Optimal Taxation from a Behavioral Viewpoint: Implications for Redistributive Policies

Jukka Pirttilä (UNU-WIDER and University of Tampere)

Winter School on Inequality and Social Welfare Theory
Canazei, January 2015

Outline

This lecture

- Deals with three of the most important advances in behavioral economics
  - relative income concerns ('relativity')
  - Prospect Theory
  - self control problems / weakness of will / present bias
- Consider their implations for tax policies (income tax in the first two cases, commodity tax for the third case)
- Before that, we start off with a brief general discussion on non-welfarist tax policies
- Some conclusions in the end

Outline

Some papers

- Kanbur, Pirttilä and Tuomala (2006) "Non-welfarist optimal taxation and behavioral public economics", Journal of Economic Surveys
- Kanbur and Tuomala (2013) "Relativity, inequality, and optimal nonlinear income taxation", International Economic Review
- Kanbur, Pirttilä and Tuomala (2008) "Moral hazard, income taxation and Prospect Theory, Scandinavan Journal of Economics
- O'Donoghue and Rabin (2006) "Optimal sin taxes", Journal of Public Economics

Outline

Non-welfarist tax policies

- Behavioral economics: people may not always behave in their own long-term interest
- Opens up a potential new role for corrective taxation
- Brings one into the realm of behavioral public economics
- Interestingly, in public economics, there is a long tradition of similar policies
  - the idea of merit goods dates back at least to Musgrave (1959)
- welfarism / non-welfarism: in the latter, consumers' preferences are not accepted as a basis for determining social welfare

A general model of non-welfarist tax policies: Kanbur et al. (2006)

- First deal with non-linear income taxation: this will serve as a basis for the analysis of relativity
- Then also consider mixed taxation (the combination of non-linear income tax and linear commodity taxes): this is the basis for sin taxation
- The model is based on Mirrlees (1971, 1976). What changes is the social welfare function
- In the Mirrlees model, asymmetric information restricts the tax policies. The government can only observe gross income, not innate abilities. Adverse selection.
- Behavioral applications follow when more structure is given to the non-welfarist social preferences

The individuals

- There is a continuum of individuals, in the basic model with no taste differences
- They have an innate ability, denoted by n, that varies according to a known distribution function $f(n)$
- The individuals max $u(x, y)$ s.t. $x = ny - T(y)$, where T stands for a general tax function
- Individual FOC: $u_x(1 - T') + u_y/n = 0$

The government

- The individual optimization condition gives the self-selection / incentive compatibility constraint for the government optimization problem
- total differentiation of utility wrt n gives

$$\frac{du}{dn} = -\frac{yu_y}{n} \equiv u_n(x, y, n)$$

- The government maximizes

$$S = \int P(x, y, n)f(n)dn$$

subject to the revenue requirement $\int T[z(n)]f(n)dn = R$

- Forming the Lagrangean, integrating by parts and differentiating with respect to u and n gives the optimality conditions

The optimal marginal tax rate

$$T'(z(n)) = \frac{P_x(s - s_p)}{\lambda} - \frac{\mu(n) u_x s_n}{\lambda f}$$

- where $\lambda$is the Lagrange multiplier of the government budget constraint, μ is the Lagrange multilplier of the incentive compatibility constraint, $s = -\frac{u_y}{n u_x}$ is the marginal rate of substitution (MRS) between x and y, and $s^P = -\frac{P_y}{n P_x}$ denotes the social (paternalistic) MRS
- The latter term at the right is standard and the same as in Mirrlees (1971)
- the first term at the right is novel: could be called the first-best motive for taxation as it corrects the individual choices to correspond to the social ones

Interpretation

- if for instance the government is "puritanical" and would like to make the individuals work more than they would like to, the first term is negative and the MTR declines to induce more work effort
- if, on the other hand, work imposes negative externality on others via relative income effects, and the government accepts these relativity considerations, the first term serves to increase the MTR

Mixed taxation

- the individual optimization as before, with the modification that income can be spent on multiple commodity goods, $qx = y - T(y)$, and x is now a vector of different commodities and q represents their consumer prices
- the producer prices are denoted as p so that $q = p + t$. The government budget constraint is now $\int \{T[z(n)] + tx(q,n)\} f(n)dn = R$.
- the social welfare function is written directly as a function of commodities $\int P[x(q,z,v,n),z]f(n)dn$, with compensated demands (the dual approach is used in the optimization)

The optimal commodity tax rule

$$t \int x_q^c f dn = -\int \frac{1}{\lambda} P_x x_q^c f dn - \int \pi(n) x_n dn$$

where $\pi > 0$ is a parameter.

