Having it all, for all: child-care subsidies and income distribution reconciled

Francesca Barigozzi* Helmuth Cremer† Kerstin Roeder‡

January 2019

Abstract

This paper studies the design of child-care policies when redistribution matters. Traditional mothers provide some informal child care, whereas career mothers purchase full time formal care in the market. The sorting of women across career paths is endogenous and shaped by a social norm about gender roles in the family. Via this social norm traditional mothers' informal child care imposes an externality on career mothers, so that the market outcome is inefficient. Informal care is too large and the group of career mothers is too small.

In a first-best, full information word equity and efficiency are separable. Redistribution is performed via lump-sum transfers and taxes which are designed to equalize utilities across all couples. The efficient reduction of gender inequality is obtained by subsidizing formal child care at a Pigouvian rate.

However, in a second-best settings, we show that a tradeoff between the reduction of gender inequality and redistributive considerations emerge. The optimal uniform subsidy is lower than the "Pigouvian" level. Under a nonlinear policy the first-best "Pigouvian" rule for the (marginal) subsidy on informal care is reestablished. While the share of high career mothers continues to be distorted downward for incentive reasons, this policy is effective in reconciling the objectives of reducing the child care related gender inequalities and achieving a more equal income distribution across couples.

***JEL*-Classification:** D13, H23, J16, J22

Keywords: Social norms, child care, women's career choices, child care subsidies, redistribution.

*University of Bologna, Italy, Email: francesca.barigozzi@unibo.it.

†Toulouse School of Economics, France, Email: helmuth.cremer@tse-fr.eu.

‡University of Augsburg, Germany, Email: kerstin.roeder@wiwi.uni-augsburg.de.

1 Introduction

While female labor force participation has been increasing steadily over the last decades (Goldin, 2006 and 2014b, Kleven and Landais, 2017) mothers continue to be the main providers of child care within the family (e.g., Paull, 2008; Ciccia and Verloo, 2012). Maternity leave and other child related career breaks or part-time work contribute in a significant way to the persistence of gender inequalities in the labor market. The so called "child penalty" appears to explain up to about 80% of the gender wage gap; see Kleven et al. (2018).

As a possible explanation for the persistence of child-care compatible (part-time) work and "child penalties", the authors point to social norms shaping women's preferences over family and career. Social norms may cause mothers who work full-time to feel guilt when delegating the care of their children to others; see, Guendouzi (2006), Rotkirch and Janhunen (2010) and Rose (2017). Thus, social norms may contribute to the differential sorting of men and women across occupations with women entering low pay occupations that allow for more flexible working hours (see Goldin, 2014 and Card, Cardoso and Kline, 2016).

During the last five decades, most developed countries have put into practice multiple child policies with various declared goals, including gender equity, higher fertility, and child development. The policies who seem to have been the most effective in reducing gender disparities are child care provision and subsidization. Evidence indicates that early childhood spending contributed substantially to enabling women to combine working life and motherhood, and to altering social norms regarding gender roles; see Olivetti and Petrongolo (2017). In OECD countries public expenditure on early childhood and educational care, in cash or in kind, represents on average 0.8 percent of GDP. It attains 2 percent in Denmark, and is above 1 percent in the rest of Scandinavia, the United Kingdom, and France. North American and Southern EU countries have the lowest rates of early childhood public spending. In the United States, early childhood public spending is 0.4 percent of GDP.[1]

Reducing gender disparities in the labor market is not the unique concern which is relevant for the design of tax and expenditure policies. Redistribution across income levels has been *the* major issue which has lead to the emergence of the concept of "welfare state" which applies to all developed countries albeit to a different degree; see Boadway and Keen (2000). Unfortunately, redistributive concerns and the objective of reducing gender disparities in the labor market may be conflicting objectives. Specifically, child care provision and subsidization may be regressive if the parents who benefit more from the policy are the ones with relatively higher income. As an example, full-time working mothers take advantage of full-time day-care relatively more

[1]See OECD, Social Expenditure database "PF3.1 Public spending on childcare and early education, in percentage of GDP" (December 2014). Available at http://www.oecd.org/els/family/database.htm and www.oecd.org/social/expenditure.htm.

than part-time working ones, but full-time working mothers generally have relatively higher income. In addition, day-care facilities may contribute to increase career prospects of full-time working mothers. Redistributive concerns may partly explain why some countries moved away from universal child care benefits and rely today, partly or exclusively, on child tax benefits; see Ferrarini *et al.* (2012).

Surprisingly, the interplay between child care provision/subsidization and redistribution has so far to a large extent been ignored in the literature.[2] We offer a fresh new look at this issue and propose a theoretical model whose crucial ingredient is an inefficient child penalty created by a gender norm. We then investigate the interaction between child penalties, child care policies and redistribution. Our research questions are the following. First, to what extent reducing the child care related gender inequalities and achieving a more equal income distribution are conflicting objectives? Second, how can this potential conflict be mitigated by an appropriate design of the child care policies?

In our setting, the sorting of women into a low- or a high-career path is endogenous and shaped by a social norm about gender roles in the family. Via this social norm a couple's informal child care imposes an externality on high-career couples, so that the *laissez-faire* is inefficient: informal child care is too large and the share of high career mothers is too small. This translates in inefficiently high child penalties. Hence, efficiency requires a reduction in child penalties or a subsidy to formal child care. But a uniform subsidy on formal care benefits full-time working mothers relatively more than part-time working mothers; and full-time working mothers belong to relatively higher-income families. As a result, the policy is regressive. With full information on families' career prospects, efficiency and equity are separable objectives. Instead, when families' career prospects are not observable, an equity-efficiency trade-off applies. We characterize the optimal non-linear child care policy that is able to reconcile, at least in part, government's objectives. Our result shows that an optimally designed child care policy performs relatively well, even when the social planner is concerned with both efficiency and redistribution.

In a recent paper Barigozzi *et al.* (2018) have examined the interplay between social norms, career choices and child-care decisions. However, redistribution across income levels is not relevant because they assume quasi-linear preferences with a constant marginal utility of income. While the excessive share of traditional couples *does* also affect the income distribution by making it more concentrated this in itself does not affect welfare in their setting. Consequently,

[2] An exception is the literature on optimal taxation with endogenous fertility: low-ability families may choose to 'specialize' in quantity, that is, to raise more children relative to higher-ability households. Child-related subsidies can, therefore, be used to enhance re-distribution: family size can be employed as an indicator for the earning capacity of the household (Cigno ??). We totally depart from that literature because the number of children is exogenous in our model. In addition, we do not solve a model of optimal income taxation, we instead design non-linear child care subsidies.

it is not surprising that the considered policies turn out to be regressive in the sense that they favor high career couples who have higher incomes than traditional couples. In particular a uniform subsidy on formal child care financed by a uniform lump-sum tax, though effective in reducing the child care related gender inequalities, implies a reverse redistribution towards couples with higher incomes.

We reconsider the design of child care subsidies when the social welfare function is utilitarian but applies a concave transformation to couples' utilities which introduces a concern for income redistribution. Couples differ in the spouses' earning opportunities in the high career path. High career mothers who work full time and provide no child care suffer a norm cost, which increases with the child care provided by low career mothers. Consequently, the child care choices of low career mothers create a negative externality for the high career mothers. This in turn implies that the *laissez-faire* is inefficient: low career mother provide too much child care and the share of high career mother is too small. Furthermore career choices exacerbate inequalities (as measured for instance by the Gini coefficient) because higher incomes are concentrated on a lower share of the population, which further decreases social welfare.

In a first-best, full information word efficiency and equity are separable. Redistribution is performed via lump-sum transfers and taxes which are designed to equalize utilities across all couples. Child care policies, on the other hand, are designed to achieve the appropriate level of informal child care and the efficient share of high career couples. Since the underlying problem is an externality, it is not surprising that the efficient policy involves a Pigouvian subsidy on market child care, which acts like a Pigouvian tax on informal care.[3] And once child care levels are efficient, the induced career choices are also efficient. However, since this policy taxes away all extra earnings of high-career couples, it is of course not incentive compatible and it cannot be implemented with the information structure we consider; recall that the spouses' earning opportunities in the high career path are not observable. This leads to the study of feasible second-best policies.

We consider two types of second-best settings. First, we assume that instruments are restricted in an *ad hoc* way to be linear. We show that a simple linear policy involves a tradeoff between child care provision and redistributive considerations. Consequently, the optimal subsidy is lower than in the pure efficiency case.

More interestingly, we then show that this tradeoff is to a large extent an artifact of the *ad hoc* restrictions imposed on the policy, namely the linearity. To make this point we consider a simple nonlinear policy under which instruments are solely restricted by the information structure. In other words, we characterize the optimal incentive compatible policy. We show that this policy

[3]This is specific to the social norm we consider in the present setting. In Barigozzi *et al.* (2018) a Pigouvian subsidy on formal child care does not restore efficiency (see comments in Section 4).

reestablishes the first-best "Pigouvian" rule for the (marginal) subsidy on informal care. In other words, even with a simple nonlinear policy there is no longer a tradeoff between child care subsidies and income redistribution. Under asymmetric information, high-career couples continue to enjoy positive rents and their share has to be reduced (compared to the FB) to mitigate these rents. Consequently the outcome remains second-best. Still the policy is effective in reconciling at least to some degree the objectives of reducing child policies and achieving a more equal income distribution across couples. Note that the subsidy on formal care can be implicit in the case where child care is provided in kind.

The information requirement to implement this policy is rather minimal. It is sufficient that career paths *or* levels of formal child care are publicly observable. Amongst these the first one appear to be the least restrictive. When consumption of formal child care is observable for each couple, "topping-up" of child care provided in kind can be prevented, which in practice may appear difficult. But our analysis shows that when career paths are observable, topping up, is not a problem anyway. High career couples will then receive full time care (in kind or subject to a non linear subsidy) and they do not want to supplement this level by care paid at full market prices anyway. And due to the implicit or explicit subsidy, low career couples consume already more formal care then they would at market prices.

From a practical perspective, the non linearity or the policy introduces a measure of means-testing into our policies because child-care fees effectively differ across income levels. Because of the information limitations, means testing remains quite basic and couples within a given career path cannot be distinguished. Still even this basic screening device has a rather dramatic impact in reconciling redistribution and child care policies.

2 The model

Consider a population of couples with children, the size of which is normalized to one. Each couple consists of a mother 'm', a father 'f', and a given number of children. Couples choose their career path, the mode of child care, and their consumption.

