The Humpty Dumpty Blues:
Disaggregation Bias in the Evaluation of Tax Systems

Edward J. McCaffery and Jonathan Baron

USC CLEO Research Paper No. C02-1
USC Law and Economics Research Paper No. 02-1

USC
THE LAW
SCHOOL

CLEO RESEARCH PAPER SERIES
LAW & ECONOMICS RESEARCH PAPER SERIES

Sponsored by the John M. Olin Foundation

University of Southern California Law School
Los Angeles, CA 90089-0071

The Humpty Dumpty Blues:

Disaggregation Bias in the Evaluation of Tax Systems

Edward J. McCaffery*
University of Southern California Law School
California Institute of Technology

Jonathan Baron†
University of Pennsylvania

January 28, 2002

* emccaffe@hss.caltech.edu.

† baron@psych.upenn.edu. Support was provided in part by NSF grant SES 9876469.

Abstract

Three experiments carried out on the World Wide Web assessed the consistency of attitudes toward various tax regimes that differed in their overall levels and degrees of tax rate graduation in the presence of framing manipulations. The regimes had two components: an income and a payroll tax. One frame involved aggregation. Subjects were asked either to design a single, global tax system or to vary one component of a tax system (payroll or income tax) with the other component held constant. The idea was to replicate the effects of income tax reform given a constant payroll tax system. Consistent with the experimental hypothesis — though not with "rational" decision making — subjects focused on the component they were asked to manipulate and did not respond fully to changes in the other part, across conditions, reflecting an underadjustment bias as well as a framing effect. The results are akin to Thaler's "mental account" model for personal financial behavior. A second manipulation involved a "metric" frame: whether putative tax burdens were given in dollars or percent terms. Once again consistent with the experimental hypothesis, subjects preferred higher rates of graduation when matters were stated in percent terms. The results point to the lability of public opinion about important questions of public finance.

Introduction

From a strictly rational point of view, and paraphrasing Gertrude Stein, a tax — at least when it has the same behavioral or price distorting properties as the other tax(es) in view — is a tax is a tax. A rational actor should care only about the total tax she pays, net of transfers. An individual concerned with evaluating the global distribution of tax burdens should aggregate all taxes having similar bases. Labels shouldn't matter.

That doesn't seem to be the way things work.

In the United States, the federal government relies on two principal sources of revenue. In the year 2000, total federal receipts were approximately 2 trillion dollars. Of this total, 950 billion, or approximately 48%, came from the individual income tax, while 650 billion, or some 33%, came from "social insurance and retirement receipts," a.k.a. payroll taxes. The next largest tax, the federal corporate income tax, accounted for less than 10% of federal revenues.[1] The individual income and payroll taxes both fall largely on individual wages. Payroll taxes systematically exempt returns to capital. Most Americans simply don't have many such returns to be taxed under the income tax, and those that do can fairly easily avoid them: the income tax is in essence a wage tax, too. (McCaffery (2000)). From a rational point of view, therefore, individuals should care about the *total* burden of payroll and income taxes, and should base their evaluation of tax systems accordingly.

But there is good reason to believe that it isn't so. The modern personal income tax was put in place in 1913; payroll taxes in 1935. But while the marginal rates under the income tax have often been cut, those under the payroll tax have *never* been. Indeed, alone among major U.S. taxes, the payroll tax has shown a monotonic increase in its levels, as

[1] Statistical Abstract, 120th edition, Table 534.

Figure 1 reveals.

Figure 1 tells a story very different from the history of marginal rates under the income tax. Because of the multiplicity of rate brackets and their changing domains, it is hard to construct a parallel figure showing the history of effective income tax rates. But a quick sketch of the highest marginal rate bracket conveys the idea: it reached a high of 94% during World War II, stayed at 90% throughout the 1950s, was cut to 70% under John F. Kennedy, then to 50% and later 28% under Ronald Reagan, and has trended up, to its current level of 39.6%, since, pending recently enacted marginal tax rate cuts (McCaffery 1994). Meantime, the overall level of the income tax — the total tax divided by gross domestic product — has remained remarkably stable, while the payroll tax keeps trending up. By the late 1990s, payroll taxes had risen to account for some 80-85% of income tax revenues and to be, for the overwhelming majority of Americans, the major tax (Mitrusi and Poterba, 2000). Yet the presidential campaign of 2000 featured a debate between Al Gore and George W. Bush about the appropriate level and type of *income* tax cuts, and Bush in fact delivered on his promise in his administration's first major piece of legislation.[2] Only by late 2001, in the face of a widening recession, was there even talk of a payroll tax cut — then in the form of a "holiday," a one-time, one month break from the tax (Stevenson 2001).

There are of course many reasons why payroll taxes might be favored over income taxes by a government that wants to raise money with the least apparent pain to taxpayers (McCaffery 1994). These include the facts that such taxes:

- are labeled "contributions";
- appear to be tied to social security and medicare benefits;

[2] H.R. 1836, The Economic Growth and Tax Relief Reconciliation Act of 2001.

