The State Intervenes in the Battle of the Sexes:

Causal Effects of Paternity Leave

Andreas Kotsadam[a,b] and Henning Finseraas[b]

Abstract

Do family policies influence attitudes and behavior or are they merely reflections of pre-existing attitudes? We consider the implementation of the Norwegian daddy quota, four weeks of parental leave reserved for the father, as a natural experiment, and examine the long-run causal effects on attitudes toward gender equality, on conflicts and sharing of household labor, and on support for public childcare. We find that respondents who had their last born child just after the reform report an 11 percent lower level of conflicts over household division of labor and that they are 50 percent more likely to equally divide the task of washing clothes than respondents who had their last child just before the reform.

[a] Department of Economics, University of Gothenburg, Sweden, Box 640, SE-405 30, Gothenburg, Sweden. Email: Andreas.Kotsadam@economics.gu.se

[b] Norwegian Social Research, P.O. Box 3223 Elisenberg, N-0208 Oslo, Norway. Email: Henning.Finseraas@nova.no

Acknowledgements: The paper has benefited from comments by seminar participants at the Institute for Social Research (ISF), the Department of Sociology and Human Geography at the University of Oslo (ISS), and at Norwegian Social Reasearch (NOVA). We would also like to thank Emilia Simeonova, Axel West Pedersen, Kristin Aarland, Niklas Jakobsson, Mette Lovgren, Kåre Vernby, and Viggo Nordvik, for useful comments.

1. Introduction

Can family policies change gender relations in a society? We examine the effect of parental leave policies on the equalization of gender relations. Scholars interested in the impact of parental leave policies on gender relations typically rely on cross-national data to get policy variation. Such studies have found cross-national differences in to what degree men participate in housework and caregiving (e.g., Fuwa and Cohen 2007; Hook 2006; 2010, Ruppanner 2010), and although it has been hypothesized that relationships between family policies and equality of gender relations might be causal (Sullivan et al. 2009), it is difficult to identify causal effects using cross-national data since neither individuals nor countries are randomly assigned to different policy regimes. As Hook (2010: 1485) readily admits, “the available data allow me to examine only whether different policy and normative contexts are associated with gender inequalities in household labor.” The complexity increases even more if one wants to assess the effects of a specific policy, since different structural forces are interwoven, making them hard to separate analytically and empirically (Knudsen and Wærness 2008).

Rather than relying on comparative data to identify variation in regulations, we study policy effects at the national level. This approach has several advantages, most notably that one can examine the impact of specific policies where program experience varies at the individual level within the same country-specific institutional setting. The obvious advantage of within-country program variation with other institutional settings held constant is that one need not worry that the program effect is spurious to a correlated but omitted institutional variable. This worry looms large in the comparative research because the low number of countries and the high correlation between potentially important institutions precludes a rigid empirical test.

We examine the impact of a Norwegian parental leave reform in 1993 with strong incentives for fathers to take parental leave on attitudes proposed to be directly affected by the reform (gender equality), on within-couple conflicts over housework, and on policy preferences assumed to be more indirectly affected by the reform (support for public childcare). We exploit the exogenous variation in the incentives to take parental leave induced by the reform to estimate the causal effects. The data allows us to conduct a quasi-experiment where we compare present attitudes and behavior of parents with children born before the implementation of the reform (the control group) with attitudes of parents with children born after the implementation of the reform (the treatment group). Our design allows us to hold the political, economic, social, and cultural contexts constant. In particular, since parents are compared at the same point in time, our groups are faced with the same labor market, childcare, and fiscal policies, which, in cross national studies, have been omitted or suffer from immense endogeneity problems.

The central finding of the paper is the identification of an effect of the daddy quota on conflicts over household labor and sharing of household work. Respondents with children born after the reform report an 11 percent lower level of conflicts over household division of labor and are 50 percent more likely to equally divide the task of washing clothes than respondents with children born before the reform. These findings are robust to a series of sensitivity tests, strengthening the argument of a causal effect. The results also indicate an effect of the reform on preferences regarding spending on childcare, although this effect is less precisely estimated. We find no effect of the reform on individual level attitudes toward gender equality.

2. The daddy quota and gender relations

Paid parental leave has a long history in Norway. A 12-week paid maternity leave was introduced as far back as in 1956, and was expanded to 18 weeks in 1977. During the 1980s and 1990s, the number of weeks was increased in several rounds, and in 2005 the leave period was 52 weeks. Since 1977, fathers could take parental leave, but at the expense of the mother's leave period, and very few fathers took parental leave. On April 1, 1993, four weeks of parental leave were added and reserved for the father to use during the infant's first year; i.e., the four weeks could not be transferred to the mother and were lost if the father did not use them. The daddy quota was further expanded to six weeks in 2006, and then to ten weeks in 2009. Yet, in this paper we focus on the introduction of four weeks in 1993. The "use it or lose it" feature of the policy obviously implies a strong incentive for families to allocate time during the first year for fathers to take parental leave. After the introduction of the daddy quota, the take-up rate for fathers rose sharply, from 4 percent before the reform to 70 percent in 1995 (Brunning and Plantenga 1999) and in 2003, 89 percent of all fathers took some parental leave (O'Brien, Brandth, and Kvande 2007). There is no obligation that the mother has to return to work when the father is home on parental leave, and although this is most often the case, a survey conducted in 1995 shows that 35% of the mothers were home (for instance by taking vacation from work) during the father's leave period (Brandt and Øverli 1998).

We analyze whether the introduction of the daddy quota influences gender relations. First, we analyze how the daddy quota affects a set of specific questions regarding how female labor force activity affects children, attitudes toward a woman having a higher wage than her partner, and whether mothers are better suited than fathers to have custody of children in the event of divorce.[1] The introduction of a daddy quota is a strong signal from the political decision makers that a father should be more actively involved in child rearing during his child's formative years and hence challenges norms of male breadwinning (Gornick and Meyers 2008; Hook 2006; 2010). Additionally, it aims to improve the bonding between father and child, which could have an impact on attitudes toward activities traditionally performed by women, and men who are exposed to non-traditional experiences are more likely to change their views on gender equality

[1]Complete question wordings are reported in the appendix.