There exist two types of career paths (indexed by j). First, a full engaging high-career path, $j = h$, where individuals who take up this career path have to work an entire day which we normalize to one. Second, a less demanding low-career path, $j = \ell$, offering flexible working hours, where individuals can freely choose how much time to spend in the labor market. The time not spent at work can be used for child care c_i, where $i = f, m$. Both jobs pay the wage rate y, but the high-career path comes with additional future earning possibilities q_i. We let $q_f \in [0, Q]$ and $q_m = \alpha q_f \in [0, \alpha Q]$, with $\alpha \in (0, 1]$. An $\alpha < 1$ captures pure discrimination: unequal pay for equally qualified workers, as it continues to be documented in nearly all developed countries.[4]

[4]The parameter α generates the unexplained component in the Oaxaca-Blinder decomposition of the GWG;

Observe that while $\alpha < 1$ adds a measure of realism to the descriptive part of our model, it will not be essential for our results that all continue to hold when $\alpha = 1$. Future revenue q_f is distributed according to the density function $f(.)$, with the cumulative distribution being $F(.)$. Future earning opportunities are perfectly correlated in a couple. Consequently, there is a single level of q_m associated with each level of q_f.[5]

Care for children provided by the spouse(s) is denoted by c_i $(i = f, m)$, while that bought in the private market is denoted by c_p. The latter costs p per unit of time. We let $p = y$, meaning that the current salary of one member in the couple exactly covers the costs of buying full-time child care on the private market.[6] The children must be taken care of for the entire day, implying $c_f + c_m + c_p = 1$. Couples in which both parents choose the high-career path thus have to fully rely on private child care. When parents enter a flexible job their salary decreases proportionally to the time devoted to care. Informal and private care constitute a family public good and its value to the parents is given by:

$$G(c_f, c_m, c_p) = v(c_f + c_m) + \beta v(c_p),$$

where $v' > 0, v'' < 0$ and $v(0) = 0$. Care provided by the father and mother are thus perfect substitutes while informal and private care are imperfect substitutes, with private care being (weakly) less welfare-enhancing than informal care, $\beta \in (0, 1]$.[7] Apart from child care, each parent derives utility from consumption of a numeraire commodity x.

Following Akerlof and Kranton (2000; 2010), individuals may suffer a disutility by deviating from the social categories that are associated with their identity (that is, an individual's sense of self), which causes behavior to conform toward those norms. We assume that individuals desire to conform to the behavior of the group they belong to, namely the behavior of women for mothers and the behavior of men for fathers. Mothers feel guilt if they provide less informal care than the average amount of care provided by woman in the society. Fathers, by contrast, suffer from social stigma when they devote more time to informal care than the average amount of time devoted to care by man in the society.

Given our assumption on the flexibility associated with the two available career paths, the social norm for mothers corresponds to the cost of the full-time job given by $\gamma_m(\max\{0; \bar{c}_m - c_m\})$, where $\bar{c}_m$ is the average time spent with children by mothers in the society. For fathers, the

see Blinder (1973) and Oaxaca (1973). Equation (4) below presents the decomposition of the GWG obtained in our model.

[5] Assortative mating is commonly observed and has been increasing over the last decades; see Schwartz and Mare (2005).

[6] This assumption is simply a normalization that has no relevance for our results. Without it we would obtain a term proportional to $(p - y)$ in the first-order conditions with respect to child care. This would affect the equilibrium levels of child care but otherwise all other results are not affected.

[7] See, for instance, Gregg *et al.* (2005), Bernal (2008), and Huerta *et al.* (2011).

social norm translates into the cost of the flexible job given by $\gamma_f(\max\{0; c_f - \bar{c}_f\})$, where $\bar{c}_f$ is the average time spent with children by fathers. The parameter $\gamma_i \in [0, 1]$, $i = f, m$, reflects the costs of norm deviations.

The timing of couples' decisions is as follows: first, parents choose their career path and then, in the second stage, they choose consumption and the amount of child care (be it formal or informal). Parents act cooperatively and maximize the sum of their utilities:

$$W = x_m + x_f + G(c_f, c_m, c_p) - \gamma_m(\max\{0; \bar{c}_m - c_m\}) - \gamma_f(\max\{0; c_f - \bar{c}_f\}). \tag{1}$$

2.1 Couple's optimization

We first analyze the choice of child care activities for a given career path. Then, by proceeding backward, we consider the choice of career path made by the couple. This allows us to determine the average child care provided in the society and thus to define the cost of the social norm both for fathers and for mothers. We consider only decisions made at the second stage by the couples that turn out to be relevant for our analysis, namely the couples where *(i)* only the father enters the high-career path while the mother enters the flexible job market (traditional couples), and those where *(ii)* both parents take up the high-career path (career couples).[8]

Traditional couple. We denote welfare of this couple by $W_{h\ell}$, where the first subscript refers to the father's career choice and the second subscript refers to the mother's career choice. Since the father took up the high-career path he is not able to take care of the children, and $c_f^* = 0$. Hence, $c_f^* - \bar{c}_f \leq 0$ and the father does not suffer any cost associated with the social norm. Noting that $c_m + c_p = 1$, the couple chooses child care private provision to maximize (1) where $x_{h\ell} = y + q$ because $p = y$. Optimal child care provision is thus implicitly determined by

$$\beta v'(c_p^*) = v'(1 - c_p^*). \tag{2}$$

First-order condition (2) has the usual interpretation: marginal utility from private child care equals the marginal benefit from informal care.

Notice that (2) implies that traditional mothers do not suffer any norm cost. To see why this is the case, consider that the level of formal and informal care only depend on the parents' career choices and not on the heterogeneous career prospect q. In different words, because preferences are linear in consumption, couples where parents choose the same job always purchase the same formal care. In addition we already know that only traditional and career couples are relevant in the analysis, hence only two levels of informal care emerge in equilibrium: traditional

[8] Only the mother in the high-career path is dominated by having both parents entering the high-career path which involves no norm costs for the father and higher future benefits. Similarly, having both parents entering the low-career path can never be optimal since then the couple forgoes future benefits q_f (see Appendix A for a formal analysis of the dominated couples' decisions).

mothers choose $c^*_{h\ell} = 1 - c^*_p$, career mothers necessarily provide $c^*_{hh} = 0$. Thus, it must be $c^*_{hh} = 0 < \bar{c} < c^*_{h\ell}$ and the social norm is not binding for traditional mothers.

The indirect utility of this $h\ell-$couple as a function of private child care c^*_p writes:

$$W^*_{h\ell} = y + q + v(1 - c^*_p) + \beta v(c^*_p)$$

High-career couple. High-career couples have no child care decision to make; they have to buy the full amount of private care on the market. Given that $c^*_f = 0$, the father does not suffer any cost associated with the social norm. Given that the social norm for fathers is never binding we can simplify the notation writing $\bar{c}_m = \bar{c}$ and $\gamma_m = \gamma$. Mothers suffer here the cost from deviating from the norm and the couple's welfare amounts to:

$$W^*_{hh} = y + q(1 + \alpha) + \beta v(1) - \gamma\bar{c}.$$

Note that high-career couples who exclusively have to rely on private child care are those with higher consumption levels, that is $x^*_{h\ell} = y + q < x^*_{hh} = y + q(1 + \alpha)$.

We are now in the position to analyze the couple's decision about the two partners' career paths. Families have to choose whether to be a high-career $hh-$couple fully relying of formal child care, or to be a traditional $h\ell-$couple where the mother provides some informal care, $c^*_{h\ell}$. A couple will become a high-career couple if it is beneficiary to do so, that is if $W^*_{hh} \geq W^*_{h\ell}$, or if

$$q \geq \hat{q}^* \equiv \frac{1}{\alpha}\left[v(1 - c^*_p) + \beta v(c^*_p) - \beta v(1) + \gamma\bar{c}\right].$$

The marginal couple $\hat{q}^*$ is the couple where parents are indifferent between belonging to a traditional and to a career couple. Given $\hat{q}^*$ we can now define average informal child care in society:

$$\bar{c} = \int_0^{\hat{q}^*} c^*_{h\ell} f(q)\mathrm{d}q = F(\hat{q}^*)c^*_{h\ell} = F(\hat{q}^*)(1 - c^*_p).$$

2.2 Market outcome

An allocation is given by the identity of the marginal couple and by the amount of child care provided by traditional couples. The following proposition characterizes the *laissez-faire* allocation.

Proposition 1 (Characterization of the *laissez-faire*) *When mothers who do not provide child care suffer from norm deviations, i.e. $\gamma > 0$, and/or the job market suffers from gender discrimination, $\alpha < 1$, then:*

(i) it is never optimal for fathers to take up the low-career path;

(ii) the marginal couple is given by

$$\hat{q}^* = \frac{1}{\alpha}\left[v(1-c_p^*) + \beta\left[v(c_p^*) - v(1)\right] + \gamma F(\hat{q}^*)(1-c_p^*)\right], \tag{3}$$

couples with future job opportunities higher or equal to the threshold $\hat{q}^*$ choose the high-career path for both parents;

(iii) private care purchased by traditional couples, c_p^*, satisfies equation (2).

There are both traditional and career couples in the economy if $\hat{q}^* \in (0, Q)$. From (3), an interior solution requires that $\hat{q}^*$ exists such that $\hat{q}^* = (1/\alpha)[v(1-c_p^*) + \beta\left[v(c_p^*) - v(1)\right] + \gamma F(\hat{q}^*)(1-c_p^*)] < Q$. Due to the concavity of $v(\cdot)$, $v(1-c_p^*) + \beta\left[v(c_p^*) - v(1)\right] > 0$ holds so that the previous inequality is always met provided that Q is sufficiently large and $F(\hat{q})$ is concave, which we assume in the remainder of the paper.

We are now in the position to define child penalties in our model. As mentioned in the Introduction, child penalties represent today the main component of the gender wage gap (up to 80% in Kleven et al. 2018). In our model, they are affected by the social norm which, in turn, generates a negative externality in the society. This is why child penalties turn out to be inefficiently high, as we will show in the next section.

The GWG is defined as the *difference in total income earned by mothers and fathers* in equilibrium and is given by:

$$\begin{aligned} GWG &= \int_0^Q [y+q]f(q)\mathrm{d}q - \left[F(\hat{q}^*)yc_p^* + \int_{\hat{q}^*}^Q [y+\alpha q]f(q)\mathrm{d}q\right] \\ &= \underbrace{F(\hat{q}^*)\left(1-c_p^*\right)y}_{\text{child penalty}} + \underbrace{\int_0^{\hat{q}^*} qf(q)\mathrm{d}q}_{\text{sorting}} + \underbrace{\int_{\hat{q}^*}^Q (1-\alpha)qf(q)\mathrm{d}q}_{\text{plain discrimination}} \end{aligned} \tag{4}$$

The GWG decomposes in the gap between the hours worked because of family duties, and in the different return to labor supplied in sectors where man and women are employed. The first term in (4) thus represents "child penalty" (see Bertrand *et al.* 2010, Goldin 2014, Kleven *et al.* 2018): mothers in traditional couples do not work full time, but spend part of their time to provide informal child care. The second term accounts for the fact that women forego the extra earning opportunities associated with the high-career path. Interestingly, this second term is affected by social norms and child care decision through $\hat{q}^*$. The model thus offers a clean explanation of how social pressure and/or persisting inequality in the labor market determines women sorting and thus their low participation in leading positions together with lower wages. Finally, the last term in (4) captures the unexplained component of the GWG of the Oaxaca–Blinder decomposition, or the plain discrimination part; it vanishes when $\alpha = 1$.