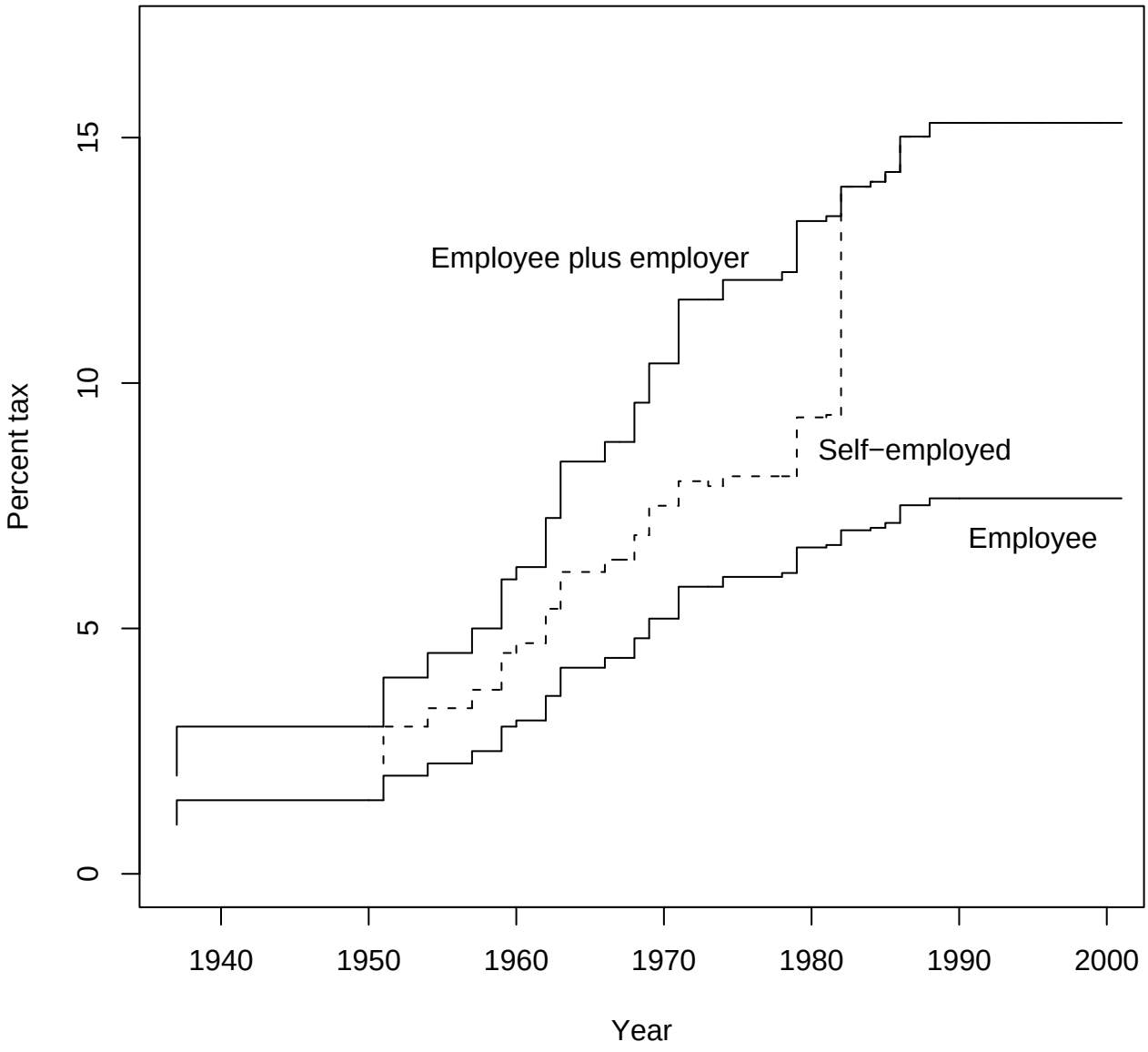


Figure 1: U.S. payroll tax over time

Source: U.S. Social Security Administration

`http://www.ssa.gov/OACT/COLA/taxRates.html`

- are matched by an employer "share," and so only appear at one-half their total level on employee pay stubs;
- require no onerous individual form completion;
- do not entail high marginal rates;
- are never manifest in a single lump sum due and payable at one time (say, April 15) to the ordinary taxpayer.

All of these explanations reflect some degree of irrationality or at least poor judgment. Thus the "contribution" label is gratuitous; the government could call income taxes "contributions," too, if it wanted. The linkage between payroll taxes and social security benefits is arbitrary; social security has long been on a "pay as you go" basis, not tethered to any particular form of government receipt. The actual incidence of the employer "share" rather patently falls on the employee, as an employee-specific cost to the employer. The relatively low marginal rates under the payroll as compared to the income tax reflect a tendency to confuse marginal and effective (or average) burdens. And so on.

In this paper, we explore another potential effect of the existence and persistence of payroll taxes: what we call a disaggregation bias. We hypothesize that, like All the King's Horses and All the King's Men, ordinary subjects will have a hard time putting back together taxes split into two. In theory, if a tax designer has full degrees of freedom over one tax system having the same base as another, fixed, component, she can replicate any ideal aggregate distribution. For example, if the fixed component has no "zero bracket" or exemption level, but rather begins (as the payroll tax in fact does) at a positive level for the first dollar, then the tax designer can create a *negative* tax bracket to offset the effect of

the fixed positive component. What we call the disaggregation bias suggests that subjects will not do this: they will have intuitions about what a "fair" tax ought to look like, and will fail to integrate existing tax systems into their appraisals of other, parallel systems. In other words, subjects will ignore or underadjust to the set aside portion of the whole. Thus, subjects will favor a total tax system that is less redistributive than they would otherwise prefer if nearly one-half of it can be removed and placed off-stage, as it were, in a relatively flat or regeressive tax. Using tightly controlled experimental settings, we were able to find this in fact to be the case.

