

Complex Tax Incentives[†]

By JOHANNES ABELER AND SIMON JÄGER*

How does 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 tax system, the other a complex one. Subjects' economic incentives are 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. The underreaction is stronger for subjects with lower cognitive ability. Contrary to predictions from models of rational inattention, subjects are equally likely to ignore large or small incentive changes. (JEL D14, H24, H31)

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, nonlinear schedules with kinks, thresholds, and various exemptions. There is growing evidence that people are not able to react optimally to such complexity: 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 et al. (2013) 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. And the United States General Accounting Office (2002) estimates that at least 2.2 million

*Abeler: University of Oxford, Department of Economics, Manor Road, Oxford OX1 3UQ, IZA, and CESifo (e-mail: johannes.abeler@economics.ox.ac.uk); Jäger: Harvard University, Department of Economics, Littauer Center, 1805 Cambridge Street, Cambridge, MA 02138, and IZA (e-mail: jaeger@fas.harvard.edu). 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. Ethical approval for the experiment was obtained from the Nottingham School of Economics Research Ethics Committee.

[†]Go to <http://dx.doi.org/10.1257/pol.20130137> to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

tax payers overpay federal taxes simply by not itemizing deductions.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 dataset 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 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.

We take the number of distinct tax rules a subject faces in the experiment as a measure of tax complexity. This is consistent with measures of complexity of real-world tax systems. The Office of Tax Simplification (OTS) in the UK, for example, uses a complexity index that includes the number of pages of legislation, the readability of the legislation, as well as the number of exemptions (OTS 2013).² Alternatively, one can take the compliance cost implied by or the ability to comply at all with the tax regulations as a measure of complexity (Slemrod 1984). As mentioned above, many tax payers are not able to optimally react to tax systems, arguably because of the inherent complexity. Direct compliance costs are also substantial in most developed countries, again implying a high level of complexity.³ In our experiment, we will be able to check how well subjects deal with the tax system and measure, e.g., the time spent deciding.

In the experiment, subjects work on a 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 between-subjects treatments, each subject facing three rounds. In the first round of the simple treatment (ST), the tax system features only two 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

¹Chetty 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 nonbunching 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. Gideon (2013) provides evidence that perceived tax rates predict savings behavior more strongly than true rates. A related and growing strand of literature documents evidence suggesting that taxpayers react to average rather than marginal tax rates (de Bartolome 1995; Liebman and Zeckhauser 2004; Feldman and Katuscak 2009); Ito (2014) documents similar results for consumer responses to a nonlinear electricity pricing scheme. Chetty, Friedman, and Saez (2013) demonstrate substantial heterogeneity in the knowledge of the EITC program across regions in the United States (see also Wuppermann, Bauhoff, and Grabka 2014 and Bhargava and Manoli 2013 for further field evidence on the response to complexity).

²Similarly, Slemrod and Kopczuk (2002) and Kopczuk (2005) characterize an income tax system as complex when it features many deductions while Wagner (1976) defines a tax system as complex when its revenue structure is dispersed. The Office of the Taxpayer Advocate within the IRS reports that the US tax code "has grown so long that it has become challenging even to figure out how long it is" (OTA 2010). Even the simplified IRS "Tax Guide 2012 for Individuals" features almost 300 pages (the Internal Revenue Code runs to about 11,000 pages, OTA 2010). The complexity of tax systems also varies strongly between countries. For instance, corporate income tax codes (for which comparable data exist, e.g., Organisation for Economic Co-operation and Development (OECD) 2007 or PricewaterhouseCoopers (PwC) 2014) differ between 700 pages in Sweden and 5,100 pages in the United States.

³Slemrod and Sorum (1985) estimate compliance costs of 5 to 7 percent of revenue raised in the United States (see also Blumenthal and Slemrod 1992); recent estimates from the Office of the Taxpayer Advocate (2010) are in a similar order of magnitude at 11 percent.