(Klein 1987: 35). Consistent with this view, Sjöberg (2004) finds that respondents score higher on a gender equality attitudes index in countries where family policies support a dual-earner family model (as the daddy quota does).

Second, we analyze whether the introduction of the daddy quota influences the reported levels of conflict over household division of labor as well as reported sharing of household work. The reported levels of conflict are interesting on its own but even more so in combination with reported sharing of household work. This addition is important since we would like to know whether the daddy quota influences behavior and not only attitudes. Even though the parental leave only covers a short period, and the father's quota an even shorter period, it is argued to have a strong impact on the division of labor within households in the long run by setting patterns since it intervenes at a critical time for renegotiating household work (e.g.Björnberg 2002; Hook 2010). Paternity leave may also raise men's skills as primary caregivers (Hook 2010). Qualitative studies of men taking parental leave find that men not surprisingly do more housework during the leave, and speculate that this might have a lasting effect due to the experiences gained during paternity leave (Lammi-Taskula 2006). Some studies have identified a cross-national variation in relative time (between men and women) spent on household work (e.g., Fuwa 2004; Geist 2005), and Hook (2006; 2010) finds that men with children who have access to paternity leave report spending more time on unpaid work. Ekberg, Eriksson, and Friebel (2005) is the only previous study on the effects of parental leave on household division of labor that use a design allowing them to make causal claims. They compare parents with children born just before and just after the introduction of the Swedish daddy quota, and find strong effects on fathers' leave taking, but no effects on subsequent leave taken for sick children. However, leave taking for sick children also involves relationship to employers and as such there may be other mechanisms (e.g. signaling) offsetting any positive effects of the daddy quota.

Third, we analyze whether the introduction of the quota affects political preferences regarding public responsibility for childcare. This is to be expected if, as argued in the welfare regime literature, there is a feedback-effect on mass political preferences; i.e., if introduction of policies that promote labor force participation of mothers (the daddy quota) increases support for other policies that promote mothers' labor force activity (public childcare). Claims of feedback effects of public policy on the mass public have been particularly popular in the (comparative) welfare state literature. Inspired by Esping-Andersen's (1990) famous claim that the institutional set-up of the welfare state influences not only voters' egotropic incentives to support the current institutions but also their support for the underlying normative principles of the current institutions, numerous journal articles have examined whether cross-national variation in support for the welfare state follows Esping-Andersen's (1990) regime-typologies (see e.g. Jæger 2006). At best, these studies found mixed support for the regime feedback hypothesis. Again, the research design of these studies is questionable since they study mass politics at one point in time – after implementation of the institutional set-up – and often use abstract indicators of public policy such as regime type dummy variables.

The present study is an important contribution to the existing literature on parental leave and gender relations since we move beyond examining mere correlations and attempt to establish whether there is a causal relationship between access to parental leave and more egalitarian gender relations. With the exception of Ekberg et al.'s (2005) working paper, none of the work reviewed above can claim to have identified a causal relationship (and none argue they have). The main reason they are unable to identify causality relates to the nature of the data they examine. The analytical approach typically used in the existing literature is to exploit comparative data with cross-country variation in parental leave arrangements. Issues related to the comparability of parental leave schemes across countries aside, the low number of macro units combined with a vast amount of competing explanations makes it difficult to conduct a rigid empirical analysis of the impact of parental leave on gender relations. Moreover, the unknown data generating process of macro data (i.e., whether a country has implemented a generous parental leave arrangement is most likely not random, even when controls are included) makes it difficult to argue that an observed correlation is causal (see Przeworski 2007 for an excellent discussion of issues regarding causal inference from comparative data). We agree with others (e.g., Campbell 2006) who argue that the national level has a crucial advantage in analyzing causality because experiences with different policies vary at the individual level. With the appropriate data, this allows for a comparison of respondents who have experienced a policy and those who have not. Some studies have examined the effect of the daddy quota using this analytical strategy (e.g. Haas and Hwang 2008, Duvander et al. 2010), but these studies do not establish causality since selection into taking parental leave is not accounted for. Our empirical strategy, explained in detail in the next section, take selection into account, and can therefore provide a more reasonable estimate of the causal effect of the daddy quota.

3. Data and method

We treat the introduction of the Norwegian daddy quota in 1993 as a policy experiment to compare present attitudes and division of household labor among those who had children before and after the introduction of the quota, respectively. The policy change creates a natural experiment that enables us to investigate the long-term individual level effect of the daddy quota. By examining attitudes for similar individuals at the same point in time, we are able to compare parents facing the same contextual conditions, including macroeconomic conditions and gender discourses.

We rely on survey data from the 2007/2008 Life cOurse, Generation and Gender (LOGG) study. LOGG is based on a nationally representative sample of 25,937 persons aged 18-84, and includes a Norwegian version of the UN-anchored Generations and Gender Survey (GGS) (Brunborg et al. 2009). Data in LOGG is collected from three sources: computer assisted telephone interviews, postal surveys, and register data. The response rate was 59.6% (n=15 140) for the telephone interviews and 72.4% for the subsequent postal surveys.

We rely on the following dependent variables. *Conflicts* over household division of labor, are measured by a survey item addressing how often the respondent and the partner disagree about the division of household labor (min=0 – never, max=10 – very often). The LOGG data also contain questions on who does most household work in the household with the responses ranging from 1 (always me) via 3 (we share it equally) to 5 (always my partner). We create an indicator variable for equal sharing that equals 1 if the respondent answers 3 (me and my partner share it equally), 0 if the respondent reports that the woman does most household work, and 2 if the man does most household work. Indicator variables are created for three specific household tasks: washing clothes, preparing food, and cleaning. These tasks are typical in that they are time-inflexible and men's participation in them is seen as especially important for gender equality (Hook 2010).

