The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

CESifo Working Papers

www.cesifo.org/wp

Complex Tax Incentives – An Experimental Investigation

Johannes Abeler
Simon Jäger

CESIFO WORKING PAPER NO. 4231
CATEGORY 13: BEHAVIOURAL ECONOMICS
MAY 2013

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 4231*

Complex Tax Incentives – An Experimental Investigation

Abstract

How does tax complexity affect people's reaction to tax changes? To answer this question, we conduct an experiment in which subjects work for a piece rate and face taxes. One treatment features a simple, the other a complex tax system. The payoff-maximizing effort level and the incentives around this optimum are, however, identical across treatments. We introduce the same sequence of additional taxes in both treatments. Subjects in the complex treatment underreact to new taxes; some ignore new taxes entirely. Contrary to predictions from models of rational inattention, subjects are equally likely to ignore large or small incentive changes.

JEL-Code: C910, D030, H310, J220.

Keywords: complexity, taxation, attention, salience, laboratory experiment.

Johannes Abeler
Department of Economics
University of Oxford
Manor Road
OX1 3UQ Oxford
United Kingdom
johannes.abeler@economics.ox.ac.uk

Simon Jäger
Department of Economics
Harvard University
1805 Cambridge Street
Cambridge, MA 02138
USA
jaeger@fas.harvard.edu

April 25, 2013
Financial support from the ESRC under grant RES-194-23-0013 is gratefully acknowledged. We thank Steffen Altmann, Raj Chetty, Richard Disney, Armin Falk, Markus Fels, Xavier Gabaix, Peter Ganong, Edward Glaeser, Tanjim Hossain, Andrea Isoni, Michael Kosfeld, Erzo Luttmer, David Laibson, Sendhil Mullainathan, Andrew Oswald, Alex Peysakhovich, Benjamin Schoefer, Andrew Schotter, Andrei Shleifer, Cass Sunstein, Roberto Weber, Matthias Wibral, and Christian Zehnder for helpful discussions. Valuable comments were also received from numerous seminar and conference participants. Francesco Fallucchi provided outstanding research assistance.

1 Introduction

We study how complexity influences choices. In particular, we analyze how the complexity of the economic decision environment influences the reaction to subsequent changes in incentives. We are motivated by the observation that existing tax and benefit systems as well as many other incentive systems and price schedules usually feature highly complex, non-linear schedules with kinks, thresholds and various exemptions. There is growing evidence that people are not able to react optimally to complex (tax) incentives: taxpayers do not bunch at kinks to the extent that would be expected if marginal incentives were fully understood and labor supply could be freely adjusted (Saez, 2010); evidence by Brown, Kapteyn, Luttmer, and Mitchell (2012) suggests that complexity thwarts individuals' ability to value annuities such as Social Security benefits; Chetty and Saez (2013) document that EITC-eligible individuals adjust their earnings if they receive personalized advice on their tax incentives.[1] Yet, there is no direct evidence on the causal effect of tax complexity. How would behavior differ in a counterfactual simple tax system that featured similar economic incentives but less complexity?

An ideal data set to study this question would contain observations of behavior in tax systems that differ only in the level of complexity. For lack of such exogenous variation in complexity in real-world tax systems we conduct a tightly controlled real-effort laboratory experiment. In the experiment, we can vary the complexity of the decision environment and introduce additional tax rules; we are therefore able to draw causal inferences about the impact of complexity on decision making.

In our experiment, subjects work on a real-effort task in a setting that mimics a progressive income tax system: subjects receive a piece rate for each unit of output produced; they also have to pay a number of taxes and receive a number of subsidies. The design features two

[1] Chetty and Saez (2013) document small effects of information provision but find significant effects on the earnings of individuals who received advice on tax incentives from tax preparers who complied with the experimental design. The interpretation of the non-bunching results in Saez (2010) is corroborated by survey evidence (Fujii and Hawley, 1988) documenting differences between self-reported and computed marginal tax rates for a substantial fraction of surveyed US taxpayers. There is also growing evidence suggesting that taxpayers react to average rather than marginal tax rates (de Bartolome, 1995; Liebman and Zeckhauser, 2005; Feldman and Katuscak, 2009); Ito (2010) documents similar results for consumer responses to a non-linear electricity pricing scheme. Chetty, Friedman, and Saez (forthcoming) demonstrate substantial heterogeneity in the knowledge of the EITC program across regions in the United States.

between-subjects treatments, each subjects facing three rounds. In the first round of the simple treatment (ST), the tax system features very few, simple rules that determine incentives. The complex treatment (CT) implements a tax system that features economic incentives that are almost identical to the ones in ST. In particular, the payoff-maximizing effort level and the payoffs around this level are identical. However, there are many more tax rules in CT—some of which partially cancel each other—leading to a tax system with a higher degree of complexity. CT mimics the large number of distinct rules that characterize real tax systems. In contrast, ST is designed to come closer to implementing the assumptions economists typically make when modelling behavioral responses to tax incentives, for instance, that individuals understand the incentives they face and know their marginal tax rate.

In the second and third round, the simple or complex tax system of the first round is again in place with one additional tax or subsidy introduced in both treatments, changing the payoff maximizing number of units of output. These additional rules are identical across the two treatments. The main focus of our study is on how subjects react to these new incentives. A priori, it is not clear whether people will react more or less strongly to new tax rules when the pre-existing incentives are more complex. If higher background complexity increases the salience of the new tax rules relative to the set of existing tax rules, subjects in CT will react more strongly to the new tax rules. In contrast, if increased background complexity takes up limited cognitive resources, subjects in CT will underreact to the new tax rules. Obviously, if subjects are fully rational, complexity will not affect behavior.

How does complexity affect behavior in our experiment? We first document that subjects in CT choose the payoff-maximizing output level less often and, more generally, choose output levels further away from this payoff-maximizing number. As a consequence, they earn about 23 percent less than subjects in ST. This shows that subjects are indeed influenced by the complexity of the decision environment. Our main result is that subjects in the complex treatment under-react to the newly introduced tax rules and do not adjust their effort provision strongly enough towards the new payoff-maximizing choice. This implies that an increase in the complexity of the decision environment lowers price elasticities. Some subjects in CT ignore the new tax rule completely and stick with their previous choice. The fraction of subjects in CT who do not adjust their decision from round to round is 8.5 to 11 percentage points higher than the corresponding fraction in ST. These subjects drive most of the underreaction in CT. The non-reacting subjects decide much faster than other subjects, suggesting that they indeed ignore the new tax rule. We also document substantial heterogeneity across subjects in the effect of complexity. Subjects who choose output levels further away from the payoff maximum in the first round are more likely to ignore newly introduced tax rules in the subsequent rounds of the experiment. Finally, by randomizing the order in which subjects face the additional tax rules (so that some subjects initially face a smaller or larger tax change) we can test whether smaller tax changes are more likely to be ignored. We cannot reject that subjects are equally likely to ignore small and large tax changes.

We discuss a conceptual framework that can help to organize our results.[2] We posit that individuals can only pay attention to a certain amount of information. We operationalize this by assuming that a subject can only take a limited number of tax rules into consideration (see, e.g., Eliaz and Spiegler, 2011, and Gabaix, 2011, for a motivation of this assumption). This simple framework predicts that choices are usually not payoff-maximizing when the number of rules a subject can take into consideration is smaller than the number of applicable tax rules. Since the complex treatment features more tax rules in the first place, the model predicts that subjects in CT are less likely to take the newly introduced rules into account. This is in line with our finding that many subjects in CT repeat their previous choice when new incentives are introduced. Overall, this framework matches the main results of the experiment well. Our results are inconsistent with models in which new tax rules become more salient relative to the existing set of rules as the complexity of the pre-existing rules increases. Moreover, the fact that the fraction of subjects ignoring an additional rule does not depend on the size of the tax change is in contrast to most models of rational inattention which predict that information is

[2] Recent theoretical papers on complexity and the closely related topics of inattention, salience, and bounded rationality include Sims (2003); Reis (2006); Kleven and Kopczuk (2011); Eliaz and Spiegler (2011); Gabaix (2011); Köszegi and Szeidl (forthcoming); Bordalo, Gennaioli, and Shleifer (2012); Persson (2012); Dahremöller and Fels (2012) and Ortoleva (forthcoming). For laboratory experiments on these issues see, e.g., Gabaix, Laibson, Moloche, and Weinberg (2006); Cheremukhin, Popova, and Tutino (2011); Fochmann and Weimann (2011); Kalayci and Serra-Garcia (2012); Crosetto and Gaudeul (2012) and Sitzia, Zheng, and Zizzo (2012).

more likely to be incorporated in decision-making if it is more costly to be ignored (e.g., Sims, 2003; Gabaix, 2011; Chetty, 2012).

A key difference between our experiment and the existing literature is that we vary the complexity of the whole tax system that subjects face; in contrast, the new tax rules that are introduced each round are identical in ST and CT and, taken in isolation, are simple and salient. Previous studies varied the salience of one tax rule or one part of a price schedule to see how salience and complexity influence decisions. Chetty, Looney, and Kroft (2009), for example, show in several ways that consumers under-react to changes in non-salient taxes (see also Goldin and Homonoff, 2013, and Feldman and Ruffle, 2012). Finkelstein (2009) shows that automating toll collection—which lowers the salience of the toll—leads to a reduction in the toll elasticity of driving.[3] Our design allows us to study how the reaction of subjects to a new tax rule depends on the complexity of the tax system in which the new rule is embedded. As our results indicate, higher levels of background complexity mute the reaction to new tax rules and lead to a higher prevalence of choice inertia. Our findings complement studies of investor behavior in DellaVigna and Pollet (2009) and Hirshleifer, Lim, and Teoh (2009), which document underreactions to earnings announcements when investors face higher information load.[4] On a more general level, our findings support the view that complexity of the decision environment is an important catalyst of behavioral anomalies such as the status-quo bias (Samuelson and Zeckhauser, 1988; Kahneman, Knetsch, and Thaler, 1991; Fleming, Thomas, and Dolan, 2010).

In a world in which individuals can react optimally to complex incentives, tax complexity gives the social planner more tax instruments and thus more degrees of freedom to maximize social welfare. In this spirit, the social planner in a Mirrleesian world (Mirrlees, 1971) or in the models used in the new dynamic public finance literature (e.g., Golosov, Tsyvinski, and Werning, 2007) can set a highly non-linear income tax schedule. Our experiment shows that introducing more complexity comes at a cost: facing a complex tax schedule, fewer

[3]For related papers studying non-salient aspects of consumer products, e.g., shipping costs on eBay, see Gabaix and Laibson (2006) and Brown, Hossain, and Morgan (2010).

[4]Relatedly, Carlin, Kogan, and Lowery (forthcoming) conduct a laboratory experiment in which increasing asset complexity leads to less trade volume in an experimental asset market.

subjects choose the payoff-maximizing level of output; moreover, the complex tax schedule mutes subjects' response to subsequent changes in incentives. Our results thus suggest that there are bounds on the level of complexity that is desirable in income tax and transfer systems. Sunstein (2011), for example, discusses several examples of recent US governmental regulation and describes how simplification can substantially improve regulatory outcomes. Our findings can also help to inform how complexity can be explicitly used as a tool by policy-makers. In the domain of taxation, a key take-away of our study is that high levels of complexity of an existing tax system reduce the responsiveness to new (tax) policies.[5] This is desirable if the goal is to shroud the economic impact of the tax, for instance if the efficiency costs of taxing a good are large due to a high price elasticity. But it could be harmful in the case of a tax on a socially undesirable activity, e.g., polluting. The fact that some subjects are particularly strongly affected by complexity in our experiment suggests that obfuscation (or simplification) could be used to design screening devices, e.g., to target welfare programs (Congdon, Kling, and Mullainathan, 2011).