ST. In particular, the payoff-maximizing output level and the payoffs around this level are identical. However, there are many more tax rules in CT (22 rules) 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 modeling 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 preexisting 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? To verify our treatment manipulation, 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 higher complexity of the decision environment. Our main result is that subjects in the complex treatment underreact to the newly introduced tax rules and do not adjust their output 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 9 percentage points higher than the corresponding fraction in ST. These subjects drive most of the underreaction in CT. The nonreacting 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. We find that subjects with higher cognitive ability—as measured by a proxy for IQ that we elicit—underreact less in CT, suggesting that the effect of complexity on the response to new taxes is particularly pronounced for individuals with lower cognitive ability. 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.⁴

⁴Note that our experiment primarily captures short-term reactions. While we did not see an increase of decision quality across rounds, the time horizon of the experiment was limited and subjects did not receive feedback on

We discuss a conceptual framework that can help to organize our results.⁵ 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 a larger fraction of subjects in CT compared to ST 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 that predict that information is 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 underreact to changes in nonsalient taxes (see also Goldin and Homonoff 2013, and Feldman and Ruffle 2015). Finkelstein (2009) shows that automating toll collection—which lowers the salience of the toll—leads to a reduction in the toll elasticity of driving.⁶ 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. More generally, our findings support the view that complexity of the decision environment is an important catalyst of behavioral anomalies, such as the status-quo bias

performance between rounds. Field evidence, e.g., Chetty, Friedman, and Saez (2013) or United States General Accounting Office (GAO) (2002), shows that even experienced tax payers are often not able to react optimally to the tax incentives they face. In addition, the learning that can occur based on net-of-tax information hinges on the informativeness of experimentation or comparison with other; if the tax environment is more complex, feedback from experimentation and comparison will be less informative. This suggests that learning might be limited in important ways and that complexity can have persistent effects on decision-making.

⁵Recent 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 (2013); Bordalo, Gennaioli, and Shleifer (2012); Persson (2012); Dahremöller and Fels (2012); Caplin and Dean (2013); and Ortoleva (2013). For laboratory experiments on these issues see, e.g., Wilcox (1993); Huck and Weizsäcker (1999); Gabaix et al. (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).

⁶For related papers studying nonsalient aspects of consumer products, e.g., shipping costs on eBay, see Gabaix and Laibson (2006) and Brown, Hossain, and Morgan (2010).

(Samuelson and Zeckhauser 1988; Kahneman, Knetsch, and Thaler 1991; Fleming, Thomas, and Dolan 2010).

A sizable literature—going back to at least Mill (1848)—analyzes how the complexity of the tax system or the salience of particular taxes affects the political economy of taxation and hypothesizes that tax complexity leads to a misperception of taxes which, in turn, affects voters' attitudes toward taxation and the appropriate size of the government.⁷ While our paper does not focus on the political economy of taxation, our experiment provides evidence that increasing the complexity of the tax code leads to less accurate perceptions of tax incentives which is in line with one of the mechanisms posited in the “fiscal illusion” literature.

At a broader level, our study adds to a growing literature documenting the effects of framing on decision-making as the change in complexity across treatments in our experiment can be construed as a change in the framing of the decision problem given that we leave the economic environment unchanged across treatments.⁸ In line with our findings, Fehr and Tyran (2001) document—in the context of an experiment on money illusion—that individuals underreact to a price change when the framing of the decision problem is in nominal terms, thus shrouding the underlying real incentives. You and Zhang (2009) provide correlational evidence that investors underreact to annual SEC reports; underreaction is stronger for more complex reports as proxied by a word-count.⁹

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 nonlinear income tax schedule. Our experiment shows that introducing more complexity comes at a cost: facing a complex tax schedule, fewer subjects choose the payoff-maximizing level of output. This implies that tax complexity has both compliance and “decision quality” costs and that these costs need to be taken into account in welfare calculations. In the political discourse, there has been an ongoing debate to reduce tax complexity in a number of countries.¹⁰

⁷In recent contributions to this literature, Cabral and Hoxby (2012) provide evidence that lower salience of the property tax is associated with higher property tax rates. Sausgruber and Tyran (2005) conduct a laboratory experiment on the effects of direct versus indirect taxation and find that less “visible,” indirect taxation leads to an underestimation of the tax burden. Wagner (1976) provides evidence consistent with the hypothesis that citizens' perception of the cost of government are distorted in more complex tax systems (see also Pommerehne and Schneider 1978).

⁸Put differently, the treatments differ primarily in “perceived” complexity (Wagner 1976) as opposed to “fundamental” complexity as CT features a complex representation of a tax system that is economically almost identical to the one in ST.

⁹For more evidence on the psychology of framing effects, see Tversky and Kahneman (1981) and Tversky and Kahneman (1986). Complementary to our findings, DellaVigna and Pollet (2009) and Hirshleifer, Lim, and Teoh (2009) document underreactions to earnings announcements when investors face higher information load. Carlin, Kogan, and Lowery (2013) conduct a laboratory experiment in which increasing asset complexity leads to less trade volume in an experimental asset market.