- the left-hand side is the compensated aggregated change in consumption. The reduction in demand is the greater, the higher is the effective tax burden on the good in question
- the second term at the right is standard. According to it, the consumption of those goods whose demand is positively related to ability should be discouraged by the tax system
- the first term, the corrective term, at the right is novel. In case of sin goods, P_x is negative and the term implies further discouragament of consumption of sinful goods.

Outline

Motivation

- Ample evidence that in addition to own absolute income, relative income concerns affect wellbeing. The evidence is surveyed by Clark, Frijters and Shields (JEL, 2006)
- the evidence mainly stems from subjective wellbeing surveys
- but there is also experimental evidence, and neuroeconomic evidence (see e.g. Dohmen et al. JPUBE 2010)
- some people also suggest that the rank in income, rather than relative income per se, could be important (Doyce et al. Psychological Science 2010)

Clark et al. (2008) conclude:

"Together, this suggests a utility function in which two-thirds of aggregate income has no effect because it is status-related, and thus disappears in a zero-sum game, and where 60 percent of the effect at the individual level evaporates within two years due to adaptation. Hence only around 13 percent of the initial individual effect will survive in the long run at the aggregate level."

Relativity and the optimal income tax

- we have already one formula that can be used to understand the implications of relativity: a corrective tax
- in principle, a difficult philosophical question whether envy or malice can be accepted as a basis for social welfare
- If it can, then the social welfare function could simply depend on the mean income / consumption in society
- Kanbur and Tuomala (2013) use weighted mean consumption $\mu = \int \omega(n)x(n)dn$, where ω gives the weights, which are subsequently set to unity

Kanbur-Tuomala (2013): The model

- Consumer utility: $u = U(x) + \psi(\mu) - V(y)$
- social welfare $\int W(u(n))f(n)dn$
- otherwise completely standard model
- they use the model to examine
  - earlier studies of 'ceteris paribus' style (i.e. for given values of other parameters). Here check how results change when numerical calculations used
  - computations also used to study the impacts on progressivity
  - and on how the impacts of relativity depend on underlying inequality

The tax rule in the general case:

- includes a corrective term, γ/λ,which equals the mean consumption evaluated as a government revenue
- the marginal tax rate tends to be high
  - when there are few income earners at that income level
  - when there are many income earners above that income level
  - when the social marginal value of depressed income is small

$$\frac{t}{1-t} = \frac{\gamma}{\lambda} + \underbrace{\left[\frac{1+E^u}{E^c}\right]}_{A_n} \underbrace{\left[\frac{[1-F(n)]}{nf(n)}\right]}_{B_n} \underbrace{\left[\frac{(1+\gamma/\lambda)U_x \int_n^\infty \left[1 - \frac{W' U_x^{(p)}}{\lambda(1+\gamma/\lambda)}\right] \frac{1}{U_x^{(p)}} f(p)\,dp}{(1-F(n))}\right]}_{C_n},$$

A special case

- Take Rawlsian social welfare function, Pareto distribution at the top and quasi-linear preferences ($U_x = 1$)
- then the asymptotic tax (the MTR for the highest incomes) can be written as

$$\frac{t}{1-t} = \phi + \left[1 + \frac{1}{\varepsilon}\right] \frac{1}{a} [1 + \phi]$$

- where ε is the elasticity of labor supply, a is the Pareto parameter and $\phi = \upsilon/(1-\upsilon)$ includes the relativity parameter, υ
- this means that the MTR is increasing in υ

Table 1

Rawlsian marginal tax rates (%) when people care about relative consumption

	$\epsilon = 1/3$	$\epsilon = 1/3$	$\epsilon = 1/2$	$\epsilon = 1/2$	$\epsilon = 1$	$\epsilon = 1$
Relative concern	$a = 2$	$a = 3$	$a = 2$	$a = 3$	$a = 2$	$a = 3$
$\nu = 0$	66.6	57	60	50	50	40
$\nu = 1/2$	83.3	78.6	80	75	75	70

Some general results

As the relativity concerns increase:

