; Wolfinger, 2005), as it is associated with a number of relationship stressors including higher incidences of poverty, substance abuse and domestic violence.

Parental preferences for their children's family formation do not only vary with education as noted above; they are also associated with parent's own family formation experiences. Several studies report that divorced parents and their children are more tolerant toward non-traditional family forms (Axinn & Thornton, 1996; Cunningham & Thornton, 2006). Although, value transmission in disrupted families is weaker than in intact families, parental preferences still affect their children's attitudes (Van Der Valk, Spruijt, De Goede, Larsen, & Meeus, 2008).

Siblings might be more similar in their family formation than unrelated persons, simply because they grow up in similar family structures and share their parents' marital history to a greater extent. Of course, there will be considerable variation in how two siblings experience events such as a parental divorce. They will generally experience divorce at different developmental stages and will have different coping resources available to them. However, compared to random unrelated persons, they are likely to share more of their parental marital history and early childhood family structure. If these factors are influential in generating sibling similarity in family formation, unrelated persons that share similar parental marital history and early childhood family structure experiences, should also be more similar in their family formation. Sibling similarity in family formation compared to unrelated persons should then be smaller once parental marital history and family structure of the family of origin are taken into account.

Parental education

Parental education can influence their children's family formation *directly* and *indirectly* through the transmission of parental education to children's education. Socialization and social control theories (Barber, 2000; Bernardi, 2003; Starrels & Holm, 2000) emphasize a *direct effect* of parental education. Following this rationale, highly educated parents will increase the similarity of their offspring's family trajectories in two regards. First, highly educated parents have higher occupational aspirations for their children and are more likely to favor delayed family formation (Barber, 2001; Plotnick, 2007; Trent & South, 1992). Through socialization parents shape their children's family plans in accordance with these preferences. Although the socialization effect is partly mediated by status transmission, several studies have shown an impact of parental education even after children's education has been taken into account. For instance, empirical evidence from Norway shows a direct influence of parents' education on children's first union formation (Blom, 1994; Wiik, 2008). Second,

parents with higher status are in a better position to exert social control. They likely have more educational and financial resources to influence their children to behave in accordance to their preferences than lower educated parents (Axinn & Thornton, 1992; Barber, 2001).

Indirect effects of parental education through intergenerational status transmission are rather obvious given the tight link between educational trajectories and family formation in modern societies. Parental education will then matter to the extent that it predicts children's own education. The prolongation of education during the last decades is seen as a major driver of postponement in the transition to adulthood (Arnett, 2000; Billari & Liefbroer, 2010; Fussell & Furstenberg, 2005). Transitions like marrying and childbearing usually take place after a person left the educational system and reached some degree of financial independence. Although higher levels of education are often linked to individualization, educational institutions also standardize early life courses to be more uniform (Brückner & Mayer, 2005; Mayer & Schoepflin, 1989; Shanahan, 2000). The standardizing effect of education continues into later adulthood for highly educated persons than for the lower educated whose lives evolve outside the grip of educational institutions at much younger ages.

Beyond structural effects, education is also related to children's own attitudes and preferences toward family formation. High education is associated with post-material values of self-realization and with liberal attitudes toward family formation (Inglehart & Baker, 2000; Lesthaeghe & Neidert, 2006; Trent & South, 1992). Highly educated young adults engage in a period of emerging adulthood in which the transitions to parenthood and marriage are postponed in favor of a life-stage of self-exploration without "enduring commitments" (Arnett, 2000). Adolescents in higher educational tracks expect and desire marriage and parenthood at older ages and see themselves at a lower risk of childbearing out of wedlock (Plotnick, 2007: 957).

Attitudinal differences by education are also important in structuring gender differences in family formation. To explain educational differences in teenage childbearing

and lone motherhood, McLanahan and Percheski (2008) propose that motherhood plays a different role as a source of identity for higher and lower educated women. Lower educated women expect motherhood to be psychologically more rewarding and view it as a more essential role in their life course. In contrast, highly educated women "have other possible sources of identity from which they can derive meaning and fulfillment." (McLanahan & Percheski, 2008: 262). In addition, the opportunity costs of parenthood are higher for highly educated women. Consequently, they more often postpone the transition to parenthood or remain childless than their lower educated counterparts.

For men on the other hand, both high and low education likely delays family formation compared to medium educated men, albeit for quite different reasons. Highly educated men delay family formation until they completed extensive education and placed themselves in an economically secure job, similar to highly educated women. In contrast, in view of the average increase in women's education, lower educated men with few resources appear as less attractive marriage partners. Struggling to find a partner they postpone marriage and are more likely to live with their parents until later ages (Carlson & England, 2011; McLanahan & Percheski, 2008).

Both indirect and direct effects of parental education suggest that siblings will be more similar in their family formation simply because they share parental education and are more likely to have similar educational trajectories themselves. If indirect and direct effects of parental education are important mechanisms generating sibling similarity in family formation, unrelated persons who share the same parental educational background should also be similar to one another in their family formation. Part of the sibling similarity in family formation should then vanish once parental education and own education of siblings and unrelated dyads are taken into account. Subsequently, we empirically test to what extent two crucial parental background characteristics - parental marital history and parental education – account for sibling similarity in family formation. While they obviously cannot capture the

full “package” of observable and unobservable shared parental background, they are solid indicators of early childhood family structure and socioeconomic background, which have been shown to correlate considerably with other possibly influential parental background factors, such as parenting styles and relationship quality (Chan & Koo, 2010; Lareau, 2003).

DATA AND METHODS

Data

The empirical analysis uses Finish register data that consists of a random sample of 1 percent of the population in 1970 (Österbacka, 2004; Statistics Finland, 1996). All subsequent family members were included in the sample. We use yearly panel data that was collected between 1987 and 2007 for the cohorts born 1969 to 1977 for which it is possible to construct family formation sequences in young adulthood from age 18 to age 30. Data requirements for the analysis are high: we need to follow individuals and their siblings over several years to reconstruct their family formation process, and establish intergenerational links to parents and between siblings. The benefits of the Finish register data for this type of analysis outweigh the censoring at the age of 30, which is arguably less problematic for studying sibling similarity in family formation. We see no obvious reason to assume systematic effects that would affect sibling similarity in family formation differently before and after age 30. In addition, the timing and sequencing of family formation in the “third decade of life” (Fussel and Furstenberg 2005) is of particular interest as the life stage when emerging adulthood is posited to unfold. Further, in view of cumulative (dis-)advantage over the life course, the occurrence or absence of specific family formation, educational and career entry events in this life stage are highly consequential for future life courses (Dannefer 2003, DiPrete and Eirich 2006). Our analysis can fully capture family formation in this third decade of life. Moreover, the Finish case allows a fairly conservative test of the impact of shared parental background characteristics. Finland, along with other Scandinavian countries represents a fairly egalitarian

social democratic welfare state in which variation in family formation with parental background can be assumed relatively small compared to liberal and conservative welfare regimes.

Only people who have at least one sibling are included. This allows us to separate the impact of having any siblings from the specific impact of siblings and thus ensures greater rigor in the sibling design. People are identified as siblings if they have the same mother. The mother is identified through living in the same household at the birth of the child. In total, we have 9,263 individuals. Among these there are 4,994 unique sibling dyads and we generate 9,263 unrelated dyads in a procedure described below. This adds up to a total of 14,259 dyads. Below we explain each of the analysis steps in detail.

Analytic strategy

To address the *first* descriptive research question, whether there is sibling similarity in family formation, we calculate the distance between the two family formation sequences in each dyad using sequence analysis (Abbott, 1995; MacIndoe & Abbott, 2004). Significantly lower sequence distances within sibling dyads than within unrelated dyads indicate sibling similarity in family formation.

To inform the *second* research question, whether parental background characteristics can account for this sibling similarity, we use the dyadic sequence distance as a dependent variable in a dyadic regression to directly measure the impact of parental background characteristics on the similarity of two people's family formation. There is a possibility that parental background characteristics affect siblings differently than unrelated persons, for example because there are mutually reinforcing effects within families or there is an interaction between parental background characteristics and (unobserved) mutual sibling influence. To take this into account we calculate fully interacted dyadic regression models by sibling status, such that each effect is given separately for unrelated dyads and for sibling

dyads. This is essentially the same as calculating separate models of a sample of unrelated dyads and a sample of sibling dyads with the added benefit that we can assess whether different effects for unrelated and sibling dyads are statistically significant within one fully interacted model.