The rest of the paper is organized as follows: the next section describes the design of the experiment. We present results in the third section. We discuss a framework to organize our results in the fourth section. The last section concludes.

2 Experimental Setup

2.1 Overview

In our experiment, subjects work on a real-effort task in a setting that mimics a progressive tax system. Subjects have to move sliders on the screen and get a piece rate for each correctly positioned slider; they also have to pay taxes and receive subsidies depending on the total number of sliders they position. We implement a real effort task that was developed by Gill and Prowse (2011, 2012). During the task, subjects see a single screen showing 48 sliders (see Figure 1). Subjects can adjust the position of each slider in a range from 0 to 100. Output

[5]See also Kleven and Kopczuk (2011) and Goldin (2012) who propose several intriguing ways in which complexity and salience can be used as policy instruments to achieve social goals.

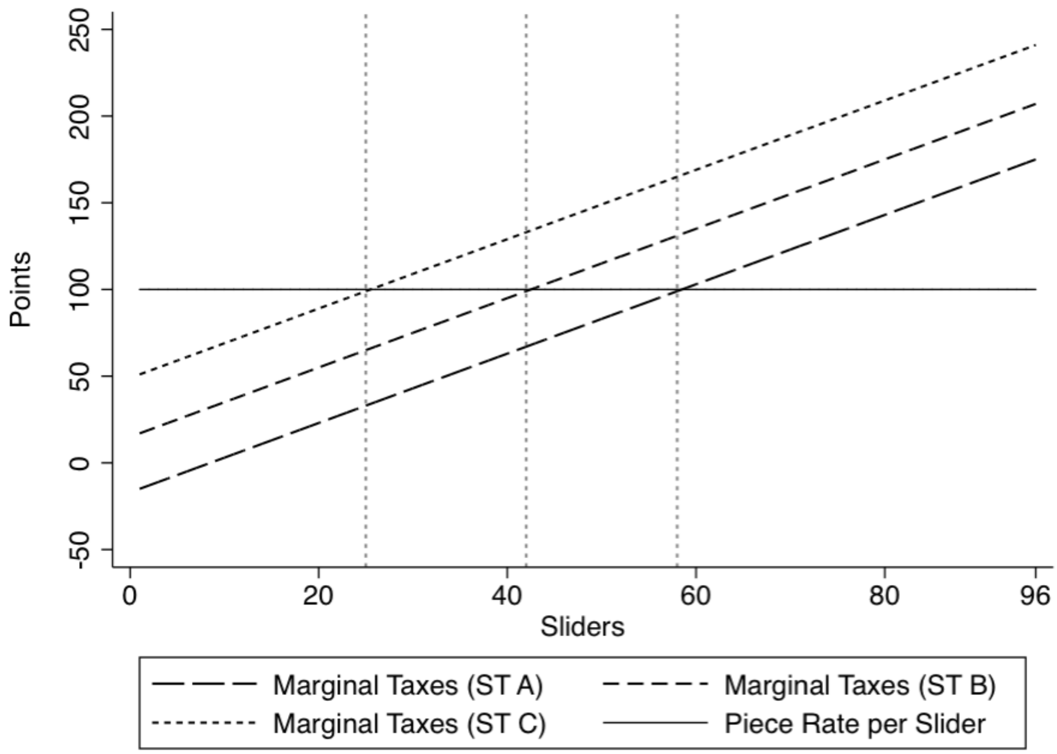
is defined as the number of sliders positioned at exactly 50. This task is attractive because it is remarkably simple and does not require pre-existing knowledge or mathematical skills. Moreover, there is little randomness in output and little room for guessing. While Gill and Prowse (2012) let their subjects only use the mouse, we also allow them to use the keyboard which reduces the real effort cost drastically (for at least 85 percent of subjects effort costs are so small that choosing the payoff-maximizing number of sliders in the main part of the experiment would be optimal, see Appendix B for details). The aim of the real-effort task is mainly to make the decision situation less abstract and psychologically more meaningful.

Figure 1: Schematic Representation of a Slider (Gill and Prowse, 2011)

(a) Initial position. (b) Positioned at 50.

In three rounds, subjects decide how many sliders they want to position: they see the tax rules that apply to their decision, commit to a number of sliders and then position the committed number. The only difference between the three rounds is that one additional tax rule is introduced after each round. All previously applicable rules remain valid. Subjects can be in one of two treatments, the simple treatment (ST) or the complex treatment (CT). The only difference between the treatments is the number of tax rules in the first round (and therefore in the following rounds). If one defines a single rule as a linear tax which applies to a sequence of adjacent sliders, the tax schedule in the first round of ST contains 2 rules; the one in CT 22. The tax schedules B and C each add one additional so-defined rule. The decision environment in CT is thus much more complex than in ST. The number of sliders that maximize payoff and the marginal payoff around this payoff optimum is, however, identical across treatments. The newly introduced tax rules are also identical across treatments.

Figure 2: Marginal Taxes for Different Tax Schedules in Simple Treatment

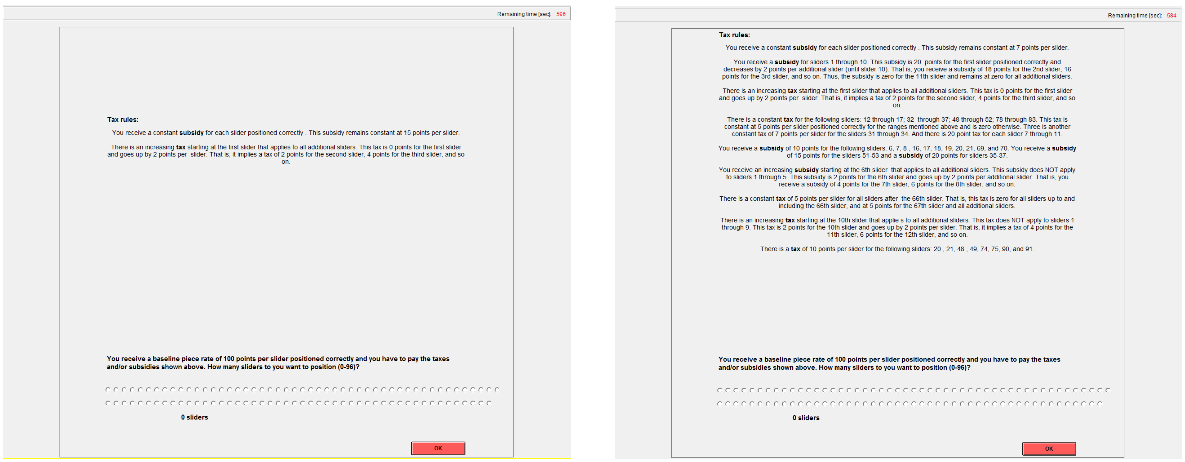


Notes: The figure displays marginal taxes and the piece rate per slider in schedules A, B, and C in the simple treatment as a function of sliders positioned correctly on the x-axis. The horizontal line indicates the piece rate per slider and the other three dashed lines denote marginal taxes under the different schedules. The dashed vertical lines indicate the number of sliders at which payoffs are maximized for a given tax schedule.

Figure 2 displays the marginal taxes subjects face in ST (see the instructions in Appendix C for the exact wording of all tax schedules). The horizontal line denotes the baseline piece rate subjects receive for each slider. The dashed line furthest to the right (long dashes) shows marginal taxes including all taxes and subsidies for the first round of ST. Monetary payoff is maximized at the point where the baseline piece rate and the schedule of marginal taxes intersect, here at 58 sliders; the net payoff for each additional slider is negative.

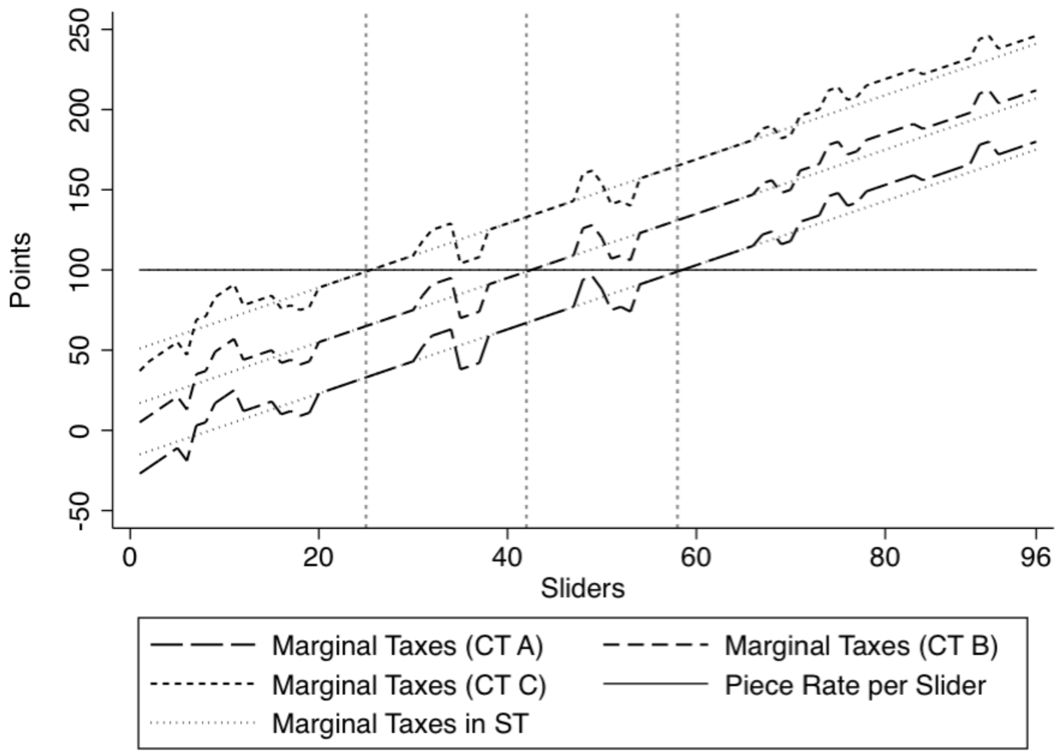
In the two subsequent rounds, additional taxes or subsidies that have constant levels per unit of output are levied while the progressivity of the tax system is not changed. This leads to parallel shifts of the marginal tax schedules. Schedules B (medium dashes) and C (short dashes) in Figure 2 display the marginal taxes that subjects in the simple treatment face in rounds 2 and 3. The number of units of output that maximize an individual's payoff are 42 for schedule B, and 25 for schedule C. All subjects in ST face schedule A in the first round. We randomize the order in which subjects in ST face schedules B and C so that half of the subjects in ST face the tax schedules in the order A-B-C and the other half in the order A-C-B.

Figure 3: Screenshot of the Tax Schedules for Round 1 in ST and CT



The key difference between the complex treatment and the simple treatment is that a number of additional tax and subsidy rules are in place, some of which are in place for a limited range of output. Figure 3 shows an overview of the tax schedules in CT and ST (see also Appendix C). Schedule A in Figure 4 displays the marginal taxes subjects in CT face in the first round. As some of the additional rules in CT cancel each other out and as Figure 4 aggregates all rules to display the overall marginal tax, the figure understates the full complexity of the tax schedule in CT as perceived by subjects.