Next, following Sjöberg (2004), we create an additive index of three strongly correlated items regarding attitudes toward *gender equality* (min=1, max=5; a high score implies a positive attitude toward gender equality).[2] The three items are whether the respondent agrees (1) that children below school age will suffer if the mother is working, (2) that children of divorced parents should live with their mother, and (3) that it is better if the man has a higher income than the woman. The respondents specify their level of agreement to the statements. Finally, we create

[2] The substantial conclusion is the same when using each item separately.

a variable of mother-work friendly policies by asking whether the respondent would like to see an increase in the level of spending on public *childcare* (0=do not support more spending, 1=do support more spending). The question wordings are reported in the appendix.

The introduction of the daddy quota affected all parents who had children after April 1, 1993 and no parents who had their lastborn child before this date. We use the implementation date to define a control group and a treatment group: The control group is those living with a partner who had their last child between April 1 1991 and March 30 1993, while the treatment group is those living with a partner who had their last child between April 1 1993 and March 30 1995. Note that we do not condition on actual take up of parental leave. In fact, we do not even have this information in our dataset. Remember that the main point of our empirical strategy is to exploit the exogenous variation induced by the reform so that we do not need to condition on actual take up (which is a main limitation in previous research on parental leave). This allows us to estimate the intention to treat effect of the reform. Had we had access to take up at the individual level, we could have estimated the treatment effect of actually using the daddy quota with the help of an instrumental variables approach. For policymakers, the intention to treat effect is probably more relevant as it answers the question with respect to the total effects of the reform. Furthermore, in terms of feedback effects, it is perfectly possible that being offered a daddy quota changes expectations and gender relations over and above the effects of actually using the leave.

The selected time window is a result of the trade-off between having as similar groups as possible (reduced time window) and as many individuals as possible (extended time window). The sensitivity analysis presents results with a time window of one year before and one year after the reform as well. The results from these regressions are less affected by the possible age-difference bias and they are also more robust to possible omitted variable biases (discussed below). However, the reduced sample size is a disadvantage.

We limit the treatment group to those having their lastborn child in order to make the comparison group more similar. This grouping may be problematic if the reform affected the total fertility rate. This is so since our sample then consists of a special type of individuals not affected in their fertility decision by the reform. The relationship between family policy, gender equality, and fertility is complex. Indeed western countries with more gender egalitarian policies have higher fertility (Aassve and Lappegård 2009; Ferrarini 2006; Gornick and Meyers 2008), although this relationship may of course be due to higher fertility rates pressuring politicians to create family-friendly policies – the literature has not established a causal effect of policies on fertility. There may also be some omitted factors causing the low fertility in more gender egalitarian countries, e.g., that the Nordic countries have low inequality and extensive safety nets, which probably reduce the risks associated with having children (e.g., Apps and Rees 2004). Similarly, extended coverage of affordable high-quality childcare is also importnat (Andersson et al. 2004). To the extent that family policies affect fertility, the high replacement rate of income in the Norwegian parental leave scheme, together with extensive public childcare, are likely to be important (and these factors are shared for our treatment and control group). Aassve and Lappegård (2009) also find that take up of cash benefits is positively correlated with fertility. Yet father involvement should not be excluded as a factor. Duvander et al. (2010) indeed find an increased probability of having another child if fathers took out parental leave for previous children. They do admit, however, that the relationship need not be causal since selection into parental leave is not accounted for.

In terms of estimation we investigate the fertility issue first by comparing *all* mothers who had a child around the reform and find no difference between mothers who had a child just before and mothers who had a child just after the reform in the number of children they had after the reform (results are available upon request). This is important since it implies that those affected by the reform are not different in their completed fertility patterns from those in the control group, a crucial feature for the internal validity of the estimation strategy. Then we run regressions where we allow the treatment group to consist of people who had not completed their fertility spell, i.e., we allow the treatment group to consist of those who had children later as well (see sensitivity analysis).

A related potential threat to the identification of causal effects is endogenous sorting, i.e., that parents could have planned the time of birth in anticipation of the policy. Births have a standard deviation of two weeks, i.e. within a two week window the timing of birth is completely random and cannot bias the treatment effect. Unfortunately, our data do not allow a proper test of endogenous sorting since we do not have that detailed information on date of birth, and since our sample size would be extremely small even if we had the information. Instead we examine the number of children born in Norway and find no spike at the month or year of the implementation of the daddy quota (Statistics Norway).

In order to know whether the parents in the two groups resemble each other on potentially confounding variables, we need to know on which characteristics they should be compared. Table 1 displays regression results for our three dependent variables when including factors that previous literature has proposed to be important for gender relations in Norway (see Jakobsson and Kotsadam 2010 for a review). Table 2 compares the mean values for the two groups of parents.

[Table 1 here]

[Table 2 here]

Although the sample is selected so as to minimize the effects of other confounding factors, Table 2 shows that we fall short with respect to age. This is important since Table 1 shows that age has a statistically significant correlation with all of our three dependent variables. In order to fully account for the effect that age may have, we present the mean values for people aged 46 and 48 years old (i.e. groups being of the mean age as those in our treatment and control groups),and we control for age in OLS regressions. As age is also predetermined, and therefore not affected by the reform, controlling for age does not introduce endogeneity problems. The fact that the treatment and control groups do not differ on any observable characteristics other than age further indicates that endogenous sorting is unlikely to be a problem.

A possible limitation of the sampling strategy is that the two groups differ by construction in the age of their youngest child, which may also be correlated with our variables of interest. In particular, it can be argued that it is not the reform that has changed the attitudes but rather that the results are driven by the age of the children. This cannot be directly tested since we need to use the variation in children's age to construct our groups. In order to explore the possible importance of children's age on gender relations, we analyze regressions including the year of the last born child as an explanatory variable in the full population. We also run placebo regressions where the initial regressions are repeated on the treatment and control groups unrelated to any actual policy change and where the two groups have equal regulations regarding the daddy quota. In particular, we pretend that the daddy quota was introduced five years earlier (1988) or five years later (1998) and hope to find no statistically significant "treatment effects".