The *third* research question, in which way siblings family formation is more similar, is addressed by calculating sequence distances between each possible pair of individual sequences and running a cluster analysis on the resulting pairwise distance matrix. The clusters represent salient substantive family formation patterns, for example a pattern of "early marriage" or of "multiple cohabitating unions and non-marital childbearing". If the members of sibling dyads have a higher probability to be in the same cluster than the members of unrelated dyads, this further supports sibling similarity in specific family formation patterns and thereby can augment the interpretation of the regression results to inform the mechanisms that potentially govern sibling similarity in family formation.

Analysis sample

As illustrated in the upper panel of figure 1, siblings naturally share the same parental background characteristics, whereas this is not the case for unrelated dyads. As a result there is heterogeneity in parental background characteristics for unrelated dyads but not for sibling dyads. This has two implications: First, it obscures the comparison of siblings and unrelated dyads, such that we are at risk of overestimating the difference between them. Second, it would lead to ambiguous reference categories in the dyadic regression for parental background characteristics: more combinations of parental background characteristics are possible for unrelated dyads than for sibling dyads. For example, siblings always share a parental divorce, whereas in unrelated dyads one person might have experienced parental divorce but the other did not. We therefore construct two analysis samples. First using random assignment, such that we randomly assign each sibling an unrelated person to form dyads of

unrelated persons (upper panel of figure 1). Second, to equalize the variation of parental background characteristics between sibling dyads and unrelated dyads, we perform a conditional assignment generating unrelated dyads by assigning two individuals to one another conditional on sharing the same parental background characteristics (lower panel of figure 1).

- Figure 1: Random and conditional assignment of unrelated dyads -

The analysis sample is constructed as follows (right panel of figure 1). To generate sibling dyads, each of the 9,263 individual focal children in our sample is matched with a sibling. For two-child families there naturally is only one possible sibling match. In families with more than two siblings (10.5% of all families), we randomly chose a sibling. To ensure that the sample only contains unique sibling dyads we exclude doublets, i.e. a match of the same two siblings once treating sibling 1 as the first dyad member and once treating sibling 2 as the first dyad member. We remain with 4,994 unique sibling dyads.

To generate unrelated dyads each focal child is matched with an unrelated child conditional on sharing the same combination of three parental background characteristics: parental education (low, medium, high), mothers age at first marriage (lowest 25 percent, middle 50 percent and top 25 percent of age distribution) and a dichotomous variable, whether the parents divorced. This yields 18 (3*3*2) possible combinations. Based on these conditions we are able to generate 9,263 unrelated dyads, i.e. there is a suitable unrelated match for each individual in the sample. Note that this conditional assignment simply equalizes the possible variance in observed parental background characteristics for sibling dyads and unrelated dyads. We later also include these parental background characteristics in the dyadic regression model to directly estimate their impact on similarity in family formation.

Methods

Sequence analysis is used to measure similarity between family formation sequences. We define ten family formation states combining residential situation, relationship status and the number of children: "parental home, single, no child" (PSNC), "own home, single, no child" (OSNC), "parental or own home, single, 1+ child" (S1C), "cohabiting, no child" (CNC), "cohabiting, 1 child" (C1C), "cohabiting, 2+ children" (C2C), "married, no child" (MNC), "married, 1 child" (M1C), "married, 2 children" (M2C), "married, 3+ children" (M3C). Those who separate from a cohabiting relationship or divorce are considered "single" again, prioritizing their residential situation over their legal marital status. Divorce before age 30 is rare in Finland and cohabitation as a substitute for marriage is relatively common. "Single" generally refers to not being in a cohabiting or married relationship, but might include other relationship forms that we are unable to identify, such as living apart together.

To measure similarity in this process optimal matching analysis, the most common form of sequence analysis, calculates pairwise distances between all sequences using two transformation operations – substitution and insertion/deletion of a state – to turn one sequence into another. The transformation operations are associated with specific costs and the distance between two sequences is given as the sum of these costs for alignment of one sequence with another. For a comprehensive introduction see MacIndoe & Abbott (2004). Because the sequences are censored at age 30 and the timing and spacing of events is crucial in family formation, we chose an algorithm that emphasizes the timing of events in determining sequence similarity but nonetheless accounts for the order of family formation states as well. This can be achieved by using only substitution operations and no insertion/deletion operations in optimal matching (Lesnard, 2010; MacIndoe & Abbott, 2004).

The substitution costs are specified to reflect both the substantive closeness of family formation states and the timing of transitions between them. Cleary, "married, 2 children" is

more similar to "married, 3+ children", than to "parental home, single, no children". Table A1 in the appendix shows the substitution cost matrix specified based on the substantive closeness of family formation states. To appropriately account for similarity in terms of timing, we weight this substitution cost matrix with the time point specific transitions probabilities between family formation states, such that the substitution of states between which transitions occur very frequently are less costly and generate less distance (Lesnard, 2010). This results in a separate substitution cost matrix for each time point (year) by weighting the substantive substitution cost matrix with the time point specific transition probabilities for each year. The approach can be considered a modified version of the Dynamic Hamming measure (Lesnard 2010) that accounts for the timing and the order of family formation events at the same time.

The output of the sequence analysis is a pairwise distance matrix that contains sequence distances for each possible combination of siblings and unrelated persons. Each of the 9,263 individual family formation sequences is compared to all other sequences in a pairwise comparison, which yields $\frac{N(N-1)}{2}$ comparisons = 42,896,953, i.e. cells in the distance matrix. This matrix is the basis for three sets of analyses that speak to the three research questions.

First, to address the descriptive question, whether siblings' family formation is more similar than for unrelated dyads, we simply extract the distance values for the sibling dyads and the randomly assigned unrelated dyads (upper panel of figure 1) from the larger distance matrix. The distribution of the sequence distances among siblings is then compared to that among randomly assigned unrelated dyads.

Second, to test whether parental background characteristics account for sibling similarity in family formation we use the dyadic distances for the siblings and the unrelated dyads as the dependent variable in a dyadic regression. The independent variables include sibling status of the dyad, gender constellation, age difference, education, parental education,

mother's age of first marriage and parental divorce. Table A2 in the appendix shows descriptive statistics for all independent variables and the dyadic outcome measure. We calculate three models. Model 1 (M1) is a baseline model using the first analysis sample of randomly assigned unrelated dyads and including only sibling status. Model 2 (M2) is the same as M1 but using the second analysis sample of conditionally assigned unrelated dyads that are equalized on parental background characteristics to resemble the sibling dyads. Model 3 (M3) is a fully interacted model by sibling status to identify sibling specific effects including the full set of covariates. M3 uses the conditionally assigned unrelated dyads that equalize variation on parental background characteristics and thus also enables to specify the same reference categories for siblings and unrelated dyads. The dyads in our analysis samples are unique but by design the same individual can occur more than once in a sibling and an unrelated dyad. We therefore calculate robust standard errors.

Third, to examine whether siblings are concentrated in specific substantive family formation patterns, we apply Ward cluster analysis using the entire distance matrix. Several cluster cut-off criteria, including Point Biserial Correlation, Average Silhouette Width, and the Calinski-Harabasz index, support a five clusters solution (Hennig and Liao 2010; Kaufman and Rousseeuw 1990; Milligan and Cooper 1985). We then examine whether siblings have a higher probability to be in the same family formation cluster than unrelated dyads, and whether this particularly is the case for specific substantive family formation patterns. The sequence analysis and the calculation of different cluster cut-off criteria were conducted using the TraMineR and the WeightedCluster packages in R (Gabadinho et al. 2011; Studer 2013).

RESULTS

Is siblings' family formation more similar (RQ 1) and what role do shared parental background characteristics play for sibling similarity (RQ 2)?