Figure 4: Marginal Taxes for Different Tax Schedules in the Complex Treatment



Notes: The figure displays marginal taxes and the piece rate per slider in schedules A, B, and C in the complex treatment (dashed lines) as a function of sliders positioned correctly on the x-axis. The horizontal line indicates the piece rate per slider; the three dashed lines denote marginal taxes under the different schedules in the complex treatment. For comparison, the three dotted lines denote marginal taxes under the different schedules in the simple treatment. The dashed vertical lines indicate the number of sliders at which payoffs are maximized for a given tax schedule.

The additional rules implemented in the second and third round of the experiment do not differ between the complex and simple treatment and only differ across subjects depending on the order of tax schedules that was assigned (like in ST, we randomize the order as A-B-C or A-C-B). This allows for an analysis of how the introduction of the same additional tax rule can have differential effects depending on the initial complexity of the tax schedule.

We have designed the tax schedules in CT and ST to be as similar as possible to each other in terms of economic incentives while still changing the level of complexity. The dotted lines in Figure 4 depict the tax schedules in ST. One can see that the tax schedules in CT are perturbations of the schedules in ST. Firstly, the number of units of output that maximize payoff in schedules A through C are identical across the complex and the simple treatment. Secondly, the total payoff generated at each respective payoff optimum is also identical in both treatments (schedule A: 3364; schedule B: 1764; schedule C: 625). Lastly, the changes in the marginal tax around the payoff-maximizing points for each schedule are also identical across treatments. As Figure 4 shows, the tax schedules in both treatments are locally identical in a neighborhood of at least four units of output around the payoff-maximizing points for each schedule (even though many more rules need to be taken into account in the complex treatment). This implies that local deviations from the payoff-maximizing choice are as costly in the simple as in the complex treatment. An overview of the tax parameters in the two treatments is given in Table 1.

Table 1: Parameters of the Tax Schedules in the Simple and Complex Treatment

Panel 1: Simple Treatment

Round	Number of Applicable Rules	Payoff-Maximizing Number of Sliders	Increase in Marginal Tax Per Slider
1	2	58	2
2	3	25 or 42	2
3	4	42 or 25	2

Panel 2: Complex Treatment

Round	Number of Applicable Rules	Payoff-Maximizing Number of Sliders	Increase in Marginal Tax Per Slider At Optimum
1	22	58	2
2	23	25 or 42	2
3	24	42 or 25	2

Notes: We define a rule as a linear tax which applies to a sequence of adjacent sliders. The indeterminacy of the payoff-maximizing number of sliders in rounds 2 and 3 is caused by the randomization of the order in which subjects face tax schedules B and C. The increase in the tax per slider in the complex treatment is measured in a neighborhood of at least 4 sliders around the payoff-maximizing number of sliders.

Our experimental design is further motivated by the choice-theoretic framework for welfare analysis for nonstandard decision-makers developed in Bernheim and Rangel (2009). The problem they address is how to infer bounds on preferences that can be used for welfare analysis if behavior is not always fully rational. Bernheim and Rangel identify assumptions ("refinements") that can make such bounds tighter when there are theoretical grounds or evidence based on which behavior in some circumstances is more informative about underlying preferences than behavior in others. In our experiment, ST offers a compelling counterfactual to what behavior and allocations of effort and money would have been in CT if decision-makers were not affected by complexity. ST can thus be seen as a benchmark for welfare analysis in CT.

2.2 Timeline

At the beginning of the experiment, subjects are familiarized with the real-effort task. They then face several control questions that test whether they understand their potential payoff for a given a number of correctly positioned sliders in several hypothetical tax and subsidy regimes. These tax regimes confront subjects with increasing marginal taxes rates (as in the main part of the experiment) and teach subjects to focus on the marginal tax when trying to find the payoff-maximizing number of sliders. This point is also reiterated in the instructions (see Appendix C). Subjects can only continue once they answered all control questions correctly.

Main Part of the Experiment

At the beginning of the main part of the experiment, subjects are informed that they have ten minutes to read the rules upon which their payment for the first round of the experiment is based and then need to explicitly commit to a number of sliders (between 0 to 96) they will position correctly in this round of the experiment. A reminder of the time limit is briefly shown after 9:30 minutes. Subjects can make a choice after the time is over but they cannot see the tax rules anymore. After subjects have committed to a number of sliders, they start working on the slider task until they reach the specificed number of sliders. Subjects do not get feedback about their resulting earnings; this happens only at the very end of the experiment. In the second and third round, subjects are informed that all rules from the previous round are still in place and that they have 4 minutes to read the one additional rule that will affect their earnings in this round and to commit to a number of sliders (a reminder is again shown 30 seconds before this time limit). All previous rules are also displayed.

Productivity Test and Questionnaire

After the final round, subjects take part in a test of their productivity on the slider task. They are paid a constant piece rate of 2 points per slider and can work for a total of up to 15 minutes without having to specify the number of sliders they will position in advance. After the productivity test, subjects are told their combined earnings from all three rounds and from the productivity test. Finally, subjects answer a brief questionnaire in which we elicit some demographic information, conduct a test of cognitive reflection (based on Frederick 2005), and ask some questions about subjects' behavior in the experiment.

2.3 Procedural Information

A total of 277 subjects participated in the experiment which was conducted at the CeDEx laboratory at the University of Nottingham. As we expected a higher level of dispersion of decisions in CT, we randomly assigned 197 subjects into CT and 80 into ST to increase the statistical power of our analysis. The experiment was implemented using z-Tree and ORSEE (Fischbacher 2007, Greiner 2004). Subjects received a show-up fee of £2.50; points earned in the experiment were converted into cash at a rate of 1p per 7 points. If the total number of points aggregated across all rounds of the experiment was negative for a participant, only the show-up fee was paid. The average payment per subject was £9.02 (approximatlely USD 14 or EUR 10.50 at the time of the experiment). The average duration of the experiment was 50 minutes.

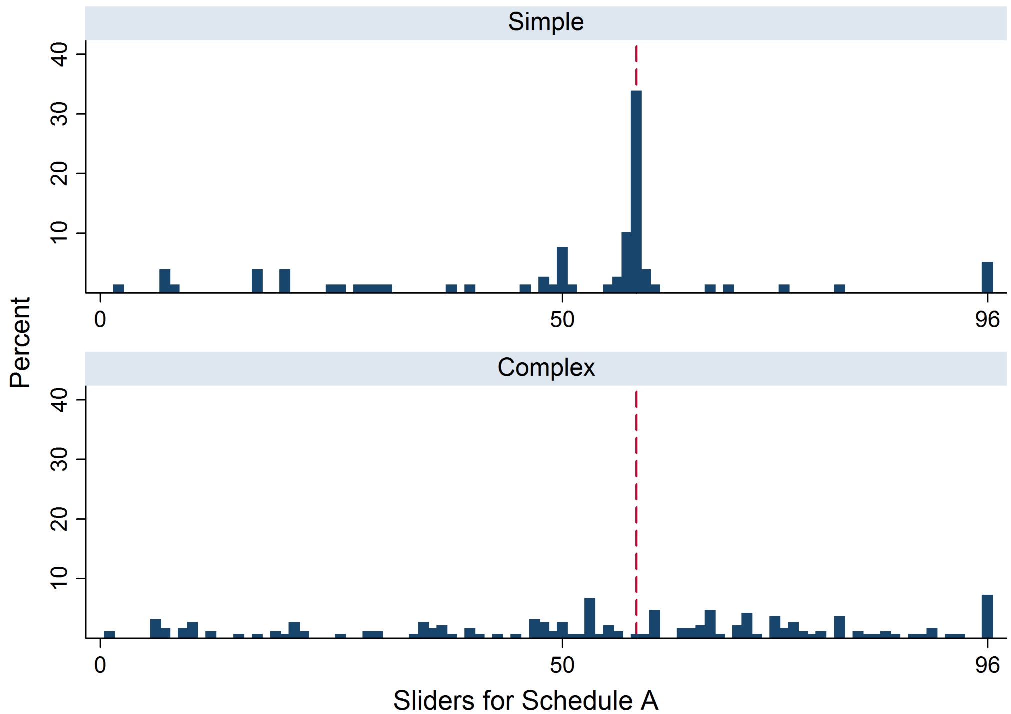
3 Results

In this section, we present results of the experiment and discuss possible explanations for the observed behavior. Before we come to the main results we document that subjects do not behave in a fully rational way.

> **Result 1:** *Subjects' choices are influenced by the level of complexity. Choices in the complex treatment are more spread out and thus, on average, further away from the payoff-maximizing choice. As a consequence, subjects in CT earn less money.*

Figure 5 depicts histograms of the choices for schedule A in the two treatments. Histograms for schedules B and C are shown in the appendix (Figures 8 and 9) and tell the same story: almost 40 percent of subjects in the simple treatment choose exactly the optimum (in the sense of payoff maximizing) marked by the dashed line in the figure. The Simple Treatment was designed to come closer to implementing the usual economic assumptions, e.g., that individuals understand the incentives they face and know their marginal tax rate. It is thus reassuring that most subjects in this treatment indeed behave more or less payoff-maximizing. In contrast, only 1.7 percent of choices in CT are optimal in the sense of maximizing payoff. This is only slightly better than random choice which would predict a success rate of about 1.0 percent. Moreover, the choices of subjects in CT are generally further away from the payoff-maximizing number. This translates into substantially lower earnings for subjects in CT as the histograms of subjects' earnings in Figure 6 document. Total profits over the three rounds are 23.0 percent lower in CT than in ST.

Figure 5: Histogram of Choices for Schedule A



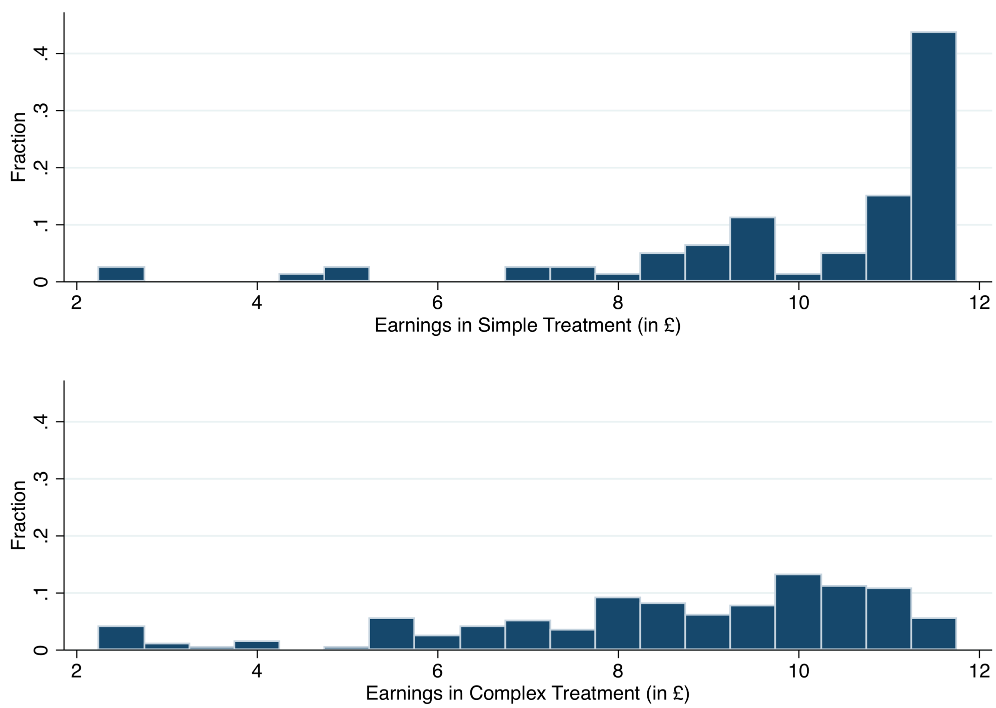
Notes: The dashed line marks the payoff-maximizing choice at 58.