4. Empirical results

The first two columns of Table 3 present mean values for our variables of interest for the group affected by the policy and for the control group. We see that parents with children born after the implementation of the reform report to have 11 percent lower level of conflicts over household work. The difference is statistically significant in a two-sided t-test at the 5 % level. Attitudes toward gender equality do not seem to differ between the two groups, but support for increased public childcare spending is 6 percentage points (18 %) higher in the group of parents affected by the reform. The last two columns present the mean values for people being of the mean age in the control and treatment group, respectively. We see that there are no statistically significant differences between the age groups, which indicate that the differences between treatment and control are most likely not driven by age differences.

[Table 3 here]

Since we know that the groups differ in age, we explicitly control for this in our main specification. The results of a regression of the dependent variables on a treatment indicator variable and age are shown in Table 4. We find two statistically significant effects of the reform: Those with children born after the implementation of the reform report a lower level of conflicts over household division of labor and are more likely to support public provision of childcare. The point estimates are not very precisely estimated, however, as the 95 percent confidence interval for the treatment coefficient on *conflicts* ranges from -.04 to -.72, while the treatment coefficient on *childcare* is significant only at the ten percent level. The results (available upon request) remain the same when also allowing for non-linear relationships between age and the outcomes of interest, i.e., the coefficient for the treatment indicator does not change when adding a squared and/or a cubic age term as additional control variables. [3]

[Table 4 here]

[3] Moreover, the treatment coefficient on *conflicts* does not change much if we control for the full range of confounding variables presented in Tables 1 and 2, but the treatment coefficient on *childcare* drops somewhat and becomes insignificant at conventional levels (results available upon request). Nonetheless, we have less faith in these coefficients as representing causal effects of the reform since the majority of these variables are most likely endogenous to the reform.

The effects identified in Table 4 are substantial: An individual of average age who was exposed to the reform has a 12.6 percent lower level of reported conflicts than an average non-exposed individual.[4] Since reform exposure is uncorrelated with all factors except age, we can compare the magnitude of the treatment indicator variable to the magnitudes of other dummy variables in the population regression (Table 1). This approach reveals that the effect of belonging to the treatment group is larger than any of the other correlations. Belonging to the treatment group also has a substantial effect for attitudes toward childcare spending: it increases the probability of supporting an increase in public childcare by 7 percentage points. Further analysis reveals that the effects of reform exposure are not significantly different between men and women (results available upon request).

4.1. Sensitivity analysis

In order to further test for the robustness of the estimated effects, we reduce the window to one year before and one year after the reform and conduct the same analysis. The results, shown in Panel A in Table 5, still point in the same direction and the magnitude for *Conflicts* is similar. The precision of the estimated effect is reduced and not significant. However, since the main reason for the lack of statistical significance in the reduced sample is the expected increase in the standard error of the estimate (due to a much smaller sample size), and not a change in the size of the coefficient, we gain confidence in the fact that the treatment coefficient is stable. The coefficient for *Childcare* is statistically significant at the ten percent level also in the reduced sample.

[4] 0.384 / (4.234 + 46*-0.0259)

[Table 5 here]

In order to explore the possible importance of children's age, which differ across our two groups, on gender relations, we present regressions including the year of the last born child as an explanatory variable in the full population sample in Table 6. We see that the year of birth of the youngest child is statistically significantly correlated with both *Conflicts* and *Childcare* when only controlling for the age of the respondents. When controlling for other potentially confounding variables, the year of birth of the lastborn child effect disappears for *Childcare*. Furthermore, the coefficient for the year of the last borne child is small and actually points in the opposite direction to what we would worry about.

We now also run placebo regressions, repeating the initial estimations but for groups having equal regulations regarding the daddy quota. In the first regression, we pretend that the daddy quota was introduced five years earlier (1988). In the next regression, we pretend that the quota was introduced five years later (1998). Panel B of Table 5 reports the results of our placebo analyses. Remember, if there is a causal effect of the reform, we should *not* find any significant treatment effects in these tables since the two groups experienced the same paternity leave regulation. Thus, the insignificant results for our variables of interest strengthen our confidence in the validity of the estimation strategy. In particular, it further reduces the worries of a spurious correlation due to age of the last child or other differences between cohorts, and shows that the results are not merely driven by time trends.

Another potential problem is if the reform affected total fertility by inducing couples to have more children. This could be so as a direct consequence of the increase in the total amount of benefits increased and/or since women perceive raising children less costly when they can share the burden to a larger extent. In our case, it is problematic if the reform affected the total fertility rate since we, in order to make the groups as similar as possible, exclude everyone who had more children after the reform. This grouping could imply that our treatment group is biased in selecting only those not affected in their fertility decisions. In order to test for this possibility, we run regressions where we allow the treatment group to consist of people who had not completed their fertility spell, i.e., we allow the treatment group to consist of those who had children later as well. Note that we cannot do the same for the control group since they would then be treated by the reform. The results are presented in Panel C in Table 5. We see that the treatment effect is robust to this change both for *Conflicts* and for *Childcare*. The extension of the treatment group also implies that our groups are less similar than in the baseline case. In particular, we note that the parents exposed to the reform now includes couples with more children ($p<0.01$). To account for this, Columns 4 to 6 also control for number of children in addition to age, and the results are robust to this control.

[Table 6 here]

The final threat to the identification of causal effects that we can come up with, is that parents of those in the treatment and the control group might differ. For instance, there has been a steady increase in the proportion of the population with higher education. This is particularly true for women, and this increase subsequently contributed to a rapid increase in female labor force participation. It is conceivable that parents with higher education promote gender equality to their children to a larger extent than parents with lower education, thus if the two groups differ in the level of parents' education, this might explain any differences in gender relations between the two groups. We are able to examine this issue as respondents are asked about the education level of their parents. We find a positive (but insignificant) correlation between treatment status and the education level of the respondent's mother. When we include mother's education level as an additional control variable, which is perfectly fine since it is strictly exogenous to our dependent variables, the treatment coefficient on conflicts actually increases (coeff = -.42, SE=.17, t=2.46).