We begin by establishing that sibling's family formation is significantly more similar than for unrelated dyads, particularly for same sex siblings. This holds true also after conditional assignment of unrelated dyads that filters out their greater heterogeneity in parental background characteristics. Figure 2 shows the distribution of the dyadic sequence distances for siblings (light grey line) and randomly assigned unrelated dyads (dark grey line) separately for the complete sample, opposite sex dyads, female dyads and male dyads. The distances are normalized between zero and 100, where zero indicates two identical sequences and 100 indicates the most dissimilar family formation trajectories. The width of the curves reflects 90 percent asymptotic confidence intervals. For all subgroups the two curves for siblings and unrelated dyads largely do not overlap indicating that in these locations of the distribution the difference between sibling dyads and unrelated dyads is statistically significant. This difference is particularly pronounced for same sex siblings, where the gap between the two curves is largest (bottom graphs in figure 2).

- Figure 2: Distribution of sequence distances among sibling dyads and randomly assigned unrelated dyads -

Family formation sequences are most different for women with an average distance of 47.3 for unrelated women and an average distance of 41.9 for sisters compared to 38.2 for unrelated men and 34.1 for brothers. Note that sisters' family formation is still more different (41.9) than for unrelated men (38.2) up to the age of 30. Nonetheless, the difference between siblings and unrelated persons is larger for sisters at 47.3 - 41.9 = 5.4 compared to 38.2 - 34.1 = 4.1 for brothers.

These findings underline the primacy of gender in structuring family formation in two regards. First, same sex siblings are much more similar to one another than opposite sex siblings. Second, there is far more variability of family formation for women until the age of 30 than for men. If most men's family formation events are simply delayed, expanding the observation window to age 40 might somewhat equalize this variation between men and women. However the significant sibling effect for brothers also until the age of 30, suggests that the censoring is not majorly distort the analysis of sibling similarity in family formation. Our findings until age 30 further highlight that women are much more actively engaged in family formation in the crucial years between age 18 and 30 that coincide with the accumulation of educational credentials and labor market entry and thus set the stage for future life-time labor market success.

Overall, the sibling effect, calculated as the difference in mean distance between siblings and randomly assigned unrelated dyads, is 44.1 – 40.1 = 4. The sibling effect after conditional assignment (not shown in figure 2) is slightly reduced at 43.7 - 40.2 = 3.5. The conditional assignment on parental background characteristics thus reduces the sibling effect by 12.5 percent, but siblings remain more similar in their family formation after conditional assignment.

The dyadic regression presented in table 2 further supports this finding. Note that negative coefficients indicate less distance and thus more similar family formation in a dyad. Model 1 (M1) shows that siblings are on average significantly more similar to one another than randomly assigned unrelated dyads by -4.10 on the distance measure ranging from zero to 100, which corresponds almost exactly to the mean difference of 4.00 established in figure 2. Model 2 (M2) shows that this effect is very similar at a difference of -3.68 between siblings and conditionally assigned unrelated dyads that have the same variation in parental background characteristics as siblings. The size of the sibling effect is thus reduced by 10.2 percent through the conditional assignment on parental background characteristics from M1

to M2, which again closely corresponds to the 12.5 reduction in effect size established in the simple mean comparison in Figure 2.

Even though many of the covariates in M3 significantly account for the variance in the dyadic distance measure and the adjusted R-square increases from 0.01 in M1 to 0.12 in M3, they have little to add to the explanation of the sibling effect. The sibling main effect remains significant when including the full set of covariates (M3), but is reduced to -2.81, which corresponds to a 31.4 percent reduction of the sibling effect compared to M1 based on randomly assigned unrelated dyads and a 23.6 percent reduction compared to M2 based on conditionally assigned unrelated dyads. We find only three significant sibling interactions: for sisters, twins and high education. Note that in M3 the coefficients for the main effects in the left column refer to the effects for unrelated dyads. The interaction effects in the right column capture the additional effects for siblings – if there are significant differences in the impact of an independent variable for unrelated dyads and siblings. Overall, the results support that sibling similarity in family formation is only moderately generated by the compositional effects of shared parental background characteristics, in terms of early childhood family structure and parental education (RQ 2). They affect similarity in family formation in the direction we would expect - but as can be seen in M3 mostly for siblings and unrelated dyads alike.

- Table 1: Ordinary Least Squares Regression predicting dyadic distance -

The results from M3 substantiate the descriptive gender findings (figure 2): women are overall most different in their family formation but sisters are relatively more similar than brothers compared to conditionally assigned unrelated dyads of the respective sex. This is visible in the positive main effect for female dyads and the significant negative interaction for sisters in M3. In contrast, unrelated men and brothers both are more similar to one another by 7.21 distance points than opposite sex dyads.

An age difference of more than three years increases sequence distance slightly by 0.93 for siblings and unrelated dyads - possibly a “mini cohort effect” for successive birth cohorts. There is a strong twin effect of -7.37 measured as siblings who were born in the same year. On the one hand, shared genetic background may account for this substantial similarity in family formation of twins. On the other hand, twins also share more of their environment and socialization experiences than siblings. There might be a culture or norm for twins to jointly step through ‘rites of passage’ in the transition to adulthood, such as leaving the parental home together or orchestrating a double marriage. Indeed, our data indicate that the probability to leave the parental home in the same year is 27 percent for twins compared to only 13 percent for siblings, 13 percent for unrelated young adults who were born in the same year, and 11 percent for unrelated young adults who were not born in the same year.

In support of the paramount importance of education in structuring early life courses and family formation, we find strong education effects. The higher the dyads combined education, the more similar is their family formation for both siblings and unrelated dyad, underlining the greater heterogeneity of family formation among the lower educated whose early life courses unfold outside of educational institutions from a much younger age. The only sibling-specific effect is that two highly educated siblings are significantly more similar in their family formation than two highly educated unrelated persons, net of controlling for parental education. Arguably there is mutual reinforcement of specific family formation patters among highly educated siblings. Additional analyses (available upon request) on the interaction of gender and education separately for siblings and unrelated dyads show that education has a much stronger effect for sisters than for brothers substantiating that women’s family formation is much more tightly interrelated with their educational and employment trajectories. Sibling status apparently reinforces the impact of gender and education in structuring family formation.

High parental education generates more similar family formation for siblings and unrelated dyads alike, net of their own education. This supports a direct influence of parental education beyond indirect effects through intergenerational status transmission as posited in socialization and social control theories. Possibly, highly educated parents make use of greater resources to guide their children's family formation, including financial support for a longer time into early adulthood. In contrast, children of lower educated parents leave the radar of parental influence and are independent at younger ages.

In line with our expectations, parental marital history as an indicator of early childhood family structure equally impacts similarity in family formation. Like education, the effects do not vary for siblings and unrelated dyads. Two members of a dyad are more similar to one another if their mother married late, likely by also delaying their own family formation. They are less similar to one another if their mother married young. This dissimilarity might be driven by intergenerational behavioral transmission: if they start family formation early themselves there is more variability early on in these sequences and the risk of future union disruption and instable family formation is higher.

In support of previous research (McLanahan & Bumpass, 1988; Wolfinger, 2000), children whose parents' divorced have more instable family formation themselves and are less similar to one another in their family formation, again within sibling and unrelated dyads alike. Parental divorce is a very different experience even for siblings in the same family depending on the age of the child, custodial arrangements and the level of parental conflict prior, during and after the divorce (Amato, 2000). Our results underline this heterogeneity in the divorce experience and do not support a uniform impact of parental divorce on siblings' family formation.

We conclude that parental background characteristics impact similarity in family behavior much in line with our expectations, but contribute surprisingly little to explaining the sibling effect in family formation. To gain a deeper insight into the nature of this sibling similarity,

we next directly examine whether siblings have a higher probability to both experience the same substantive family formation patterns than unrelated dyads.

In which way is siblings' family formation similar (RQ 3)?

Figure 3 shows sequence index plots (Scherer 2001) of the five family formation patterns derived with cluster analysis. Each horizontal line in the graph represents one individual sequence, with different colors indicating different family formation states. Since it is impossible to plot all 9, 263 sequences due to over-plotting in the graph, figure 3 shows a random selection of five percent of the sample to represent the substantive pattern of each cluster. The clusters are ordered ascending according to level of education in each group. The sequences are sorted by age of first cohabitation within each cluster. Table 2 shows descriptive statistics for all clusters.