Before turning to the design of child policy aimed at increasing gender equity, we define the social planner's objective function and the optimal allocation.

3 The optimal allocation

The social planner is interested both in efficiency and in redistribution. Specifically, the social welfare function is assumed to be a concave transformation, $\Psi(\cdot)$, of the families' welfare functions in order to capture inter-family inequality aversion. Thus, a first-best (fb) allocation is defined by aggregate consumption levels $x_{h\ell}^{fb}(q)$ and $x_{hh}^{fb}(q)$, by the indifferent couple, $\hat{q}^{fb}$ (which determines the share of female participation in the high-career path), and by the level of formal childcare provided by mothers in traditional couples, $c_p^{fb}(q)$ for $q < \hat{q}^{fb}$ (recall that, by definition, $c_p^{fb}(q) = 1$ for $q \geq \hat{q}^{fb}$).

Specifically, the social planner chooses $\{x_{hh}(q), x_{h\ell}(q), c_p(q), \hat{q}\}$ to maximize the following welfare function:

$$
\begin{aligned}
SW = & \int_0^{\hat{q}} \Psi\big(x_{h\ell}(q) + v(1 - c_p(q)) + \beta v(c_p(q))\big) f(q) \mathrm{d}q \\
& + \int_{\hat{q}}^{Q} \Psi\big(x_{hh}(q) + \beta v(1) - \gamma \bar{c}\big) f(q) \mathrm{d}q
\end{aligned} \tag{5}
$$

subject to the budget constraint:

$$
y + \int_0^Q q f(q) \mathrm{d}q + \int_{\hat{q}}^Q \alpha q f(q) \mathrm{d}q = \int_0^{\hat{q}} x_{h\ell}(q) f(q) \mathrm{d}q + \int_{\hat{q}}^Q x_{hh}(q) f(q) \mathrm{d}q, \tag{6}
$$

where $\bar{c} = \int_0^{\hat{q}} (1 - c_p(q)) f(q) \mathrm{d}q$.

In Appendix A.2 we derive the optimal allocation that is characterized as follows.

Welfare is constant irrespective of the couple's career path and their future earning possibilities:

$$
W_{h\ell}^{fb}(q) = W_{hh}^{fb}(q) \quad \forall q;
$$

Formal child care is such that $c_p^{fb}(q) = c_p^{fb}\ \forall q$ and is implicitly given by:

$$
v'(1 - c_p^{fb}) = \beta v'(c_p^{fb}) + [1 - F(\hat{q}^{fb})]\gamma. \tag{7}
$$

The left-hand side denotes the social marginal benefit of informal child care provision while the right-hand side denotes its social marginal cost. Note again that the above equation is independent of a traditional couple's q. Compared to the *laissez-faire* described in (**??**), the social marginal cost contains an additional term $[1 - F(\hat{q}^{fb})]\gamma$ which reflects the negative externality of informal care provision on type-hh couples whose share is $1 - F(\hat{q}^{fb})$. Informal child care is thus inefficiently high in the *laissez-faire*, that is $c_p^* < c_p^{fb}$. Not surprisingly c_p^{fb} and $\hat{q}^{fb}$ do not depend on the social welfare function Ψ. This is due to the quasi-linearity of preferences. All Pareto-efficient allocations imply the same levels of c_p and $\hat{q}$, but *may* differ in consumption levels. But since we use a symmetric social welfare function any concave Ψ implies that in the

FB utility levels are equalized. However, the degree of concavity will matter in the second-best settings considered below.

It is interesting to observe that the consumption of couples is constant in each career-path: $x_{h\ell}^{fb}(q) = x_{h\ell}^{fb}$ and $x_{hh}^{fb}(q) = x_{hh}^{fb}$ $\forall q$. This implies that:

$$x_{hh}^{fb} - x_{h\ell}^{fb} = v(1 - c_p^{fb}) + \beta[v(c_p^{fb}) - v(1)] + \gamma F(\hat{q}^{fb})(1 - c_p^{fb}) > 0 \tag{8}$$

The above expression shows that high-career couples do not get higher consumption because of their higher q (as it was the case in the *laissez-faire*), but because the government compensates them for their utility loss due to full private care and due to their cost of the social norm. Finally,

$$\alpha\hat{q}^{fb} f(\hat{q}^{fb}) = f(\hat{q}^{fb})[v(1-c_p^{fb}) + \beta(v(c_{h\ell}^{fb}) - v(1)) + \gamma F(\hat{q}^{fb})(1-c_p^{fb})] - \gamma[1 - F(\hat{q}^{fb})](1-c_p^{fb}) f(\hat{q}^{fb}) \tag{9}$$

so that

$$\hat{q}^{fb} \equiv \frac{1}{\alpha}\{[v(1 - c_p^{fb}) + \beta(v(c_{h\ell}^{fb}) - v(1))] + \gamma F(\hat{q}^{fb})(1 - c_p^{fb}) - \gamma[1 - F(\hat{q}^{fb})](1 - c_p^{fb})\} \tag{10}$$

Comparing (3) and (10) and recalling that $c_p^* < c_p^{fb}$, we observe that $\hat{q}^* > \hat{q}^{fb}$, that is the share of high-career couples is inefficiently low in the *laissez-faire*.

Expression (9) has a simple interpretation in terms of cost and benefits of *decreasing* $\hat{q}$ (that is moving $f(\hat{q})$ couples from traditional to high-career). The LHS measures the marginal benefits in terms of extra future earnings. In the RHS, the first two terms in brackets represent the net lost utility from informal care and the norm cost, respectively. The last term is the Pigouvian term which is negative because the externality imposed on all high-career couples decreases because the average informal care falls. Formally, we have $\partial\bar{c}/\partial\hat{q} = (1 - c_p)f(\hat{q})$. Since a negative cost is effectively a benefit this term could have been moved to the LHS, but since the interpretation of (9) also shows that of (10) this presentation is more telling.[9]

Observe that $\hat{q}^{fb}$ does not depend on Ψ; it is the same in *all* Pareto efficient allocations. The first-best level $\hat{q}^{fb}$ is set purely on efficiency grounds—to maximize the size of the cake which is then redistributed according to social preferences (which in our case involves equalization of utilities).

The following propositions characterizes the optimal allocation:

Proposition 2 (The optimal allocation) *The optimal allocation* $\{x_{hh}^{fb}, x_{h\ell}^{fb}, c_p^{fb}, \hat{q}^{fb}\}$ *maximizes the social welfare function (5) subject to the budget constraint (6) and is characterized as follows:*

[9] Similarly, multiplying both sides of (9) by -1, would be more in line with the original FOC, because it then measures the cost and benefits (reversed from the interpretation discussed) of increasing $\widehat{q}$.

(i) All couples receive the same welfare, no matter the career choice of the mother and the value of future earnings: $W_{h\ell}^{fb}(q) = W_{hh}^{fb}(q)\ \forall\ q$. High-career couples get higher consumption because they are compensated for their utility loss due to full private care and due to their cost of the social norm.

(ii) Formal child care $c_p^{fb}(q) = c_p^{fb}$ is the same for all traditional couples and satisfies (7). It is chosen such that the negative externality induced by the social norm is fully internalized.

(iii) The share of high-career couples is given by $1 - F(\hat{q}^{fb})$ where the marginal couple $\hat{q}^{fb}$ is defined in (10).

(iv) The optimal level of the GWG entails a value of child penalties and a sorting differential equivalent to $F(\hat{q}^{fb})\left(1 - c_p^{fb}\right) y$ and $\int_0^{\hat{q}^{fb}} qf(q)dq$, respectively.

Point (iv) directly follows from substituting $(c_p^{fb}, \hat{q}^{fb})$ into equation (4).

3.1 Welfare analysis of the *laissez-faire* allocation

By comparing the optimal allocation and the market outcome we can establish in which sense the *laissez-faire* allocation is inefficient.

Proposition 3 (Welfare analysis of the *laissez-faire*) *In the laissez-faire allocation:*

(i) in each career path welfare is increasing in couples' career prospect q. Welfare is also generally different among couples belonging to different career paths.

(ii) Formal child care, c_p^*, is inefficiently low and informal care, $c_{h\ell}^*$, is too high. This is due to the negative externality that informal care exerts on high-career mothers through the social norm.

(iii) Female participation in the high-career path is inefficiently low, $\hat{q}^{fb} < \hat{q}^*$.

(iv) In the GWG, both child penalties and adverse sorting are inefficiently high.

In the FB all couples receive the same welfare and high-career couples get higher consumption because they are compensated for their utility loss due to full private care and to their cost of the social norm. Point *(i)* of the proposition shows in the laissez-faire, welfare is increasing in q both among traditional couples and among career couples. Thus, welfare is equalized neither among couples belonging to different career paths nor among couples within the same career path.

Point *(ii)* show that the negative externality translates into underconsumption of formal child care by traditional couples in *laissez-faire* ($c_p^{fb} > c_p^*$). Point *(iii)* concerns the share of women entering the high-career path which is always inefficiently low in *laissez-faire*. When the negative externality is internalized, formal child care increases and the cost of the social norm falls. As a result the high-career path becomes more attractive, implying $\hat{q}^{fb} < \hat{q}^*$.

Finally, point *(iv)* requires some explanations. For any given q, in the *LF*, the female spouse's earnings are less than or equal to her FB earnings. Specifically, child penalties are lower in the FB because women's labor income is higher due to the higher formal child care ($c_p^{fb} > c_p^*$). In addition, adverse sorting is lower because more women enter the high-career path and benefit from future prospects ($\hat{q}^{fb}$ is strictly lower than $\hat{q}^*$).

Note that child penalties $F(\hat{q}^*)\left(1 - c_p^*\right) y$ are directly driven by the average informal child care provision appearing in the norm cost $\gamma\bar{c} = \gamma F(\hat{q}^*)(1-c_p^*)$. The optimal level of child penalty is thus obtained when the negative externality exerted by traditional mothers on career mothers is taken into account. This explains why, in the model, efficiency is reached via the appropriate reduction of child penalties.

