

CHAPTER 14

POLICY IMPACT

KAREL VAN DEN BOSCH
BEA CANTILLON

1. INTRODUCTION

At a certain level, questions about the impact of policy are easy to answer. Consider the two Korea states, North and South. Fifty years ago, ravaged by war, both were dirt-poor, both had few natural resources, and their prospects were bleak. The North and the South followed policies which were almost diametrically opposed. The former adopted the centralized economic policies of China and the Soviet Union. The latter pursued policies that were more free-market oriented (though certainly not completely *laissez-faire*), and more open to the outside world. Now, the South is a prosperous country, after nearly a half-century of unprecedented growth (in the context of development since 1950, the economic crisis in 1997 was only a minor setback), while the North is one of the poorest countries on earth, suffering regular famines.

That policy can make a difference is therefore clear. Certainly, mistaken policies can have disastrous results. But the example of the two Koreas also raises two questions of a general nature. The first is: did policy makers really have a choice? Or were policies largely dictated by circumstances, in this case in particular by the cold war and international power relations? Secondly, which South Korean policies were key to the economic success? Or did the precise policies not matter much, as long as they did not impede private enterprise? Both questions ask: do politics matter? but in different ways. The first question does so in the spirit of Castles and McKinlay (1997), who enquire whether policy makers can make real choices, or

* The authors thank the editors of the *Handbook* for very helpful comments, Joanna Geerts and members of the Centre for Social Policy for useful references, and Mieke Augustyns for efficient research assistance.

whether their actions are largely determined by social and economic forces beyond their control (and perhaps even beyond their consciousness). The second question asks whether the policies that are enacted (irrespective of how they are arrived at) make a difference for persons' actual circumstances of living. It is the second question with which we will be concerned in this chapter.

This is of course a very large question, which we cannot possibly do justice to in a short chapter. Let us note the main limitations. In order to maintain coherence, we focus our review on the impact of *public income transfer* programs, mainly because that is the area of research with which we are familiar. However, we believe that at least some of the points made also apply to the study of other areas of public policy. Even in this domain we must be selective as regards topics and studies. We do not even claim that the studies quoted are in some sense the best or the most interesting; we use them to make the points we want to make, with a certain preference for cross-national analyses. While we would have liked to concentrate on the impacts itself, methodological discussions cannot be avoided, as different approaches (sometimes) come up with different answers.

The chapter proceeds as follows. The next section reviews a number of approaches than can be taken in the study of policy impacts. In the third section we look at the impact of tax-and-transfer systems on income inequality and poverty. Though the reduction of inequality and the relief of poverty are not the only explicit goals of public transfer systems, and perhaps not even the main ones (Barr 1992), most of the actual goals would imply some redistribution, and therefore "it seems reasonable to assess welfare state policies in terms of their redistributive impact" (Sefton, this volume). The following section considers the impact of public transfers on various activities, in particular labor market participation and informal care. These are both areas where, it has been argued, welfare state programs have unwanted effects, discouraging people from working, and crowding out informal care by relatives and friends. We will see what the evidence in this regard says. The final section has some concluding remarks.

2. METHODS TO ASSESS POLICY IMPACT

Analysts use a variety of approaches to assess policy impact. Often, *social experiments* are seen as the ideal way to evaluate policies. In such experiments, persons are randomly assigned either to a "treatment" group, which receives the benefits or services of a certain program, or to a "control group," which does not. Program impacts are measured as the difference between outcome variables (e.g. income labor market participation, skill level) before and after the "treatment," after adjusting for the results in the control group, which are supposed to capture the effects of all other factors apart from the program which might influence the outcomes. Despite their clear attractiveness, social experiments have serious limitations, as emphasized by

Heckman, Lalonde, and Smith (1999). First, they are much better suited for evaluating new measures that are not yet implemented than for ongoing programs. Secondly, social experiments are inevitably limited in scope, in time, and geographically; and subjects are aware of this. Thirdly, while people can be excluded from programs, participation is generally by and large voluntary, so that the “treatment” group is often self-selected to some extent, introducing bias into the impact estimates. Finally, experiments are expensive and time intensive, and put heavy demands on program administrators and fieldworkers; the requirement for rigorous randomization may conflict with the professional attitude of the latter.

A second approach is the *difference-in-difference approach*. Here, outcomes for persons who get some benefit or service in an actual program are compared with those for otherwise similar persons who do not participate in the program. This approach therefore is similar to the experimental method, with the important difference that it concerns actual programs, implying that the researcher has no say in the assignment of cases to the program. The main problem of this approach is of course to find a suitable comparison group. By definition, persons in the comparison group cannot be completely identical to persons in the “treatment” group—if they were, they would also be eligible for the program in question. Sometimes the assumption is made that the control group is not really comparable, but that any developments apart from the introduction of the program would affect both groups equally, so that any difference in outcomes between the groups can be attributed to the program. Thus, Francesconi and Van der Klaauw (2004) use single women without children as a control group in their evaluation of the impact of the Working Families Tax Credit on single mothers. Schoeni and Blank (2000) compare the labor market participation rates of educated women with those of less educated women to assess the impact of welfare reforms in the USA, arguing that those reforms will have little impact on the first group of women. The approach can also be used on cases at a higher level of aggregation, e.g. states in the USA. When some states implement a measure while others do not, or (more often) do so at different times, outcome variables on the state level can be used to gauge the aggregate impact of the program, assuming that state effects are constant across years, and that any period effects are common to all states. The worry of course is that those assumptions are violated. Additional difficulties are that states often do not enact exactly the same program, or that all states implement them at nearly the same time (Blank 2002).