- Figure 3: Sequence Index Plot of five family formation clusters -

- Table 2: Description of clusters -

- Figure 4: Probability of two dyad members to be in the same cluster -

The highest educated cluster (66 percent medium or high education, table 2) at the top of figure 3 follows a "traditional" pathway of leaving the parental home, cohabitating, marrying in their late twenties and having one or two children by age 30. In the middle of the educational distribution we find two groups that are characterized by childlessness: a pattern we name "home stayers" since they remain in the parental home until their late twenties and a "living alone/transitional cohabitation" pattern. The latter consists of young adults who establish their own households in their early twenties and have several transitional periods of cohabitation throughout their twenties. The "home stayers" are primarily men. Finally, we find two family formation patterns at the bottom of the educational distribution that are

characterized by early parenthood and high parities either within marriage or - for the lowest educated - out of wedlock in cohabiting unions and as single parents. They are primarily experienced by women (70 and 69 percent). The clusters underline the stratification of family formation by education and gender visible in the dyadic regression (table 1): low education goes along with more eventful and variable family formation trajectories of high parities and parenthood out of wedlock. High education is associated with a "traditional" timing and sequencing of family formation and establishing a one- or two-child family secured in marriage by age 30. Another pattern that is characteristic for high and medium education is a relatively uniform delay of family formation (table 2).

Figure 4 shows the probability of the focal person's dyad partners – siblings and conditionally assigned unrelated persons – to be in the same cluster as the focal persons. If two members of a sibling dyad have a higher probability to be in the same cluster than two members of an unrelated dyad this indicates that specific family formation patterns are concentrated within families and offers insights into the nature of sibling similarity in family formation. The line in the middle of figure 4 and the corresponding percentages show the overall probability to be in the same cluster as the focal person conditional on the focal person's cluster membership. The graph sets this line to zero to illustrate deviations from the mean for siblings and unrelated persons. First, across all clusters siblings have a higher probability to be in the same cluster with the focal person (solid lines on the right of figure 4) than unrelated persons (dashed lines on the left). For instance, the mean probability to be in the "parenthood out of wedlock" group – given that the focal person is in this cluster – is 23 percent for siblings but only 14 percent for unrelated persons. This corresponds to a 35 percent deviation from the mean, i.e. a 35 percent higher probability for a sibling to be in the same cluster as the focal person than for the population average.

The clusters in figure 4 are ordered descending according to the degree of sibling concentration within each group. Siblings particularly have a higher probability to be in the

same cluster in the family formation patterns that are associated with educational disadvantage ("parenthood out of wedlock" and "early high parities") and in the "home stayers" group. Those associated with educational disadvantage are primarily experienced by women, and the home stayer group by men. These findings support that specific gendered family formation patterns and economic disadvantage are encapsulated within families and thus contribute to the reproduction of inequality (McLanahan & Percheski, 2008). We conclude that sibling effects play a rather minor role in stratifying family formation relative to education and gender that work in similar ways for siblings and unrelated persons alike. Yet family internal dynamics seem to reinforce this stratification of family formation by education and gender, such that siblings have a higher probability to experience a similar family formation pattern, particularly in patterns of early high parities and parenthood out of wedlock that go along with educational disadvantage.

DISCUSSION

This paper proposes a novel research design using sibling comparisons and sequence analysis to study family of origin effects on holistic family formation trajectories. We seek to contribute to the literature in three regards. First, going beyond a focus on sibling similarity in isolated fertility transitions in previous research, we establish that siblings are moderately but significantly more similar to one another in holistic family formation trajectories than unrelated persons. This is particularly the case for same sex siblings. We thereby acknowledge the interrelation of different family formation events. Second, we systematically scrutinize the role of shared parental background for generating this sibling similarity in family formation. We find that equalizing siblings' and unrelated dyads' parental background characteristics decreases the sibling effect moderately by about 10 percent. Further the dyadic regression showed that parental background largely affects similarity in family formation for siblings and unrelated dyads in the same way. We conclude that these shared parental

background characteristics only play a relatively marginal role in generating sibling similarity in family formation, compared to the stratifying forces of gender and education. Apparently, there are other family internal dynamics e.g. latent parent characteristics such as parent-child relationship quality and mutual sibling influence that generate the remaining unexplained sibling similarity. To gain a deeper insight into the nature of sibling similarity in family formation, we subsequently showed that substantive family formation patterns that are associated with economic disadvantage are concentrated within families. Brothers tend to share a pattern of 'home staying', whereas sisters are concentrated in family formation patterns of 'early high parities' and 'parenthood out of wedlock'. These results suggest that siblings reinforce one another in following family formation patterns stratified by education and gender.

To further improve our understanding of sibling similarity in family formation, future research should address several issues. First, lacking data on the relationship quality between parents and their children as well as among siblings, we were unable to account for possible moderating effects of these psychological characteristics. Close emotional relationships can function as 'transmission belts' (Schönpflug, 2001) and could be one explanation for the remaining unexplained sibling effect we observe. Second, more detailed indicators of parents' and siblings' social and occupational status would be desirable to assess family background effects in greater detail. Yet, it will remain impossible to account for all relevant observable and unobservable parental background characteristics in any model. In this study, we were able to rigorously scrutinize the impact of two crucial background factors that have been shown to correlate with many other potentially relevant parental background characteristics, such as parenting styles (Chan & Koo, 2010; Lareau, 2003). Third, we had no information on interactions between siblings and could not directly measure mutual sibling influence in family behavior - an important alternative explanation for the sibling effect in family

formation. Finally, the analysis of more complete family formation trajectories until the age of 40 and beyond, would allow a more comprehensive examination of these processes.

A few considerations on the Finish context are in order. Given that the overall variation in family formation in Finland is relatively low in international comparison, any effects we find likely are lower than in other countries. Also, the impact of shared parental background might be larger in countries with closed stratification systems in which intergenerational status transmission is stronger than in Finland. Systematic comparisons with countries representing other social mobility and welfare state regimes would be useful.

In sum, our findings show that there is significant sibling similarity in family formation, but shared parental education and parental marital history add little to explaining this effect. Instead education and gender stand out as major stratifying forces of family formation. Sibling status rather seems to reinforce the impact of education and gender, which is visible for instance in a high probability for sisters to experience family formation patterns of high parities and parenthood out of wedlock that are both associated with educational disadvantage. As our cluster results show, scrutinizing the nature and driving forces of sibling similarity in family formation can potentially improve our insight into the link between family behavior and the reproduction of social inequality. Combining a sibling design and sequence analysis offers insights both into the amount of similarity in the full family formation trajectories of siblings and unrelated dyads and into the substantive content of this similarity in the family formation clusters. This approach is in principle easily transferable and might yield promising results when applied to other research questions, such as sibling similarity in educational and employment trajectories, or health trajectories.

REFERENCES

Abbott, A. (1995). Sequence Analysis: New Methods for Old Ideas. *Annual Review of Sociology*, *21*(1), 93–113. doi:10.1146/annurev.so.21.080195.000521

Amato, P. R. (1996). Explaining the Intergenerational Transmission of Divorce. *Journal of Marriage and Family*, *58*(3), 628–640.

Amato, P. R. (2000). The Consequences of Divorce for Adults and Children. *Journal Of Marriage And The Family*, *62*(November), 1269–1287.

Amato, P. R., & DeBoer, D. D. (2001). The Transmission of Marital Instability across Generations: Relationship Skills or Commitment to Marriage? *Journal of Marriage and Family*, *63*(4), 1038–1051. Retrieved from http://www.jstor.org/pss/3599812

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist*, *55*(5), 469–480. doi:10.1037//0003-066X.55.5.469

Axinn, W. G., Clarkberg, M. E., & Thornton, A. (1994). Family Influences on Family Size Preferences. *Demography*, *31*(1), 65–79.

Axinn, W. G., & Thornton, A. (1992). The influence of parental resources on the timing of the transition to marriage. *Social Science Research*, *21*(3), 261–285.