4 Decentralizing the first-best allocation

Decentralization of the first-best solution requires a subsidy s on formal child care and individualized lump-sum taxes or transfers $T_{h\ell}(q)$ and $T_{hh}(q)$. When a subsidy s is in place, the net price of private child care is $p^n = p - s = y - s$, and a couple's optimal child care decision solves:

$$v'(1 - c_p) - s = \beta v'(c_p). \tag{11}$$

Comparing (11) with (7) shows that a subsidy of

$$s^{fb} = [1 - F(\hat{q}^{fb})]\gamma \tag{12}$$

implements the FB level of child care. Since formal and informal care sum up to one, a subsidy on market care is effectively a tax on informal care. According to equation (12) s^{fb} corresponds to a Pigouvian *tax* on informal child care; it equals the marginal social cost of the externality informal care imposes on high-career couples.

The lump-sum transfers $T_{h\ell}(q)$ and $T_{hh}(q)$ must be chosen such that welfare levels between all couples are equalized, that is

$$\begin{aligned} W_{h\ell}(q) &= y + q + s^{fb}(c_p^{fb}) + v(1 - c_p^{fb}) + \beta v(c_p^{fb}) + T_{h\ell}(q) = \\ W_{hh}(q) &= y + (1+\alpha)q + s^{fb} + \beta v(1) - \gamma\bar{c} + T_{hh}(q) \end{aligned}$$

Does the government also need to distort career choices? Assume not, then by definition

$W_{h\ell}(\hat{q}) = W_{hh}(\hat{q})$ and $T_{h\ell}(\hat{q}) = T_{hh}(\hat{q})$, so that we have

$$\begin{aligned}
&y + q + s^{fb}(c_p^{fb}) + v(1 - c_p^{fb}) + \beta v(c_{h\ell}^{fb}) \\
&= y + (1+\alpha)q + s^{fb} + \beta v(1) - \gamma \bar{c} \\
&\Leftrightarrow \quad \hat{q}^{fb} = \frac{1}{\alpha}[v(1 - c_p^{fb}) + \beta \left(v(c_p^{fb}) - v(1) \right) + \gamma \bar{c} - s^{fb}(1 - c_p^{fb})]
\end{aligned} \tag{13}$$

With (12) we can see that (13) and (10) coincide once formal child care is subsidized at the right rate so that no distortion concerning the career choice is needed.

Hence, with sufficiently powerful instruments efficiency and redistribution can be addressed separately: the Pigouvian subsidy s^{fb} on private child care restores efficiency in informal child care provision while the two sets of non-linear transfers $T_{h\ell}(q)$ and $T_{hh}(q)$ assure equal welfare to all couples. Note that the individualized transfers redistribute from high to low q couples but also compensate the high-career couples for their utility losses due to full private care and to their cost of the social norm.[10]

We now turn to the study of second-best policies.

5 Linear policy

First, we consider a simple policy under which instruments are restricted in an *ad hoc* way. In other words, we remain agnostic about the information structure. We assume that the instruments necessary to implement the first-best are not available (specifically the individualized transfers) and consider a simple policy which is empirically appealing and effectively used in practice.

The considered policy consists of a uniform (linear) subsidy s on market child care, financed by a uniform lump-sum tax τ. The government's budget constraint is then given by

$$\tau = sF(\hat{q}(p^n))c_p(p^n) + s[1 - F(\hat{q}(p^n))],$$

where $p^n = p - s = y - s$ is the net, after subsidy, price of market care and $c_p^s = c_p(p^n)$ is implicitly determined by:

$$v'(1 - c_p^s) - s = \beta v'(c_p^s) \tag{14}$$

[10] In Barigozzi *et al.* (2018), the social norm is determined by child-care decisions made by the median couple of the preceding generation. With this different modeling strategy it turns out that a Pigouvian subsidy does not restore efficiency but reduces informal care too much. Hence the optimal subsidy must be set below the Pigouvian rule.

The social welfare function can be written as:

$$SW(s,\tau) = \int_0^{\hat{q}(p^n)} \Psi\big(y + q + sc_p(p^n) - \tau + v(1 - c_p(p^n)) + \beta v(c_p(p^n))\big) f(q)\mathrm{d}q$$
$$+ \int_{\hat{q}(p^n)}^{Q} \Psi\big(y + (1+\alpha)q + s - \tau + \beta v(1) - \gamma\bar{c}(p^n)\big) f(q)\mathrm{d}q, \tag{15}$$

where $\hat{q}^s = \hat{q}(p^n)$ and $\bar{c}(p^n) = F(\hat{q}^s)(1 - c_p(p^n))$. The FOC wrt τ is given by:

$$\lambda = \int_0^Q \Psi'(q) f(q)\mathrm{d}q \equiv E[\Psi'], \tag{16}$$

where E is the expectation operator and where $\Psi(q)$ is defined as $\Psi(W_{h\ell}(q))$ for $h\ell$ couples and as $\Psi(W_{hh}(q))$ for hh couples. This equation has a familiar flavor from linear taxation models, in particular Sheshinski (1972). It states that the social marginal cost of raising an additional dollar, λ, should be equal to its social marginal benefit, $E[\Psi']$. Now define:

$$E_{h\ell}[\Psi'] \equiv \frac{\int_0^{\hat{q}^s} \Psi'(q) f(q)\mathrm{d}q}{F(\hat{q}^*)} \quad \text{and} \quad E_{hh}[\Psi'] \equiv \frac{\int_{\hat{q}^s}^{Q} \Psi'(q) f(q)\mathrm{d}q}{1 - F(\hat{q}^*)}, \tag{17}$$

which represent the average marginal utilities of income by traditional and high-career couples respectively.

The FOC with respect to s is given by:

$$F(\hat{q}^s)E_{h\ell}[\Psi']c_p^s + (1 - F(\hat{q}^s))E_{hh}[\Psi']\left[1 + \gamma F(\hat{q}^s)\frac{\mathrm{d}c_p^s}{\mathrm{d}p^n} - \gamma(1 - c_p^s) f(\hat{q}^s)\frac{\mathrm{d}\hat{q}^s}{\mathrm{d}p^n}\right]$$
$$- \lambda\left[F(\hat{q}^s)c_p^s - sF(\hat{q}^s)\frac{\mathrm{d}c_p^s}{\mathrm{d}p^n} + s(1 - c_p^s) f(\hat{q}^s)\frac{\mathrm{d}\hat{q}^s}{\mathrm{d}p^n} + 1 - F(\hat{q}^s)\right] = 0. \tag{18}$$

Noting that $E[c_p^s] = F(\hat{q}^s)c_p^* + 1 - F(\hat{q}^s)$ we show in Appendix A.4 that the optimal subsidy on informal child care, s^o, amounts to:

$$s^o = \gamma\frac{(1 - F(\hat{q}^s))E_{hh}[\Psi']}{E[\Psi']} - \frac{\operatorname{cov}[\Psi', c_p^s]}{E[\Psi']\frac{\partial E[c_p^s]}{\partial p^n}} \tag{19}$$

The first expression is the Pigouvian term and the second term is the redistributive term. When $\Psi'' = 0$ so that social welfare is not concave and there is no concern for redistribution and the above expression reduces to $s^o = [1 - F(\hat{q}^s)]\gamma$, which is the first-best Pigouvian rule. From expression (13) this also yields $\hat{q} = \hat{q}^s$ so that we return to the first-best allocation. When the social welfare function is concave, we have $\operatorname{cov}[\Psi', c_p^s] < 0$ since families with higher formal care have a higher welfare. In the Appendix we show that $\partial E[c_p^s]/\partial p^n < 0$ so that the second term on the RHS in expression (19) is negative (a positive fraction is preceded by a negative sign). Redistributive concerns thus decrease optimal child care subsidies since it is mainly the high-career couples who profit from such subsidies. Furthermore, we have $E_{hh}[\Psi'] < E[\Psi']$ so that the Pigouvian term is also reduced compared to its first-best counterpart. This is because the externality affects high career-couples who in the second-best have a lower social marginal utility. The marginal social damage of the externality is determined by converting their (marginal) utility into social (marginal) utility, which is achieved by the term $E_{hh}[\Psi']/E[\Psi']$.[11] Consequently, we have $s^o < s^{fb}$; see Appendix A.4 for the formal proof.

Proposition 4 (Linear child care subsidy) *The optimal linear policy when redistribution is relevant ($\Psi'' > 0$) implies:*

(i) $s^o < s^{fb}$ *because it is mainly the high-career couples who profit from this policy. Thus, formal child care purchased by traditional mothers,* c_p^s*, is inefficiently low* $\left(c_p^{fb} > c_p^s\right)$*;*

(ii) and $\hat{q}^s > \hat{q}^{fb}$ *so that there are more traditional couples in the second best than in the FB. The marginal couple is distorted upwards to reduce the share of high career couples receiving the subsidy for full-time formal care which improves redistribution.*

(iii) The GWG is inefficiently high.

As expected, the linear subsidy mitigates the inefficiency of the laissez-faire informal care provision but does not fully restore efficiency. To understand (*iii*) the effect of the linear subsidy on the GWG consider that, for any given q, with the linear subsidy, the female spouse's earnings are less than or equal to her FB earnings. Specifically, the component of the GWG due to child penalties is higher with the linear policy because women's labor income falls due to the increase in informal child care ($c_{hl}^{fb} < 1 - c_p^s$). In addition, the GWG from adverse sorting increases because, with the linear policy, less women enter the high-career path and benefit from future prospects ($\hat{q}^s > \hat{q}^{fb}$). However, welfare is obviously higher with the linear policy than in the LF.

6 Nonlinear policy

Now, we take a different approach and assume that the available policies are not restricted in an *ad hoc* way. Instead, we study the design of the best policy that is available *given the information structure.* This is not just a matter of theoretical interest. The important underlying practical question is whether the distortions characterized in the previous section are unavoidable once redistribution under asymmetric information is involved, or whether they are simply artifacts of the linearity of the considered policy.

Under full information this approach yields the first-best, but this supposes that all relevant variables, including a couple's high-career earning opportunities q are publicly observable. We shall now assume that q is not publicly observable but that both the career path and the level of market care *are* observable at the individual (couple's) level. The government can then

[11] In the FB, utilities are equalized so that this term is equal to one.

offer two contracts conditioned on the *reported* type $\tilde{q}$ denoted by $\{J(\tilde{q}), c_p^g(\tilde{q}), T(\tilde{q})\}$, where $J \in \{h\ell, hh\}$ indicates the career path. T is the transfer that households have to pay and $c_p^g(\tilde{q})$ is the amount of formal child care provided by the government. Since $c_p^g(\tilde{q})$ is observable at the couple's level, the distinction between in-kind provision and a nonlinear taxation of market care is not relevant; see Cremer and Gahvari (1997). To be more precise, this is simply a matter of practical implementation of the underlying optimal contract. This implies, in particular, that when $c_p^g(\tilde{q})$ is interpreted as in-kind provision, topping up is not possible.[12] As usual we shall, without loss of generality, concentrate on incentive compatible contracts.