¹⁰See Rohaly and Gale (2004) and Gravelle and Hungerford (2012) for a discussion of proposals to simplify the tax code in the United States. Sunstein (2011) discusses several examples of recent US governmental regulation and describes how simplification could improve regulatory outcomes. Our paper implies that a simplification of the tax code will affect the behavioral response to taxation (and increase the response to tax reforms) beyond any direct effects on compliance costs which are typically the primary focus of analyses of tax simplification. Attempts to simplify the tax code, e.g., by eliminating deductions, are politically challenging as groups who benefit from particular

In addition to documenting the costs of tax complexity, our findings can also help to inform how complexity can be explicitly used as a tool by policymakers.¹¹ 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. Reducing the response to taxation 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.¹² But it could be harmful in the case of a tax on a socially undesirable activity, e.g., polluting. As some subjects are particularly strongly affected by complexity in our experiment and cognitive ability predicts underreaction to new incentives in CT, 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. Section I describes the design of the experiment. We present results in Section II. We discuss a framework to organize our results in the Section III. The last section concludes.

I. Experimental Setup

A. 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 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 is defined as the number of sliders positioned at exactly 50. This task is attractive because it is remarkably simple and does not require preexisting 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 A for details). The aim of the task is mainly to make the decision situation less abstract and psychologically more meaningful.

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

provisions have strong incentives to oppose simplification. An analysis of previous attempts to simplify the tax code shows only limited success (e.g., Blumenthal and Slemrod 1992 and Slemrod 1992).

¹¹ 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.

¹² More generally, this is in line with the theory of the second best (Lipsey and Lancaster 1956): If there exists one distortion in the economy, for instance, a distortive income tax system, it can be optimal to introduce an additional distortion, such as tax complexity, even when that distortion comes itself at a cost.

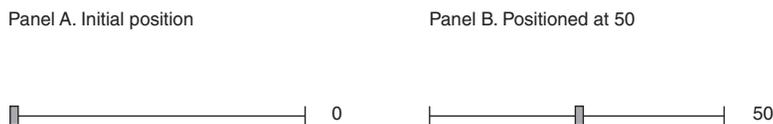


FIGURE 1. SCHEMATIC REPRESENTATION OF A SLIDER (GILL AND PROWSE 2011)

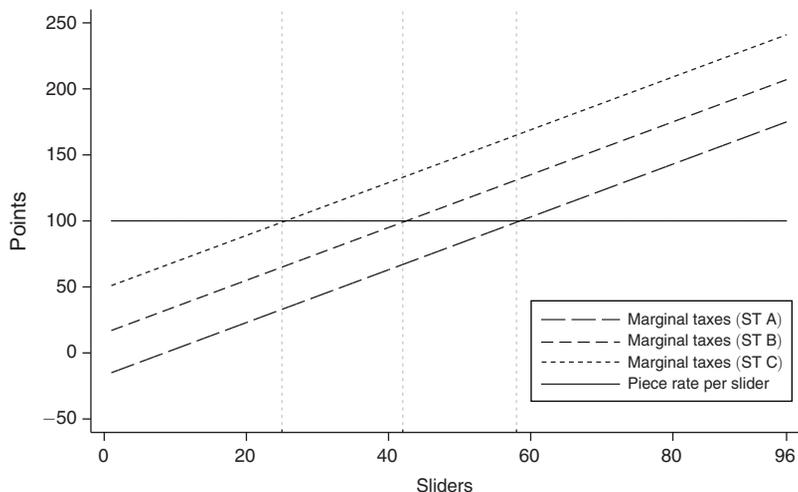


FIGURE 2. MARGINAL TAXES FOR DIFFERENT TAX SCHEDULES IN SIMPLE TREATMENT

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.

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 that applies to a sequence of adjacent sliders, the tax schedule in the first round of ST contains two rules; the one in CT contains 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 displays the marginal taxes subjects face in ST (see Figure 4 and online 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.

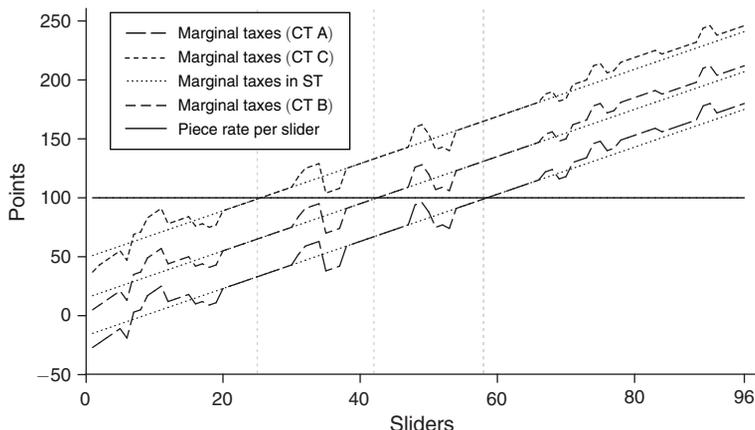


FIGURE 3. MARGINAL TAXES FOR DIFFERENT TAX SCHEDULES IN THE COMPLEX TREATMENT

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.