Perhaps the most basic strategy is to compare outcome variables *before and after* the introduction or administration of a benefit or service. If data are available for a number of periods, one can control for other trends such as changes in the unemployment rate when evaluating labor market participation-enhancing programs. While intuitively plausible, the method can be misleading. On the micro level there is the possibility that entry into a program can be the result of a temporary setback, which would be remedied even without the program (the “Ashenfelter dip;” see Heckman, Lalonde, and Smith 1999). A person may become unemployed, take part in a job-search program, and find work again, but the last event may not be the result of the program. On the aggregate (state or country) level, the introduction of a

program can be endogenous: measures may be enacted precisely because the situation calls for them.

The complement to the before–after approach is the *cross-sectional* method. On the micro level it compares the outcomes for participants with those for non-participants in a program. It can be regarded as a curtailed version of the difference-in-difference method, and given what has been said above, the limitations of this approach are obvious, and need not be spelled out. On the macro level of societies, this approach enjoys great popularity, especially in political science, under the label of the *comparative method* (see e.g. Ragin 1987). The method is plagued by the so-called degrees of freedom problem: while societies differ from each other in innumerable respects, the small number of cases (at best a few dozen, often much less, in most studies) prevents researchers from taking account of more than a few.

All approaches reviewed above have in common that they compare outcomes after a program has been implemented or administered with a situation that existed or had existed in the real world—either the situation of other comparable cases at the same moment who did not participate in the program, or the situation of the same cases before they took part in it. In *model-based* evaluations the comparison is made not with a really existing state, but with a hypothetical or simulated counterfactual one. In this approach researchers use a model to predict the impact of the introduction or administration (or, alternatively, the absence) of a program with particular features on subjects such as persons or organizations. For instance (and to make the abstract description more concrete), Blundell et al. (2000) use survey data, a tax and benefit simulation model, and a labor market behavioral model to predict the impact of the Working Families Tax Credit in the UK on hours of work and labor market participation. The validity of such predictions depends of course crucially on the quality of the data and on, in particular, that of the model and its parameters. Typically for behavioral models, these parameters are estimated using survey data, which makes them subject to sampling variability, and more importantly, to specification error. Moreover, model parameters estimated on the whole population or a large group may not always be applicable to the rather specific groups on which many real-world programs focus.

A particular kind of model is presented by tax and benefit models. These models incorporate, in as much detail as possible, the tax and benefit rules existing in a country, and can calculate disposable income out of gross income or market income for households in a micro database (Sutherland 2001). More interestingly, one can replace some existing rules with alternative ones, and compare the resulting income distribution with the current one, providing a very detailed picture of the impact of the alternative rule. Typically, such models do not incorporate behavioral reactions, and therefore provide only a first-order approximation of the true impact. However, for many purposes this is quite informative.

Independent of these methods, a useful distinction can be made between studies which look at the social impact of large institutions, such as the welfare state as a whole, and research which tries to identify the effects of particular measures or policy

reforms. The first kind is often rather academic in nature, while the latter tends to be more policy oriented. “Holistic” studies are generally cross-national, comparing aggregate indicators of programs and society-wide indicators of social outcomes. “Particular” studies are more limited in scope, often considering only one country.

Finally, all methods reviewed only help to discover impacts that the researcher is looking for. Yet, there may be a host of unintended effects that we just have not thought about.¹ Theory and previous studies might help in thinking of unintended consequences, but otherwise it is just a matter of imagination.

3. THE IMPACT OF PUBLIC TAX-AND-TRANSFER SYSTEMS ON INCOME INEQUALITY AND POVERTY

In this section we will review two “holistic” approaches to the study of the impact of the public tax-and-transfer system on income inequality and poverty, namely the “pre-post taxes and transfers” method, and the (truly) comparative approach. In the third section we look at the impact of US welfare reforms in the Clinton era on a number of outcomes.

3.1 The “Pre-post” Approach

The standard method to assess the degree of redistribution effected by taxes and transfers is to compare the distributions of income “pre taxes and transfers,” i.e. income when taxes have not been subtracted and without transfers, and “post taxes and transfers,” i.e. disposable income. Income “pre taxes and transfers” is variously called market income, factor income, private income, or original income, depending on what is precisely included in transfers.² In terms of Section 2, the method can be seen as a rather crude instance of the model-based approach to the measurement of policy impacts. An important element of the standard method is that income is measured on the household level, not on the individual level. The idea is that members of one household pool their resources, so that economic well-being is produced on the household level and equally shared among its members. Of course,

¹ For instance, Peltzman (1975) shows that seat belts saved lives of passengers in cars, but (because drivers felt safer and hence free to drive more carelessly) cost about an equal number of lives among pedestrians.

² In the literature, the words “before” and “after” are often used instead of “pre” and “post.” However, since the former terms inappropriately suggest a temporal order, these are avoided here.

larger households need more income than smaller ones to achieve the same level of economic well-being, although they profit from economies of scale in the consumption of housing, heating, and such items. An equivalence scale is therefore used to adjust household incomes.

A fairly large number of studies have employed the standard approach, e.g. Ringen (1989), Mitchell (1991), Deleeck, Van den Bosch, and De Lathouwer (1992). A fairly comprehensive study is provided by Mahler and Jesuit (2004), using data from the Luxembourg Income Study, and covering twelve OECD countries (including the main Anglo-Saxon countries, as well as Scandinavian and northern European nations) for the period 1981–2000. Their main results are consistent with previous studies. First of all, the measured overall impact of taxes and transfers on inequality is large. The Gini coefficient, a commonly used measure of income inequality, is nearly halved in Sweden, and even the limited American welfare state (at least in terms of cash transfers) achieves a reduction of 23 per cent. The impact on income poverty (using a poverty line set at 50 per cent of national median equivalent income) is even more impressive. Pre taxes and transfers between 24 and 32 per cent of all households are in poverty, while “post-government,” poverty rates vary between 5 and 17 per cent; on average across countries about two-thirds of market income poor households are lifted above the poverty line by taxes and transfers.