- marginal tax rates increase at all levels of income
- the drop off in the MTR for higher income levels is mitigated
- the redistribution measure, $\dfrac{dt(n)}{dF(n)}$, (the steepness of t) increases

TABLE 3

RAWLSIAN MARGINAL TAX RATES (%) WITH THE LOGNORMAL DISTRIBUTION

	$\epsilon = 1/3$ $\nu = 0$	$\epsilon = 1/3$ $\nu = 1/2$	$\epsilon = 1/3$ $\nu = 0$	$\epsilon = 1/3$ $\nu = 1/2$	$\epsilon = 1$ $\nu = 0$	$\epsilon = 1$ $\nu = 1/2$	$\epsilon = 1$ $\nu = 0$	$\epsilon = 1$ $\nu = 1/2$
$F(n)$	$\sigma = 0.7$	$\sigma = 0.7$	$\sigma = 0.39$	$\sigma = 0.39$	$\sigma = 0.7$	$\sigma = 0.7$	$\sigma = 0.39$	$\sigma = 0.39$
0.10	93.4	96.7	89.4	94.7	87.8	93.9	80.8	90.4
0.20	87.6	93.8	79.1	89.6	78.0	89.0	65.5	82.7
0.50	77.7	88.9	65.9	82.9	63.6	81.8	49.1	74.6
0.75	68.2	84.1	55.0	77.5	51.7	75.8	37.9	68.9
0.90	61.5	80.7	46.9	73.5	44.4	72.2	30.7	65.9
0.95	59.5	78.7	43.8	71.9	30.2	71.1	28.1	64.0
0.99	51.7	75.8	36.9	68.5	34.9	67.4	22.6	61.3
0.999	47.5	73.8	31.7	65.9	31.1	65.6	18.8	59.4

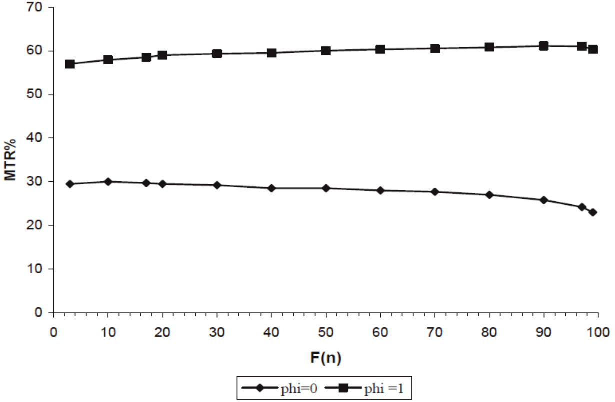


FIGURE 2

($\sigma = 0.5$) MARGINAL TAX RATE CURVES

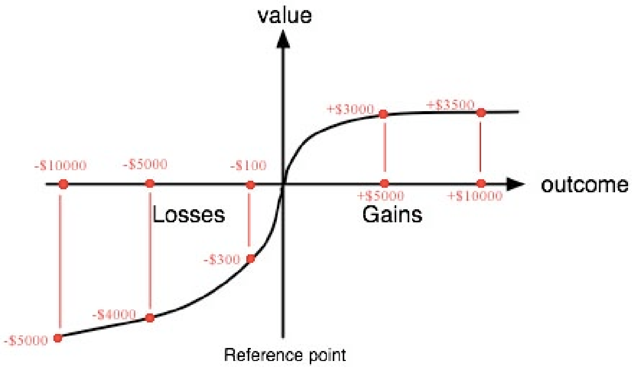
Relativity and pre-existing income inequality

- Few general results
- but with quasi-linear preferences, Rawlsian preferences, and Pareto distribution: The higher is inequality, the lower is the effect of relativity in raising the marginal tax rate
- a likely reason: the MTR is already high

Outline

Prospect Theory

- Developed by Kahneman and Tversky (E'metrica 1979)
- the main alternative for expected utility in modeling decision making under uncertainty
- three tenets
  - reference dependence: rather than levels, welfare depends on gains and losses relative to a reference point
  - loss aversion: losses matter more than gains of equal size
  - diminishing sensitivity: value function is concave for gains and convex for losses


value
outcome
+$3000
+$3500
-$10000
-$5000
-$100
+$5000
+$10000
Losses
Gains
-$300
-$4000
-$5000
Reference point

Evidence that loss aversion matters

- Boyce et al. (Psycholocial Science, 2013) use British and German data to test whether people are more sensitive to losses than gains in individual income
- De Neve et al. (LSE WP 2014) use Gallup and Eurobarometer data to do the same but for macro fluctuations
- both find evidence for loss aversion