Our predictions about disaggregation are related to theories of mental accounting proposed by Thaler (1980, 1985, 1999). Thaler shows many ways in which people keep separate mental accounts in their personal financial judgments and decision making, leading to anomalies when viewed from an integrated (aggregated) perspective. Folowing Thaler's lead, Camerer (2000) and Read, Loewenstein and Rabin (1999) have argued that many biases in decision making stem from (or at least require) an kind of isolation, in which people fail to view decisions in the context of a set of similar decisions. Our results extend these analyses to matters of public finance.

Experiments

We ran three experiments, each of which involved two different framing manipulations. We define a framing manipulation as a formal presentation where the relevant information is held constant across two or more modes of description or presentation. The experiments were carried out on the World Wide Web, using paid subjects and interactive questionnaires designed to simulate realistic issues facing real-world tax system designers. The

subjects were given detailed instructions that made all framing manipulations potentially transparent.

The primary framing manipulation involved the aggregation or separation of two taxes. In one experimental condition, we presented income and payroll taxes separately. In the other, we presented total taxes alone, where the total tax was simply the sum of the income and payroll taxes. Subjects compared tax policies to a standard point of comparison in these two forms of presentation, "segregated" and "total." We call this the "aggregation frame." We expected that the properties of the income or payroll tax — such as the rate of graduation — would be more salient in the segregated frame. Subjects would react only to the tax(es) they were asked to evaluate, ignoring the rest.

We did not impose a revenue-neutral constraint in these experiments; we allowed the total level of taxation to vary. We thus expected the aggregation frame to influence the perception of three different variables: the extent to which the total tax is graduated, the level or magnitude of the income tax, and the level or magnitude of the payroll tax.

A second framing variable was whether we presented tax burdens in terms of dollars or percent. We call this the "metric frame." Consistent with the prior literature, we expected people to favor more steeply graduated taxes when they were presented in the metric of percent (McCaffery and Baron, 2001; Roberts and Hite, 1994). When presented in the metric of dollars, it looks like people with high incomes pay more, even if the percent is constant. In other words, a dollars frame makes a flat-rate tax *look* progressive — generating a "progressivity illusion" (McCaffery and Baron, 2001) that seems to satisfy subjects' requirements for fairness in allocation of tax burdens.

Combined, we expected these frames to make subjects inconsistently favor a range of distributions of tax burdens, from relatively flat where subjects were asked to set income

tax rates in the segregated/dollars frame to relatively progressive where they were asked to set total taxes in the percent frame.

Experiment 1

In our first experiment, we presented people with one of the taxes and asked about the other, or about the total tax (combining the two). Our main hypothesis was that people would underadjust for the tax they were given: that the aggregation frame would have real effects. (See Chapman & Johnson, 1994, for a review of the literature on anchoring and underadjustment.) We hypothesized that the level of the total tax would be higher when the given tax was higher; that is, that subjects would underadjust or ignore the given tax. Similarly, the extent to which the given tax is graduated will not be fully taken into account in the judgments about the other tax or the total tax: a more progressive (regressive) fixed tax will lead to a more progressive (regressive) total one.

Method

Forty-six subjects completed a questionnaire on the World Wide Web, and were paid $3 each. Their ages ranged from 20 to 67 (median 33.5); 28% were male; 15% were students. The questionnaire began:

> In the U.S., most federal taxes come from two sources: income taxes and what we call payroll taxes. Payroll taxes are deducted from a paycheck; income taxes require taxpayers to complete annual forms, though most income tax is also withheld from paychecks. Historically, payroll taxes have been used to pay for certain benefits to individuals, such as government pensions, disability and unemployment insurance, and medical care for seniors. Income taxes have been

used for other functions of government, such as national defense, law enforcement, scientific research, transportation projects, and so on. The link between tax and government use of the money is arbitrary, however: payroll taxes can be used for other functions, and income ones for individual benefits.

We are interested in your attitudes towards the mix of taxes, payroll and income ones.

In the cases presented here, we ask your views about fair income, payroll, and total taxes. (Total taxes are the sum of income and payroll tax.) Sometimes we list the payroll taxes and ask what you think about income taxes; sometimes we do the reverse. Sometimes we list one tax and ask your opinion about the total tax.

In all cases, suppose that these taxes are for individuals who earn income, and all their income comes from salary. You will see the taxes for five different income levels ranging from $20,000 per year to $320,000.

If one of the taxes is 0 for everyone, that means that you should imagine that this tax has been abolished and all revenue is collected through the other tax.

In all cases, you can respond with negative amounts. Negative taxes amount to money returned to taxpayers by the government. If you think that one of the taxes is too high for some group, you can respond with a negative amount in the other tax for that group. You can also respond with total taxes that are less than one of the two taxes. Again, this would mean that you think that the other tax should be negative.

An example of one of the 32 screens is:

> Suppose that the payroll tax for in individual is as follows:
>
> For a person making $20,000, the payroll tax is $0
>
> For a person making $40,000, the payroll tax is $2,000
>
> For a person making $80,000, the payroll tax is $8,000
>
> For a person making $160,000, the payroll tax is $24,000
>
> For a person making $320,000, the payroll tax is $64,000
>
> (These numbers are included in red, below, for your convenience.)
>
> Fill in the amount you think is fair for each level of income for the income tax.
>
> Answer in dollars, without the $ sign:
>
> For a person making $20,000, a fair income tax is $__ ($0)
>
> For a person making $40,000, a fair income tax is $__ ($2,000)
>
> For a person making $80,000, a fair income tax is $__ ($8,000)
>
> For a person making $160,000, a fair income tax is $__ ($24,000)
>
> For a person making $320,000, a fair income tax is $__ ($64,000)

Note that these instructions drew attention to the equivalence of payroll and income taxes as sources of government revenue by informing subjects that the use of the funds was independent of their source, and abstracted from capital income by stating that all income comes from salary. Subjects were transparently told that "total taxes" were simply the sum of income and payroll ones. Subjects were also told that they could choose negative taxes, an important point of real-world tax policy design.