Table 6 in the appendix shows that these differences in behavior are highly significant in a regression framework with a set of control variables. We take as dependent variables a dummy which equals 1 if the subject chose the payoff-maximizing effort level in a given round; the absolute distance to the payoff-maximizing choice for each decision; and subjects' earnings from each decision. We regress these variables on a dummy for being in CT and add in subsequent regressions controls for the choice order (A-B-C or A-C-B), for an interaction of CT and choice order and for age, gender, and IQ.[6] Each subject enters the regression three times but we do not assume that an individual's decisions are independent: standard errors

[6]We combine three sets of information to derive our measure of IQ. We know subjects' math grade in their final high school year. It has been shown that math grade correlates highly with Spearman's g, the quantity that IQ-tests aim to measure (Deary, Strand, Smith, and Fernandes, 2007). Subjects also complete the Cognitive Reflection Test which also correlates with IQ (Frederick, 2005). Finally, subjects answer a set of questions to test their financial numeracy (similar to the ones in Gerardi, Goette, and Meier, 2010). Our measure of IQ is the principal factor of a factor analysis of these three variables, standardized to have mean 0 and standard deviation 1.

computed by block bootstrap clustered at the subject level are shown in the table. The results are remarkably robust across dependent variables and specifications: subjects in CT are less likely to choose the optimum, are generally further away from the optimum and earn less money.[7] All p-values of the treatment dummy are below 0.001. These results also hold for the first round, i.e., tax schedule A, in isolation.[8]

Figure 6: Histogram of Earnings by Treatment



[7]To take account of the fact that the earnings are capped at the maximal level, one could use tobit regressions with an upper limit at this maximal earnings level (Table 6 reports OLS estimates). Since this level varies between the three rounds, we recode the earnings variable as maximal earnings – earnings and run tobit regressions with a lower limit at 0. All p-values of the CT dummy stay below 0.001 and the point estimates increase strongly to around 600. Moreover, if the total earnings of a subject was negative, it was set to zero. This happened for 4.3 percent of subjects. Results do not change if we recode earnings according to realized final earnings.

[8]At the same time, subjects are not indifferent about their slider choice and decisions in CT are not completely random. While having difficulties to find the optimum exactly, most subjects change their choice between rounds in the direction of the new payoff maximum (71 percent in CT compared to 87 percent in ST), though the adjustment does often not go far enough, see below. As a consequence, average profit is well above what one would expect under uniformly random choice (t-tests, all $p < 0.001$). Since the real effort costs of moving the sliders are positive for most subjects, subjects who are indifferent with respect to the monetary rewards should choose 0. This happens only rarely; actually, the maximum of 96 is chosen more often than 0.

Subjects in CT realize that they have more difficulty in dealing with the more complex situations. In the post-experimental questionnaire, they agree significantly more than subjects in ST with the statements "Figuring out how many sliders I wanted to position was very stressful." and "I would have needed more time to figure out how many sliders to position." and agree less with the statement "I think that I chose the number of sliders that maximized my monetary payoff."[9]

These results validate our treatment manipulation and also confirm that subjects in general are influenced by the complexity of the decision environment. Next, we explore how subjects try to cope with the complex environment.

> **Result 2:** *Complexity attenuates subjects' reaction to changes in incentives: Subjects in the complex treatment take their previous round's decision as a point of departure and do not adjust their choice as much in reaction to new incentives as subjects in the simple treatment.*

In Table 2, columns 1–3, we take the change in effort level from the previous round, i.e., from the first to the second and the second to the third, as dependent variable. To make the effort changes comparable across treatments, the effort change in the second decision in the choice order A-C-B is multiplied by -1. It is thus always optimal to *reduce* the effort from stage to stage (given a payoff-maximizing choice in the previous stage). We regress this variable on a dummy for being in the complex treatment and the various control variables mentioned above. Since effort choices and also changes in effort choices are spread out over the whole range of possible choices (see Figures 5 and 8–11) we use median regressions to limit the influence of extreme outliers. We include both effort changes that a subject makes in the regression and use bootstrapped standard errors clustered at the individual level.

We find that subjects in CT react less strongly to new incentives. The median change in effort by subjects in the simple treatment is -17 which is actually the optimal reaction if

[9] On a scale from 1 "Disagree completely" to 6 "Agree completely", subjects in CT average 4.4 in the first question and 5.0 in the second, while subjects in ST average 3.2 and 3.0. The third question was asked separately for the three rounds; for each round, subjects in CT report a lower agreement (3.2, 3.3, 3.7) than subjects in ST (4.4, 4.4, 4.7). These differences are highly significant in tobit regressions with and without the controls used in Table 6 (all p<0.001).

Table 2: Change in effort level from previous round

Dependent variable:	Change in effort level from previous round			1 if subject chose same effort as in previous round		
	(1)	(2)	(3)	(4)	(5)	(6)
1 if CT	5.000***	6.000**	7.514**	0.085***	0.111***	0.112***
	(1.525)	(2.718)	(3.274)	(0.018)	(0.026)	(0.027)
1 if Choice Order A-B-C		1.000	2.881		0.012	0.013
		(1.846)	(2.845)		(0.013)	(0.014)
CT * A-B-C		-3.000	-4.367		-0.052	-0.053
		(3.347)	(3.816)		(0.035)	(0.035)
Age			1.462**			0.003
			(0.723)			(0.008)
1 if Female			1.486			-0.017
			(1.669)			(0.025)
IQ Measure			-0.725			-0.008
			(1.092)			(0.012)
Constant	-17.000***	-17.000***	-48.018***	0.006	0.000*	-0.058
	(0.447)	(1.833)	(15.222)	(0.006)	(0.000)	(0.168)
N.Obs.	554	554	554	554	554	554

Notes: Quantile (median) regression (columns 1–3) and OLS estimates (columns 4–6). The dependent variable in columns 1–3 is the change in effort level from the previous round; thel effort changes in the second decision of the choice order A-C-B are multiplied by -1 to make them comparable; coming from a payoff-maximizing effort level, an effort decrease is always optimal. The dependent variable in columns 4–6 is a dummy equaling 1 if the subject chose the same effort as in the previous round. Each subject enters the regression twice, standard errors computed by block bootstrap clustered at the subject level are in parentheses. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

they started from a payoff-maximizing choice. The median change in effort by subjects in the complex treatment is only -12. The positive coefficient for the CT dummy shows that subjects in CT react *less* to the new incentive. This difference is highly statistically significant and remains significant when we add controls for the choice order and an interaction of CT and choice order in column 2 and when we additionally add controls for age, gender and IQ in column 3.

We also explore whether complexity affects subjects' tendency to completely ignore the change in incentives across rounds.

> **Result 3:** *Subjects in the complex treatment are less likely to react to the change in incentives at all and more likely to leave their previous effort choice unchanged.*

In Table 2, columns 4–6, the dependent variable is a dummy which equals 1 if a subject chose the same effort as in the previous round. The estimates show that subjects in ST almost never choose the same effort as in the previous round. This happens only once out of the 160 effort changes in ST (0.6 percent of effort changes). In contrast, 9.1 percent of effort changes in CT equal zero. These treatment differences are highly significant and robust to the inclusion of the aforementioned control variables. Figures 10 and 11 in the appendix depict histograms of the change of effort from round to round. One can clearly see the spike at zero in the complex treatment.

Note that our design makes it hard to detect such an effect: choosing the same level of output as in the previous round cannot be driven by a mechanical default effect as subjects had to actively enter a choice. The high frequency of this extreme form of underreaction also speaks against a simple decision-error explanation for the attentuated reaction to incentives (Result 2), e.g., some subjects in CT deciding randomly each round which would also attenuate the average reaction to incentive changes. Our results show that increased background complexity can trigger status-quo effects (e.g., Samuelson and Zeckhauser 1988).

In the experiment, the order of tax changes subjects face was randomized: half of the subjects face an additional tax of 66 points per sliders in the second round (choice order A-C-B); the other half face an additional tax of 32 points per slider in this round (A-B-C). Therefore, the monetary costs of ignoring the tax rule introduced in the second round differ depending on the choice order (1087 points for ignoring the larger tax vs. 256 points for ignoring the smaller tax). Most theories of bounded rationality feature some variant of rational inattention implying that information is more likely to be ignored if it is less relevant for decision-making (e.g., Sims, 2003; Gabaix, 2011; Chetty, 2012). Our design allows us to shed some light on this prediction by studying whether the smaller tax change is more likely to be ignored than the larger one. Surprisingly, we do not find evidence for this.

> **Result 4:** *Subjects in the complex treatment are equally likely to ignore small and large incentive changes.*

In Table 3, we regress a dummy which equals 1 if a subject chose the same effort in the second and the first round of the experiment on a dummy indicating whether the subject faced a large tax change in the second round, i.e., whether the subject was in choice order A-C-B. We restrict the sample to CT as almost no subject in ST left their effort level unchanged. We find a positive point estimate of 5.0 percentage points implying that the *larger* tax change is ignored more often. The estimate is, however, not significant and stays insignificant when we add the controls mentioned above so that we cannot reject that both tax changes are equally likely to be ignored.

Table 3: Effect of the magnitude of incentive change

Dependent variable: 1 if subject chose same effort in first and second round		
	(1)	(2)
1 if Choice Order A-C-B	0.050	0.053
	(0.040)	(0.039)
Age		0.009
		(0.013)
1 if Female		-0.074*
		(0.039)
IQ Measure		-0.023
		(0.019)
Constant	0.061**	-0.087
	(0.024)	(0.264)
N.Obs.	197	197

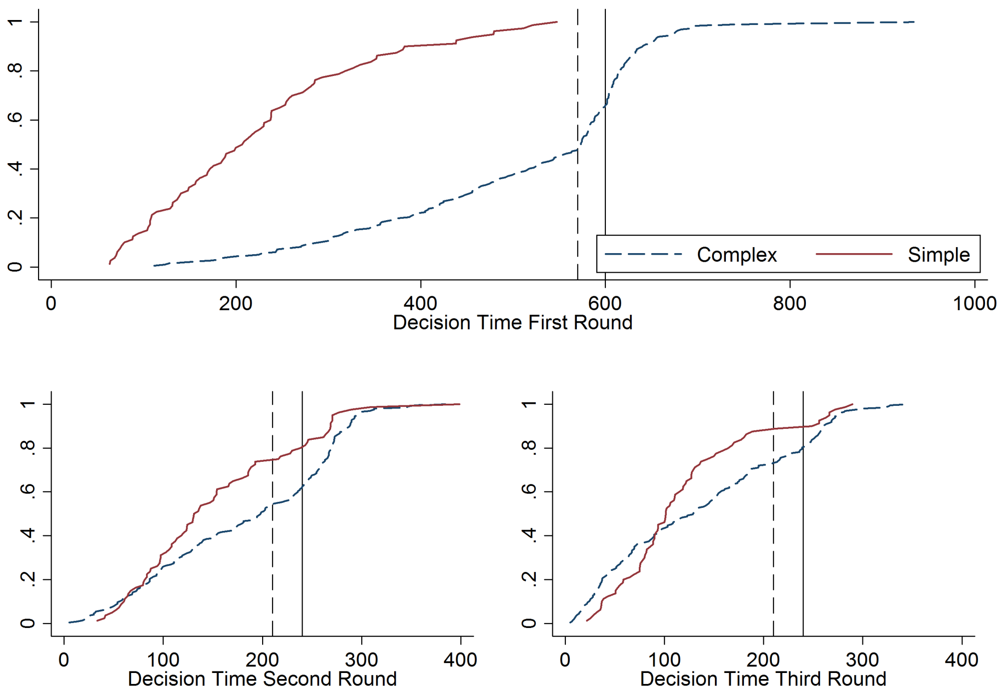
Notes: OLS estimates. The sample is restricted to the complex treatment. The dependent variable is a dummy equaling 1 if the subject chose the same effort as in the previous round. Bootstrapped standard errors are in parentheses. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

In addition to the actual choice, we also measured how long subjects needed for their decision.