Given that no topping up is possible it must be $c_p^g(q) = 1$ for all hh couples. In addition, given that, conditional on the career path, all families have the same preferences for child care, it is impossible to separate families according to q once the career path has been assigned. Hence, the government offers only two contracts: $\{T_{h\ell}, c_p^g\}$ for $h\ell$-couples and $\{T_{hh}, 1\}$ for hh-couples. In other words, all traditional couples consume the same level of market care and face the same tax or transfer. The same is true for all high-career couples.[13]

The average externality now is $\bar{c} = F(\hat{q}^g)\left(1 - c_{h\ell}^g\right)$, where $\hat{q}^g$ indicates future prospects of the marginal couple, or the couple such that welfare is the same in the two career paths, $\hat{q}^g : W_{hh}(\hat{q}^g) = W_{h\ell}(\hat{q}^g)$.

The government maximizes the following welfare function:

$$\max_{T_{h\ell}, c_p^g, T_{hh}, \hat{q}^g} SW = \int_0^{\hat{q}^g} \Psi(\underbrace{y + q + c_p^g y - T_{h\ell} + v\left(1 - c_p^g\right) + \beta v\left(c_p^g\right)}_{W_{h\ell}}) f(q)\,\mathrm{d}q$$

$$+ \int_{\hat{q}^g}^{Q} \Psi(\underbrace{2y + (1+\alpha)\,q - T_{hh} + \beta v(1) - \gamma F(\hat{q}^g)\left(1 - c_p^g\right)}_{W_{hh}}) f(q)\,\mathrm{d}q \tag{20}$$

subject to the budget constraint

$$F(\hat{q}^g)\,T_{h\ell} + [1 - F(\hat{q}^g)]\,T_{hh} - p\left[F(\hat{q}^g)\,c_p^g + 1 - F(\hat{q}^g)\right] \geq 0, \tag{21}$$

and subject to the following incentive constraint:

$$\begin{aligned} &2y + (1+\alpha)\,\hat{q}^g - T_{hh} + \beta v(1) - \gamma F(\hat{q}^g)\left(1 - c_p^g\right) \\ &- \left(y + \hat{q}^g + c_{h\ell}^g y - T_{h\ell} + v\left(1 - c_p^g\right) + \beta v\left(c_p^g\right)\right) = 0. \end{aligned} \tag{22}$$

Since there is pooling in both groups, incentive compatibility requires simply that $\hat{q}^g$ is indifferent between the two career paths. This follows because $\partial W_{hh}(q)/\partial q = 1 + \alpha > \partial W_{h\ell}(q)/\partial q = 1$

[12] With the considered information structure it can be prevented and nothing can be gained by allowing it.

[13] This is a well known property in contract theory and we skip the proof. To establish the results formally one has to maximize social welfare subject to the budget and incentive constraints. A simple first-order approach will show that the solution involves pooling within each career group.

so that W_{hh} increases faster in q than $W_{h\ell}$. Consequently, condition (22) ensures that no high-career couple with future earnings $q \geq \hat{q}^g$ should have an incentive to mimic a traditional couple, that is $W_{hh}(q) \geq W_{h\ell}(q)\ \forall\ q \in [\hat{q}^g, Q]$. Similarly, it implies that no traditional couple wants to mimic a high career couple. Denote the Lagrangian multipliers associated with the budget constraint and the incentive constraint $\widehat{\lambda}$ and μ respectively.

Using the expectation operators defined in (17) we can write the FOCs with respect to the transfers $T_{h\ell}$ and T_{hh} as:

$$
\begin{aligned}
&- F(\hat{q}^g)E_{h\ell}[\Psi'']F(\hat{q}^g) + \mu + \widehat{\lambda}F(\hat{q}^g) = 0 && (23)\\
&- (1 - F(\hat{q}^g))E_{hh}[\Psi''] [1 - F(\hat{q}^g)] - \mu + \widehat{\lambda} [1 - F(\hat{q}^g)] = 0 && (24)
\end{aligned}
$$

Combining (23) and (24) and rearranging yields:

$$
\widehat{\lambda} = \int_0^{\hat{q}^g} \Psi'(\cdot) f(q)\mathrm{d}q + \int_{\hat{q}^g}^{Q} \Psi'(\cdot) f(q)\mathrm{d}q = E\left[\Psi'\right]. \tag{25}
$$

This equation simply states that the marginal cost of raising additional revenue, $\widehat{\lambda}$, must be equal to its marginal social benefit, $E[\Psi']$. The FOC with respect to formal child care for traditional couples, c_p^g, is given by:

$$
\begin{aligned}
&\int_0^{\hat{q}^g} \Psi'(\cdot)\left[y - v'\left(1 - c_p^g\right) + \beta v'\left(c_p^g\right)\right] f(q)\,\mathrm{d}q + \int_{\hat{q}^g}^{Q} \Psi'(\cdot)\gamma F(\hat{q}^g) f(q)\mathrm{d}q\\
&- \widehat{\lambda} pF(\hat{q}^g) + \mu\left[\gamma F(\hat{q}^g) - y + v'\left(1 - c_p^g\right) - \beta v'\left(c_p^g\right)\right] = 0 && (26)
\end{aligned}
$$

In Appendix A.5 we show that by using (24) and (25) the (26) reduces to:

$$
v'(1 - c_p^g) + \beta v(c_p^g) = [1 - F(\hat{q}^g)]\gamma. \tag{27}
$$

Comparing this expression to (11) shows that the level of child care c_p^g can be decentralized by a subsidy on market care given by:

$$
s^g = [1 - F(\hat{q}^g)]\gamma. \tag{28}
$$

Consequently, the public provision of c_p^g corresponds to an implicit subsidy on market care which is set according to the Pigouvian *rule* defined by (12). In other words, it reflects the marginal social damage which is here measured by the extra norm cost imposed on all career couples. This is an interesting result because it implies that the downward distortion on s implied by the redistributive bias obtained in the previous section indeed appears to be an artifact of the *ad hoc* restrictions imposed on the policy, namely its simple linear specification. When the policy is constrained only by the information structure this distortion vanishes. However, while s^g is set according to the first-best Pigouvian *rule*, its actual *level* will differ from s^{fb}, unless $\hat{q}^g = \hat{q}^{fb}$.

This brings us to the next question namely the comparison between $\hat{q}^g$ and $\hat{q}^{fb}$. This amounts to studying whether the solution under asymmetric information involves a distortion on the marginal couple and if yes in which direction.

The FOC with respect to $\hat{q}^g$ can be written as:

$$\begin{aligned}&\Psi(W_{h\ell}(\hat{q}^g)) - \Psi(W_{hh}(\hat{q}^g)) - \gamma f(\hat{q}^g)(1-c_p^g)(1-F(\hat{q}^g))E_{hh}[\Psi'] \\ &+ \widehat{\lambda}\left[f(\hat{q}^g)(T_{h\ell}-T_{hh}) + pf(\hat{q}^g)\left(1-c_p^g\right)\right] + \mu\left[\alpha - \gamma f(\hat{q}^g)\left(1-c_p^g\right)\right] = 0,\end{aligned} \tag{29}$$

where the first two terms vanish because of the incentive constraint.

The approach is to evaluate the FOC for $\hat{q}^g$ at $\hat{q}^{fb}$ while adjusting all the other endogenous variables according to their respective FOCs.[14] When $\hat{q}^g = \hat{q}^{fb}$ we have from (27) that $c_p^g = c_p^{fb}$. In Appendix A.6 we show that

$$T_{h\ell} - T_{hh} = \gamma[1-F(\hat{q}^{fb})](1-c_p^g) - y(1-c_p^g). \tag{30}$$

Solving (24) for μ and inserting (30) in (29), we have:

$$\begin{aligned}\left.\frac{\partial\mathcal{L}}{\partial\hat{q}^g}\right|_{\hat{q}^g=\hat{q}^{fb}} =& - E_{hh}[\Psi'](1-F(\hat{q}^g))\gamma f(\hat{q}^{fb})(1-c_p^g) \\ &+ E[\Psi']\left[f(\hat{q}^{fb})(-y(1-c_p^g)+\gamma(1-F(\hat{q}^{fb}))(1-c_p^g)) + f(\hat{q}^{fb})y(1-c_p^g)\right] \\ &+ [-E_{hh}[\Psi'](1-F(\hat{q}^g)) + E[\Psi'](1-F(\hat{q}^{fb}))][\alpha - \gamma f(\hat{q}^{fb})(1-c_p^g)] \\ =&(1-F(\hat{q}^{fb}))(E[\Psi'] - E_{hh}[\Psi'])\alpha > 0.\end{aligned} \tag{31}$$

So that we have $\hat{q}^g > \hat{q}^{fb}$. In words, the second-best solution implies an upward distortion of the marginal couple $\hat{q}^g$. Consequently, there are more traditional couples in the second-best solution than in the FB.

To understand this expression note that a couple with $q \geq \widehat{q}$ enjoys an informational rent of $\alpha(q-\widehat{q}) = W_{hh}(q) - W_{hh}(\widehat{q})$. Total rents are thus given by:

$$R = \int_{\widehat{q}}^{Q} \alpha(q-\widehat{q})f(q)\mathrm{d}q$$

and we have:[15]

$$\frac{\partial R}{\partial\widehat{q}} = -\alpha\int_{\widehat{q}}^{Q} f(q)\mathrm{d}q = -\alpha[1-F(\widehat{q})].$$

[14] If the other variables were held constant the sign of the derivative would be inconclusive. However, adjusting all the other variables in an optimal way reduces the problem to a single dimension so that the derivative is informative. As an example, consider the maximization of $f(x,y)$ and denote the solution (x^*,y^*). Showing that at any given point (x,y), $\partial f/\partial x > 0$ is not enough to show that $x > x^*$. However, by using the FOC for y we reduce the problem to the maximization of $f(x,y^*(x))$ and the derivative of this expression allows us to compare x and x^*, as long as the problem is concave which we have to assume anyway.

[15] Note that the derivative wrt the lower bound is zero.

Under full information these rents can be extracted and redistributed. Under asymmetric information they cannot because of the incentive constraint. As $\widehat{q}$ increases the extra amount $\alpha[1 - F(\widehat{q})]$ can be extracted and redistributed which implies a social benefit of $(E[\Psi'] - E_{hh}[\Psi'])\alpha(1 - F(\hat{q}^{fb}))$. In words, the second-best solution involves an upward distortion in the marginal couple in order to reduce "informational rents" of the high-career couples. This means that by increasing the level of q of the marginal couple more tax revenue can be extracted from the high-career couple and redistributed to the traditional couples with lower income, so that welfare increases.