Secondly, although the impact of government income redistribution through taxes and transfers is large in all countries, the variation across welfare states is important. Scandinavian and the Benelux countries achieve the largest reductions in measured inequality: between 40 and 50 per cent. Germany and France score somewhat lower, around 39 per cent, while taxes and transfers in the UK, Australia, and Canada reduce inequality by around 30 per cent. The reduction is smallest in the USA, only 23 per cent. A study by Immervoll et al. (2004) using data from the European Community Household Panel and national data-sets complements this picture, as it provides results for a number of European countries which are not (well) represented in the LIS database, in particular the southern European countries. They find that the tax-benefit system is highly distributive in a number of Scandinavian and European continental countries. Most southern European countries on the other hand have a low degree of redistribution (about 30 per cent reduction in the Gini). Ireland, the UK, and also Spain form a middle group.

Thirdly, most of the redistribution is achieved through transfers—on average across countries they account for 73 per cent of the overall reduction, while taxes account for only 27 per cent. While there is considerable variation across countries in the relative importance of taxes and transfers in fiscal redistribution, the maximum share of taxes is 44 per cent—in the USA. The main factor explaining this variation appears to be the aggregate share of transfers in total household income (or what one could call the size of the overall transfer budget); where this is large, taxes account for only a small part of total redistribution; where this is small, as in the USA, Australia, and Canada, taxes are more important.

The empirical finding that taxes are less redistributive than transfers might be considered surprising, as in many countries most transfers are not explicitly means

tested, while tax systems in all OECD countries are to some extent progressive, meaning that as income rises taxes paid as a proportion of income increase. However, this progressivity is relatively limited in countries with the highest average tax rates, such as Sweden and Denmark (Wagstaff et al. 1999). When progressivity is zero, taxes are proportional to income, and do not effect any reduction in income inequality (as it is commonly understood and measured). Conversely, several countries with a rather progressive tax structure, such as France and Germany, tend to enjoy low average tax rate. In those countries, the relatively limited overall size of the tax intake prevents it from having an important impact on the overall income distribution. There appears to be some sort of a trade-off between progressivity and the average tax rate (Verbist 2004). The reason for this trade-off could be that as the government has to increase taxes to cover its expenses, it becomes increasingly difficult, politically and economically, to put most of the burden on the highest incomes, and everyone has to take up their share in the total cost of government activities. On the other hand, even though in most countries most public *transfers* are not means tested, they still tend to go to households with no or little other income, thus considerably reducing measured inequality and income poverty. This point applies in particular to pensions.

The standard “pre-post” method has a number of shortcomings and problems. The first is that, as it is commonly applied, it takes only account of cash transfers, and not of transfers in kind, such as (most importantly) health care and education. This point is addressed in a paper by Garfinkel, Rainwater, and Smeeding (2004). They find that “full income,” which includes the cash value of in-kind benefits, is less unequally distributed than disposable income. The difference is largest among English-speaking nations, especially the USA. After taking account of in-kind benefits (as well as the taxes required to finance them), these countries still have the most unequal distributions of income, but the differences from the northern continental European countries and Scandinavia are narrowed substantially. The reasons for this shift are: first, that some nations, in particular the USA, that spend relatively little on cash transfers, devote more of their resources to in-kind benefits; and secondly, that the big spending welfare states rely more heavily on indirect taxes and taxation of cash benefits than e.g. the USA.

As Garfinkel et al. themselves note, there remain a number of conceptual and empirical problems in this type of analysis, regarding the incidence and the valuation of in-kind benefits. One problem is that the equivalence scales typically used are designed for consumption that is paid out of disposable income. For the analysis of “full income,” a different equivalence scale might be needed, which would reflect the greater needs of children for education, and of the elderly for health care.

A second problem of the standard method (again, as it is typically applied) is that the income accounting period is usually only one year. But a large part of social security can be considered as an institution that forces people to make transfers across the life cycle (forced savings), rather than between-person or between-household transfers; this point applies of course in particular to pensions. Actually, in all countries a large part of the measured reduction in overall inequality is due to

pensions (Mahler and Jesuit 2004). One way to address this point is to look only at the non-elderly (although social insurance systems for sickness, invalidity, and unemployment also incorporate intraperson transfers). The figures of Mahler and Jesuit (2004) indicate that among households headed by persons at working age (25–59), the equalizing impact of public transfers is considerably lower, though still respectable: on average 26 per cent instead of 37 per cent among the population as a whole. (Yet, disposable income inequality among this group is smaller than among the population as a whole.) Moreover, countries that score high on redistribution among the total population are not necessarily those that achieve a large equalizing effect among those at working age.

Unfortunately, data that permit us to analyze the equalizing effect of social transfers on a lifetime basis do not seem to exist. The next best thing is to construct a model, using data from panel surveys, to construct estimates of lifetime earnings and transfers. As data requirements are high, and the construction of such models involves a great deal of researcher time, energy, and intelligence, few such models have been constructed. Nelissen (1993) for the Netherlands and Falkingham and Harding (1996) for Australia and Britain are some of the few. Nelissen (1993, 236) reports that the social security system reduces lifetime income inequality by about 26 per cent in the oldest cohorts studied (born 1930–45), and somewhat less for younger cohorts. Most of the reduction is due to public flat-rate pensions and invalidity benefits; semi-public earnings-related additional pensions actually *increase* lifetime inequality. Falkingham and Harding (1996, 254) find that the net effect of the tax/transfer system in Britain is to reduce the Gini coefficient by 0.082; in Australia the effect is greater, at 0.097. In percentage terms the reduction in inequality represents 25 per cent and 26 per cent. The authors conclude that the primarily social assistance-based system of Australia, with its emphasis on poverty alleviation, in conjunction with a more progressive tax system, results in a greater degree of interpersonal income equalization, while the primarily social insurance-based system of Britain achieves a greater degree of intrapersonal redistribution (Falkingham and Harding 1996, 264). While the figures just quoted cannot be directly compared with the annual redistribution results discussed above, they do indicate that a substantial amount of income redistribution from high- to low-income persons occurs even in a lifetime perspective.