Result 5: *Subjects in CT take longer for the first decision than subjects in ST. This difference is much smaller for the second decision and vanishes for the third.*

Figure 7 shows the cumulative distributions of decision times of subjects in CT and ST. The top panel shows the time in the first round. In this round, subjects had 600 seconds to make a choice; a reminder of this time limit was briefly shown after 570 seconds. Subjects could make a choice after the time was over but they could not see the tax rules anymore. The lower panel shows decision times in the second and third round. Here, the rules were masked after 240 seconds and a reminder was shown after 210 seconds.

Figure 7: Decision Times in the Three Rounds (in Seconds)



Notes: The solid vertical lines mark the time limit after which the individual tax rules were masked in each round. Subjects could still make their choice after the time limit. They were reminded of the time limit 30 seconds before the limit (dashed line).

One would expect that subjects react to a more complex environment by thinking longer about their decision. This is indeed the case for the first round. As one can see from Figure 7, subjects in CT take on average more than twice as long as subjects in ST (511 seconds vs. 225 seconds). Table 4 shows that this difference is highly significant (column 1) and remains significant if we control for the additional variables described above (columns 2 and 3). This understates the true underlying need for additional time, as many subjects in CT are forced to shorten their deliberation time and to make a choice once they reach the end of the allotted time.[10] In contrast, average decision times in the second and third round are much closer

[10]Subjects in CT do not deliberate longer in the first round because they are slower in general. Since subjects were allocated randomly to treatments, there should be no difference in innate characteristics of subjects across the treatments. And indeed, subjects do not differ in the speed with which they tackle decisions and tasks that do not differ across treatments. For example, they take the same time to answer the control question (p=0.667) and also take similar times to move the sliders in the real-effort task (p=0.097, p=0.287, and p=0.678 for first,

together. The difference in the second round is still significant (columns 4–6); the difference in the third round is yet smaller and not significant anymore when including controls (columns 7–9).

To shed light on what mechanisms could underlie the effects of complexity that we have documented, we analyze whether there are some subjects who are more strongly affected by complexity than others.

> **Result 6:** *Subjects in CT who choose effort levels further away from the payoff optimum are more likely to ignore tax changes in the following rounds. Subjects who ignore the first tax change are more likely to also ignore the second change.*

Since the treatment effect between ST and CT is to a large extent driven by a higher share of subjects who do not react to new incentives, we test whether this measure of behavior is correlated with other behaviors of subjects. Such no-reaction occurs extremely rarely in ST, and we therefore confine our analysis to subjects in the complex treatment. In columns 1 and 2 of Table 5, we regress a dummy variable indicating whether a subject chose the same effort level as in the previous round on a measure of within-round decision quality. We use the distance to the payoff-maximizing choice in the first round as a measure of decision quality. The regressions reveal that subjects who choose effort levels further away from the payoff optimum in the first round are more likely to ignore tax changes in the following rounds. A one standard deviation increase in the distance to the payoff optimum in the first round leads to a 3.8 percentage point increase in the likelihood of ignoring a tax change in one of the following rounds (coefficient of 0.0025, sd of 15.22), about a third of the overall treatment effect.

We also observe within-subject persistence in the proclivity to ignore tax changes: in columns 3 and 4 of Table 5, we regress a dummy variable indicating inaction between rounds 2 and 3 on a dummy for inaction between rounds 1 and 2. We find that subjects who ignored the first tax change entirely were more likely to also ignore the tax change in the last round of the experiment: 41.2 percent of the subjects who stick to their choice from the first to the

second and third round, respectively).

Table 4: Decision Time

Dependent variable:	Decision time first round			Decision time second round			Decision time third round		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1 if CT	285.490***	273.140***	267.326***	30.274***	34.483**	33.927**	17.365*	6.762	4.672
	(16.419)	(22.058)	(23.057)	(9.893)	(15.576)	(15.860)	(9.323)	(14.847)	(14.841)
1 if Choice Order A-B-C		-24.761	-21.688		-3.887	-2.817		-24.724*	-22.882
		(26.771)	(28.297)		(17.146)	(17.391)		(14.305)	(14.573)
CT * A-B-C		24.700	21.052		-8.480	-11.209		21.188	18.063
		(33.886)	(34.819)		(21.751)	(21.751)		(19.704)	(19.713)
Age			1.169			-2.692			0.372
			(4.456)			(3.385)			(3.593)
1 if Female			-18.336			-17.861*			-22.538**
			(16.498)			(10.449)			(10.594)
IQ Measure			23.075***			2.646			7.657
			(8.555)			(5.415)			(5.167)
Constant	225.201***	237.582***	227.515**	153.066***	155.009***	218.161***	117.777***	130.139***	135.754*
	(13.577)	(17.161)	(92.476)	(8.445)	(13.269)	(67.693)	(7.251)	(11.868)	(71.868)
N.Obs.	277	277	277	277	277	277	277	277	277

Notes: OLS estimates. The dependent variable is the time subjects took to decide in the three rounds of the experiment (measured in seconds). Standard errors computed by block bootstrap clustered at the subject level are in parentheses. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

second round also do so for the third round. Of the subjects who change their decision from first to second round, only 6.7 percent leave their choice unchanged for the third round. These results suggest that there is indeed heterogeneity across subjects in the degree to which they are affected by complexity.

Finally, we assess the validity of our interpretation that subjects who do not adjust their effort levels across rounds "ignore" the new tax rule. To do so, we regress the time subjects needed for their decision-making in a given round on a dummy variable indicating whether a subject chose the same effort level as in the previous one. As columns 5 and 6 of Table 5 reveal, subjects who did not adjust their effort choices across rounds took on average more than one minute less to make their decision than subjects who changed their effort in response to a change in taxes. This suggests that the status-quo effect we document in the complex treatment is driven by subjects ignoring the new rule.[11]

4 Discussion

One way of conceptualizing how people deal with complex decision environments is to posit that they only pay attention to a subset of the available information. Below, we outline a very simple framework that is able to capture this intuition and is in line with our data. For our experiment, it is most natural to define complexity as a function of the number of tax rules in place. To focus ideas, we define a single rule as a linear tax or subsidy which applies to a sequence of adjacent sliders (the idea of our framework does not depend on the exact definition of a rule, as long as CT retains more rules than ST). Under this definition, tax schedule A of ST consists of two rules and tax schedule A of CT consists of 22 rules. Tax schedules B and C each add one additional rule.

We assume that a subject has a capacity to pay attention to k tax rules out of the total of N rules. All of these k rules are used in a fully rational way when making the decision. The

[11] Our measure of IQ also correlates with behavior in the experiment. Low-IQ subjects in CT react significanlty less strongly to new taxes. They are more likely to leave their choice unchanged between rounds but this latter effect is not significant. Low-IQ subjects in CT also spend significantly less time deciding in all three rounds.

Table 5: Within-Subject Correlation of Behavior in the Complex Treatment

Dependent Variable	1 if subject chose same effort as in previous round		1 if subject chose same effort in rounds 2 and 3		Decision time	
	(1)	(2)	(3)	(4)	(5)	(6)
Distance to payoff-maximizing choice (Round 1)	0.0025**	0.0025**				
	(0.0012)	(0.0013)				
1 if subject chose same effort in rounds 1 and 2			0.3451***	0.3583***		
			(0.1214)	(0.123)		
1 if subject chose same effort as in previous round					-63.7110***	-65.3692***
					(15.2165)	(14.8722)
Constant	0.0389	0.0272	0.0667***	0.1566	165.0621***	241.2201***
	(0.0255)	(0.2018)	(0.0187)	(0.1744)	(5.7604)	(70.4217)
Control Variables	No	Yes	No	Yes	No	Yes
N.Obs.	394		197		394	

Notes: OLS estimates. The sample is confined to individuals in the complex treatment. Control variables are choice order, age, gender, and IQ. Robust standard errors are in parentheses. Each subject enters the regression twice for columns 1, 2, 5 and 6; standard errors computed by block bootstrap are in parentheses. The dependent variable in columns 5 and 6 is the decision time (measured in seconds) in rounds 2 and 3. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

remaining $N - k$ rules are completely ignored. This sharp drop of attention at k is similar to models of consideration sets (Eliaz and Spiegler, 2011) or sparsity (Gabaix, 2011). We cannot observe how the rules are chosen; this could be done according to many criteria, for example, according to the salience of particular rules.[12] Every criterion induces an ordering on the N rules. The top k rules of the ordering form the consideration set. If a new rule is introduced, it is integrated into the ordering; again, the top k rules are then chosen.

If we assume that $4 < k_i < 22$ for all subjects i and that all possible orderings are chosen with positive probability in the group of subjects, this framework is in line with the main results of our experiment. $k_i > 4$ ensures that subjects in ST always choose the payoff-maximizing effort level as the tax schedules in ST contain a maximum of four rules. Subjects in ST indeed fare quite well at finding the payoff-maximizing effort level and most of the subjects who are not exactly at this effort level choose effort levels close by. Assuming that $k_i < 22$ and that all possible orderings are chosen with positive probability ensures that at least some participants in CT (if the number of subjects is large enough) will not choose the payoff-maximizing effort level.[13] Subjects in CT will thus be on average further away from the payoff maximum and earn less profit than subjects in ST. Moreover, if a new rule is not included in the consideration set, the choice in this round will be identical to the choice in the previous round. Our framework therefore offers a natural way to generate a cluster of subjects who do not react at all to the new rule if the number of available rules exceeds the size of their consideration set (and the new rule is not included in the top k rules). Our empirical evidence shows that the difference between ST and CT is indeed mainly driven by inertia, i.e., a larger share of subjects who do not change their behavior compared to the previous round. The within-person correlation of inertia that we document is also consistent with this framework if there is heterogeneity in k: subjects with small consideration sets will be unlikely to include the new rule in the second round (as they are unlikely to include any rule) and also unlikely

[12]This could include the length of each rule's text, the size of the taxes levied, some sort of cost-benefit analyses, etc. For recent economic models of salience, see, e.g., Bordalo, Gennaioli, and Shleifer (2012) or Köszegi and Szeidl (forthcoming). The ordering of rules could also be randomly chosen, but would then be fixed.

[13]We need the second assumption because potentially only a subset of rules influences the marginal payoff around the payoff maximum. For example, rules that only affect payoffs above this level could be ignored—once the maximum is identified.

to include the new rule in the third round.