5. Division of household labor

We have seen that the daddy quota considerably reduces the level of conflicts over household labor within couples and we argue that the most likely reason for this is a more equal sharing of housework in the group of parents exposed to the reform. In order to strengthen the interpretation of the result, we analyzed other variables in the data and find assurance in the fact that reform exposure is not correlated with having more conflicts in general or with having more conflicts over, e.g., money or sex (results are available upon request). The best way to see the effects of policy on actual sharing of household work would be to use time-use data combined with specific national policies (Sullivan et al., 2009). Under ideal settings, we would exploit a time-use dataset with a large enough sample of individuals who had their last child in a time period around the reform to investigate the actual sharing, but no such dataset exists. In this section, we complement the analysis by investigating the effects of the reform on equal sharing of household work. In particular, we examine the probability of equally sharing three specific time-inflexible tasks: washing clothes, preparing food, and cleaning. First we examine mean differences between treatment and control group in share of respondents that share tasks equally, next we estimate multinomial logit regressions where we distinguish between respondents reporting that the women does most of the household work, respondents reporting that they share the tasks equally, and respondents reporting that the man does most household work.

[Table 7 here]

Panel A of Table 7 shows the mean values for the three household tasks for the treated and untreated individuals, respectively. Cleaning is the most equally shared task and washing clothes is the least shared task, with preparing food placing in between. The difference between the treated and control individuals with respect to equal sharing is only statistically significant for the task of washing clothes. For this task, the magnitude of the difference is very large: the probability that a couple divides the task equally is 50 percent higher in the treated group than in the control group. This mean difference is robust to the same sensitivity checks as above.

Panel B of Table 7 presents the results from the multinomial regressions, and Panel C the sensitivity checks. The reference category is that the woman does most of the household work, i.e. we expect positive treatment coefficients. As evident, treated respondents report a more equal sharing of the task of washing clothes. The predicted probability that a treated respondent reports an equal sharing of this task is seven percentage points higher compared to an untreated respondent. Note further that there is a positive treatment effect also on the "man does most work" versus "woman does most work", which is also significant at the ten percent level for the cleaning variable.

Why do we find strong effects only on the task of washing clothes? Measurement issues aside, we note that we find a pronounced effect for the least equally shared task. Our best guess is that conflicts over household labor today are reduced in the treatment group since the increased participation of fathers in the early years of their children's life has changed the gendered dynamics within the household in the direction of more equal sharing of household work and that we are only able to capture this in our dataset for the area where the inequality was the largest.

6. Concluding discussion

The question of whether policies influence attitudes or change gender relations has received a lot of scholarly attention. The research designs applied in previous work have mainly been carried out without sufficient attention toward finding suitable control groups, and the questions of causality have therefore not been resolved. This paper investigates the introduction of the daddy quota in Norway, taking the issue of causality as a key objective. Treating the reform as a natural experiment – a treatment group of people who had their last child just after the reform is contrasted to individuals who had their last child just before the reform – enables us to compare the long-run effects of the reform on the treated individuals' attitudes and division of household labor.

Did the daddy quota durably affect the division of labor within the treated households? We find robust support for the notion that the reform made treated individuals report much fewer conflicts over household work compared to non-treated individuals. We find that this difference is not merely driven by a time trend or by the age differences between the groups. We also investigate whether treated individuals were more likely to report equal sharing of certain household tasks and find that for the most unequally shared task (washing clothes), there are large and robust effects. People who had their lastborn child after the reform were 50 percent more likely to report equal sharing of washing clothes. In sum, we are confident that we have identified a durable effect of the daddy quota on the division of household labor.

Did the daddy quota durably affect individual-level attitudes toward gender equality? We find no support for the hypothesis that the treated individuals have different attitudes toward gender equality than those in the control group. While it may very well be the case that the introduction of a daddy quota made people in general more gender egalitarian in their attitudes, we cannot conclude that those who were actually treated were more affected. Data from the World Values Survey shows a continued change toward gender equality over the period investigated: for instance, while 89 percent of the respondents disagreed or strongly disagreed with the statement that university education is more important for a boy than for a girl in 1996, the proportion was 97 percent in 2007. We suspect that any effect on the individuals treated by the reform may have leveled out over time due to a societal trend in attitudes toward gender equality. Nonetheless, the fact that no treatment effect on the treated is identified is an important finding.

Did the daddy quota produce durable feedback effects? There are indications of feedback effects of the daddy quota since treated individuals are more in favor of increased childcare spending than their non-treated counterparts. Our interpretation of this finding is that the daddy quota, which supports mothers' employment opportunities, had a feedback effect on other policies that increase mothers' employment. We do not believe that policies that promote female employment in general are affected by the daddy quota since we find no treatment effect on the support for increased eldercare spending (results are available upon request), which is another area with well documented effects on female employment (e.g., Kotsadam 2011). Another interpretation is that personal childcare experience increases the understanding of the importance of publicly provided alternatives, but since we do not find any difference in treatment effect between men (who should be the ones gaining more experience) and women (who now share the responsibility of childcare to a larger extent), this is an unsatisfactory interpretation.

The research design of the study has strong internal validity, i.e., strong validity of inferences about whether any observed covariation between reform exposure and outcome reflects a causal relationship from the former to the latter. One drawback, however, is the insufficient focus on understanding causal mechanisms, i.e., why did the daddy quota produce the observed effects? As we show, the division of household labor has been affected, possibly by entrenched patterns imposed when the last child was born, but we have not investigated this particular explanation and hence do not know for sure whether this is the actual mechanism. Another drawback of the study is its limited external validity, i.e., limited validity of inferences about the causal relationship in other settings, with other individuals and other policies (Shadish et al. 2002). We studied a specific family policy in a specific country at a specific time. It is therefore difficult to generalize the findings to effects of family policies in general. Thereby, even if we had isolated clear mechanisms it is not obvious that they would carry over to other contexts since this would require implementation-specific counterfactuals (Cartwright 2007). Nonetheless, we can conclude that even relatively small reforms are able to change gender relations, even at the personal level over a period as long as 15 years. To gain further knowledge on the effects of family policies, we urge future studies to look for similar natural experiments in other settings, with other types of family policies, and during other time periods. The resulting cumulative knowledge growth could then lead to a type of generalized causal inference (Shadish et al. 2002), which in turn could increase our understanding of the effects of family policies and the underlying causal mechanisms.