The design was 2 x 2 x 8. In half of the 32 cases, the subject responded in percent; in the other half of the cases, in dollars (as here). In each of these halves, the subject responded

either with the "other" tax (income tax if payroll given, as here, or vice versa), or with the total tax. Each of these four conditions (other vs. total, dollars vs. percent) used eight different sets of given rates, including one in which the given rate was set at 0. These varied in the total level of tax and in the extent of rate graduation. The specifics are shown in the table below.

Results

Table 1 shows the mean total tax rates that the subjects provided, in percent. The four columns under "Response" indicate whether the subject responded in dollars or in percent and whether as the "other tax" (e.g., payroll tax, if income tax was fixed) or as the total tax. The figures we list in the table are always the total tax in percent. To get the total tax in the cases where only the "other" tax was asked about, we simply added the given tax shown in the left side of the table to the responses. Likewise, we converted responses given in dollars to percent.

If the subjects were fully rational and consistent — if the two frames and the varying starting points did not matter — each cells in the table would be identical. They are not: the frames and starting points matter. The metric frame shows up as a difference between the Dollar columns and the Percent columns. The aggregation frame shows up as a difference between the Total and Other columns. The underadjustment bias shows up as a difference between rows.

Table 1 reveals that both frames and the starting points mattered. The overall level or magnitude of taxation was, as hypothesized, higher when responses were in terms of the other tax than when they were in terms of the total tax ($t_{45} = 7.42$, $p = .0000$), with the interesting exception of the case in which the given tax was set at 0: the aggregation

Table 1: Total tax rates in percent, Experiment 1.

Given rates on:					Response: Dollars		Percent		Mean
$20k	$40k	$80k	$160k	$320k	Total	Other	Total	Other	
Payroll tax given, Income tax response									
0	0	0	0	0	14.97	14.60	17.56	16.90	16.01
0	5	10	15	20	14.89	21.13	17.60	23.68	19.32
5	10	15	15	15	15.25	21.68	17.20	24.28	19.60
10	10	10	5	5	15.28	18.84	17.55	22.51	18.54
Income tax given, Payroll tax response									
0	0	0	0	0	15.66	13.24	17.02	16.15	15.52
0	5	10	15	20	15.44	20.35	17.13	22.01	18.73
0	8	16	24	32	16.00	24.13	17.79	27.36	21.32
10	10	10	10	10	14.75	18.71	16.92	22.11	18.12
				Mean:	15.28	19.09	17.35	21.87	

frame mattered. (Ten subjects did not respond differently at all when they were asked for total tax or the other tax. The results were essentially the same when these subjects were removed from the analysis.) The level of taxation was also higher when responses were in percent than in dollars ($t_{45} = 4.22$ $p = .0001$): the metric frame mattered.

Subjects were insufficiently responsive to changes in the given rates. They anchored on whatever rates they were given and did not adjust enough to make all the rows the same. In particular, total taxes were lower when the given rate was zero than when it was not (for the first and fifth row vs. the mean of the others, $t = 6.65$, $p = .0000$; and t was almost as high when the ten non-responders were removed).

Table 2 shows graduation, which we define as the slope of the percent tax as a function of income step, with each income step (i.e., each doubling of income) defined as one unit. This is logically independent of the level of taxation, shown in the prior table.

Once again, the frames mattered. Subjects could have — and to be consistent, "should" have — adjusted what they could to produce the same level of graduation in each instance. They did not. Graduation rates were higher for percent than for dollars ($t_{45} = 5.78$, $p = .0000$), showing the effect of the metric frame. Subjects were also, as hypothesized, insufficiently sensitive to the extent to which the given, "other," tax was graduated: the aggregation frame mattered. Subjects appeared to focus only on what they were asked to judge. The clearest comparison to illustrate this effect is between the 6th and 8th rows of the table, where the overall rate of the given income tax was the same, despite the difference in its graduation (and compare Table 1, where, even after adjustment, the level of taxation in these two conditions is about the same); yet subjects favored a far less graduated overall tax system when the given income tax was flat, in Row 8, then graduated, in Row 6 ($t_{45} = 5.77$, $p = 0.0000$).

Table 2: Graduation (tax change for each step) as a function of aggregation frame, Experiment 1.