If deliberation time increases with the number of new rules considered, then deliberation time in the first round will be longer for subjects in CT than for subjects in ST. This is again in line with our data. Deliberation time in the second and third round will be much more similar as subjects in both treatment will at most consider one new rule. We find in the experiment that deliberation times get more similar across treatments in the later rounds. Subjects who do not change their behavior across rounds decide particularly fast. Finally, if the criteria that subjects use to rank rules are not built on some sort of cost-benefit analysis (but rather related to, e.g., salience), the potential payoff consequences of considering a new rule will have no influence on the likelihood of its inclusion in the consideration set. Our finding that the magnitude of non-reaction to small and large taxes is similar points to the importance of this type of criteria. At the same time, this finding is in contrast with most models of choice under complexity which posit at least some relation to incentives (e.g., Sims, 2003; Gabaix, 2011; Chetty, 2012).

This is a very simple way to think about how subjects make choices that can still organize large parts of our data well. At the same time, other models will also be able to explain some of our data even though our results are inconsistent with models in which new rules become more salient relative to the existing set of rules as the complexity of the decision environment increases.[14] The aim of this paper is, however, not to test the many different models of choice under complexity against each other but rather to investigate how background complexity influences the reaction to newly introduced incentives.

5 Conclusion

We conducted a lab experiment to test how the complexity of pre-existing incentives influences the reaction to changes in economic incentives. Subjects participated in one of two treatments

[14]For example, one could assume that subjects have a stock of mental resources that depletes when mental effort is exerted (see, e.g., Baumeister, Bratslavsky, Muraven, and Tice, 1998). The lower the mental reserves are, the higher is the marginal cost of thinking. Such a model could generate the lower responsiveness to new rules under complexity, the patterns in decision time and within-person correlation of behavior, though it would have difficulties generating the complete non-response to new rules that we observe frequently.

which confronted them with either a simple or complex tax system. The same sequence of additional tax rules was then introduced in both treatments. We find that subjects in the complex treatment react less strongly to the newly introduced incentives. This is driven by a larger share of subjects who do not react at all. A simple framework based on consideration sets matches our results well.

In a world with fully rational decision makers allowing for a complex tax system can be optimal as it gives the social planner more tax instruments to tailor individual behavior to maximize social welfare. In contrast, our findings suggest that introducing complexity comes at a cost: boundedly rational decision makers will not only fail to choose the payoff-maximizing response to non-salient tax rules (as in Chetty, Looney, and Kroft, 2009) but will also under-react to salient rules if the existing system is too complex.

Clearly, our results—taken at face value—cannot be quantitativley translated into policy-relevant elasticities. So what implications can be drawn from our lab experiment? First, we document that incentive complexity can be an important trigger of status-quo effects. This reveals a mechanism through which increasing tax complexity lowers elasticities. Second, our experiment shows that there is substantial heterogeneity in the effect of complexity: some individuals are particularly affected by increasing tax complexity. This heterogeneity can be used in the design of optimal policy as it can help to target policies more precisely. Third, we find no evidence that larger tax changes are less likely to be ignored. This casts some doubt on the applicability of rational inattention models in the domain of taxation; more research is needed to understand how individuals allocate their attention.

In future work, our experimental design could be extended to assess whether the effects of complexity persist if individuals can learn over time. Decision quality did not improve across rounds in our experiment; the time horizon of the experiment was, however, rather limited. If learning new rules takes more time in initially complex tax systems, diff-in-diff studies aimed at estimating taxable income elasticities based on changes in the tax code will underestimate 'true' elasticity parameters when the underlying tax code is more complex. In addition to investigating the nexus between learning and complexity, it would be interesting to study the effects of complexity in an environment in which individuals can hire the service of sophisticated agents, such as tax consultants. While field evidence suggests that complexity may also affect experienced individuals (see Chetty, Looney, and Kroft, 2009) with access to sophisticated advice (Chetty and Saez, 2013)[15], understanding the role of learning and advice would be crucial for identifying ways to improve decision making.

References

Baumeister, R., E. Bratslavsky, M. Muraven, and D. Tice (1998): "Ego depletion: is the active self a limited resource?," *Journal of personality and social psychology*, 74(5), 1252.

Bernheim, B. D., and A. Rangel (2009): "Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics," *The Quarterly Journal of Economics*, 124(1), 51–104.

Bordalo, P., N. Gennaioli, and A. Shleifer (2012): "Salience theory of choice under risk," *The Quarterly Journal of Economics*, 127(3), 1243–1285.

Brown, J., T. Hossain, and J. Morgan (2010): "Shrouded attributes and information suppression: Evidence from the field," *The Quarterly Journal of Economics*, 125(2), 859–876.

Brown, J., A. Kapteyn, E. Luttmer, and O. Mitchell (2012): "Do Consumers Know How to Value Annuities? Complexity as a Barrier to Annuitization," *mimeo.*

Carlin, B., S. Kogan, and R. Lowery (forthcoming): "Trading complex assets," *The Journal of Finance.*

Cheremukhin, A., A. Popova, and A. Tutino (2011): "Experimental evidence on rational inattention," *Federal Reserve Bank of Dallas Working Paper 1112.*

[15]The field experiment by Chetty and Saez (2013) demonstrates that individuals who hire professional tax preparers adjust their earnings when given salient information about the marginal incentives they face. This suggests that even individuals with access to professional tax advice may not fully understand the marginal tax incentives they face.

Chetty, R. (2012): "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply," *Econometrica*, 80(3), 969–1018.

Chetty, R., J. Friedman, and E. Saez (forthcoming): "Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings," *American Economic Review*.

Chetty, R., A. Looney, and K. Kroft (2009): "Salience and Taxation: Theory and Evidence," *The American Economic Review*, 99(4), 1145–1177.

Chetty, R., and E. Saez (2013): "Teaching the tax code: Earnings responses to an experiment with EITC recipients," *American Economic Journal: Applied Economics*.

Congdon, W., J. Kling, and S. Mullainathan (2011): *Policy and choice: public finance through the lens of behavioral economics*. Brookings Institution Press.

Crosetto, P., and A. Gaudeul (2012): "Do consumers prefer offers that are easy to compare? An experimental investigation," *SSRN Working Paper 1943603*.

Dahremöller, C., and M. Fels (2012): "Product Lines, Product Design and Limited Attention," *SSRN Working Paper 1993128*.

de Bartolome, C. (1995): "Which tax rate do people use: Average or marginal?," *Journal of Public Economics*, 56(1), 79–96.

Deary, I., S. Strand, P. Smith, and C. Fernandes (2007): "Intelligence and educational achievement," *Intelligence*, 35(1), 13–21.

DellaVigna, S., and J. Pollet (2009): "Investor inattention and Friday earnings announcements," *The Journal of Finance*, 64(2), 709–749.

Eliaz, K., and R. Spiegler (2011): "Consideration sets and competitive marketing," *The Review of Economic Studies*, 78(1), 235–262.

Feldman, N., and P. Katuscak (2009): "Effects of predictable tax liability variation on household labor income," *SSRN Working Paper 1130300*.

Feldman, N., and B. Ruffle (2012): "The Impact of Tax Exlusive and Inclusive Prices on Demand," .

Finkelstein, A. (2009): "E-ZTax: Tax Salience and Tax Rates.," *Quarterly Journal of Economics*, 124(3), 969–1010.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10(2), 171–178.

Fleming, S., C. Thomas, and R. Dolan (2010): "Overcoming status quo bias in the human brain," *Proceedings of the National Academy of Sciences*, 107(13), 6005–6009.

Fochmann, M., and J. Weimann (2011): "The effects of tax salience and tax experience on individual work efforts in a framed field experiment," Discussion paper, IZA Discussion Paper.

Frederick, S. (2005): "Cognitive reflection and decision making," *The Journal of Economic Perspectives*, 19(4), 25–42.

Fujii, E., and C. Hawley (1988): "On the accuracy of tax perceptions," *The Review of Economics and Statistics*, pp. 344–347.

Gabaix, X. (2011): "A sparsity-based model of bounded rationality," Discussion paper, National Bureau of Economic Research.

Gabaix, X., and D. Laibson (2006): "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *The Quarterly Journal of Economics*, pp. 505–540.

Gabaix, X., D. Laibson, G. Moloche, and S. Weinberg (2006): "Costly information acquisition: Experimental analysis of a boundedly rational model," *The American Economic Review*, pp. 1043–1068.

Gerardi, K., L. Goette, and S. Meier (2010): "Financial literacy and subprime mortgage delinquency: Evidence from a survey matched to administrative data," *Federal Reserve Bank of Atlanta Working Paper Series*, (2010-10).

Gill, D., and V. Prowse (2011): "A novel computerized real effort task based on sliders," .

——— (2012): "A structural analysis of disappointment aversion in a real effort competition," *The American Economic Review*, 102(1), 469–503.

Goldin, J. (2012): "Optimal Tax Salience," *SSRN Working Paper 2009108*.

Goldin, J., and T. Homonoff (2013): "Smoke Gets in Your Eyes: Cigarette Tax Salience and Regressivity," *American Economic Journal: Economic Policy*, 5(1), 302–336.

Golosov, M., A. Tsyvinski, and I. Werning (2007): "New dynamic public finance: a user's guide," in *NBER Macroeconomics Annual 2006, Volume 21*, pp. 317–388. MIT Press.

Greiner, B. (2004): "An Online Recruitment System for Economic Experiments," in *Forschung und wissenschaftliches Rechnen*, ed. by K. Kremer, and V. Macho, GDWDG Bericht 63, pp. 79–93. Ges. für Wiss. Datenverarbeitung, Göttingen.

Hirshleifer, D., S. Lim, and S. Teoh (2009): "Driven to distraction: Extraneous events and underreaction to earnings news," *The Journal of Finance*, 64(5), 2289–2325.

Ito, K. (2010): "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing," *Energy Institute at Haas Working Paper*, 210.

Kahneman, D., J. Knetsch, and R. Thaler (1991): "Anomalies: The endowment effect, loss aversion, and status quo bias," *The Journal of Economic Perspectives*, pp. 193–206.

Kalayci, K., and M. Serra-Garcia (2012): "Complexity and Narrow Bracketing in Credit Choice," *Discussion Paper, University of Munich*.

Kleven, H., and W. Kopczuk (2011): "Transfer Program Complexity and the Take Up of Social Benefits"," *American Economic Journal: Economic Policy*, 3, 54–90.

Kőszegi, B., and A. Szeidl (forthcoming): "A Model of Focusing in Economic Choice," *The Quarterly Journal of Economics.*

Liebman, J., and R. Zeckhauser (2005): "Schmeduling," *mimeo Harvard University.*

Mirrlees, J. (1971): "An exploration in the theory of optimum income taxation," *The Review of Economic Studies*, pp. 175–208.

Ortoleva, P. (forthcoming): "The Price of Flexibility: Towards a Theory of Thinking Aversion," *Journal of Economic Theory.*

Persson, P. (2012): "Attention Manipulation and Information Overload," *Working Paper, Columbia University.*

Reis, R. (2006): "Inattentive consumers," *Journal of Monetary Economics*, 53(8), 1761–1800.

Saez, E. (2010): "Do Taxpayers Bunch at Kink Points?," *American Economic Journal: Economic Policy*, 2(3), 180–212.

Samuelson, W., and R. Zeckhauser (1988): "Status quo bias in decision making," *Journal of Risk and Uncertainty*, 1(1), 7–59.

Sims, C. (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50(3), 665–690.