Appendix

Question wordings for the dependent variables.

Conflicts: ”Couples might disagree over different aspects in life. On a scale from 0 to 10 where 0 means ‘never’ and 10 means ‘very often,’ how often over the last year have you and your [cohabitant/spouse/partner] disagreed over housework?”

Gender equality: Additive index from 1 to 5 where a high score implies more favorable attitudes toward gender equality. The index is constructed of the following survey items: ”Children below school-age will suffer if the mother is working (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree),” ”If parents divorce, it is best for the child to live with her/his mother (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree),” “It is better if the man has a higher income than the woman (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree)”.

Childcare: ”Regarding public spending, do you think public spending in the following areas should increase, decrease, or stay the same? Public childcare”.

References

Aassve, A. and Lappegård, T. 2009. Childcare Cash Benefits and Fertility Timing in Norway. European Journal of Population 25, 67-88.

Andersson, G., Duvander, A-Z., and Hank, K. 2005. Do child-care characteristics influence continued childbearing in Sweden? An investigation of the quantity, quality, and price dimension. Journal of European Social Policy 14, 407-418.

Apps, P. and R. Rees. 2004. Fertility, Taxation and Family Policy. Scandinavian Journal of Economics. 106:4, 745-763.

Bernhardt,E., Noack T., and Hovde Lyngstad, T. 2008. Shared Housework in Norway and Sweden: Advancing the Gender Revolution. Journal of European Social Policy18 (3): 275-288.

Björnberg, U. 2002. Ideology and choice between work and care: Swedish family policy for working parents. Critical Social Policy 22, 33-52.

Brandth, B., Kvande, E. 2009. Gendered or Gender-Neutral Care Politics for Fathers? The ANNALS of the American Academy of Political and Social Science, 624, 177-189.

Brandth, B. and B. Øverli. 1998. Omsorgspermison med "kjærlig tvang". En kartlegging av fedrekvoten [Parental leave with "caring force". A survey of the daddy quota]. ALLFORSK Working Paper, NTNU.

Brunborg, H., B. Slagsvold and T. Lappegård. 2009. LOGG 2007 – en stor undersøkelse om livsløp, generasjon og kjønn [LOGG 2007 – A large study of the life-course, generation, and gender]. Samfunnsspeilet23 (1), 2-8.

Brunning, G. and J. Plantenga. 1999. Parental Leave and Equal Opportunities: Experiences in Eight European Countries. Journal of European Social Policy, 9, 195-209.

Campbell, A. L. 2006. Policy Feedbacks and the Political Mobilization of Mass Publics. manuscript, MiT.

Cartwright, N. 2007. Hunting Causes and Using Them: Approaches in Philosophy and Economics. Cambridge University Press, Cambridge, U.K. 2007.

Duvander, A.-Z., Lappegård, T., and Andersson, G. 2010. Family policy and fertility: Fathers' and mothers' use of parental leave and continued childbearing in Norway and Sweden. Journal of European Social Policy20(1): 45-57.

Ekberg, J., Eriksson, R., and Friebel, G. 2005. Parental Leave – A Policy Evaluation of the Swedish ”Daddy-Month” Reform, IZA Discussion Paper No. 1617 2005.

Esping-Andersen, G. 1990.The Three Worlds of Welfare Capitalism. Cambridge: Polity.

Ferrarini, T. 2006. Families, States and Labour Markets : Institutions, Causes and Consequences of Family Policy in Post-War Welfare States. Cheltenham: Edward Elgar.

Fuwa, M. 2004. Macro-Level Gender Inequality and the Division of Household Labor in 22 Countries. American Sociological Review 69 (6): 751-767.

Fuwa, M. and Cohen, P. N. 2007. Housework and Social Policy. Social Science Research 36 (2): 512-530.

Geist, C. 2005. The Welfare State and the Home: Regime Differences in the Domestic Division of Labour.European Sociological Review 21:23-41.

Gornick, J. and M. Meyers. 2008. Creating Gender Egalitarian Societies: An Agenda for Reform.Politics & Society, 36, 313-349.

Haas, L. and Hwang, C. P. 2008. The Impact of Taking Parental Leave on Fathers' Participation In Childcare And Relationships With Children: Lessons from Sweden. Community, Work & Family 11(1): 85-104.

Hook, J. L. 2006. Care in Context: Men’s Unpaid Work in 20 Countries, 1965-2003.American Sociological Review71 (4):639-660.

Hook, J. L. 2010. Gender Inequality in the Welfare State: Sex Segregation in Housework, 1965-2003.American Journal of Sociology115, 5, 1480-1523.

Jakobsson, N., and A. Kotsadam. 2010. Do Attitudes Toward Gender Equality Really Differ Between Norway and Sweden? Journal of European Social Policy 20(2), 142-159.

Jæger, M. M. 2006. Welfare Regimes and Attitudes Towards Redistribution: The Regime Hypothesis Revisited. European Sociological Review 22 (2): 157-170.

Klein, E. 1987. The Diffusion of Consciousness in the United States and Western Europe, in Mary Katzenstein and Carol Mueller (eds.) The Women’s Movements in the United States and Western Europe. Philadelphia: Temple University Press.

Knudsen, K. and Wærness, K. 2001. National Context, Individual Characteristics and Attitudes on Mothers’ Employment: A Comparative Analysis of Great Britain, Sweden and Norway. Acta Sociologica 44 (1): 67-79.

Kotsadam, A. 2011. Does Informal Eldercare Impede Women's Employment? The Case of European Welfare States. Forthcoming in Feminist Economics 2011. DOI: 10.1080/13545701.2010.543384

Lammi-Taskula, J. 2006. Nordic men on parental leave: can the welfare state change gender relations, in Anne Lise Elingsæter and Arnlaug Leira (eds.) Politicizing parenthood in Scandinavia: gender relations in welfare states. Bristol: The Policy Press.