Given rates on:					Response: Dollars		Percent		Mean
$20k	$40k	$80k	$160k	$320k	Total	Other	Total	Other	
Payroll tax given, Income tax response									
0	0	0	0	0	3.73	4.47	5.99	5.99	5.05
0	5	10	15	20	3.89	7.38	5.85	9.20	6.58
5	10	15	15	15	3.83	5.75	6.03	7.16	5.69
10	10	10	5	5	3.80	2.70	6.05	5.43	4.50
Income tax given, Payroll tax response									
0	0	0	0	0	4.46	3.74	6.11	5.61	4.98
0	5	10	15	20	4.26	6.53	5.85	8.33	6.24
0	8	16	24	32	4.30	9.20	5.76	10.95	7.55
10	10	10	10	10	3.76	3.31	5.67	5.68	4.60
				Mean:	4.00	5.39	5.91	7.30	

Experiment 2

In our second experiment, we tested to see how the metric and aggregation frames affected the comparison between pairs of proposed tax systems: rather than the quasi-mathematical task of filling in amounts, we asked subjects for qualitative opinions about several tax rates relative to a standard. The use of the standard provided a constant reference point, so that each subject did not need to change her criteria for fairness as the experiment progressed. Proposed tax rates differed in the overall magnitude of the income tax; the extent to which the income tax was graduated; whether the taxes were expressed in percent or dollars (the metric frame); and whether the income and payroll taxes were aggregated into one total tax or segregated (the aggregation frame). The graduation variable had three levels: a flat tax, a graduated tax, and a more steeply graduated tax with a negative tax for the lowest income. Each subject saw all conditions in a different random order.

We did not have hypotheses about how subjects would evaluate the graduation of taxes in general: prior literature has shown that subjects tend to fall into "camps" of roughly equal size supporting flat, moderately progressive, and highly progressive taxes (Hite & Roberts 1991; see also McCaffery & Baron 2001). Nor did we have a hypothesis over the preferred size or magnitude of the total tax burden, though we suspected a general tendency towards tax aversion would lead subjects to favor less over more (McCaffery & Baron 2001). Rather, our hypotheses centered on the interaction of the aggregation and metric frames with these other two variables. We expected subjects to be less tax averse when taxes are segregated. We also expected subjects to be more (less) supportive of overall graduation when the given tax was more (less) graduated: that is, we expected subjects to underadjust to what was given, and therefore make the overall system more like its arbitrarily fixed

component. We also suspected, as a feature of the aggregation frame and the salience of falling below zero, that a negative tax would draw more disapproval in the segregated frame. This is noteworthy in part because the initial impetus of the earned income tax credit under the income tax was in fact to offset the burden of payroll taxes among the working poor (McCaffery, 1999). Thus the "negative" graduated tax in this experiment is a realistic feature of actual contemporary tax law design.

Method

Fifty-four subjects completed a questionnaire on the World Wide Web, for $3 each. They ranged in age from 16 to 72 (median 35); 72% were female; and 13% were students. The introduction to the questionnaire read:

> This questionnaire has been designed to help us learn more about what people think of as a fair tax rate system.
>
> In the U.S., taxes are roughly of two types, income taxes and what we shall call payroll taxes. Payroll taxes are usually a flat percentage of earnings (up to some maximum). Income taxes are more graduated. The percent tax is higher for people with higher income.
>
> In the examples to follow, you will make judgments of the fairness of tax systems. In half of the cases, there is no payroll tax.
>
> Also, half of the time you will see taxes as percents, half of the time as dollars.
>
> Sometimes income taxes are negative. That means that the government pays the taxpayer, instead of the taxpayer paying the government. This is a way to help poor people.

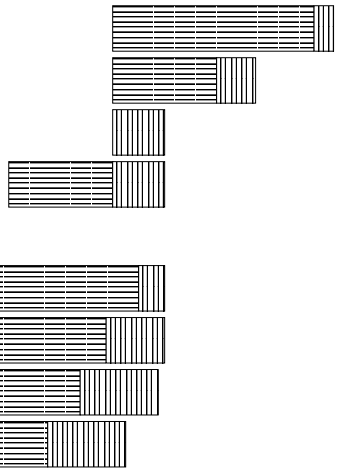
Suppose that these are for married couples who both earn income, and all their income comes from salary.

There are 48 screens. In each screen you see a proposal for taxes, and a standard. In each case you will compare the proposal to the standard for overall fairness. Please look carefully at all the differences between the proposal and the standard.

The instructions continued with illustrations of the dollar and percent formats.

The following shows an example of one of the conditions. Specifically, this is the condition with the negative tax, low scale of income taxes, segregated payroll and income, and percent. The figure illustrated the tax rates, but the rates were also presented numerically, at four different levels of income.

Income	**Income tax**	**Payroll tax**
$160,000	38.4%	3.7%
$80,000	19.2%	7.5%
$40,000	0,0%	10.5%
$20,000	−19.2%	10.5%
Standard for comparison		
$160,000	30.0%	5.6%
$80,000	24.0%	11.2%
$40,000	18.0%	15.0%
$20,000	12.0%	15.0%



The screen also included at the top a color key and scale (marking off percent) for the bars (which were colored in the original) and a question at the bottom, “How fair is the

proposal (top) compared to the standard?" The choices, on buttons, were: Very much more fair, Much more fair, More Fair, Less fair, Much less fair, Very much less fair.

In addition to the variables of dollars vs. percent and total vs. segregated, the trials varied in payroll tax level, income tax level, and graduation.

The income tax, as a proportion of income, was determined by the formula

$$M + G \cdot log_2(Y/20000)$$

where Y is income, M is the minimum tax, and G is the rate of graduation. For the standard case, M was 12% and G was 6%. Thus, the income tax on \$20,000, \$40,000, \$80,000, and \$160,000 was, respectively, 12%, 18%, 24%, and 30%. This was the "graduated" tax.

For the "flat-tax," M was 24% and G was 0.

For the tax we refer to as "negative," M was -24% (a 24% negative income tax on \$20,000) and G would be 24% (so that the income tax on \$40,000 was 0).

Before the rates were displayed, they were all multiplied either by 1.2 (high scale) or 0.8 (low), except for the standard case, which stayed the same.