The most basic problem of the “pre-post” method, as many authors have observed, is the assumption that benefits, taxes, and contributions have no feedback effect on the pre-tax, pre-transfer distribution of “market” incomes. This assumption is of course quite unrealistic: without a system of benefits and taxes people would change their work, saving, and family formation behavior. These second-order effects, as well as any macroeconomic “third-order” effects, are disregarded in the standard “pre-post” method. The direction of the resulting bias in the estimate of pre-transfer market income is theoretically indeterminate (Danziger, Haveman, and Plotnick 1981, 979). In the next section we will discuss behavioral responses regarding labor supply; it will turn out that transfer programs are expected to reduce labor supply, especially if they are means-tested. However, the theoretical effect of taxes is

ambiguous. Economic theory also cannot predict the direction of the private savings response to transfer programs (Danziger, Haveman, and Plotnick 1981, 982). People may reduce life-cycle and precautionary saving when they can expect pay-as-you-go old-age pensions or unemployment benefits. However, economists have identified a number of other possible mechanisms, making the net result of transfers on saving behavior uncertain. Little theoretical effort appears to have been spent on the effect of public transfers on household formation. Youngsters may leave the parental home earlier if they are eligible for some benefit when they live on their own. Such benefits may also induce more frequent divorce. Conversely, lacking an old-age pension, many elderly persons might choose (or be forced) to live with their children. These examples suggest that a generous system of public transfers will lead to family dissolution, in the sense that the total population will be spread out across a larger number of families of smaller size. However, the net effect of this on pre-transfer income inequality is hard to establish.

Despite these theoretical ambiguities, it seems likely that in the absence of transfers and taxes, income would be less unequally distributed than measured “pre-taxes-and-transfers” income is now. A large proportion of households now have little or no income except from public benefits, especially but not exclusively among the elderly, and this pushes up observed “pre-taxes-and-transfers” income inequality. Obviously, such households would need some form of non-public income if public benefits were abolished. A confirmation of this hunch can be found in the results of Mahler and Jesuit (2004). Observed “pre-taxes-and-transfers” income inequality is actually higher in generous welfare states such as Sweden, the Netherlands, and Belgium than it is in the USA and Australia. Given what we know about these societies (e.g. the fact that wage inequality is relatively low in the Scandinavian and Benelux countries), it appears highly unlikely that market income inequality in the absence of public transfers would be as high as it would be in the United States. The implication of this is that the “pre-post” method almost certainly overstates the equalizing effect of the public tax-and-transfer system. Another implication concerns the general finding reported above that taxes appear to be less equalizing than transfers. This result might well be biased, as the distribution of taxes is compared with the distribution of gross income, which includes transfer payments, and is therefore less unrealistic than the distribution of “pre-tax-and-transfer” incomes (Ringen 1989, 179).

Above we have discussed possible changes in private behavior that would occur if public transfers did not exist. However, it is probable that the institutional context would also be different (Danziger, Haveman, and Plotnick 1981, 979). Employees that cannot look forward to public pensions would demand (larger) company pensions. Perhaps mutual insurance companies would spring up (again). Last (but not least, although rarely mentioned), there would also be political reactions, one of which would be a probably irresistible demand for the reinstatement of public transfers. The last sentence points to the most fundamental problem of the “pre-post” method: we cannot really envisage what a developed democratic society without public transfers would look like. After all, no such society exists, and if any country tried to totally abolish public transfers, it might well prove economically and politically

unsustainable. This implies that the question, “what is the impact of public transfers on income inequality,” is fundamentally unanswerable, as the proper counterfactual cannot be established (West-Pedersen 1994; Barr 1992, 745). The implication of this is that we cannot measure the impact of any welfare state in an absolute sense; what we could possibly do is to compare the effects of different welfare states.

Given this basic change of strategy, one might try to put the “pre-post” method into a comparative framework. Instead of looking only at one country at a time, one might compare the difference in inequality between pre- and post-transfer distributions across a number of countries. However, the necessary assumption for this approach is that second-order effects are constant across countries, or at least not systematically related to the various systems of public transfers, and this is unlikely to be the case (West-Pedersen 1994, 9). Generous systems will have other effects than strict ones; people will behave differently in response to selective benefits than to universal ones. Therefore, it is at best uncertain whether the cross-national variation in the inequality-reducing effects as measured by the “pre-post” method tells us much about the true comparative redistributive impact of different of tax-and-transfer systems. Given the available data as reviewed above, it seems likely that the inequality-reducing effect of large welfare states is overstated relative to those of smaller welfare states.

3.2 The (Truly) Comparative Approach

We turn now to studies where outcomes of different welfare states are compared with each other, instead of with a hypothetical situation. An obvious but not trivial requirement of comparative studies into the impact of tax-and-transfer systems is to characterize the welfare states one wants to study. Several approaches exist. *First*, international reference works such as MISSOC (Mutual Information System on Social Protection in European Union Member States, as well as other European countries; European Commission 2004), enable one to compare particular welfare arrangements, such as the eligibility rules of particular social security benefits. However, one tends to lose sight of the forest because of the trees. A *second* way is the model family method, following which net incomes under a given tax-and-transfer system are calculated for a set of hypothetical families (Bradshaw and Finch 2002; OECD 2002). This approach therefore reflects the fact that household incomes are always income packages, composed of various sources of income and benefits, which may interact in complicated ways. Thus, they can reveal the real net minimum income guarantee available to families. While the results cannot be regarded as indicators of real-world impacts, they can be informative in that they only reflect (explicit or implicit) policy choices. For this reason they can be used to evaluate trends in government policies regarding minimum incomes and replacement rates, and also to compare policies across welfare states. *Third*, analysts

(Titmuss 1974; Esping-Andersen 1990; and many others) have produced social security and welfare state typologies, which depart from institutional characteristics and not from data on outcomes; see below and Sefton, this volume. Yet, many studies prefer a *fourth* approach, and use total expenditure on welfare state arrangements as a proxy for welfare state effort.