Axinn, W. G., & Thornton, A. (1996). The Influence of Parents' Marital Dissolutions on Children's Attitudes Toward Family Formation. *Demography*, *33*(1), 66. doi:10.2307/2061714

Barber, J. S. (2000). Intergenerational Influences on the Entry into Parenthood: Mothers' Preferences for Family and Nonfamily Behavior. *Social Forces*, *79*(1), 319–348. doi:10.1093/sf/79.1.319

Barber, J. S. (2001). The Intergenerational Transmission of Age at First Birth among Married and Unmarried Men and Women. *Social Science Research*, *30*(2), 219–247.

Bernardi, L. (2003). Channels of Social Influence on Reproduction. *Population Research and Policy Review*, *22*(5/6), 527–555. doi:10.1023/B:POPU.0000020892.15221.44

Billari, F. C., & Liefbroer, A. C. (2010). Towards a new pattern of transition to adulthood? *Advances in Life Course Research*, *15*(2-3), 59–75. doi:10.1016/j.alcr.2010.10.003

Blom, S. (1994). Marriage and Cohabitation in a Changing Society: Experience of Norwegian Men and Women Born in 1945 and 1960. *European Journal of Population*, *10*(2), 143–73.

Brückner, H., & Mayer, K. U. (2005). The De-Standardization of the Life Course: What It Might Mean and If it Means Anything Whether It Actually Took Place. In R. Macmillan (Ed.), *The structure of the life course. Standardized? Individualized? Differentiated?* (pp. 27–54). Amsterdam: Elsevier.

Carlson, M. J., & England, P. (Eds.). (2011). *Social Class and Changing Families in an Unequal America*. Stanford, CA: Stanford University Press.

Chan, T. W., & Koo, A. (2010). Parenting Style and Youth Outcomes in the UK. *European Sociological Review*, *27*(3), 385–399. doi:10.1093/esr/jcq013

Conley, D. (2008). Bringing Sibling Differences In: Enlarging Our Understanding of the Transmission of Advantage in Families. In A. Lareau & D. Conley (Eds.), *Social Class: How Does It Work?* (pp. 179–200). New York: Russell Sage Foundation.

Cunningham, M., & Thornton, A. (2006). The Influence of Parents' Marital Quality on Adult Children's Attitudes Toward Marriage and Its Alternatives: Main and Moderating Effects. *Demography*, *43*(4), 659–672. doi:10.1353/dem.2006.0031

Diekmann, A., & Engelhardt, H. (1999). The Social Inheritence of Divorce: Effects of Parent's Family Type in Postwar Germany. *American Sociological Review*, *64*(6), 783–793.

East, P. L. (1998). Impact of Adolescent Childbearing on Families and Younger Sibling: Effects that Increase Younger Siblings' Risk for Early Pregnancy. *Applied Developmental Science*, *2*(2), 62–74. doi:10.1207/s1532480xads0202_1

East, P. L., & Jacobson, L. J. (2001). The Younger Siblings of Teenage Mothers: A Follow-Up of Their Pregnancy Risk. *Developmental Psychology*, *37*(2), 254–264.

Elzinga, C. H., & Liefbroer, A. C. (2007). De-standardization of Family-Life Trajectories of Young Adults: A Cross-National Comparison Using Sequence Analysis. *European Journal of Population Revue européenne de Démographie*, *23*(3-4), 225–250. Retrieved from http://www.springerlink.com/index/10.1007/s10680-007-9133-7

Erola, J., Härkönen, J., & Dronkers, J. (2012). More Careful or Less Marriageable? Parental Divorce, Spouse Selection and Entry into Marriage. *Social Forces*, *90*(4), 1323–1345. doi:10.1093/sf/sos073

Fasang, A. & Raab, M. (2013). Intergenerational Patterns of Family Formation. Paper presented at the 2013 Annual Meeting of the Population Association of America, New Orleans.

Feng, D., Giarrusso, R., Bengtson, V. L., & Frye, N. (1999). Intergenerational Transmission of Marital Quality and Marital Instability. *Journal of Marriage and Family*, *61*(2), 451–463.

Fussell, E., & Furstenberg, F. F. (2005). The Transition to Adulthood During the Twentieth Century. In R. A. Settersten, F. F. Furstenberg, & R. G. Rumbaut (Eds.), *On the Frontier of Adulthood* (pp. 29–75). Chicago & London: The University of Chicago Press.

Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1–37.

Haurin, R. J., & Mott, F. L. (1990). Adolescent sexual activity in the family context: the impact of older siblings. *Demography*, *27*(4), 537–57.

Hennig, C., & Liao, T. F. (2010). Comparing latent class and dissimilarity based clustering for mixed type variables with application to social stratification. London. Retrieved from http://www.ucl.ac.uk/statistics/research/pdfs/rr308.pdf

Inglehart, R., & Baker, W. E. (2000). Modernization, Cultural Change, and the Persistence of Traditional Values. *American Sociological Review*, *65*(1), 19–51. doi:10.2307/2657288

Jennings, E. A., Axinn, W. G., & Ghimire, D. J. (2012). The Effect of Parents' Attitudes on Sons' Marriage Timing. *American Sociological Review*, *77*(6), 923–945. doi:10.1177/0003122412464041

Kaufman, L., & Rousseeuw, P. J. (1990). F*inding Groups in Data. An Introduction to Cluster Analysis.* Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9780470316801

Kuziemko, I. (2006). Is Having Babies Contagious? Estimating Fertility Peer Effects Between Siblings. Working Paper.

Lareau, A. (2003). *Unequal Childhoods. Class, Race, and Family Life*. Berkeley: University of California Press.

Lesnard, L. (2010). Setting Cost in Optimal Matching to Uncover Contemporaneous Socio-Temporal Patterns. *Sociological Methods & Research*, *38*(3), 389–419. doi:10.1177/0049124110362526

Lesthaeghe, R. J., & Neidert, L. (2006). The Second Demographic Transition in the United States: Exception or Textbook Example? *Population and Development Review*, *32*(4), 669–698. doi:10.1111/j.1728-4457.2006.00146.x

Liefbroer, A. C., & Elzinga, C. H. (2012). Intergenerational transmission of behavioural patterns: How similar are parents' and children's demographic trajectories? *Advances in Life Course Research*, *17*(1), 1–10. doi:10.1016/j.alcr.2012.01.002

Lyngstad, T. H., & Prskawetz, A. (2010). Do siblings' fertility decisions influence each other? *Demography*, *47*(4), 923–934.

MacIndoe, H., & Abbott, A. (2004). Sequence Analysis and Optimal Matching Techniques for Social Science Data. In A. Bryman & M. Hardy (Eds.), *Handbook of Data Analysis* (pp. 387–406). London: SAGE Publications.

Mayer, K. U., & Schoepflin, U. (1989). The State And The Life Course. *Annual Review of Sociology*, *15*, 187–209.

McLanahan, S., & Bumpass, L. (1988). Intergenerational Consequences of Family Disruption. *The American Journal of Sociology*, *94*(1), 130–152.

McLanahan, S., & Percheski, C. (2008). Family Structure and the Reproduction of Inequalities. *Annual Review of Sociology*, *34*(1), 257–276. doi:10.1146/annurev.soc.34.040507.134549

McLanahan, S., & Sandefur, G. (1994). *Growing Up with a Single Parent: What Hurts, What Helps?* Cambridge, MA: Harvard Univ. Press.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2), 159–179. doi:10.1007/BF02294245

Murphy, M. (1999). Is the relationship between fertility of parents and children really weak? *Biodemography and Social Biology*, *46*(1), 122–145. doi:10.1080/19485565.1999.9988991

Murphy, M., & Knudsen, L. B. (2002). The Intergenerational Transmission of Fertility in Contemporary Denmark: The Effects of Number of Siblings (Full and Half), Birth Order, and Whether Male or Female. *Population Studies*, *56*(3), 235–248.

Österbacka, E. (2004). *It Runs in the Family. Empirical Analyses of Family Background and Economic Status*. Åbo: Åbo Akademis förlag.

Plotnick, R. D. (2007). Adolescent expectations and desires about marriage and parenthood. *Journal of Adolescence*, *30*(6), 943–63. doi:10.1016/j.adolescence.2007.01.003

Popenoe, D. (1993). American Family Decline, 1960-1990: A Review and Appraisal. *Journal of Marriage and Family*, *55*(3), 527–542.