The payroll tax was (as in real life) regressive. It was a fixed proportion P of income, up to some cutoff C, after which the tax did not increase with income at all. For the standard case, P was 15% and C was \$60,000. For other cases, the rate was 10.5%.

In the segregated condition, the two taxes appeared separately. In the total condition, the total tax increased by the amount of the payroll tax, and the payroll tax was 0.

In sum, the design was 3 x 2 x 2 x 2: three levels of graduation for the income tax; two levels of scale for the income tax; percent vs. dollars; and total vs. segregated presentation. Because of a programming error, each of the 24 conditions was repeated twice for each subject. The 48 cases were presented in a different random order for each subject.

Results

We normalized the 6 possible fairness responses around 0, with "very much less fair" set at negative 2.5, and "very much more fair" set at positive 2.5. As we suspected, though we were not explicitly testing for it, there was a general tax aversion: subjects thought that low taxes were fairer. The high scale had a mean rating of -0.58 and the low scale had a mean rating of 0.19 ($t_{53} = 9.39$, $p = .0000$, for the difference).

Consistent with prior literature, there was no significant linear trend on attitudes towards graduation (Hite & Roberts, 1991). Subjects preferred the middle level — a moderately graduated tax without a negative tax — to the extremes of a flat tax and a more steeply graduated tax with a negative tax component. The mean ratings were: flat tax -0.49, graduated 0.42, negative -0.51.

Turning to our primary hypotheses, once more we found that the metric frame mattered. Subjects thought that taxes were more fair when presented as dollars than as percent (-0.32 vs. -0.07, $t = 3.39$, $p = .0013$). There was also evidence for the disaggregation effect: subjects tended to prefer the segregated condition, even though it had two taxes rather than one (-0.11 vs. -0.28, $t = 1.99$, $p = .0523$).

The interaction of Percent and Graduation was significant ($t_{53} = 7.19$, $p = .0000$). Table 3 shows the mean ratings. The flat tax seems much less fair when shown as a percent than when shown as dollars.

The interaction of Segregation and Graduation was also significant ($t = 3.45$, $p = .0011$, for negative vs. flat; when graduation was coded as a categorical variable, $F_{2,2533} = 4.34$, $p = .0131$). Table 4 shows the mean ratings. Although the segregated condition yielded higher fairness ratings overall, the aggregation frame changed the rank ordering of fairness.

Table 3: Mean rating as a function of graduation and metric frame, Experiment 2.

	Graduation		
	Negative	Graduated	Flat
Dollars	−0.57	0.36	−0.01
Percent	−0.42	0.48	−1.01

The negative tax was rated lowest in the segregated condition, below the flat tax, but the flat tax was rated lowest in the integrated condition.

Table 4: Mean rating as a function of graduation and aggregation frame, Experiment 2.

	Graduation		
	Negative	Graduated	Flat
Integrated	−0.50	0.35	−0.69
Segregated	−0.49	0.48	−0.33

Finally, the scale of income taxes interacted with both the metric and the aggregation frames. Higher taxes seemed fairer in the percent presentation ($t = 3.84$, $p = .0003$) and, to some extent, in the segregated presentation ($t = 1.84$, $p = .0721$).

In sum, the two framing manipulations affected judgments of the fairness of taxes. Presenting information as percent made taxes seem fairer overall, made the flat tax seem less fair, and made subjects less bothered by high levels of taxes in general. Presenting information about two taxes in a segregated form made people think taxes were somewhat fairer on the whole. But segregation also made a negative tax seem less fair — and thus,

implicitly, a positive tax on the lowest income levels more fair — and improved the relative standing of a flat tax.

Experiment 3

Our final experiment, while continuing our use of the two basic frames, metric and aggegration, asked the subjects to manipulate properties of taxes in a way that held constant the total tax burden but allowed for changing the mix and/or level of graduation between or within the tax(es). Once again we had a hypothesis about the effects of the metric frame: when taxes are presented in a percent frame, people prefer more graduation. In the dollar frame, a flat percentage tax already appears to be graduated because the tax goes up with income.

To invoke the aggregation frame, we varied what subjects were asked about as opposed to what they were not asked about. In some items, we asked subjects to adjust the payroll tax, and, in others, the income tax. The total tax was simply the sum of the two, and subjects could completely compensate a change in one with a change in the other in our experiments. Yet we hypothesized once again that subjects would focus on the tax they were asked to adjust and under-compensate for changes in the other tax.

This third experiment thus gave a good look at the anchoring and underadjustment bias. We provided a starting point and asked subjects to make adjustments. They were free to arrive at the same endpoint in many different cases. A question was whether they would adjust enough to do so, or if the "anchor" provided by the starting point would shape the endpoint, as we hypothesized.

Method

Eighty-eight subjects completed a questionnaire on the World Wide Web, for $3 each. They ranged in age from 18 to 57 (median 36); 77% were female; and 16% were students. The questionnaire began as follows:

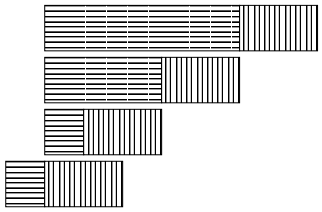
> This questionnaire has been designed to help us learn more about what you think of as a fair tax system.
>
> In the United States, most taxes come from two sources: income taxes and what we call payroll taxes. Payroll taxes are currently a flat percentage of earnings up to some maximum. Income taxes are currently graduated: the percent of income paid in taxes is higher for people with higher income. In this study, though, we ask you to imagine that payroll taxes can be more like income taxes some times.
>
> In the examples to follow, we ask you to adjust taxes to make them what you consider fair. The screens will differ. Each screen will show you the taxes paid by a household with incomes of $160,000, $80,000, $40,000, and $20,000. You can suppose that each household consists of a married couple where both spouses earn income, and that all of their income comes from wages subject to a possible payroll tax. In some of the cases, however, there will be no payroll tax, just an income tax. Also, half of the time you will see taxes given as percentages, and half of the time as dollars.
>
> We will ask you to adjust either the level or the "graduation" of one or the other tax. When we ask you about the level, this means that increasing that tax will decrease the other tax; decreasing the identified tax will increase the other.