Studies using the last method have now established that there is a strong and negative relationship between social expenditure and income poverty (as well as income inequality) (cf. Bradbury and Jäntti 2001; Cantillon, Marx, and Van den Bosch 2003). Scandinavian countries spend the most, and have the lowest levels of poverty; the Anglo-Saxon countries, as well as southern European nations, spend much less, and poverty is much higher in those societies. As Oxley et al. (2001, 392–6) show, some countries achieve better “efficiency” in terms of child poverty reduction (i.e. poverty is reduced more for each euro or dollar spent) through targeting more on low-income groups. However, “effort” and “targeting” are negatively related, and thus “countries with higher ‘efficiency’ due to targeting have traded a good part of this away by reducing ‘effort.’”

Incontrovertible and important though this relationship is, it raises a number of questions. Welfare states differ in more respects than the size of total expenditures and the degree of targeting. If those were the only important characteristics, the policy recommendation would be simple: increase expenditure (and/or improve targeting for those countries which already spend a lot). However, if proof were needed that things are not that simple, it is given in a paper by Van den Bosch (2002). Using cross-country micro-data, he simulated an across-the-board increase in benefits within existing systems, such that all countries would spend the same proportion of aggregate income on social transfers. Surprisingly, such a move would *not* lead to a convergence in poverty rates, but rather the reverse, as poverty would increase in some European countries where it is already high.

Also, *societies* which sustain well-developed social support systems are likely to be different from those with smaller welfare states. It is suggestive (as well as perhaps surprising) that across OECD countries social expenditure and the incidence of low pay are strongly negatively related (Cantillon, Marx, and Van den Bosch 2003). Alvarez (2001) calls the finding that wage-egalitarian societies present the highest levels of welfare effort and redistribution “the puzzle of egalitarianism.” Part of the reason for this puzzle may be that generous benefits reduce labor supply among those commanding low wages, while the high taxes needed to pay them discourage high wage earners from putting in many hours, leading to a more condensed wage distribution, both from above and from below. But, as Atkinson (1999, 67–8) suggests, another reason may be that some countries are characterized by notions of equity that at the same time support pay norms, collective agreements, and adequate minimum wages, as well as quasi-universal and generous benefits. Politically, such countries could be characterized by strong labor unions (West-Pedersen 1994).

Analysts, especially those favoring the welfare state-type approach, have emphasized a number of methodological shortcomings of total expenditure as a proxy for

welfare state effort. They argue that a euro spent on an earnings-related civil servant pension does not represent the same degree of welfare state effort as a euro spent on social assistance. Another simple but important drawback of this line of comparative research of welfare states is that total expenditure is not really an input indicator, certainly not a policy-input indicator, but at best an intermediate indicator. Governments after all do not each year set down the total budget for welfare state expenditure; social security budgets tend to be open ended. Total expenditure is the result of incremental policy making in the past, as well as social and economic developments on which the government has little influence.

Esping-Andersen (1990), Korpi and Palme (1998), and others have tried to characterize welfare states by way of a typology. Having collected a smaller or larger number of indicators of welfare state characteristics, they try to capture similarities and differences into a limited number of types. Mostly this is done analytically, i.e. the authors formulate a number of ideal types, and typecast actual welfare states according to how closely they resemble one of those types. Alternatively, De Beer, Vrooman, and Willeboer Schut (2001) follow an empirical strategy, investigating whether fifty-eight institutional characteristics of welfare states cluster together to form distinct types (though they use indicators that other researchers would regard as outcomes, such as labor market participation rates). While different typologies employ different names, and produce somewhat different country groupings, the basic pattern is always the same; see Sefton, this volume for a description of Esping-Andersen's (1990) typology.

Korpi and Palme (1998, 675) find the expected relation between welfare state type and budget size (which is here regarded as an outcome of institutions, not as a characteristic): welfare states that rely heavily on means testing or on flat-rate benefits tend to have smaller total expenditure levels than welfare states where earnings-related benefits play a larger role. For this reason, the former perform worse in terms of the impact on income inequality and poverty. This leads the authors to formulate the "Paradox of redistribution:" "The more we target benefits at the poor and the more concerned we are with creating equality via equal public transfers to all, the less likely we are to reduce poverty and inequality" (Korpi and Palme 1998, 661).

This being said, welfare state types are not always very distinguishable as regards their impact. Even the correlation between welfare state type and budget size of which Korpi and Palme (1998, 675) make so much is not very strong, and "some countries in the basic security [mainly Anglo-Saxon] and corporatist [mainly European continental] categories have total expenditures levels approximating those in the encompassing group [Scandinavia]." De Beer, Vrooman, and Willeboer Schut (2001, 5) find that "the liberal welfare states perform consistently worse on the indicators for income levelling, income (in)equality and poverty ... There is however no consistent difference between the social-democratic countries and the corporatist countries. [Both] achieve roughly comparable results in terms of income protection by using quite different institutions." The qualification "in terms of income protection" is important here; as regards labor market outcomes social

democratic welfare states radically differ from corporatist ones: whereas the former are characterized by high labor market participation, in particular of women, the opposite is true of the latter.

3.3 The Impacts of US Welfare Reforms

As each year brings a few or more, smaller or larger, changes in the institutions of each welfare state, and many of these are evaluated in some way, it is impossible and probably fruitless to attempt a review of all “particularistic” studies of separate measures, programs, and reforms. In this section we focus on one particular reform, namely the US social policy reforms during the Clinton presidency in the years after 1993. The reason for this choice is that this reform was radical, wide ranging, and has been well studied, and is therefore a good case to illustrate a number of points. An implication is that we will not only review the impact on poverty and income distribution, since other outcome variables were equally, if not more, important for this reform.