Przeworski, A. 2007. Is the Science of Comparative Politics Possible?, in Carles Boix and Susan C. Stokes (eds.), Oxford Handbook of Comparative Politics. New York: Oxford University Press.

Ruppander, L. 2010. Cross-national reports of housework: An investigation of the gender empowerment measure. Social Science Research 39(6):963-975.

Shadish, W, T. Cook and D. Campell 2002. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin Company, Boston, 2002.

Sjöberg, O. 2004. The Role of Family Policy Institutions in Explaining Gender-Role Attitudes: A Comparative Multilevel Analysis of Thirteen Industrialized Countries.Journal of European Social Policy 14 (2): 107-123.

Stortingsmelding (1991-1992). Likestillingspolitikk for 1990-åra [Gender Equality Policies for the 1990s], No.70 (1991-1992).

Sullivan, O., S. Coltrane, L. Mcannally, and E. Altinas 2009. Father-Friendly Policies and Time-Use Data in a Cross-National Context: Potential and Prospects for Future Research. The ANNALS of the American Academy of Political and Social Science 624: 234-254.

Table 1: Factors proposed to be important for gender relations.
Column 1 and 2: OLS. Column 3: Marginal effects after probit regression.

	(1)	(2)	(3)
	Conflicts	Gender Eq.	Childcare
Woman	-0.0166	0.229***	0.106***
	(0.0626)	(0.0154)	(0.0111)
Age	-0.0403**	0.0104***	-0.0105***
	(0.0158)	(0.00323)	(0.00235)
Age-sq	0.0139	-0.0226***	0.00719***
	(0.0152)	(0.00330)	(0.00241)
High education	0.166*	0.304***	-0.0714***
	(0.0880)	(0.0220)	(0.0155)
Medium education	-0.0269	0.132***	-0.0639***
	(0.0824)	(0.0204)	(0.0145)
Married	-0.139	-0.0129	-0.0386**
	(0.0961)	(0.0261)	(0.0186)
Widow	-0.929	0.0774*	0.0163
	(0.604)	(0.0432)	(0.0313)
Divorced	-0.0747	-0.00741	0.0282
	(0.227)	(0.0314)	(0.0225)
Children	0.386***	-0.0429**	0.107***
	(0.0789)	(0.0201)	(0.0145)
Income (ln)	0.0154	0.00805***	-0.00160
	(0.0101)	(0.00242)	(0.00179)
Partner’s income	0.0710***	0.00670***	0.00379**
	(0.0168)	(0.00208)	(0.00153)
Part time work	-0.218***	-0.0814***	-0.0581***
	(0.0780)	(0.0197)	(0.0142)
Labor force part.	0.208*	0.0648**	0.0419**
	(0.119)	(0.0276)	(0.0202)
Public sector	0.124*	0.0472***	-0.00393
	(0.0686)	(0.0177)	(0.0129)
Religious	-0.0972	-0.187***	-0.0585***
	(0.0618)	(0.0166)	(0.0119)
Constant	2.857***	3.363***	
	(0.399)	(0.0666)	
Observations	6735	9672	9741
R-squared	0.081	0.180	

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 2: Mean values of relevant control variables in the treatment group and the control group.

	Control n=490	Treated n=495
Age	48.11	46.16***
Number children	2.42	2.36
Woman	0.55	0.54
High education	0.34	0.37
Low education	0.19	0.20
Ln(Income)	11.76	11.89
Married	0.68	0.68
Widow	0.01	0.01
Divorced	0.19	0.18
Partner income	9.94	9.92
Part time	0.19	0.17
Labor force partic.	0.91	0.91
Public sector	0.38	0.39
Religious	0.23	0.25

*** $p<0.01$, ** $p<0.05$, * $p<0.1$ (p-value in a two-sided t-test of the difference between groups).

Table 3: Mean values of the dependent variables in the treatment group and the control group

Dependent variables	Control n=490	Treated n=495	People aged 48	People aged 46
Conflicts	2.98	2.65**	2.51	2.68
Gender Equality	3.81	3.88	3.82	3.81
Childcare	0.34	0.40*	0.38	0.41

*** p<0.01, ** p<0.05, * p<0.1 (p-value in a two-sided t-test of the difference between groups).

Table 4: Main results. Two year window. Columns 1 and 2 present results from OLS regressions and Column 3 presents marginal effects after probit regressions.

	(1) Conflicts	(2) Gender Eq.	(3) Childcare
Treatment	-0.384**	0.0396	0.0619*
	(0.174)	(0.0547)	(0.0363)
Age	-0.0259	-0.0145***	-0.000213
	(0.0168)	(0.00470)	(0.00326)
Constant	4.234***	4.515***	
	(0.830)	(0.232)	
Observations	763	728	733
R-squared	0.008	0.014	

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 5: Sensitivity analysis, validity tests, and robustness checks

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Panel A: Reduced sample. One year window.

	(1) Conflicts	(2) Gender Eq.	(3) Childcare
Treatment	-0.355	0.0281	0.0842*
	(0.236)	(0.0794)	(0.0511)
Age	-0.0233	-0.0142*	0.000717
	(0.0235)	(0.00723)	(0.00473)
Constant	4.100***	4.509***	
	(1.146)	(0.347)	
Observations	361	350	353
R-squared	0.009	0.012	

Panel B: Placebo regressions. Two year window.