(If you think it doesn't matter, then leave each one where it starts, but you must try to see what it looks like if you change it.) When we ask you about the graduation, "more" means that the chosen tax will become more progressive — that is, relatively more of that tax will come from the higher income households — and "less" means that the tax will be "flatter," or less progressive.

Finally, note that sometimes taxes are negative. That means that the government pays the taxpayer, instead of the taxpayer paying the government.

The introduction ended with illustrations of the type of display used. A typical screen was as follows:

Income	**Income tax**	**Payroll tax**
$160,000	37.5%	15.0%
$80,000	22.5%	15.0%
$40,000	7.5%	15.0%
$20,000	−7.5%	15.0%



As before, this display was preceded by a key and scale, and following by a set of response buttons, labeled with what the subject was to adjust and with the amount and direction of possible adjustments, for example:

You adjust:

Amount of payroll-tax graduation:

Much less	Less	More	Much more

Each subject saw 32 screens, in random order. The design was 2 x 2 x 2 x 4. One manipulation was percent vs. dollars. The screen shown here is in the percent condition. Two other manipulations were the starting points for the graduation of the income tax (.05

or .15, see below) and the starting points for graduation of the payroll tax (.0 or .1, see below). The four conditions were:

1. Integrated: there was only an income tax (and what would otherwise be the payroll tax was added to the income tax).
2. Adjustment of the income tax graduation.
3. Adjustment of the payroll tax graduation.
4. Adjustment of the relative size of income tax relative to payroll tax. (We do not report the results of this manipulation here. They did not bear on our major hypotheses.)

The rates presented assumed that a doubling of income was associated with a constant change in the tax rate in percent, for both income and payroll taxes. Both taxes were determined by the following formula:

$$\text{Rate} = \text{Mid} + \text{Grad} \cdot (log_2(\text{Income}/\$20{,}000) - 1.5)$$

Rate is the rate in percent, which was converted to dollars in the dollars condition. Mid was the rate at the midpoint of the income range (between $40,000 and $80,000). The tax rate pivoted around this point as graduation changed. Grad was the change in percent for each level shown ($20,000 to $40,000, etc.). Grad was multiplied by the number of levels from the midpoint: -1.5 for $20,000, -.5 for $40,000, and so on.

The starting points for Mid were .15 for both taxes. The starting points for the graduation rates were .05 (5%) and .15 for the income tax and 0 and .10 for the payroll tax. (Notice that, when Grad was .15, the tax on $20,000 was $.15 - 1.5 \cdot .15$ or –.075.) In the conditions in which graduation was adjusted, the “Much more” and “Much less” buttons

changed Grad by .10 up or down. The "More" and "Less" buttons changed Grad by .02 up or down.

In the last condition, when the subject adjusted the relative size of the income tax, "Much more" raised the income tax by .10 and lowered the payroll tax by the same amount, so that the total was constant. Our hypothesis here was that subjects who (for example) oppose graduation will be more willing to lower the income tax when it is more graduated, even though the graduation itself is unaffected by their adjustment.

Subjects had to make at least one adjustment in each screen before going on. They could undo that change if they were happy with the starting point.

Results

We first consider the results for the graduation rates (Grad). When the subject adjusted the graduation rates, the mean ending rate was .107 for the dollar condition and .134 for the percent condition. As hypothesized, graduation was higher for the percent condition ($t_{87} = 7.74$, $p = 0.0000$). This difference was essentially the same for all conditions, so we collapse over percent and dollars henceforth.

Table 5 shows the rates for the conditions in which the subject adjusted the graduation rate, collapsing over percent and dollar frames. The first two columns show the starting values. The next two columns show what the subject adjusted. The final column shows the total combined rate of graduation, considering both taxes after subject adjustments. If the subjects had a fixed view about fairness in graduation, the numbers in the far right column should all be the same.

The first three rows are the integrated condition, in which what would otherwise be the payroll taxes were added to the income taxes. This resulted in three conditions rather

than four, since the condition with graduation rates of .00 (payroll) and .15 (income) was indistinguishable from the condition with rates of .10 and .05.

Table 5: Graduation responses as a function of aggregation frame and starting point.

	Starting:		After adjustment:			
	Payroll	Income	Payroll	Income	Total	
Integrated	.00	.05		.074	.074	
	.00	.15		.107	.107	
	.00	.25		.145	.145	
Adjust income	.00	.05		.061	.061	
	.00	.15		.107	.107	A
	.10	.05		.036	.136	B
	.10	.15		.082	.182	
Adjust payroll	.00	.05	.018		.068	
	.00	.15	.001		.151	C
	.10	.05	.067		.117	D
	.10	.15	.042		.192	

It is apparent from Table 5 that responses changed as a function of the starting point. As the starting point went up, the final response went up too. All effects of the starting point were highly significant. The same results were found for subjects who made many adjustments before stopping, as opposed to subjects who made few adjustments. (Note, though, that it was possible to make large adjustments with very few button clicks.)