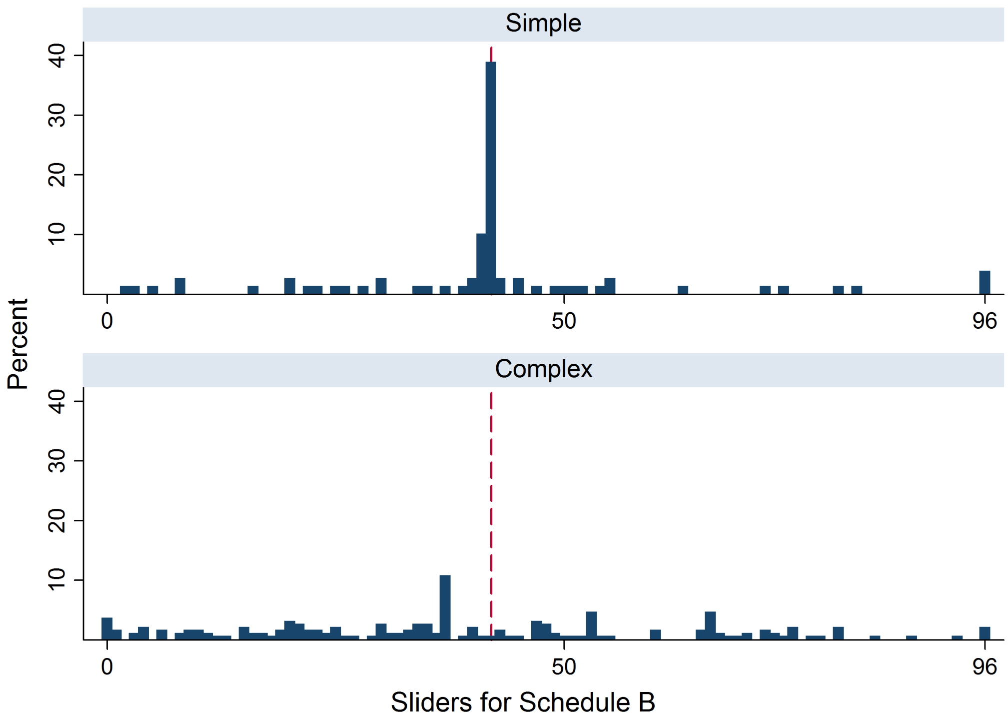
Sitzia, S., J. Zheng, and D. Zizzo (2012): "Complexity and Smart Nudges with Inattentive Consumers," *SSRN Working Paper 2180956.*

Sunstein, C. (2011): "Empirically informed regulation," *U. Chi. L. Rev.*, 78, 1349.

Appendix

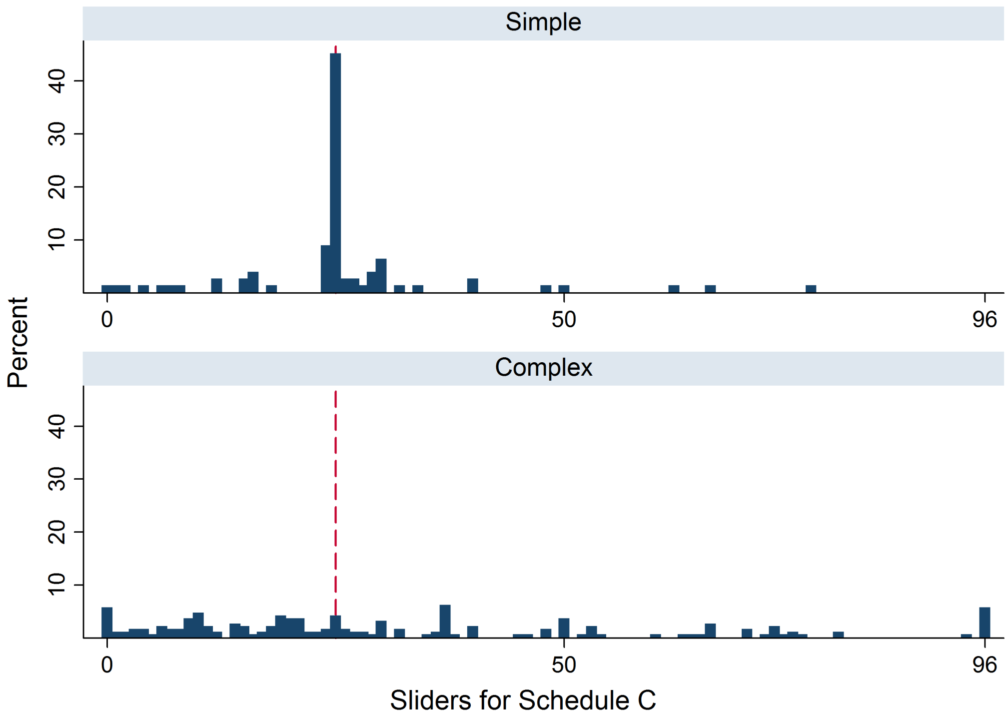
A Additional Figures and Table

Figure 8: Histogram of Choices for Schedule B



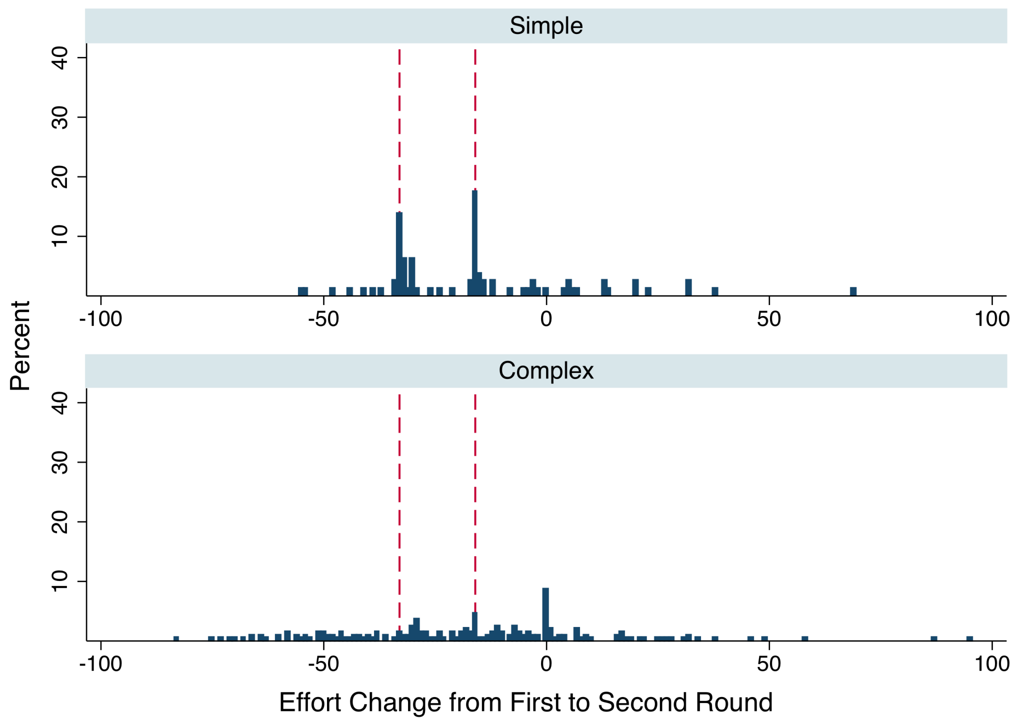
Notes: The dashed line marks the payoff-maximizing choice at 42.

Figure 9: Histogram of Choices for Schedule C



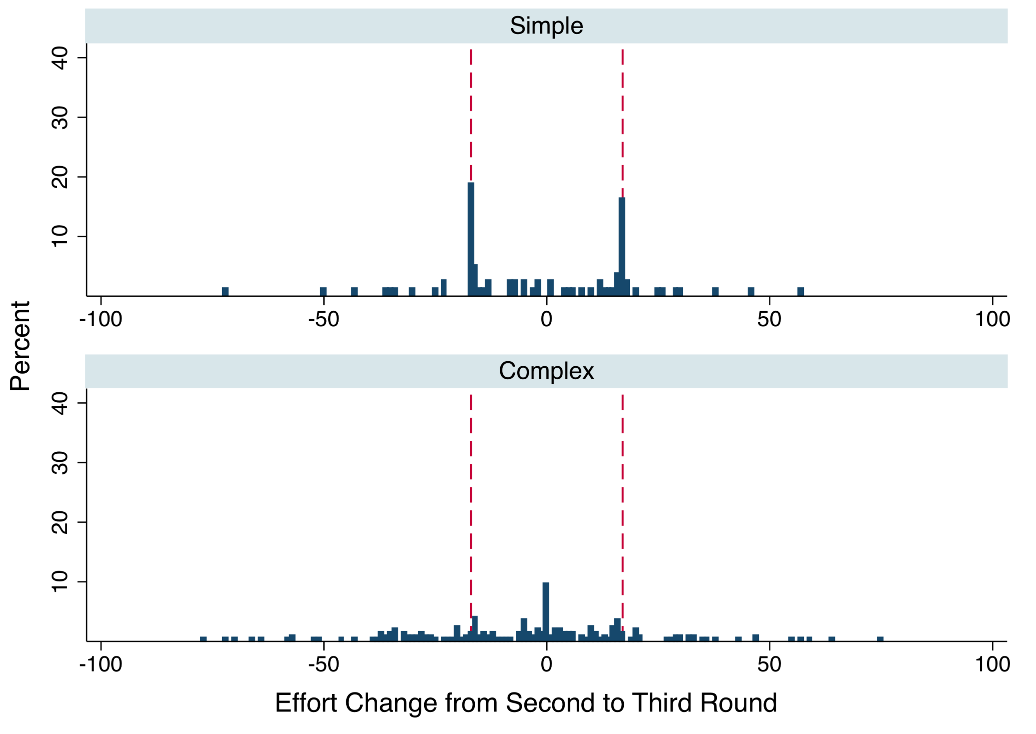
Notes: The dashed line marks the payoff-maximizing choice at 25.

Figure 10: Histogram of Effort Changes from First to Second Round



Notes: The dashed line marks the optimal change in effort coming from a payoff-maximizing effort in the previous round. Which effort change is optimal depends on the choice order.

Figure 11: Histogram of Effort Changes from Second to Third Round



Notes: The dashed line marks the optimal change in effort coming from a payoff-maximizing effort in the previous round. Which effort change is optimal depends on the choice order.