		Pre-treatment placebo			Post-treatment placebo	
	(1) Conflicts	(2) Gender Eq.	(3) Childcare	(4) Conflicts	(5) Gender Eq.	(6) Childcare
Placebo treatment	0.212	0.0216	-0.0172	-0.110	-0.0275	0.0289
	(0.181)	(0.0636)	(0.0412)	(0.161)	(0.0528)	(0.0369)
Age	-0.00642	-0.0147**	-0.00675*	-0.0267*	-0.00857*	-0.00962***
	(0.0175)	(0.00623)	(0.00391)	(0.0146)	(0.00495)	(0.00353)
Constant	2.551***	4.528***		4.074***	4.277***	
	(0.938)	(0.337)		(0.640)	(0.218)	
Observations	609	579	587	865	739	751
R-squared	0.003	0.014		0.004	0.004	

Panel C. Extended samples. Accounting for possible effects on total fertility:

VARIABLES	(1) Conflicts	(2) Gender Eq.	(3) Childcare	(4) Conflicts	(5) Gender Eq.	(6) Childcare
Treatment	-0.372**	0.0112	0.0604*	-0.365**	0.0296	0.0653*
	(0.164)	(0.0513)	(0.0342)	(0.165)	(0.0520)	(0.0345)
Age	-0.0248*	-0.0108**	-0.00331	-0.0238	-0.00822*	-0.00267
	(0.0144)	(0.00421)	(0.00297)	(0.0148)	(0.00431)	(0.00304)
Number of children				-0.0360	-0.0729***	-0.0198
				(0.0880)	(0.0246)	(0.0180)
Constant	4.181***	4.336***		4.219***	4.387***	
	(0.713)	(0.208)		(0.712)	(0.208)	
Observations	965	883	891	965	883	891
R-squared	0.007	0.008		0.007	0.017	

Table 6: The partial correlation between the year of the lastborn child and the dependent variables. OLS regressions, except Columns 2 and 5 where marginal effects after probit regressions are presented.

VARIABLES	(1) Conflicts	(2) Gender Eq.	(3) Childcare	(4) Conflicts	(5) Gender Eq.	(6) Childcare
Year of last born child	0.0270***	0.000831	-0.00210**	0.0246***	0.000871	0.000793
	(0.00450)	(0.00149)	(0.00105)	(0.00560)	(0.00160)	(0.00115)
Age	-0.0239***	-0.0166***	-0.00815***	-0.0135**	-0.00983***	-0.00519***
	(0.00462)	(0.00153)	(0.00108)	(0.00595)	(0.00169)	(0.00122)
Woman				0.0340	0.219***	0.0990***
				(0.0667)	(0.0185)	(0.0132)
High education				0.199**	0.302***	-0.0606***
				(0.0927)	(0.0259)	(0.0181)
Medium education				-0.0106	0.121***	-0.0596***
				(0.0862)	(0.0237)	(0.0168)
Married				-0.275***	0.00810	-0.0897***
				(0.106)	(0.0297)	(0.0214)
Widow				-1.375**	0.0825	-0.0451
				(0.615)	(0.0524)	(0.0373)
Divorced				-0.196	0.00463	-0.0570**
				(0.242)	(0.0403)	(0.0281)
Income (ln)				0.0117	0.0108***	-0.00365*
				(0.0102)	(0.00270)	(0.00204)
Partner’s income				0.0613***	0.00471*	0.000991
				(0.0197)	(0.00285)	(0.00208)
Part time				-0.252***	-0.114***	-0.0485***
				(0.0815)	(0.0230)	(0.0165)
Labor force partic.				0.178	0.116***	0.0170
				(0.123)	(0.0329)	(0.0246)
Public sector				0.0948	0.0595***	-0.0160
				(0.0722)	(0.0204)	(0.0147)
Religious				-0.0938	-0.182***	-0.0514***
				(0.0643)	(0.0187)	(0.0134)
Constant	-49.97***	2.888		-46.53***	1.972	
	(9.165)	(3.043)		(11.38)	(3.253)	
Observations	8700	7971	8027	6052	7345	7403
R-squared	0.085	0.103		0.088	0.172	

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 7: Division of household labor.

Panel A: Mean values for equalization of household work in the treatment group and the control group

	Control	Treated	N
Cooking	0.31	0.28	393/371
Wash clothes	0.14	0.21**	392/372
Cleaning	0.38	0.38	390/368

*** p<0.01, ** p<0.05, * p<0.1 (p-value in a two-sided t-test of the difference between groups)

Panel B: Treatment effect (two year window) on equalization of household work.
Multinomial regressions. Reference category is "women does most household work".

	(1) Cooking		(2) Wash clothes		(3) Cleaning	
	Ref. vs. Equal sharing	Ref. vs. Man does most	Ref. vs. Equal sharing	Ref. vs. Man does most	Ref. vs. Equal sharing	Ref. vs. Man does most
Treatment	-0.0972	0.2391	0.5439***	1.1886**	-0.0058	0.8248*
	(0.1670)	(0.2370)	(0.1960)	(0.4624)	(0.1539)	(0.4313)
Age	-0.0049	0.0309	0.0139	-0.0231	-0.0252*	0.0524
	(0.0156)	(0.0213)	(0.0180)	(0.0460)	(0.0147)	(0.0331)
Constant	-0.3999	-3.1720***	-2.4639***	-2.7481	0.7491	-5.8321***
	(0.7602)	(1.0658)	(0.8801)	(2.2516)	(0.7133)	(1.6970)
Observations	764		764		758	

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Panel C: Sensitivity checks. Treatment effect on *wash clothes*.
Multinomial regressions. Reference category is "women does most household work".

	(1) Total fertility rate		(2) Pre-treatment placebo		(3) Post-treatment placebo	
	Ref. vs. Equal sharing	Ref. vs. Man does most	Ref. vs. Equal sharing	Ref. vs. Man does most	Ref. vs. Equal sharing	Ref. vs. Man does most
Treatment	0.5520***	1.0812**	0.1599	-0.1590	-0.0599	-0.0537
	(0.1838)	(0.4653)	(0.2485)	(0.4868)	(0.1903)	(0.3680)
Age	0.0048	0.0064	0.0173	-0.0009	0.0198	-0.0455
	(0.0158)	(0.0364)	(0.0239)	(0.0409)	(0.0171)	(0.0383)
Constant	-2.0229***	-4.1611**	-2.7659**	-3.1602	-2.3927***	-1.3123
	(0.7727)	(1.8431)	(1.3012)	(2.2039)	(0.7677)	(1.5493)
Observations	964		611		857	

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.