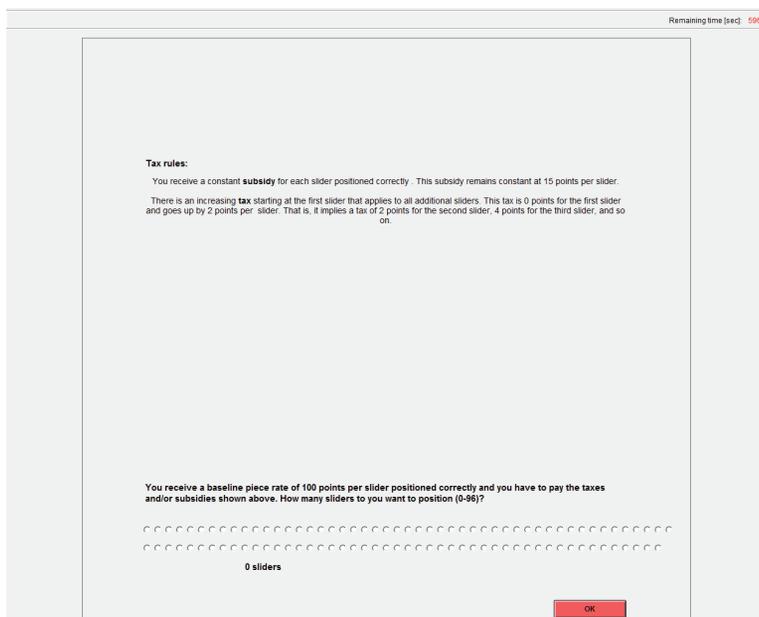


FIGURE 4. SCREENSHOT OF THE TAX SCHEDULES FOR ROUND 1 IN ST

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

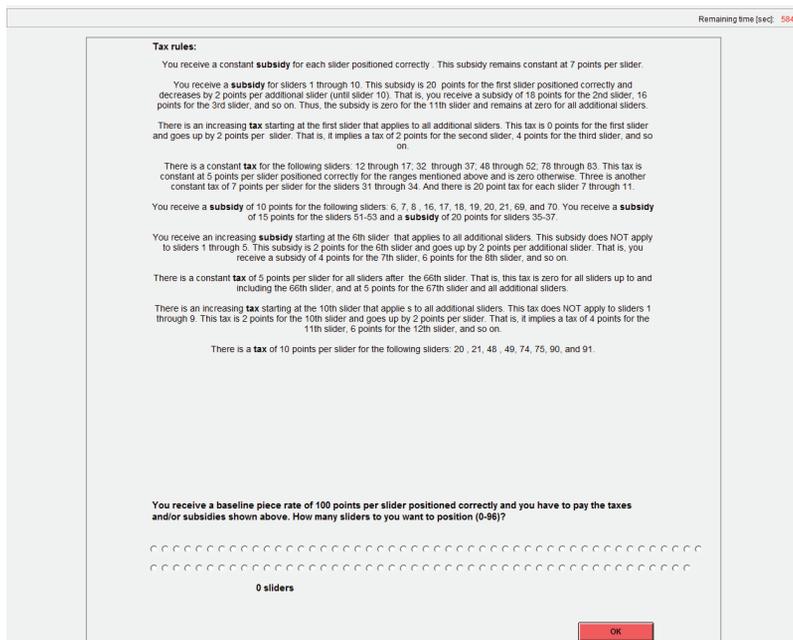


FIGURE 5. SCREENSHOT OF THE TAX SCHEDULES FOR ROUND 1 IN CT

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.

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 (see Figure 5 and online Appendix C). Schedule "CT A" in Figure 3 displays the marginal taxes subjects in CT face in the first round.¹³

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 3 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

¹³ We chose a relatively high level of complexity in CT to approximate the highly complex tax systems in the real world. Moreover, we wanted a notable difference in complexity between ST and CT to eliminate any doubt whether CT is indeed more complex than ST. Result 1, is a further "first-stage" check whether our treatment manipulation of the level of complexity worked as intended.