Objectives of the Clinton reform included “to make work pay,” and to get people out of welfare and into work. To this end the Earned Income Tax Credit program was greatly expanded. This program provides persons with children who are working with a refundable tax credit for each dollar earned up to a maximum, thereby in effect topping up low earnings. (A refundable tax credit is not just subtracted from taxes to be paid, but actually paid out to households when no taxes are due.) Furthermore, among other reforms, a lifetime limit of five years was set on federal-funded welfare. For further detail, we refer to Blank and Ellwood (2001). The budget implications of the reform were huge: between 1992 and 1999, annual real federal spending on new or expanded programs increased by over \$30 billion, which is nearly twice as much as total spending on Aid to Families with Dependent Children (AFDC), the main pre-reform welfare program. As a result, the net gain from working for single mothers on welfare dramatically increased (Blank and Ellwood 2001, 7).

It is instructive to compare the Clinton welfare reform with a simple earnings disregard program, where welfare recipients can keep part of their benefit up to a point if they start earning. This does have the desired effect of creating financial incentives for non-working welfare recipients to enter the labor market, but also creates unwanted incentives for current non-recipients to reduce their work effort (Blank, Card, and Robins 1999, 12). This appears to be one of the key reasons for the disappointing results of the negative income tax experiments of the 1970s. By contrast, the Clinton welfare reforms contained a number of provisions to limit this unwanted side effect, including eligibility restrictions that target benefits to long-term welfare recipients, and hours restrictions that limit benefits to full-time workers (Blank, Card, and Robins 1999, 40).

What was the impact of those changes? Perhaps surprisingly, given the scale and size of the reforms, this question is not easy to answer. Certainly, at the end of Clinton's second term, the number of people on welfare had more than halved compared with the start of his first term. Labor force participation among single women with children increased by more than 10 percentage points in this period. Poverty fell significantly. However, at the same time the US economy went through a period of strong growth and labor force expansion. It turns out to be quite difficult to disentangle the impact of policies from the effects of the booming economy. As Blank and Ellwood (2001, 31) write, it is relatively easy to document that outcomes changed at the same time as policy. To establish causality is another matter.

Researchers have spent considerable effort on doing just that, using a variety of methods and data, but relying mostly on difference-in-difference studies on the state level (see Section 2). These studies indicate that policy changes were important in getting people off welfare. Regarding labor market participation, researchers tend to agree that the Clinton policy changes dramatically increased work by single parents, though it is less clear what was the relative contribution of EITC and other work supports versus welfare reform (Blank and Ellwood 2001, 39).

The focus on labor market participation entails a danger of increased poverty, if earnings are no greater than the welfare income they replace, and if some persons are taken off the welfare books without any alternative source of income. Overall, however, the net effect of the policy reforms appears to be positive: poverty declined, and the income of female-headed families with children rose. At the same time, some single-mother families at the very bottom probably became worse off. The most serious question concerns what will happen if the economy stops growing (Blank and Ellwood 2001, 53–4). The policy changes are such that the welfare system is most effective during an economic upturn (when people find it easy to find a job); how it will perform during a recession remains to be seen.

4. THE IMPACT OF INCOME TRANSFERS ON ACTIVITY

It is often alleged that the welfare state, while perhaps a good thing in principle, has a number of unwanted side effects, which reduce its real impact. The perverse effects of welfare state programs haven been most forcefully put forward by Murray (1984). He argues that in the USA, the numbers of poor stopped shrinking in the early 1970s, and then began growing, despite the combination of economic growth and huge increases in expenditures on the poor. Other basic indicators of well-being also took a turn for the worse in the 1960s, most consistently and most drastically for the poor.

The reason for this turn of events, according to Murray, was precisely the huge expansion of welfare state programs, which encouraged behavior that perpetuated the state of poverty, through early school drop-out, weak attachment to the labor market, and family break-up. These failures were then masked through too generous transfers. While many analysts have argued that Murray's thesis does not fit the facts (e.g. Jencks 1992), much time and energy have been devoted to identifying the possible perverse side effects of welfare state programs. In this section we will look at two such side effects, namely discouraging people from working, and crowding out informal care by relatives.

4.1 Impact on Labor Supply

The impact of welfare state programs on labor market participation is the subject of an enormous literature, often of great technical complexity, which is impossible to do justice to in one section of a short chapter. Below, we present certain highlights which give some impression of the variety of issues and results.

The standard economic textbook model (Danziger, Haveman, and Plotnick 1981, 979; Atkinson 1993a) is that persons trade off work against leisure, and that *ceteris paribus* they will prefer leisure over work. Under these assumptions, transfer programs that provide income support without requiring work will unambiguously reduce labor supply through the income effect, that is, people will use the extra income to "buy" extra leisure time. Some persons will work fewer hours, and others will stop working altogether. Transfers that are means tested will have an additional labor supply reducing effect, as for each euro or dollar earned a part of the benefit is withdrawn. The effect of taxes is ambiguous: the fact that taxes reduce net earnings may induce persons either to work more to make up for the lost earnings (income effect), or to work less, as each hour worked brings in less in net earnings (substitution effect).

This bare-bones economic textbook model ignores many dimensions of work and labor supply, as explained by Atkinson (1993a). One is the assumption that people are completely free to choose their hours of work, implying that there is no involuntary unemployment, or compulsory early retirement. Another is the disregard for the institutional context of labor supply decisions, e.g. the presence of collective bargaining, restrictions on laying-off employees, or the fact that real-world tax systems often produce non-linear budget constraints. Income-tested benefits moreover may imply that the budget constraint is non-convex, and effective marginal tax rates may be higher at low earnings than higher up the scale. People living on social assistance may even find themselves in a so-called "poverty trap," as any effort to obtain additional earnings may not bring them any advance in net-income terms. Furthermore, labor market decisions are not made individually, but within families, which may be taxed

jointly, and where there is also unpaid but essential household production work to be done. The trade-off is therefore not simply one between net income and leisure, but between consumption goods bought in the market and having more time for household activities, and also between the incomes and non-working time of husband and wife. Moreover, lifetime considerations may be important, as people may work hard during their prime-age years to provide for their (early) retirement.