Table 1. Results of Multilevel Regressions Showing the Effects of Income Changes on Life Satisfaction in the German Socio-Economic Panel (N = 163,000) and on General Psychological Disorder in the British Household Panel Survey (N = 110,079)

	Regression 1		Regression 2		Regression 3		Regression 4	
Predictor	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Subjective well-being at $T-1$ (β_1)	0.282	0.003**	0.227	0.003**	0.227	0.004**	0.219	0.004**
Change in log-transformed income from $T-1$ to T (β_2)	0.049	0.013**	0.053	0.012**	0.023	0.008**	0.014	0.008
Income-loss dummy (β_3)	–0.019	0.005**	–0.014	0.005**	0.006	0.006	0.001	0.006
Negative change in log-transformed income from $T-1$ to T (β_4)	0.081	0.019**	0.045	0.018*	–0.073	0.012**	–0.045	0.012**

Note: T = a particular year; $T-1$ = the year before T. Regressions 1 and 2 used the life-satisfaction variable from the German Socio-Economic Panel (GSOEP), and Regressions 3 and 4 used the General Health Questionnaire (12-item scale) from the British Household Panel Survey (BHPS). Both measures of well-being were standardized (M = 0, SD = 1) across the respective samples. No additional controls were included in Regressions 1 and 3. Regressions 2 and 4 included the following control variables: year dummy variables, age, gender, current education level (number of years in GSOEP, highest academic qualification in BHPS), current marital status, square root of current household size, current health, whether there were children present in the household, current disability status, current employment status, and changes from one year to the next (i.e., $T-1$ to T) in education level, marital status, the square root of household size, health (health satisfaction in GSOEP, subjective health status in BHPS), parental status, disability status, and employment status. We recoded any missing values for the control variables with sample-wide averages and included a dummy variable to indicate that a variable with a previously missing value had been recoded in this way.
*$p < .05$. **$p < .01$.

Table 4: Economic Growth and Subjective Well-Being

	Eurobarometer		Gallup World Poll		BRFSS	
	(1)	(2)	(3)	(4)	(5)	(6)
Economic Growth	0.017***		0.005***		0.005***	
	(0.005)		(0.002)		(0.001)	
Negative Growth		-0.041**		-0.008***		-0.008***
		(0.017)		0.003		(0.002)
Positive Growth		0.007		0.004*		0.001
		(0.004)		(0.002)		(0.003)
Loss aversion ratio		6		2		8
N_{macro}	493	493	815	815	1222	1222
N_{micro}	1,049,972	1,049,972	1,008,073	1,008,073	2,283,596	2,283,596
R^2	0.005	0.006	0.001	0.001	0.001	0.001

Clustered standard errors in parentheses

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

Dependent variable: life satisfaction (standardised)

Country and year dummies included in models 1-4, state and season dummies in 5-6

Economic growth variables are the absolute value of negative or positive growth, 0 otherwise

Taxation and income uncertainty:

- model proposed by Mirrlees (1974)
- in contrast to Mirrlees (1971), people are identical ex ante, but they face different shocks. There are income differences ex post.
- this is the same as the standard moral hazard framework
- the government only observes gross income, not effort and luck separately. Designs an optimal redistributive tax / insurance system taking into account individuals' incentive compatibility constraint

The model in the conventional case:

- the worker does not know what income (z) he will receive for each possible effort, y.
- Denote the distribution for z given y as $F(z,y)$ and its density as $f(z,y)$
- utility is $\int v(x)f(z,y)dz - y$
- the consumer is risk averse, $v' > 0, v'' < 0$
- the FOC is $\int v(x)f_y dz - 1 = 0$

The FOA:

- the individual optimization constraint serves as an incentive constraint in the government optimization
- this is the FOA (first-order approach)
- the FOA is valid if two conditions hold
  - MLRC: income is increasing stochastically in effort, $\partial g / \partial y > 0$, where $g = f_y / f$ is the likelihood ratio
  - CDFC (diminishing returns to effort), $F_{yy} > 0$

The optimal tax:

- the government max the expected value of individual utility, given its budget constraint, $\int(z-y)f(z,y)dz = 0$, (with the Lagrange multiplier λ) and the IC constraint (with the Lagrange multiplier α
- leads to the well-known formula that balances provision of insurance and incentives for effort (where δ is the coefficient of absolute risk aversion)

$$MTR = 1 - x' = 1 - \frac{\alpha v' g'}{\lambda \delta}$$

Enter Prospect Theory (Kanbur et al. 2008):

- individual welfare now depends on $\int e(c)f(z,y)dz - y$, where $c = x - \bar{x}$
- $e(c)$ has the following properties
  - $e' > 0$
  - $e'(-c) > e'(c)$ (loss aversion)
  - $e'' > 0$ for $c < 0$ and $e'' < 0$ for $c > 0$ (diminishing sensitivity)
- the formulation could allow for a kink at $c = 0$

Implications

- The optimal tax has the same structure with welfarist government, but with v' replaced by e'
- FOA is still valid for income above the reference point but not for income below the reference point
- the reason is that the individual is risk loving in the loss area
- FOA is perhaps valid in the non-wefarist case where social welfare depends on $\int v(x)f(z,y)dz - y$ and reference-dependence only enters via the incentive constraint

Implications for the optimal tax

- The optimal tax has three components
  - randomization between minimum consumption and the reference consumption in areas below the reference point
  - an area of full insurance near the reference consumption (since the individual is extremely risk averse here, e'' very large inside δ)
  - conventional partial insurance in the area above reference consumption

Results from simulations

- Again, numerical simulations needed to gain more understanding for the shape of the MTR
- In the paper, a computational example with CRRA type of utility and gamma distribution of income
- results for two different reference point, 30% of mean income ($\bar{x} = 0.1$) and 50% of mean income ($\bar{x} = 0.23$)
- According to the results,
  - the flat segment covers both minimum income and reference income
  - the MTR is increasing in reference income

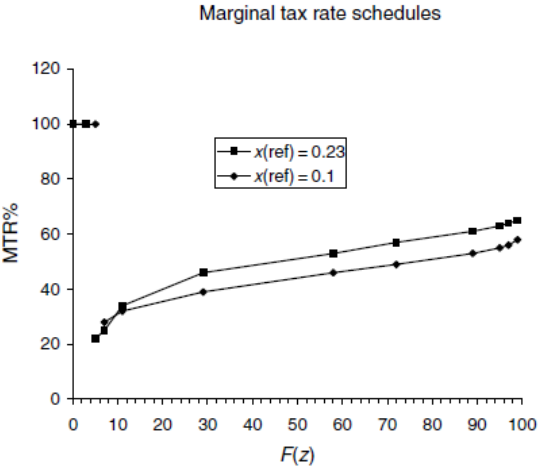


Fig. 2. Marginal tax rates with two different reference consumption levels

Outline

Background

- It has become clear that some (many? most?) of us have problems in self control
- This means that decisions made in the short run are not necessarily in line with our long-term interests
- Typical examples include smoking, drinking and consumption of junk food
- In these cases, it would be in our own interest to ask the government to intervene
- Typical modeling of the present bias via hyperbolic discounting

A model of sin taxes (O'Donoghue and Rabin 2006)

- The consumer's instantatenous utility in period t takes the form $u_t = v(x_t;\rho) - c(x_{t-1};\gamma) - z_t$
- here, x refers to the consumption of the unhealthy good, z to the consumption of a composite good and ρ and γ are parameters that capture taste heterogeneity
- the costs of harmful consumption are increasing in x but can be either convex or concave
- Following the idea of hyperbolic discounting, introduce an additional parameter, $\beta \leq 1$. This captures the short-sightedness of the consumer
- β is allowed to vary across consumers. If harmful consumption is taxed, leads to redistribution between consumers

Chosen consumption versus optimal consumption

- the person effectively faces a series of independent decisions -> drop the time index
- The consumer makes decisions based on $u^* = v(x;\rho) - \beta c(x;\gamma) - z$
- whereas his optimal long-term choice should depend on $u^{**} = v(x;\rho) - c(x;\gamma) - z$
- in the absence of any intervention (in the form of taxes), the chosen x^* is greater than the optimal x^{**} for people with $\beta < 1$.