We can now also return to the *levels* of the implicit subsidy implied by the policy. Equation (12) and (28) together with $\hat{q}^g > \hat{q}^{fb}$ imply $s^g < s^{fb}$, so that asymmetric information leads to a lower implicit subsidy on formal care. Intuitively, the strict Pigouvian rule applies in both cases but with $\hat{q}^g > \hat{q}^{fb}$ the group of high-career couples affected by the externality is smaller so that its marginal social damage is also smaller. Consequently, using (11) we'll also have $c_p^g > c_p^{fb}$.(*check*) As in the linear case all these results emerge as long as $\Psi'' < 0$ so that social welfare is concave and there is a concern for redistribution. When $\Psi'' = 0$ we return to the FB solution.

To sum up, while the nonlinear policy brings us back to the first-best Pigouvian rule for the marginal subsidy, it continues to imply a downward distortion on informal care and there will be more traditional couples than efficient. Consequently, the potential conflict between child care provision and redistribution does not solely arise with linear instruments.

Finally, let us revisit the underlying information structure. We have assumed for simplicity that a couple's formal care and career path are observable. We have made this assumption for the ease of exposition, but the arguments and results we presented make clear that the observability of the career path is effectively not necessary. The policy we characterize here can be implemented as long as a couple's level of formal care is observable. This is because high-career couples need full-time care so that their choice of child care would reveal any attempt to mimic a traditional couple. Similarly, a traditional couple mimicking a high-career one would have to choose full-time day care so that mimicking involves the same consumption bundle with or without observable career paths.

The main results of this section are summarized in the following proposition.

Proposition 5 *Assume that couples' formal child care is observable and can be provided publicly at level $c_p^g(q)$ or subject to a nonlinear tax or subsidy. The optimal incentive compatible policy when redistribution is relevant ($\Psi'' > 0$) implies:*

(i) that there is pooling within the traditional and the high career couples groups: all traditional couples receive the same level of formal care and pay the same tax and similarly for all high career couples.

(ii) that high-career couples receive full-time formal care while the level of c_p^g implies an implicit marginal subsidy which is determined by the Pigouvian rule: it equals $s^g = [1 - F(\hat{q}^g)]\gamma$ which reflects the marginal social damage represented by the norm cost imposed on the high-career couples.

(iii) $\hat{q}^g > \hat{q}^{fb}$ so that there are more traditional couples in the second best than in the FB. The marginal couple is distorted upwards to reduce the high-career couples' informational rents which improves redistribution.

(iv) $s^g < s^{fb}$; while both levels are set according to the Pigouvian rule, the inequality follows because there are less high-career couples in the second best so that the marginal social damage of the norm cost is smaller.

(v) that the GWG is inefficiently high.

Again, the policy mitigates the inefficiency of the *laissez-faire* informal care provision but does not fully restore efficiency. Welfare is obviously higher with the nonlinear policy than with the linear one because the distortions imposed by the nonlinear policy are relatively lower. The effect of the nonlinear policy on the GWG is as in Proposition 4 before: for any given q, with the nonlinear subsidy, the female spouse's earnings are less than or equal to her FB earnings. Specifically, the component of the GWG due to child penalties is higher with the nonlinear policy because women's labor income falls due to the increase in informal child care $\left(c_p^g > c_p^{fb}\right)$. In addition, the GWG from adverse sorting increases because, with the nonlinear policy, less women enter the high-career path and benefit from future prospects ($\hat{q}^g > \hat{q}^{fb}$). We expect the GWG to be higher with the linear than with the nonlinear policy because departure from the FB is lower under the nonlinear policy.

7 Conclusion

We have studied the design of childcare policies when women's career choices are endogenous. High career mothers suffer from a norm cost caused by "mothers' guilt". Through their child care choices low career mothers create a negative externality via the norm cost. Consequently, the LF solution is inefficient; it implies too much informal child care and a share of high-career mothers which is too low.

Childcare policies are effective in enhancing efficiency and reducing gender inequalities. However, since they provide larger benefits to high income couples, they tend to be regressive. Under full information, this effect can be offset by lump-sum transfers and the optimal policy is a Pigouvian subsidy on formal child care. A uniform subsidy, on the other hand, involves a tradeoff between efficiency and redistribution (across couples) and should be set below the Pigouvian level.

Under a nonlinear policy the first-best "Pigouvian" rule for the (marginal) subsidy on informal care is reestablished. While the share of high career mothers continues to be distorted downward for incentive reasons, this policy is effective in reconciling the objectives of reducing the child care related gender inequalities and achieving a more equal income distribution across couples.

From a practical perspective this policy can be implemented either through in-kind provision of childcare, at different levels, depending on the mothers career path, and financed with non-linear taxes. Alternatively non-linear subsidies of market care can be used. Either way, day-care fees should be contingent on the amount of time children spend in the facility. Providing free child care to all couples, on the other hand, is never optimal.

References

[1] **Akerlof, G.A., and R.E. Kranton**, "Economics and identity," *The Quarterly Journal of Economics*, 2000, 715–753.

[2] **Akerlof, G.A., and R.E. Kranton**, "Identity economics," *The Economists' Voice*, 2010, *7* (2).

[3] **Aronsson, T., and D. Granlund**, "Gender norms, work hours, and corrective taxation," *Journal of Behavioral and Experimental Economics*, 2015, *56*, 33–39.

[4] **Averett, S.L., Peters, E.H., and D.M. Waldman**, "Tax credits, labor supply, and child care," *Review of Economics and Statistics*, 1997, *79* (1), 125–135.

[5] **Barigozzi, F., Cremer H., and K. Roeder**, "Women's Career Choices, Social Norms and Child Care Policies", *Journal of Public Economics*, 168, 162-173, 2018.

[6] **Bénabou, R., and J. Tirole**, "Identity, morals and taboos: beliefs as assets," *Quarterly Journal of Economics*, 2011, *126*, 805–855.

[7] **Bernal, R.**, "The effect of maternal employment and child care on children's cognitive development," *International Economic Review*, 2008, *49* (4).

[8] **Bertrand, M., Goldin C., and L.F. Katz**, "Dynamics of the gender gap for young professionals in the financial and corporate sectors." *American Economic Journal: Applied Economics*, 2010, *2* (3), 228–55.

[9] **Bertrand, M.**, "New perspectives on gender," *Handbook of Labor Economics*, 2011, *4*, 1543–1590.

[10] **Bertrand, M., Black, S.E., and A. Lleras-Muney**, "Breaking the glass ceiling? The effect of board quotas on female labor market outcomes in Norway,"*NBER WP 20256*, 2014.

[11] **Bertrand, M., P. Cortes, C. Olivetti and J. Pan**, "Social norms, labor market opportunities, and the marriage gap for skilled women", 2016, *NBER WP 22015*.

[12] **Bettio, F., and J. Plantenga**, "Comparing care regimes in Europe." *Feminist Economics,* 2004, *10* (1), 85–113.

[13] **Bisin, A., and T. Verdier**, "Cultural transmission, marriage, and the evolution of ethnic and religious traits,"*The Quarterly Journal of Economics*, 2000, *115* (3), 955–988.

[14] **Bisin, A., and T. Verdier**, "The economics of cultural transmission and the dynamics of preferences,"*Journal of Economic Theory*, 2001, *97* (2), 298–319.

[15] **Blau, F.D., and L.M. Kahn**, "The US gender pay gap in the 1990s: slowing convergence,"*Industrial and Labor Relations Review*, 2006, *60* (1), 45–66.

[16] **Blau, F.D., and L.M. Kahn**, "The gender wage gap: extent, trends, and explanations,"*Journal of Economic Literature*, 2017, *55* (3), 789–865.

[17] **Blinder, A.**, "Wage discrimination: reduced form and structural estimates," *Journal of Human Resources*, 1973, *8*, 436–55.

[18] **Boadway, R. and M. Keen,** "Redistribution," in Atkinson A., and and F. Bourguignon, eds, *Handbook of income distribution*, North-Holland, 2000.

[19] **Buechel B., T. Hellmann, M.M. Pichler,** "The dynamics of continuous cultural traits in social networks", *Journal of Economic Theory*, 2014, *154*, 274–309.

[20] **Bursztyn L., Fujiwara T. and Pallais A.**, "'Acting wife': marriage market incentives and labor market investments", *American Economic Review*, 2017, *107* (11), 3288–3319.

[21] **Card, D., A.R. Cardoso, and P. Kline**, "Bargaining, sorting, and the gender wage gap: quantifying the impact of firms on the relative pay of women." *Quarterly Journal of Economics*, 2016, *131* (2), 633–86.

[22] **Ciccia, R., and M. Verloo**, "Parental leave regulations and the persistence of the male breadwinner model: using fuzzy-set ideal type analysis to assess gender equality in an enlarged Europe,"*Journal of European Social Policy*, 2012, *22* (5), 507–528.

[23] **Del Rey, E., M. Racionero and J.I. Silva**, "On the effect of parental leave duration on unemployment and wages", *Economics Letters*, 2017, 158 14–17.

[24] **European Commission,** "More women in senior positions", Directorate-General for Employment, Social Affairs and Equal Opportunities, January 2010, available at http://ec.europa.eu/danmark/documents/alle_emner/beskaeftigelse/more_women_in _senior_positions.pdf

[25] **Farré, L., and F. Vella**, "The intergenerational transmission of gender role attitudes and its implications for female labour force participation,"*Economica*, 2013, *80* (318), 219–247.

[26] **Felfe, C., and R. Lalive**, "Does early child care help or hurt children's development," *Journal of Public Economics*, 2018, 159, 33-53.

[27] **Fernandez, R.**, "Alfred Marshall lecture: women, work, and culture," *Journal of the European Economic Association*, 2007, *5* (2–3), 305–332.

[28] **Fernandez, R., and A. Fogli**, "Culture: an empirical investigation of beliefs, work, and fertility," *American Economic Journal: Macroeconomics*, 2009, *1* (1), 146–177.

[29] **Fernandez, R., Fogli, A., and C. Olivetti**, "Mothers and sons: preference formation and female labor force dynamics," *The Quarterly Journal of Economics*, 2004, 1249–1299.

[30] **Fortin, N.M.**, "Gender role attitudes and the labour-market outcomes of women across OECD countries," *Oxford Review of Economic Policy*, 2005, *21* (3), 416–438.

[31] **Gallen, Y.**, "Motherhood and the gender productivity gap," The University of Chicago Harris School of Public Policy, Working Paper, January 26, 2018.

[32] **Garcia, I., J.A. Molina, and V.M. Montuenga,** "Gender differences in childcare: time allocation in five European countries", *Feminist Economics*, *17* (1), January 2011, 119–150.