Powers, D. A. (2001). Unobserved Family Effects on the Risk of a First Premarital Birth. *Social Science Research*, *30*(1), 1–24. doi:10.1006/ssre.2000.0682

Powers, D. A., & Hsueh, J. C.-T. (1997). Sibling models of socioeconomic effects on the timing of first premarital birth. *Demography*, *34*(4), 493–511. doi:10.2307_3038305

Scherer, S. (2001) Early Career Patterns: A Comparison between Great Britain and West Germany. *European Sociological Review* 17, 119-114.

Schönpflug, U. (2001). Intergenerational Transmission of Values: The Role of Transmission Belts. *Journal of Cross-Cultural Psychology*, *32*(2), 174–185. doi:10.1177/0022022101032002005

Shanahan, M. J. (2000). Pathways to Adulthood in Changing Societies: Variability and Mechanisms in Life Course Perspective. *Annual Review of Sociology*, *26*, 667–692.

Sigle-Ruston, W., & McLanahan, S. (2004). Father Absence and Child Wellbeing: A Critical Review. In D. P. Moynihan, L. Rainwater, & T. Smeeding (Eds.), *The Future of the Family* (pp. 116–155). New York: Russel Sage Foundation.

South, S. J. (2001). The Variable Effects of Family Background on the Timing of First Marriage: United States, 1969–1993. *Social Science Research*, *30*(4), 606–626. doi:10.1006/ssre.2001.0714

Starrels, M. E., & Holm, K. E. (2000). Adolescents' Plans for Family Formation: Is Parental Socialization Important? *Journal of Marriage and Family*, *62*(2), 416–429.

Statistics Finland. (1996). *The Finnish Longitudinal Census Data File*. Helsinki: Statistics Finland.

Studer, M. (2013). WeightedCluster Library Manual: A practical guide to creating typologies of trajectories in the social sciences with R. *LIVES Working Papers*, 24.

Teachman, J. (2003). Childhood Living Arrangements and the Formation of Coresidential Unions. *Journal of Marriage and Family*, *65*(3), 507–524. doi:10.1111/j.1741-3737.2003.00507.x

Trent, K., & South, S. J. (1992). Sociodemographic Status, Parental Background, Childhood Family Structure, and Attitudes toward Family Formation. *Journal of Marriage and the Family*, *54*(2), 427. doi:10.2307/353074

Van Der Valk, I., Spruijt, E., De Goede, M., Larsen, H., & Meeus, W. (2008). Family traditionalism and family structure - Attitudes and intergenerational transmission of parents and adolescents. *European Psychologist*, *13*(2), 83–95. doi:10.1027/1016-9040.13.2.83

Van Poppel, F., Monden, C., & Mandemakers, K. (2008). Marriage Timing over the Generations. *Human Nature*, *19*, 7–22. doi:10.1007/s12110-008-9033-5

Widmer, E. D. (1997). Influence of Older Siblings on Initiation of Sexual Intercourse. *Journal of Marriage and Family*, *59*(4), 928–938.

Wiik, K. A. (2008). "You"d Better Wait!'--Socio-economic Background and Timing of First Marriage versus First Cohabitation. *European Sociological Review*, *25*(2), 139–153. doi:10.1093/esr/jcn045

Wolfinger, N. H. (2000). Beyond the Intergenerational Transmission of Divorce: Do People Replicate the Patterns of Marital Instability They Grew Up With? *Journal of Family Issues*, *21*(8), 1061–1086. doi:10.1177/019251300021008006

Wolfinger, N. H. (2003). Parental Divorce and Offspring Marriage: Early or Late? *Social Forces*, *82*(1), 337–353. doi:10.1353/sof.2003.0108

Wolfinger, N. H. (2005). *Understanding the Divorce Cycle: The Children of Divorce in Their Own Marriages*. New York: Cambridge University Press.

Wu, L. L., & Li, J.-C. A. (2005). Historical Roots of Family of Diversity. Marital and Childbearing Trajectories of American Women. In R. A. Settersten, F. F. Furstenberg, & R. G. Rumbaut (Eds.), *On the Frontier of Adulthood* (pp. 110–149). Chicago & London: The University of Chicago Press.

Table 1: OLS Regression predicting combined (Dynamic Hamming + own subcosts) dyadic distance

	M1 (random assignment)	M2 (conditional assignment)	M3 (conditional assignment)	
			Main Effects	*Interaction Effects*
Sibling indicator	-4.10***	-3.68***	-2.81***	
	(0.28)	(0.28)	(0.83)	
Gender Constellation (ref.: Opposite sex)				
Both female			2.86***	-2.49***
			(0.43)	(0.73)
Both male			-7.21***	-1.24
			(0.41)	(0.68)
Age Difference (ref.: 1 to 3 years)				
Born in same year			0.17	-7.52***
			(0.56)	(1.79)
More than 3 years			0.93*	0.23
			(0.38)	(0.62)
Educational level (ref.: both low)				
Low-high			-4.95***	-0.50
			(0.68)	(1.32)
Medium-low			-3.11***	-0.20
			(0.48)	(0.82)
Both medium			-7.18***	0.13
			(0.58)	(0.92)
Medium-high			-8.39***	-1.18
			(0.66)	(1.07)
High-high			-9.52***	-3.05*
			(0.97)	(1.46)
Parental educational level (ref.: low)				
Medium			-1.55***	0.71
			(0.42)	(0.65)
High			-3.79***	1.96
			(0.68)	(1.02)
Mother‘s age at marriage (ref.: average age)				
Late marriage			-2.01***	0.20
			(0.44)	(0.70)
Early marriage			2.33***	0.05
			(0.42)	(0.67)
Experienced parental divorce			2.38***	0.08
			(0.45)	(0.73)
Constant	45.94***	45.62***	50.65***	
	(0.18)	(0.18)	(0.50)	
Observations	14,259	14,259	14,259	
Adjusted R-squared	0.01	0.01	0.12	

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.0

Table 2: Description of Clusters

	Late marriage, few children	Home stayer	Living alone, transitional cohabitation	Early marriage, many children	Parenthood out of wedlock	Total
Parental background						
Parental Education						
Low	0.56	0.59	0.59	0.72	0.75	0.62
Medium	0.32	0.31	0.30	0.25	0.22	0.29
High	0.12	0.09	0.11	0.03	0.03	0.09
Mother's marriage age						
Early marriage	0.27	0.23	0.29	0.39	0.36	0.28
Marriage at average age	0.51	0.49	0.48	0.46	0.47	0.49
Late marriage	0.23	0.28	0.23	0.15	0.17	0.23
Experienced parental divorce	0.16	0.16	0.23	0.19	0.25	0.19
Children's characteristics						
Gender: Female	0. 53	0.34	0.49	0.70	0.69	0.50
Educational level						
Low	0.34	0.42	0.45	0.57	0.70	0.46
Medium	0.44	0.41	0.41	0.36	0.28	0.40
High	0.22	0.17	0.14	0.08	0.02	0.15
Children's demographic behavior						
Median age at						
Leaving parental home	22	25	21	20	20	22
First Cohabitation	23	26	22	20	21	23
First Marriage	26			23		-
First Birth	27			23	24	-
Number of kids at age 30						
Childless	0.26	0.85	0.86	0.02	0.06	0.56
Average number	1.15	0.19	0.16	2.46	1.72	0.78
(Standard Deviation)	(0.84)	(0.51)	(0.44)	(0.82)	(0.89	(1.02)
Total	0.21	0.32	0.26	0.09	0.12	1.00
(Number of observations)	(1982)	(2939)	(2395)	(844)	(1103)	(9263)