Thus, economic theory, certainly when some model assumptions are relaxed, cannot provide a clear-cut answer as regards the direction of the effect of real world tax-and-transfer systems, and moreover, theory is silent on the *magnitude* of the effects, which is as important as the direction. Empirical studies only can provide useful answers. There are several approaches in this domain. One is to use real-world socioeconomic experiments, of which the best-known example is probably the New Jersey negative income tax experiment (Pechman and Timpane 1975). The broad conclusion from this and other similar experiments was that there was a noticeable but not massive reduction in work effort (Atkinson 1993a, 43). Yet, although the evidence produced by such experiments is unique, it cannot be regarded as conclusive, for the reasons set out in Section 2. Other studies have followed the before-after method, or the modeling approach outlined in Section 2.

Atkinson (1993b, 297), reviewing a number of such studies, concludes that, overall, “a number of the effects that have been identified are relatively small in size,” and “there are relatively few situations in which a disincentive effect has been clearly established.” There is evidence that taxation causes married women to work less, but little evidence of a negative response by prime-age male workers. There is also little clear evidence that *benefits* represent a major discouragement to take up work. One reason for this is that, though the tax-and-transfer system in many countries creates a poverty trap, this may affect relatively few people. Also, transfers may have a positive impact (the so-called entitlement effect), as people keep working or looking for work in order to become or remain eligible for benefits.

Another group for which tax-and-transfer arrangements may have an important effect on labor market participation (apart from married women) is men aged 50–64. In many countries participation rates for this group have fallen drastically during the last four decades. Gruber and Wise (1998) show that, across a number of OECD countries, labor force participation of older persons is strongly related to the implicit social security tax on work. This implicit tax arises because in many countries, staying on for one more year in the labor force for older persons implies a reduction in the present discounted value of total pension benefits during the remaining lifetime. In some cases, this reduction is even larger than the net wages earned during the extra period in work! The “tax force to retire” is especially strong in Italy, Belgium, the Netherlands, Germany, and France. However, as Gruber and Wise note, in some countries (e.g. Belgium) the reduction in labor market participation of older persons was not an unwanted side product; rather, encouraging older workers to leave the labor force was an explicit goal, with a view to easing labor market tension and reducing unemployment among younger workers.

Welfare state arrangements and even public transfers can also help to keep persons *in* work. This was after all one of the objectives of the Clinton social policy reforms discussed in Section 3.3. Another illustration is provided by an interesting cross-national study by Gornick, Meyers, and Ross (1996) on the employment of mothers with young children. Gornick et al. note that easier (cheaper) access to child care will increase mothers' employment rate, either (and equivalently) because it reduces the value of time spent at home, or because it increases the net wage mothers can earn. The effect of paid maternity leave cannot be predicted unambiguously—on the one hand it may strengthen mothers' attachment to paid work, on the other it may induce some women to stay at home (temporarily) who would otherwise have kept on working. The direction and especially the magnitude of these effects is therefore an empirical matter. Gornick et al. look at what they call the "child penalty:" the decrease in the probability of employment of mothers, given the presence of young children, all else equal. Compared with an analysis of employment rates per se, this has the advantage that all kinds of institutional and macroeconomic variables are implicitly controlled, insofar as it can be assumed that these other factors affect mothers of young children and other women, e.g. mothers of teenage children, equally. Gornick et al. compare the "child penalty" with a pair of indices that integrate a range of measures of public support for child care and parental leave. They find that these two are strongly related—in some countries which do not strongly support maternal employment the "child penalty" is as large as 35 (Australia) or 45 percentage points (UK), while in Sweden there appears to be no "child penalty" whatsoever.

4.2 The Impact of Welfare State Provisions on Family Care

Some observers maintain that the welfare state not only carries an economic cost in lost hours of work, but also crowds out compassion and activity from private life (Burenstam Linder 1970, quoted in Ringen 1989, 119). One relationship that should be particularly sensitive to such perverse influences is that between the elderly and their children. Formal, social, and emotional ties are less strong than they are between spouses, and between parents and young children within the nuclear family. Old-age care is generally seen as more burdensome than child care (Ringen 1989, 129–30). So what is the evidence as regards the effect of increasing the supply of public old-age care on family care? According to Ringen (1989, 134) "informal care in the family sector is still the dominant form of old-age care." "There are no signs . . . of a decline in family activity, of less vitality or compassion in the sensitive relationships between the elderly and younger family members." However, since Ringen wrote those conclusions, much new research on this topic has been published.

Many writers on this topic take the position that family care and public provisions, far from being substitutes, are actually complements. Several arguments are

advanced in this regard. Families will be more willing to provide help when burdens are not too heavy. Also, generous pensions enable the older generation to reciprocate support from the younger generation. Public services may allow families to specialize in psycho-social support rather than instrumental help (Daatland 2001, 18–19). Three kinds of evidence can be called upon to determine whether the substitution or the complement effect predominates. First of all, there are cross-country differences. These indicate that substitution effects are likely, as countries with the highest level of services seem to have the lowest level of family care (Daatland 2001, 19). However, these differences may be due to the more familistic culture of Germany and Italy (which may be associated with both less public care and more private care), compared with the (allegedly) more individualistic societies of Scandinavia. Secondly, there are cross-sectional studies which investigate whether elderly people tend to receive help from one source only, or whether public services and family help appear together. Such studies typically suggest that family care and public provisions are indeed complements, as many elderly persons use both even when controlling for need (e.g. Künemund and Rein 1999, in a five-country study). In a literature review with a focus on longitudinal studies, Penning and Keating (2000) conclude that the findings suggest that formal services are not used to displace or substitute for informal care but rather, that formal services tend to be used to supplement and complement the care provided by the informal network.