[33] **Gelbach, J.B.**, "Public schooling for young children and maternal labor supply," *The American Economic Review*, 2002, *92* (1), 307–322.

[34] **Gregg, P., E. Washbrook, C. Propper, and S. Burgess**, "The effects of a mother's return to work decision on child development in the UK," *The Economic Journal*, 2005, *115*, 48–80.

[35] **Goldin, C.**, "The quiet revolution that transformed women's employment, education, and family," *The American Economic Review*, 2006, *96* (2), 1–21.

[36] **Goldin, C.**, "A pollution theory of discrimination: male and female differences in occupations and earnings human capital in history", in The American Record, Leah Platt Boustan, Carola Frydman, and Robert A. Margo, Editors, University of Chicago Press, 2014a, 313–348.

[37] **Goldin, C.** "A grand gender convergence: its last chapter." *American Economic Review*, 2014b, *104* (4), 1091–119.

[38] **Guendouzi, J.**, "'The guilt thing': balancing domestic and professional roles," *Journal of Marriage and Family*, 2006, *68* (4), 901–909.

[39] **Huerta, M.d.C., W. Adema, J. Baxter, M. Corak, M. Deding, M.C. Gray, W.-J. Han, and J. Waldfogel**, "Early maternal employment and child development in five OECD countries," *OECD Social, Employment and Migration Working Papers*, 2011, *118.*

[40] **Kimmel, J.**, "Child care costs as a barrier to employment for single and married mothers," *Review of Economics and Statistics*, 1998, *80* (2), 287–299.

[41] **Kleven, H., and C. Landais**, "Gender inequality and economic development: fertility, education and norms", *Economica,* 2017, 84, 180–209

[42] **Kleven, H., C. Landais, and J.E. Søgaard**, "Children and gender inequality: evidence from Denmark", American Economic Journal: Applied Economics, forthcoming, 2018.

[43] **Lalive, R., and J. Zweimüller**, "How does parental leave affect fertility and return to work? Evidence from two natural experiments" *The Quarterly Journal of Economics*, 2009, 1363–1402.

[44] **Lalive, R., and J. Zweimüller**, "Parental leave and mothers' careers: the relative importance of job protection and cash benefits" *The Review of Economic Studies*, 2014, *81*, 219–265.

[45] **Lindbeck, A., S. Nyberg and J. Weibull**, "Social norms and economic incentives in the welfare state," *The Quarterly Journal of Economics*, 1999, *114* (1), 1–35.

[46] **Lindbeck, A., S. Nyberg and J. Weibull**, "Social norms and welfare state dynamics", *Journal of the European Economic Association*, 2003, *1* (2–3), 533–542.

[47] **Matsa, D.A., and A.R. Miller**, "A female style in corporate leadership? Evidence from quotas," *American Economic Journal: Applied Economics*, 2013, *3* (3), 136–69.

[48] **OECD**, "Doing better for families," *OECD Employment Database*, 2010.

[49] **Olivetti, C., E. Patacchini, and I. Zenou**, "Mothers, peers and gender identity," *Mimeo*, 2016.

[50] **Olivetti C. and B. Petrongolo**, "The economic consequences of family policies: lessons from a century of legislation in high-income countries" *Journal of Economic Perspective*, 2017, *31* (1), 205–230.

[51] **O'Neill, J.**, "The gender gap in wages, circa 2000," *The American Economic Review*, 2003, *93* (2), 309–314.

[52] **Oaxaca, R.,** "Male–female wage differences in urban labor markets", *International Economic Review*, 1973, *14*, 693–709.

[53] **Panebianco, F.**, "Socialization networks and the transmission of interethnic attitudes", *Journal of Economic Theory* 2014, 150, 583–610

[54] **Paull, G.**, "Children and women's hours of work," *The Economic Journal*, 2008, *118*, F8–F27.

[55] **Rose, J.**, "Never enough hours in the day: employed mother's perceptions of time pressure," *Australian Journal of Social Issues*, 2017, *52*, 116–130.

[56] **Rotkirch, A., and K. Janhunen**, "Maternal guilt," *Evolutionary Psychology*, 2010, *8* (1), 90–106.

[57] **Ruhm, C.J.**, "The economic consequences of parental leave mandates: lessons from Europe," *The Economic Journal*, 1998, *113* (1), 285–317.

[58] **Schwartz, C.R., and R.D. Mare**, "Trends in educational assortative marriage from 1940 to 2003," *Demography*, 2005, *42* (4), 621–646.

[59] **Sheshinski, E.**, "The Optimal Linear Income Tax", *Review of Economic Studies*, 1972, 39, 297-302.

Appendix

A.1 Couples' optimization

A.1.1 Only the mother enters the high-career path

Since the mother is in the high-career path, she is not able to take care of the children, and $c_m^* = 0$. Hence, $\bar{c}_m - c_{h\ell} > 0$ and the mother suffers the cost of not conforming to the norm. If the father provides some child care he suffers because of the norm too, so that here both social norms are potentially binding.

Welfare of this couple is denoted by $W_{\ell h}$. Noting that $c_{\ell h} + c_p = 1$ and $p = y$, the couple chooses $c_f = c_{\ell h}$ to maximize:

$$\max_{c_{\ell h}} \quad W_{\ell h} = y + \alpha q + v(c_{\ell h}) + \beta v(1 - c_{\ell h}) - \gamma_f(\max\{0; c_{\ell h} - \bar{c}_f\}) - \gamma_m \bar{c}_m.$$

Optimal child care provision, $c^*_{\ell h}$, is implicitly determined by:

$$v'(c^*_{\ell h}) = \beta v'(1 - c^*_{h\ell}) + I\gamma_f$$

where I is an indicator function which takes value 1 when the social norm for fathers is binding, namely when $c^*_{\ell h} > \bar{c}_f$, and 0 otherwise.

Indirect welfare $W^*_{\ell h}$ writes:

$$W^*_{\ell h} = y + \alpha q + v(c^*_{\ell h}) + \beta v(1 - c^*_{\ell h}) - \gamma_f(\max\{0; c^*_{\ell h} - \bar{c}_f\}) - \gamma_m \bar{c}_m.$$

A.1.2 Both couples enter the low-career path

Here again, if the father provides some child care, he suffers because he deviates from the norm. Both social norms are potentially binding. Welfare of this couple is denoted by $W_{\ell\ell}$. Noting that $c_m + c_f + c_p = 1$ and $p = y$, the couple chooses $c_{\ell\ell} = c_m + c_f$ to maximize:

$$\begin{aligned}\max_{c_m;c_f} \quad W_{\ell\ell} =& (1 - c_f)\, y + (1 - c_m)\, y - p(1 - c_{\ell\ell}) \\ &+ v(c_{\ell\ell}) + \beta v(1 - c_{\ell\ell}) - \gamma_f(\max\{0; c_{\ell h} - \bar{c}_f\}) - \gamma_m(\max\{0; \bar{c}_m - c_m\}) \\ =& y + v(c_{\ell\ell}) + \beta v(1 - c_{\ell\ell}) - \gamma_f(\max\{0; c_{\ell h} - \bar{c}_f\}) - \gamma_m(\max\{0; \bar{c}_m - c_m\})\end{aligned}$$

Optimal child care provision, $c^*_{\ell\ell} = c^*_m + c^*_f$, is implicitly determined by the two conditions:

$$\begin{aligned}v'(c^*_f) &\leq \beta v'(1 - c^*_{\ell\ell}) + I\gamma_f \\ v'(c^*_m) &\leq \beta v'(1 - c^*_{\ell\ell}) + I\gamma_m\end{aligned}$$

Welfare $W^*_{\ell\ell}$ now is:

$$W^*_{\ell\ell} = y + v(c^*_{\ell\ell}) + \beta v(1 - c^*_{\ell\ell}) - \gamma_f(\max\{0; c^*_f - \bar{c}_f\}) - \gamma_m(\max\{0; \bar{c}_m - c^*_m\}).$$

A.2 The optimal allocation

Denoting λ the Lagrangean multiplier with respect to the budget constraint, the FOCs of (5) with respect to the couples' consumption levels can be rewritten as:

$$\begin{aligned}\frac{\partial SW}{\partial x_{h\ell}(q)} &= \Psi'(W_{h\ell}(q))f(q) - \lambda f(q) = 0 \quad \forall q \leq \hat{q} \\ \frac{\partial SW}{\partial x_{hh}(q)} &= \Psi'(W_{hh}(q))f(q) - \lambda f(q) = 0 \quad \forall q > \hat{q}.\end{aligned}$$

so that:

$$\Psi'(W_{hh}^{fb}(q)) = \Psi'(W_{h\ell}^{fb}(q)) = \lambda \quad \Leftrightarrow \quad W_{h\ell}^{fb}(q) = W_{hh}^{fb}(q) \quad \forall q.$$

Equalizing welfare levels across career paths, we can write:

$$x_{h\ell}^{fb}(q) + v(1 - c_p^{fb}(q)) + \beta v(c_p^{fb}(q)) = x_{hh}^{fb}(q) + v(1) - \gamma F(\hat{q}^{fb})(1 - c_p^{fb}(q)) \quad \forall q. \tag{A.1}$$

We now consider the point-by-point derivative of the social welfare with respect to $c_p(q)$. Given that $c_p(q)$ exerts a negative effect on all $hh-$couples we have:

$$\Psi'(W_{h\ell}^{fb}(c_p^{fb}(q)))\frac{-\partial W_{h\ell}^{fb}(c_p^{fb}(q))}{\partial c_p^{fb}(q)} f(q) + \int_{\hat{q}}^{Q} \Psi'(W_{hh}^{fb}(\varepsilon))\frac{\partial W_{hh}^{fb}(\varepsilon)}{\partial \bar{c}}\frac{-\partial \bar{c}}{\partial c_p^{fb}(q)} f(\varepsilon)\,\mathrm{d}\varepsilon = 0$$

which gives:

$$\Psi'(W_{h\ell}^{fb})\left[v'(1 - c_p^{fb}(q)) - \beta v'(c_{hp}^{fb}(q))\right] f(q) + \int_{\hat{q}}^{Q} \Psi'(W_{hh}^{fb})\left(-\gamma f(q)\right) f(\varepsilon)\,\mathrm{d}\varepsilon = 0$$

Considering that $W_{h\ell}^{fb} = W_{hh}^{fb}$, we can simplify the previous equation as follows:

$$v'(1 - c_p^{fb}(q)) - \beta v'(c_p^{fb}(q)) - \gamma \int_{\hat{q}}^{Q} f(\varepsilon)\,\mathrm{d}\varepsilon = 0$$

showing that it must be $c_p^{fb}(q) = c_p^{fb}$ $\forall q$. Rearranging, the above equation we obtain (26) in the main text.