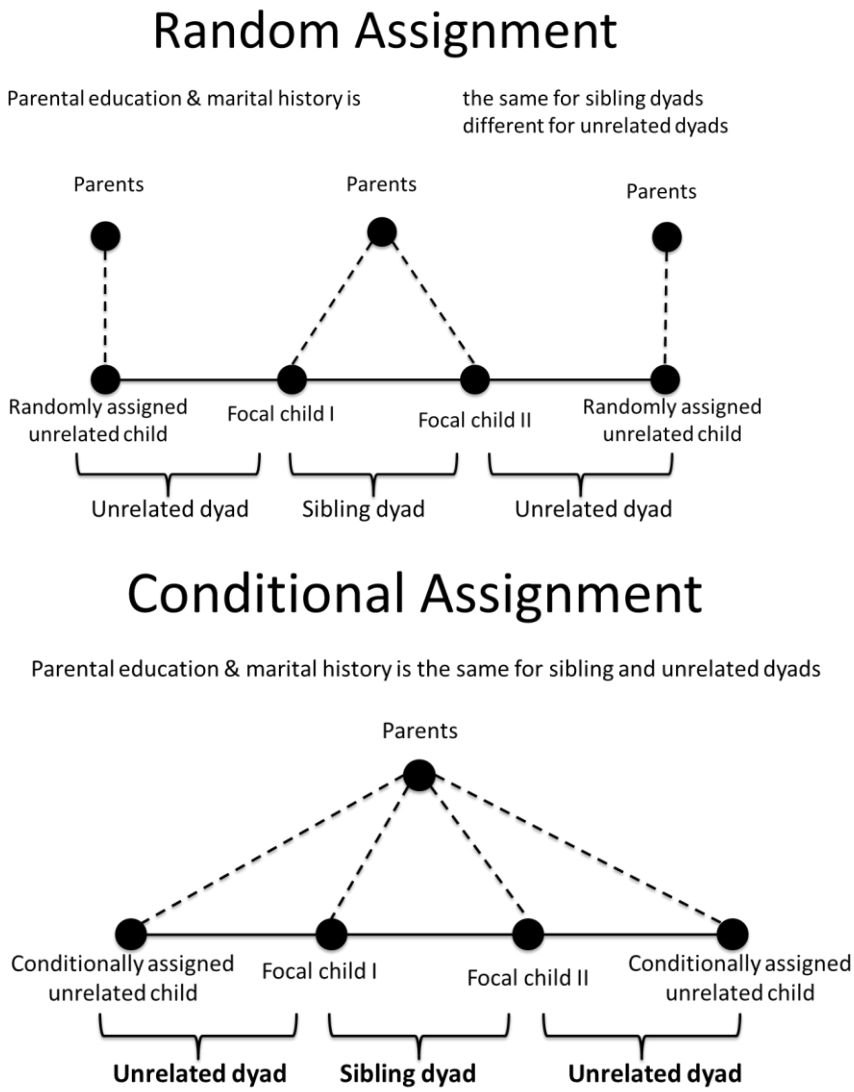
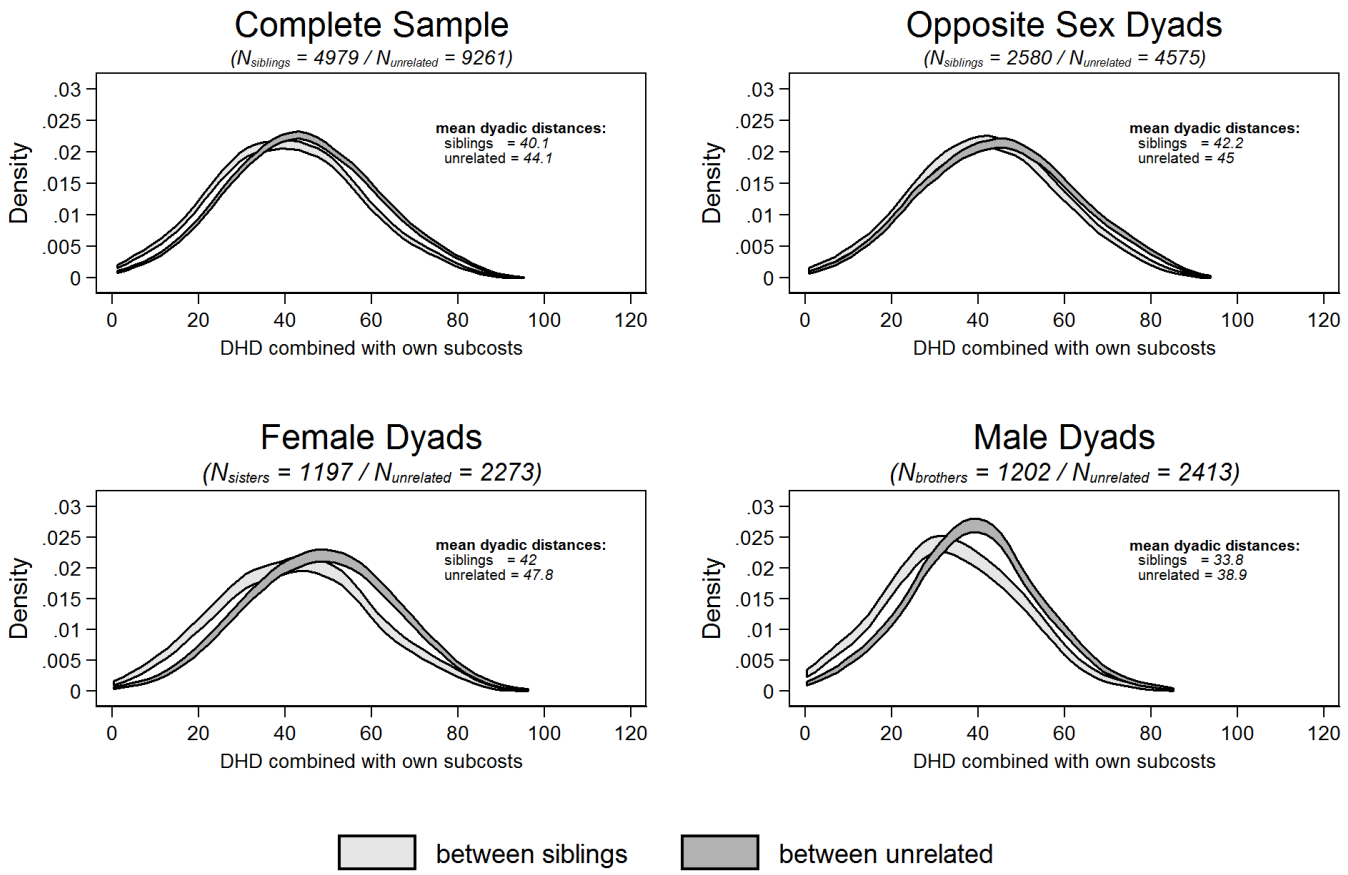


Figure 1: Random and conditional assignment of unrelated dyads



Note: Fixed-bandwidth kernel density with asymptotic confidence intervals

Figure 2: Distribution of sequence distances among sibling dyads and randomly assigned unrelated dyads