Government problem

- the government levies a tax on the consumption of x and returns the tax revenue using a lump-sum transfer b
- the government is utilitarian (gives equal weight to all) and maximises $\sum u^{**}$
- this would otherwise be a fairly standard many-person Ramsey problem, but it is made more complicated by the inclusion of taste heterogeneity
- the first result is that when some consumers have $\beta < 1$, the optimal tax is positive (whereas it would be zero in the absence of self-control problems)

Some further results

- Is the tax always the higher, the more prevalent self-control problems are?
  - Not necessarily. One also needs a condition that the consumers with low β also have a sufficiently large sensitivity to taxes (i.e. their ${}^{du}/_{dt}$ is large enough)
  - Their Proposition 2
- Can sin taxes be Pareto efficient?
  - according to the first intuition yes, since consumers with self-control issues could benefit, and those without benefit as they get a lump-sum transfer.
  - this is true in the model with no taste differences if those with low self control are again sufficiently responsive to tax changes
  - with taste differences, an additional requirement needs to be imposed

What about quantitative importance?

- For this end, the authors present a simulation example and vary the size and the prevalence of self-control problems
- the key here is going to be the severity of health costs (relative to the production costs of x).
- The back-of-the-envelope calculations by Gruber and Koszegi (2004, JPUBE) suggest that for cigarattes this ratio could be as high as 10.
- the authors were not aware of estimates for unhealthy food, but they also experiment with 2 (plausible for junk food)
- the results show that the magnitude of the tax can be sizable
- what about introducing income differences? While the relative consumption of sin goods among low SES groups is high, the health benefits are also higher for them (Harkanen et al. 2014, Food Policy). Regressivity should, therefore, not seen as an argument against sin taxes

Table 1
Optimal taxes for different populations

Health cost and elasticity	Proportion of population with:				r	t^* (%)	$\underline{t}$ (%)
	β=1	β=0.99	β=0.95	β=0.9			
γ=10 and ε_D=−0.5	1/2	1/2	0	0	0.18	5.15	5.00
	1/2	0	1/2	0	0.19	28.53	24.81
	1/2	0	0	1/2	0.19	63.71	48.48
	1/4	1/4	1/4	1/4	0.19	49.26	39.26
	1/2	1/4	1/8	1/8	0.18	29.26	23.20
γ=10 and ε_D=−1.0	1/2	1/2	0	0	0.09	5.28	4.99
	1/2	0	1/2	0	0.09	31.68	24.34
	1/2	0	0	1/2	0.10	72.72	45.83
	1/4	1/4	1/4	1/4	0.10	56.41	38.01
	1/2	1/4	1/8	1/8	0.09	37.58	24.55
γ=2 and ε_D=−0.5	1/2	1/2	0	0	0.67	1.01	1.00
	1/2	0	1/2	0	0.68	5.21	5.00
	1/2	0	0	1/2	0.69	10.82	9.97
	1/4	1/4	1/4	1/4	0.69	8.52	7.98
	1/2	1/4	1/8	1/8	0.68	4.65	4.36
γ=2 and ε_D=−1.0	1/2	1/2	0	0	0.33	1.02	1.00
	1/2	0	1/2	0	0.33	5.34	4.99
	1/2	0	0	1/2	0.34	11.31	9.93
	1/4	1/4	1/4	1/4	0.34	8.84	7.96
	1/2	1/4	1/8	1/8	0.34	4.90	4.43

Outline

Summary

- In many different ways, individual preferences and decision making often divert from the assumptions commonly made by economists
  - these deviations are systematic and they also matter quantitatively
- the government is often in a unique position to try to influence behavior. Taxes can be used to influence relative income 'rat races' or to direct consumption choices. They are heavily used to direct savings choices (via mandatory pension payments), an issue which we have not discussed
- But notice that this often calls for non-welfarist interventions
- clearly, Amartya Sen is a frontrunner in propising non-welfarism. Lately also Nick Stern has argued that economists are, because of their training, oversensitive towards paternalism

Some other concluding thoughts

- It is easy to agree with Diamond and Gruber who, in their answers to the NSF question about long-term research agendas, ephasized the seeking for correct normative models for behavioral economics.
  - Gruber: How much choice should be limited? Should one simply restrain the number of potential choices or provide a better choice framework?
- Perhaps advances possible in redistribution from the behavioral development economics angle. The fascinating new research by Mani et al. (2013, Science) suggests that poverty traps could be alleviated also indirectly via transfers, but should these be in-kind or in cash?
- Reference dependence and taxation an area that is understudied. How should optimal taxes vary over (individual and macro) income fluctuations? Back to Musgrave where fiscal policy was part of public economics.