Finally, one can follow developments over time: when public services expand, does family care go down, and vice versa? Here the available evidence is mixed. A study by Lingsom (1997, quoted in Daatland 2001) for Norway suggests that this does not happen. Families were not crowded out, nor did they withdraw, when alternative sources of help were available. On the other hand, Johansson et al. (2003) claim that results show that relatives more often provided care to older people half a century ago than in contemporary Sweden. More recently, cutbacks in public services in Sweden have led to a substantial reversal in care patterns. Increased input from families matches the decline of public services. A positive reading of these results would be that even in individualistic Sweden the welfare state has not destroyed the bonds between elderly persons and their children: when needed (again), the latter are ready to provide help.

5. CONCLUSION

Since this chapter as a whole is fairly short and rather synthetic in nature, it hardly needs summary. However, we would like to make some general points, first on methodological issues and then on substantive ones.

First, a methodological point that is perhaps rather uncontroversial, but still worth making. Theory, certainly economic theory, is in general insufficient to predict the

impact of policies. Theory can guide us as to what to look for, but often the direction of the effects, and almost always their magnitude, can only be established empirically. Often, effects that loom large in the theoretical literature turn out to be insubstantial in the real world.

A second, perhaps less obvious point is that, even though the tool kit of policy analysts contains a variety of methods, it is often very hard to identify, let alone quantify the impact of particular policies with a reasonable degree of accuracy. Even the consequences of the US welfare reform under Clinton turned out to be hard to pinpoint, despite the scope of the reforms, and the wealth of data seemingly available. Social experiments are perhaps inherently the most powerful method, but they are suitable only for programs that are not yet in place, and that can be enacted on a small scale. For larger and existing programs the difference-in-difference method is perhaps the most valid and convincing way to measure policy impacts, whenever it can be applied. The problem of finding a suitable comparison group is often not trivial, though. The fundamental problem seems to be that the impacts of policy changes are often small compared with those of exogenous social and economic developments. It then becomes difficult to tease out the message from the noise.

Thirdly, macro-social comparative studies, which look at large institutions such as welfare states as a whole, have given us important new insights in the past decades. However, the fact that multivariate analysis is nearly impossible with fifteen or twenty cases (rich democratic nations) limits crucially the power of this approach. It therefore has no answer to the basic fact that each welfare state is embedded in a different society, making it very difficult to distinguish impact from association. Welfare state typologies are very useful to get some grasp on the otherwise bewildering variety of institutional characteristics, but appear to have limited potential as predictors of impacts. Perhaps the most fruitful approach is represented by comparative studies which look at the impact of policy packages offered by different welfare states to particular groups, such as mothers with young children, or males at pre-retirement ages. At this middle-of-the-road level, policies can be described, or even quantified with a fair degree of precision; there is often more variety in outcomes; and the relationship between policies and outcomes is more easily established, and easier to interpret.

The main *substantive* conclusion we can draw from the material presented above (despite some methodological reservations) is that policies do have an impact, in the sense of making a difference to people's actual living circumstances. There can be little doubt that large welfare states are more equalizing than smaller welfare states, although it is probable that large welfare states can only flourish in societies that are rather egalitarian in the first place. Their impact is not entirely frittered away through unintended side effects. The experience of US welfare reform under Clinton indicates that a well-designed package of programs can induce people to move off welfare rolls and into work. Comparative research shows that older people retire early when pension and other benefit systems contain clear incentives to do so. Studies suggest

strongly that mothers with young children continue working, or return to the labor force after a time, if a package of benefits and services is in place that helps them to do so.

Secondly, the examples just quoted suggest that a large policy impact requires a large program—or package of programs. Measures need to be well designed, well funded, and sustained over time. Attempts to get results “on the cheap” can backfire. The largest example of this is perhaps the “paradox of redistribution” (Korpi and Palme 1998). Welfare states that attempt to target resources onto the poor tend to have lower redistributive budgets, resulting ultimately in more poverty and more income inequality, compared with welfare states that rely on more universal benefits.

The third conclusion is an instance of the previous one, but worth mentioning in its own regard: people react to incentives, provided these are clear and large. Welfare mothers in the USA move back to work if it is made clearly worth their while to do so. Older men in some continental welfare states retire early in great numbers, when the rules of existing pension and other benefit systems minimize the gains of continuing to work (calculated on a lifetime basis).

Fourthly, we do not intend to imply that getting a large impact is just a matter of spending a large amount of money. In all of the examples just quoted the impact was produced by a package of programs, not by just a single measure. Such a package needs to be well designed, so that the different parts work together towards the same objectives. The comparison of the complicated welfare reforms under Clinton with the rather simple negative income tax proposals indicates that real-world policy packages are often quite complex and detailed, and need to be so, in order to contain unwanted side effects, and to keep costs in check.

REFERENCES

- ALVAREZ, P. 2001. *The Politics of Income Inequality in the OECD: The Role of Second Order Effects*. Luxembourg Income Study Working Paper No. 284. Syracuse, NY: Syracuse University.
- ATKINSON, A. B. 1993a. Work incentives. Pp. 20–49 in *Welfare and Work Incentives: A North European Perspective*, ed. A. B. Atkinson and G. V. Morgensen. Oxford: Clarendon Press.
- 1993b. Conclusions. Pp. 289–97 in *Welfare and Work Incentives: A North European Perspective*, ed. A. B. Atkinson and G. V. Morgensen. Oxford: Clarendon Press.
1999. The distribution of income in the UK and OECD countries in the twentieth century. *Oxford Review of Economic Policy*, 15 (4): 56–75.
- BARR, N. 1992. Economic theory and the welfare state: a survey and interpretation. *Journal of Economic Literature*, 30: 741–803.
- BLANK, R. 2002. Evaluating welfare reform in the United States. *Journal of Economic Literature*, 40: 1105–66.
- CARD, D., and ROBINS, P. 1999. *Financial Incentives for Increasing Work and Income among Low Income Families*. NBER Working Paper No. 8437. Cambridge, Mass.: National Bureau of Economic Research.