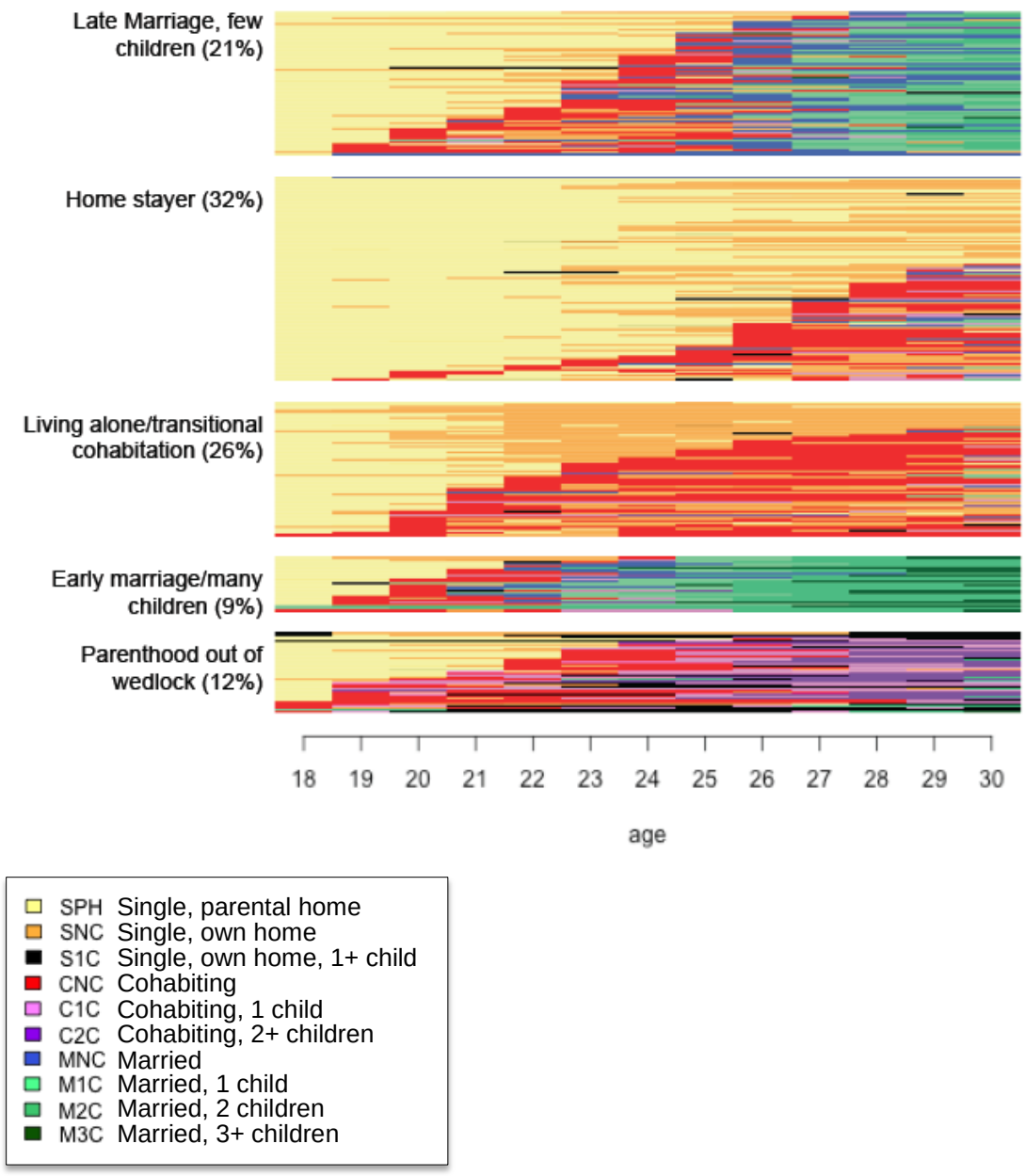


Figure 3: Sequence Index Plots of the five family formation clusters (random selection of five percent of the full sample) with clusters sorted ascending according to highest educational level (view in color)

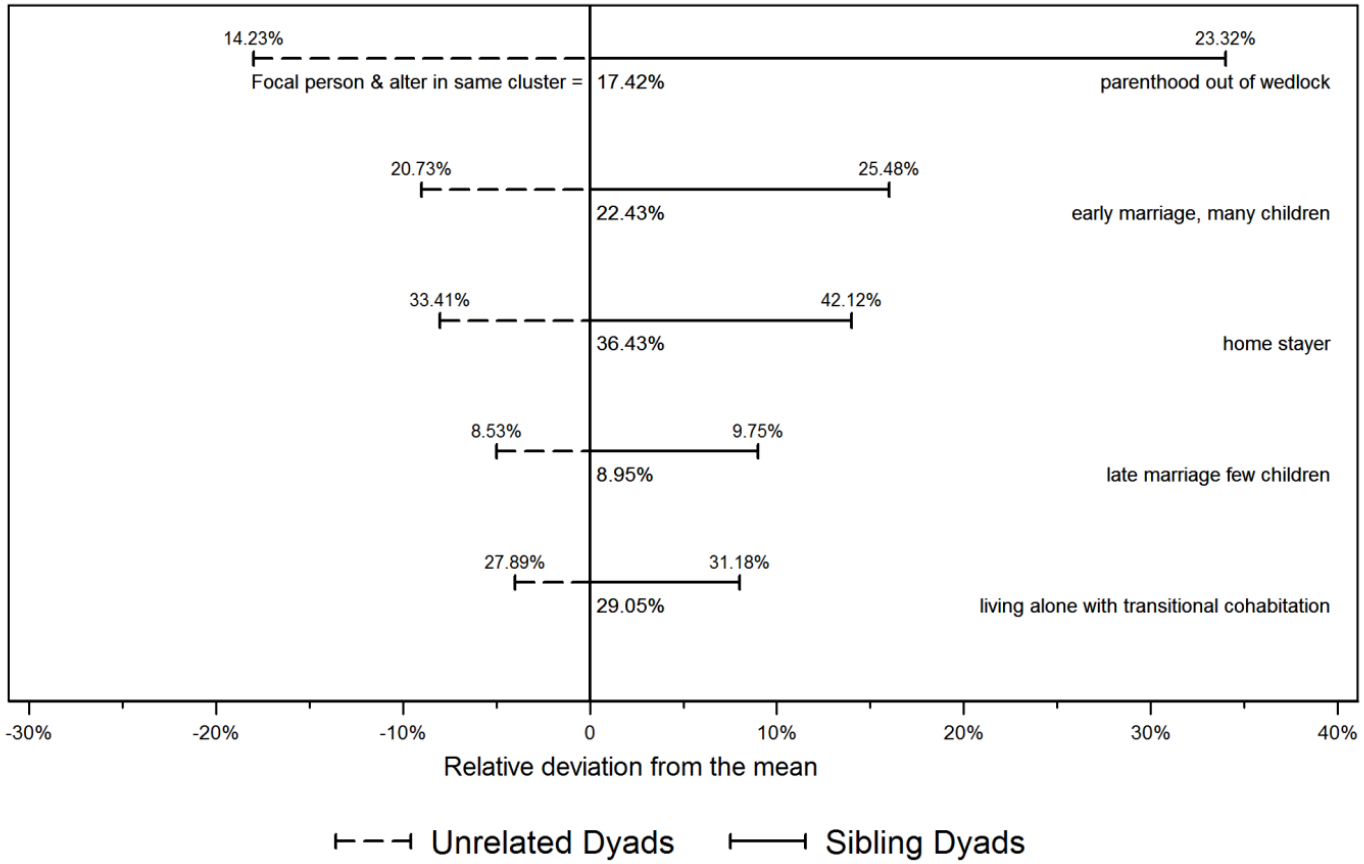


Figure 4: Focal person & alter in the same cluster: Conditional probabilities and relative deviations from overall mean

Appendix

Table A1: Substitution cost matrix

	SPH	SNC	CNC	MNC	SWC	C1C	M1C	C2C	M2C	M3C
SPH	0									
SNC	2	0								
CNC	4	2	0							
MNC	6	4	2	0						
SWC	8	6	5	5	0					
C1C	7	5	3	4	4	0				
M1C	8	6	4	2	5	2	0			
C2C	9	7	5	6	6	2	4	0		
M2C	10	8	6	4	7	4	2	2	0	
M3C	11	9	7	5	8	5	3	3	1	0

Table A2: Descriptive Results

	Sibling dyads	Unrelated dyads	Total
Parental background ***(fixed for sibling and unrelated dyads)***			
Parental Education			
Low			0.62
Medium			0.29
High			0.09
Mother's marriage age			
Early marriage			0.23
Marriage at average age			0.49
Late marriage			0.28
Experienced parental divorce			0.19
Children's characteristics			
Gender constellation			
Opposite sex	0.52	0.49	0.50
Both female	0.24	0.25	0.24
Both male	0.24	0.26	0.25
Age difference			
Born in same year	0.03	0.11	0.08
One to three years	0.62	0.54	0.57
More than three years	0.35	0.35	0.35
Educational level			
Low-high	0.06	0.10	0.08
Both Low	0.28	0.24	0.26
Low-medium	0.30	0.34	0.33
Both medium	0.18	0.16	0.17
Medium-high	0.12	0.12	0.12
Both high	0.06	0.04	0.04
Dyadic Distance	41.94	45.62	44.33
(SD in brackets)	(18.03)	(17.51)	(17.78)
Number of dyads	4996	9263	14259