Table 6: Baseline Effect of Complexity

Dependent variable:	1 If payoff-maximizing choice			Distance to payoff-maximizing choice			Earnings		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1 if CT	-0.375***	-0.340***	-0.354***	14.930***	13.613***	14.256***	-370.389***	-336.206***	-338.801***
	(0.046)	(0.064)	(0.060)	(2.068)	(3.047)	(3.041)	(68.592)	(107.05)	(108.390)
1 if Choice Order A-B-C		0.083	0.092		-3.146	-3.571		100.6	103.876
		(0.091)	(0.081)		(3.941)	(3.821)		(106.815)	(110.384)
CT * A-B-C		-0.070	-0.082		2.68	3.31		-68.2	-80.748
		(0.091)	(0.082)		(4.092)	(3.983)		(135.075)	(138.320)
Age			0.004			0.245			-35.993
			(0.006)			(0.458)			(27.479)
1 if Female			-0.082***			2.791*			-40.494
			(0.024)			(1.504)			(73.885)
IQ Measure			0.053***			-2.007***			23.336
			(0.013)			(0.737)			(32.635)
Constant	0.392***	0.350***	0.327**	5.138***	6.687**	-0.094	1611.358***	1561.058***	2299.397***
	(0.046)	(0.064)	(0.129)	(1.865)	(2.866)	(9.504)	(51.221)	(89.469)	(549.261)
N.Obs.	831	831	831	831	831	831	831	831	831

Notes: OLS and tobit estimates. The dependent variable is, in Columns 1–3, a dummy for choosing the payoff-maximizing effort level (OLS); in columns 4- 6, the distance to the payoff-maximizing choice (tobit with lower limit at 0); and in Columns 7–9, the earnings from the decision measured in British pence per round (OLS). Each subject enters the regression three times, standard errors computed by block bootstrap clustered at the subject level are in parentheses. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

B Real-effort Costs

The idea behind having a real-effort task linked to the effort choice was to make the experiment less abstract and more psychologically meaningful. At the same time, such a real-effort task introduces an effort cost of actually moving the sliders which might be heterogenous across subjects and might influence the experimental results. It turns out, however, that in our case the real-effort costs are not big enough to overturn the experimental incentives, partly because subjects could use the mouse and the keyboard to move the sliders which reduced the required effort compared to, e.g., Gill and Prowse (2012). After the main part of the experiment, subjects faced an additional phase in which they had the opportunity to move as many sliders as they wanted, up to a maximum of 144 sliders, for a piece rate of 2 points each. 2 points was the marginal incentive around the optima in the main part of the experiment. In this free-choice phase, 82 percent of subjects work to the maximum, i.e., they move 144 sliders. Another 3 percent try but don't manage to finish in the alloted time (15 minutes). The real-effort cost must therefore be below 2 points per slider for the vast majority of subjects and should not hinder subjects from choosing the (financially) optimal effort level. Moreover, the behavior of subjects in CT and ST does not differ in this phase (p=0.205) and subjects in CT and ST do not differ in their agreement or disagreement with the statement "After having decided on the number of sliders, actually positioning the sliders was very stressful." (p=0.299). We conclude that the real-effort cost is too small to have influenced the effects of the treatment.

C Instructions (Online Appendix)

[Screen 1]

Thank you for participating in this experiment. For your arrival on time, you will receive £2.50 that will be paid to you at the end of the experiment in addition to all earnings from the experiment. If you use the computer in an improper way you will be excluded from the experiment and from any payment. Please turn off mobile phones now and leave them turned off throughout the experiment. If you have a question during the course of the experiment, please raise your hand and we will come to your place and answer your question in private. Please refrain from communicating with other participants throughout the experiment.

During the experiment you can earn points; points will be converted into pence at a rate of 7 to 1; that means that you receive 1p per 7 points. The sum of all points that you earned will be paid out to you at the end of the experiment.

The experiment consists of four parts in which you will position sliders on the screen and a questionnaire after the main part of the experiment.
Only once the experimenter tells you so, press OK to proceed to the next screen.

[Screen 2]

In the four parts of the experiment, your task will consist of positioning sliders on screens containing 48 sliders. Each slider is initially positioned at 0 and can be moved as far as 100. Each slider has a number to its right showing its current position. You can use the mouse or keyboard (arrow keys) in any way you like to move each slider. You can readjust the position of each slider as many times as you wish.

Your payment depends on the number of sliders positioned at exactly 50. We will call a slider positioned at 50 a "correct slider". You will always get a piece rate per correct slider but you will also have to pay taxes and/or receive subsidies depending on the number of correct sliders.

Do you have any questions at this point?

Before the main experiments starts, please answer a couple of example questions on the next screens.

Please press OK now to proceed to the next screen.

[Control Question 1]

Consider the following example:

You receive a piece rate of 20 points per correctly positioned slider.

A constant **tax** of 5 points is levied on each slider. That means that for each correct slider, you will have to pay a tax of 5 points.

Please answer the following questions and click OK. If one of your answers turns out to be not correct, please try again or ask the experimenter for help.

- Suppose you had positioned 3 sliders correctly. What is the total amount in piece rates that you would receive for the three sliders together?
- Suppose you had positioned 3 sliders correctly. What amount of taxes do you have to pay for the third slider?
- Suppose you had positioned 3 sliders correctly. What are your net earnings (piece rate - taxes) for the third slider?

[Control Question 2]

Consider the following example:

You receive a piece rate of 20 points per correctly positioned slider.

There is an increasing **tax** per slider. This tax is 2 points for the first slider and goes up by 2 points per additional slider. That is, it implies a tax of 4 points for the second slider, 6 points for the third slider, and so on.

Please answer the following questions and click OK. If one of your answers turns out to be not correct, please try again or ask the experimenter for help.

- Suppose you had positioned 4 sliders correctly. What amount of taxes do you have to pay for the fourth slider?
- Suppose you had positioned 4 sliders correctly. What are your net earnings (piece rate - taxes) for the fourth slider?

[Control Question 3]

Consider the following example:

You receive a piece rate of 20 points per correctly positioned slider.

There is an increasing **subsidy** per slider that you get on top of the piece rate. This subsidy is 2 points for the first slider and goes up by 2 points per slider. That is, it implies a subsidy of 4 points for the second slider, 6 points for the third slider, and so on.

- Suppose you had positioned 4 sliders correctly. What amount of subsidies do you receive for the fourth slider?
- Suppose you had positioned 4 sliders correctly. What are your net earnings (piece rate + subsidies) for the fourth slider?

[Control Question 4]

Consider the following example:

You receive a piece rate of 20 points per correctly positioned slider.

There is an increasing **tax** per slider. This tax is 6 points for the first slider and goes up by 6 points per additional slider. That is, it implies a tax of 12 points for the second slider, 18 points for the third slider, and so on.

- How many sliders should you position correctly to maximize your financial earnings?

Hint: Your payment is maximized if you position sliders correctly up to just before the point at which the taxes per slider are higher than the piece rate.

[Control Question 5]

Please position the three sliders exactly at 50.

[Instructions – Round One]

The main part of the experiment starts now.

In this part of the experiment, you will work on the slider task. You will first learn about the piece rate and the particular taxes and subsidies that are relevant for this stage. You will then decide how many sliders you want to position, taking the level of the piece rates and all taxes/subsidies into account.

You have 10 minutes to read these rules and decide on the number of sliders before the actual slider task begins.

Once you have decided how many sliders you want to position, there is no time constraint for actually positioning the sliders.

You receive a baseline piece rate of **100 points** per slider positioned correctly. But you also have to pay a number of taxes and can receive a number of subsidies depending on the number of sliders positioned correctly. All taxes and subsidies are added together.

Your payment is maximized if you position sliders correctly up to the point at which the taxes per slider are higher than the piece rate plus the potential subsidies.

If you have any questions, please raise your hand. If not, please press OK to see the taxes and subsidies for this stage.

[Tax Rules – Round One (Complex Treatment)]

You receive a constant **subsidy** for each slider positioned correctly. This subsidy remains constant at 7 points per slider.

You receive a **subsidy** for sliders 1 through 10. This subsidy is 20 points for the first slider positioned correctly and decreases by 2 points per additional slider (until slider 10). That is, you receive a subsidy of 18 points for the 2nd slider, 16 points for the 3rd slider, and so on. Thus, the subsidy is zero for the 11th slider and remains at zero for all additional sliders.

There is an increasing **tax** starting at the first slider that applies to all additional sliders. This tax is 0 points for the first slider and goes up by 2 points per slider. That is, it implies a tax of 2 points for the second slider, 4 points for the third slider, and so on.

There is a constant **tax** for the following sliders: 12 through 17; 32 through 37; 48 through 52; 78 through 83. This tax is constant at 5 points per slider positioned correctly for the ranges mentioned above and is zero otherwise. There is another constant tax of 7 points per slider for the sliders 31 through 34. And there is a 20 point tax for each slider 7 through 11.

You receive a **subsidy** of 10 points for the following sliders: 6, 7, 8, 16, 17, 18, 19, 20, 21, 69, and 70. You receive a **subsidy** of 15 points for the sliders 51-53 and a **subsidy** of 20 points for sliders 35-37.

You receive an increasing **subsidy** starting at the 6th slider that applies to all additional sliders. This subsidy does NOT apply to sliders 1 through 5. This subsidy is 2 points for the 6th slider and goes up by 2 points per additional slider. That is, you receive a subsidy of 4 points for the 7th slider, 6 points for the 8th slider, and so on.

There is a constant **tax** of 5 points per slider for all sliders after the 66th slider. That is, this tax is zero for all sliders up to and including the 66th slider, and at 5 points for the 67th slider and all additional sliders.

There is an increasing **tax** starting at the 10th slider that applies to all additional sliders. This tax does NOT apply to sliders 1 through 9. This tax is 2 points for the 10th slider and goes up by 2 points per slider. That is, it implies a tax of 4 points for the 11th slider, 6 points for the 12th slider, and so on.

There is a **tax** of 10 points per slider for the following sliders: 20, 21, 48 , 49, 74, 75, 90, and 91.

[Tax Rules – Round One (Simple Treatment)]

You receive a constant **subsidy** or each slider positioned correctly. This subsidy remains constant at 15 points per slider.

There is an increasing **tax** starting at the first slider that applies to all additional sliders. This tax is 0 points for the first slider and goes up by 2 points per slider. That is, it implies a tax of 2 points for the second slider, 4 points for the third slider, and so on.

[Instructions for Slider Task – Round One]

You have decided to position ___ sliders correctly. You will have to do so on the next screens. The number of sliders you have already positioned and the remaining time are shown on the top of the screen. Once you have positioned ___ sliders correctly, press OK to proceed to the next stage of the experiment.

Note: For a slider to count as "correct slider" it has to be positioned exactly at 50.

If you decided to position more than 48 sliders, please position all 48 sliders on the screen and press OK. Then a second screen will be shown with the remaining sliders.

Please click OK now to start positioning the sliders.

[New Instructions – Round Two]

In the next stage of the experiment, all tax and subsidy rules from the previous stage are still applicable. However, there is an additional rule applicable and this rule also determines your payment.

You have 4 minutes to read these rules and decide on the number of sliders before the actual slider task begins.

Once you have decided how many sliders you want to position, there is no time constraint for actually positioning the sliders.

[New Tax Rules – Round Two]

[ABC] In addition to the previous rules, there is a constant **tax** for each slider. This tax remains constant at 32 points per slider.

[ACB] In addition to the previous rules, there is a constant **tax** for each slider. This tax remains constant at 66 points per slider.

[Instructions for Slider Task – Round Two]

[identical to Instructions for Slider Task – Round One]

[New Instructions – Round Three]

[identical to New Instructions – Round Two]

[New Tax Rules – Round Three]

[ABC] In addition to the previous rules, there is a constant **tax** for each slider. This tax remains constant at 34 points per slider.

[ACB] In addition to the previous rules, there is a constant **subsidy** for each slider. This subsidy remains constant at 34 points per slider.

[Instructions for Slider Task – Round Three]

[identical to Instructions for Slider Task – Round One]

[Instructions Productivity Test]

In this last part of the experiment, you will work on the slider task again. In contrast to before, you do not have to commit in advance to how many sliders you will do.

You will be paid 2 points per slider positioned at exactly 50. **There are no taxes or subsidies in this stage.**

You can work for a maximum of 15 minutes on this task but even if you have time left, you can only do up to 144 sliders and earn the respective piece rates.

Please click OK to start this stage of the experiment.

[Feedback]

The main part of the experiment is now over. Your total earnings from the four stages is ___ points.

We would now like to ask you to fill in a short questionnaire. After the questionnaire, the experiment is over and you will be paid.