Taking the derivative of the social welfare function with respect to the marginal couple $\hat{q}$ and rearranging, yields:

$$\alpha \hat{q} f(\hat{q}^{fb}) = f(\hat{q}^{fb})\left[x_{hh}^{fb}(\hat{q}^{fb}) - x_{h\ell}^{fb}(\hat{q}^{fb}) - \gamma[1 - F(\hat{q}^{fb})](1 - c_p^{fb})\right]. \tag{A.2}$$

Given that $c_p(q) = c_p$ $\forall q$, we observe that $x_{h\ell}^{fb}(q) = x_{h\ell}^{fb}$ and $x_{hh}^{fb}(q) = x_{hh}^{fb}$ $\forall q$. Hence, equation (**??**) can be rewritten as:

$$x_{hh}^{fb} - x_{h\ell}^{fb} = v(1 - c_p^{fb}) + \beta[v(c_p^{fb}) - v(1)] + \gamma F(\hat{q}^{fb})(1 - c_p^{fb}) > 0 \tag{A.3}$$

With (A.3) we can rewrite (A.2) as (9) in the main text.

A.3 Comparative statics

Child care, c_p, and the marginal couple, $\hat{q}$, are implicitly determined by the following two equations:

$$f_1(c_p, \hat{q}, p^n) \equiv y - p^n - v'(1 - c_p) + \beta v'(c_p) = 0$$
$$f_2(c_p, \hat{q}, p^n) \equiv y - \alpha \hat{q} + c_p y + p^n(1 - c_p) + v(1 - c_p) + \beta[v(c_p) - v(1)] + \gamma_m F(\hat{q})(1 - c_p)$$

When we want to know the effect in price changes of formal child care, we have to solve:

$$\begin{bmatrix} \frac{\partial f_1}{\partial c_p} & \frac{\partial f_1}{\partial \hat{q}} \\ \frac{\partial f_2}{\partial c_p} & \frac{\partial f_2}{\partial \hat{q}} \end{bmatrix} \begin{bmatrix} \mathrm{d}c_p \\ \mathrm{d}\hat{q} \end{bmatrix} = - \begin{bmatrix} \frac{\partial f_1}{\partial p^n} \\ \frac{\partial f_2}{\partial p^n} \end{bmatrix} \mathrm{d}p^n.$$

Inserting the derivatives and inverting the first matrix, we have:

$$\begin{bmatrix} \mathrm{d}c_p \\ \mathrm{d}\hat{q} \end{bmatrix} = \frac{1}{D} \begin{bmatrix} -\alpha + \gamma f(\hat{q})(1-c_p) & 0 \\ \gamma F(\hat{q}) & v''(1-c_p) + \beta v''(c_p) \end{bmatrix} \begin{bmatrix} 1 \\ -(1-c_p) \end{bmatrix} \mathrm{d}p^n,$$

where $D = [-\alpha + \gamma f(\hat{q})(1-c_p)][v''(1-c_p) + \beta v''(c_p)] > 0$. We thus have:

$$\frac{\mathrm{d}c_p}{\mathrm{d}p^n} = \frac{1}{v''(1-c_p) + \beta v''(c_p)} < 0 \tag{A.4}$$

$$\frac{\mathrm{d}\hat{q}}{\mathrm{d}p^n} = \frac{-[v''(1-c_p) + \beta v''(c_p)](1-c_p) + \gamma F(\hat{q})}{[-\alpha + \gamma f(\hat{q})(1-c_p)][v''(1-c_p) + \beta v''(c_p)]} > 0 \tag{A.5}$$

A.4 Uniform subsidies

The FOC wrt s can be written as

$$E[\Psi' c_p^*] + (1 - F(\hat{q}^*))E_{hh}[\Psi']\gamma \left[F(\hat{q}^*)\frac{\mathrm{d}c_p^*}{\mathrm{d}p^n} - (1-c_p^*)f(\hat{q}^*)\frac{\mathrm{d}\hat{q}^*}{\mathrm{d}p^n} \right] - E[\Psi']E[c_p^*]$$
$$- E[\Psi']s \left[-F(\hat{q}^*)\frac{\mathrm{d}c_p^*}{\mathrm{d}p^n} + (1-c_p^*)f(\hat{q}^*)\frac{\mathrm{d}\hat{q}^*}{\mathrm{d}p^n} \right] = 0,$$

where $E[c_p^*] = F(\hat{q}(p^n))c_p(p^n) + 1 - F(\hat{q}(p^n))$. Noting that

$$\frac{\partial E[c_p^*]}{\partial p^n} = F(\hat{q}^*)\frac{\mathrm{d}c_p^*}{\mathrm{d}p^n} - (1-c_p^*)f(\hat{q}^*)\frac{\mathrm{d}\hat{q}^*}{\mathrm{d}p^n} < 0$$

and $\mathrm{cov}[\Psi', c_p^*] = E[\Psi' c_p^*] - E[\Psi']E[c_p^*]$, we can write

$$\frac{\partial SW}{\partial s} = \mathrm{cov}[\Psi', c_p^*] - (1 - F(\hat{q}^*))E_{hh}[\Psi']\gamma\frac{\partial E[c_p^*]}{\partial p^n} + E[\Psi']s\frac{\partial E[c_p^*]}{\partial p^n}. \tag{A.6}$$

Setting this expression equal to zero and s olving for s yields equation (19). Further evaluating (A.6) at the Pigouvian level $s^{fb} = [1 - F(\hat{q}^{fb})]\gamma_m$, which from (13) implies $\hat{q}^* = \hat{q}^{fb}$ yields

$$\left.\frac{\partial SW}{\partial s}\right|_{s=s^{fb}} = \mathrm{cov}[\Psi', c_p^*] - (1-F(\hat{q}^{fb}))E_{hh}[\Psi']\gamma\frac{\partial E[c_p^*]}{\partial p^n} + E[\Psi'][1-F(\hat{q}^{fb})]\gamma_m\frac{\partial E[c_p^*]}{\partial p^n} = \mathrm{cov}[\Psi', c_p^*] < 0 \tag{A.7}$$

so that assuming concavity we must have $s^o < s^{fb}$.

A.5 Proof of equation (27)

The FOC wrt $c_{h\ell}^g$ is given by:

$$\int_0^{\hat{q}^g} \Psi'(\cdot)\left[y - v'\left(1 - c_{h\ell}^g\right) + \beta v'\left(c_{h\ell}^g\right)\right] f(q)\,\mathrm{d}q + \int_{\hat{q}^g}^{Q} \Psi'(\cdot)\,\gamma F(\hat{q}^g)\, f(q)\mathrm{d}q$$
$$- \widehat{\lambda} p F(\hat{q}^g) + \mu\left[\gamma F(\hat{q}^g) - y + v'\left(1 - c_{h\ell}^g\right) - \beta v'\left(c_{h\ell}^g\right)\right] = 0.$$

With equations (24) and (25) and the following definitions:

$$E_{h\ell}[\Psi'] = \frac{\int_0^{\hat{q}^g} \Psi'(\cdot)f(q)\mathrm{d}q}{F(\hat{q}^g)} \quad \text{and} \quad E_{hh}[\Psi'] = \frac{\int_{\hat{q}^g}^{Q} \Psi'(\cdot)f(q)\mathrm{d}q}{1-F(\hat{q}^g)}$$

we can rewrite the above FOC as:

$$\begin{aligned}&E_{h\ell}[\Psi']F(\hat{q}^g)[y - v'(1-c^g_{h\ell}) + \beta v'(c^g_{h\ell})] + \gamma F(\hat{q}^g)E_{hh}[\Psi'](1-F(\hat{q}^g)) - E[\Psi']yF(\hat{q}^g)\\ &+ [-E_{hh}[\Psi'](1-F(\hat{q}^g)) + E[\Psi'](1-F(\hat{q}^g))]\left[\gamma F\left(\hat{q}^g\right) - y + v'\left(1-c^g_{h\ell}\right) - \beta v'\left(c^g_{h\ell}\right)\right] = 0.\end{aligned}$$

Noting that $E_{h\ell}[\Psi']F(\hat{q}^g) + E_{hh}[\Psi'](1-F(\hat{q}^g)) = E[\Psi']$, we can write:

$$\begin{aligned}&E[\Psi'][y - v'(1-c^g_{h\ell}) + \beta v'(c^g_{h\ell})] - E[\Psi']yF(\hat{q}^g)\\ &+ E[\Psi'](1-F(\hat{q}^g))[\gamma F\left(\hat{q}^g\right) - y + v'\left(1-c^g_{h\ell}\right) - \beta v'\left(c^g_{h\ell}\right)] = 0\end{aligned}$$

which reduces to:

$$[1-F(\hat{q}^g)]\gamma - v'(1-c^g_{h\ell}) + \beta v'(c^g_{h\ell}) = 0.$$

A.6 Proof of equation (30)

Solving the IC constraint for $T_{h\ell} - T_{hh}$ yields

$$T_{h\ell} - T_{hh} = -y - \alpha\hat{q}^g - \beta v\left(1\right) + \gamma F\left(\hat{q}^g\right)\left(1-c^g_{h\ell}\right) + c^g_{h\ell}y + v\left(1-c^g_{h\ell}\right) + \beta v\left(c^g_{h\ell}\right)$$

From (13) we have the first-best marginal couple:

$$\hat{q}^{fb} \equiv \frac{1}{\alpha}\left[v(c^{fb}_{h\ell}) + \beta v(1-c^{fb}_{h\ell}) - \beta v(1) + \gamma F(\hat{q}^{fb})c^{fb}_{h\ell} - \gamma[1-F(\hat{q}^{fb})]c^{fb}_{h\ell}\right]$$

We now substitute $c^g_{h\ell} = 1 - c^{fb}_{h\ell}$ and $\hat{q}^g = \hat{q}^{fb}$:

$$\begin{aligned}T_{h\ell} - T_{hh} = &-y - v(1-c^g_{h\ell}) - \beta v(c^g_{h\ell}) + \beta v(1) - \gamma F(\hat{q}^{fb})(1-c^g_{h\ell}) + \gamma[1-F(\hat{q}^{fb})](1-c^g_{h\ell})\\ &- \beta v\left(1\right) + \gamma F(\hat{q}^{fb})\left(1-c^g_{h\ell}\right) + c^g_{h\ell}y + v\left(1-c^g_{h\ell}\right) + \beta v\left(c^g_{h\ell}\right).\end{aligned}$$

The above equation simplifies to:

$$T_{h\ell} - T_{hh} = \gamma[1-F(\hat{q}^{fb})](1-c^g_{h\ell}) - y(1-c^g_{h\ell}).$$