Except for the lowest rates of Grad, most subjects adjusted downward. The starting

points we chose were, on the whole, too graduated for our subjects. The mean downward revision was .030, and 73% of the subjects made downward revisions on the average.

Given that most adjustments were downward, our hypothesis predicts greater downward adjustments when the subject is adjusting the more graduated of the two taxes. This is because subjects focus on what they are asked to adjust. Consider the two rows labeled A and B. In both, the subject is adjusting the income tax. In condition A, the income tax has a .15 rate of graduation, compared to .05 in B. And so, consistent with our hypothesis, the downward adjustment in A is far greater than that in B, and we end up with a less graduated overall tax — or, conversely, a tax more like the fixed component, here the payroll tax. This effect was significant ($t_{87} = 4.94$, $p = 0.0000$). Likewise, the rows C and D show the same effect, where it is the payroll tax being adjusted: a greater downward adjustment in the condition where the payroll tax is more graduated (D). ($t_{87} = 5.19$, $p = 0.0000$). Of course, A and C, and B and D, have identical starting points, and the interaction (B–A–D+C) was also significant ($t_{87} = 5.31$, $p = 0.0000$). In sum, subjects paid more attention to what they were adjusting, even though the net effect of their adjustment was apparent visually. In real life, this effect is very likely even larger because the net effects are often "off the screen."

Conclusion

In public finance as in life, it seems, it matters what one is looking at. In particular, a tax system split in two is not easily put back together again, and very different overall patterns of taxation can emerge because of this.

In particular, we found an aggregation framing effect in all three experiments. When people are asked to adjust a tax so that they approve of the total tax system (Experiments

1 and 3), they do not use the tax they can change to compensate for what they see as undesirable features of some other tax. Subjects seem instead simply to focus on what they are asked to judge. Once again, this is akin to Thaler's "mental accounts" model, whereby subjects treat different "pots" or "pools" of money differently, notwithstanding money's inherent fungibility, and also akin to the idea of "choice bracketing" as described by Read et al. (1999).

Experiment 2 showed further, perhaps unintended, aspects of tax disaggregation. When subjects evaluate taxes split in two, they are less bothered by high taxes overall. But any negative tax stands out as unfair, even when it simply offsets a positive tax burden in another tax — and even when subjects favor an overall low or zero level of taxation on the lowest income class. So unfair does this negative tax seem (so little are subjects able to integrate tax systems), that a flat tax, not especially attractive in an integrated setting, becomes preferable.

In real life the disaggregation bias could lead to various impediments to reform. For example, in the U.S., it is analytically possible to use more progression in the income tax to compensate for less progression in the payroll tax. But public political arguments about the income tax typically present the tax as if it had to be fair on its own, viewed independently of whatever other taxes might exist. Meanwhile, the relatively regressive payroll tax remains unchanged.

A second result is the metric framing effect. People prefer more steeply graduated taxes when tax rates are expressed in percent rather than in dollar terms. This is not surprising, and, indeed, we have found it before (McCaffery & Baron, 2001).

We might imagine still other frames, such as varying the denominator against which tax burdens are measured: as a percent of disposable income, consumption, and so on. We

would expect that public judgments about fairness in tax could be manipulated by choice of frame. Further experiments might ascertain what happens when people are confronted with the inconsistencies that arise from metric framing effects. In the meantime, any particular way of eliciting public judgments should not be accepted as the last word on what the public wants.

Finally, we found an underadjustment effect. The starting point affects the end point, when people adjust taxes to make up an ideal tax. Again, this is not surprising. And, again, it causes us to be concerned about the lability of public judgments. In this case, the effect would lead to smaller reforms than might otherwise be acceptable, because citizens can only consider one tax at a time.

References

Camerer, C. F. (2000). Prospect theory in the wild: Evidence from the field. In D. Kahneman & A. Tversky (Eds.), *Choices, values and frames*, pp. 288–300. New York: Cambridge University Press.

Chapman, G. B., & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making, 7*, 223–242.

Hite, P.A. and Roberts, M.L. (1991). An experimental investigation of taxpayer judgements on rate structure in the individual income tax system. *Journal of the American Taxation Association, 13*, 47-63.

McCaffery, E.J. (1994). Cognitive theory and tax. *UCLA Law Review, 41*, 1861-1947.

McCaffery, E.J. (2000). A Voluntary Tax? revisited. *National Tax Association Proceedings*, 268-74.

McCaffery, E.J., and Baron, J. (2001). Framing the Family: Normative Evaluation of Tax Policies Involving Household Composition. *USC Olin Research Paper No. 00-18.*

Mitrusi, A. and Poterba J. (2000). The Distribution of Payroll and Income Tax Burdens, 1979-1999. *National Tax Journal, 53* 765-94.

Read, D., Loewenstein, G., & Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty, 19*, 171–197.

Roberts, M.L. & Hite, P.A. (1994). Progressive Taxation, Fairness, and Compliance. *Law & Policy 16*, 27-48.

Stevenson, R.W. (2001). 38 Billion Tax Hiatus Proposed. *The New York Times*, November 28, 2001, Section B, page 8, col. 5.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization, 1*, 39–60.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science, 4*, 199–214.

Thaler, R. (1999). Mental accounting matters. *Journal of Behavioral Decision Making, 12*, 183–206.