TABLE 1—PARAMETERS OF THE TAX SCHEDULES IN THE SIMPLE AND COMPLEX TREATMENT

Round	Number of applicable rules	Payoff-maximizing number of sliders		Increase in marginal tax per slider
		Order of schedules		
		A-B-C	A-C-B	
<i>Panel A. Simple treatment</i>				
1	2	58	58	2
2	3	42	25	2
3	4	25	42	2
<i>Panel B. Complex treatment</i>				
1	22	58	58	2
2	23	42	25	2
3	24	25	42	2

Notes: We define a rule as a linear tax that applies to a sequence of adjacent sliders. The increase in the tax per slider in the complex treatment is measured in a neighborhood of at least four sliders around the payoff-maximizing number of sliders.

the complex and the simple treatment. Secondly, the total payoff generated at each respective payoff optimum is also identical in both treatments (schedule A: 3,364; schedule B: 1,764; 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 3 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.

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, we can use two methods to uncover the “true” preferences for welfare analysis. Since we know the payoff function, we can take the theoretically possible earnings as benchmark to calculate the forgone earnings in CT. We could also take realized choices in ST as benchmark treating it as a measure of what a subject can realistically be expected to achieve in this setting.

B. Timeline

At the beginning of the experiment, subjects are familiarized with the task. They then face several control questions that test whether they understand their potential payoff for a given number of correctly positioned sliders in several hypothetical tax and subsidy regimes. These tax regimes confront subjects with increasing marginal tax 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. 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 10 minutes to read the rules on 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 nine and a half 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 specified 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 four 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.

C. 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 (subjects were not told this last piece of information explicitly). The average payment per subject was £9.02 (approximately USD 14 or EUR 10.50 at the time of the experiment). The average duration of the experiment was 50 minutes.

II. Results

In this section, we present results of the experiment and discuss possible explanations for the observed behavior. Before we discuss the main results, we verify

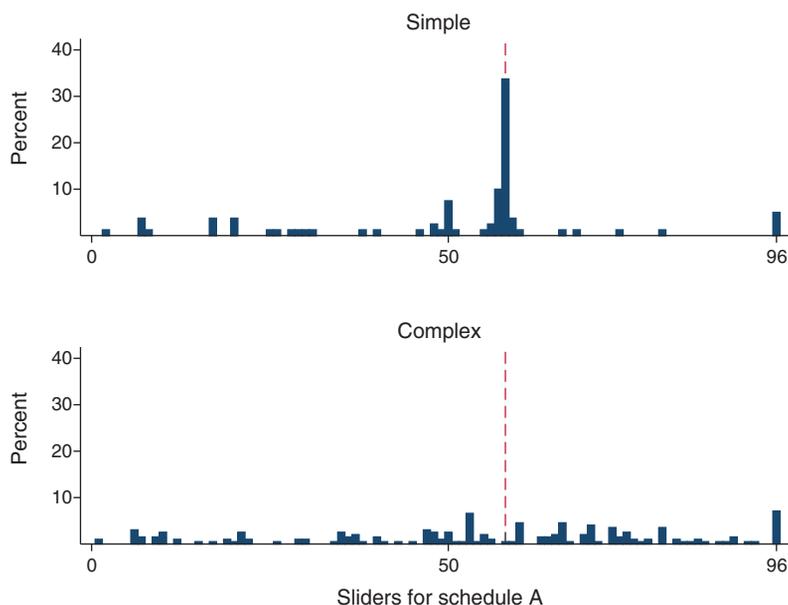


FIGURE 6. HISTOGRAM OF CHOICES FOR SCHEDULE A

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

whether our treatment manipulation was able to effectively increase the complexity of the situation.

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 6 depicts histograms of the choices for schedule A in the two treatments. 34 percent of subjects in ST choose exactly the payoff-maximizing output level. In contrast, only 1.7 percent of choices in CT are payoff-maximizing. 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 7 document. The median subject in ST earns 98 percent of maximally attainable earnings and total earnings are 23.0 percent lower in CT than in ST (we provide more summary statistics about performance by treatment in Table 6 in online Appendix B).

Table 7 in the online Appendix shows that these differences in behavior are highly significant in a regression framework with a set of control variables and the following dependent variables: a dummy which equals 1 if the subject chose the payoff-maximizing output level in a given round; the absolute distance to the payoff-maximizing choice for each decision; and subjects' earnings from each decision. Subjects in CT are less likely to choose the optimum, are generally further away from the

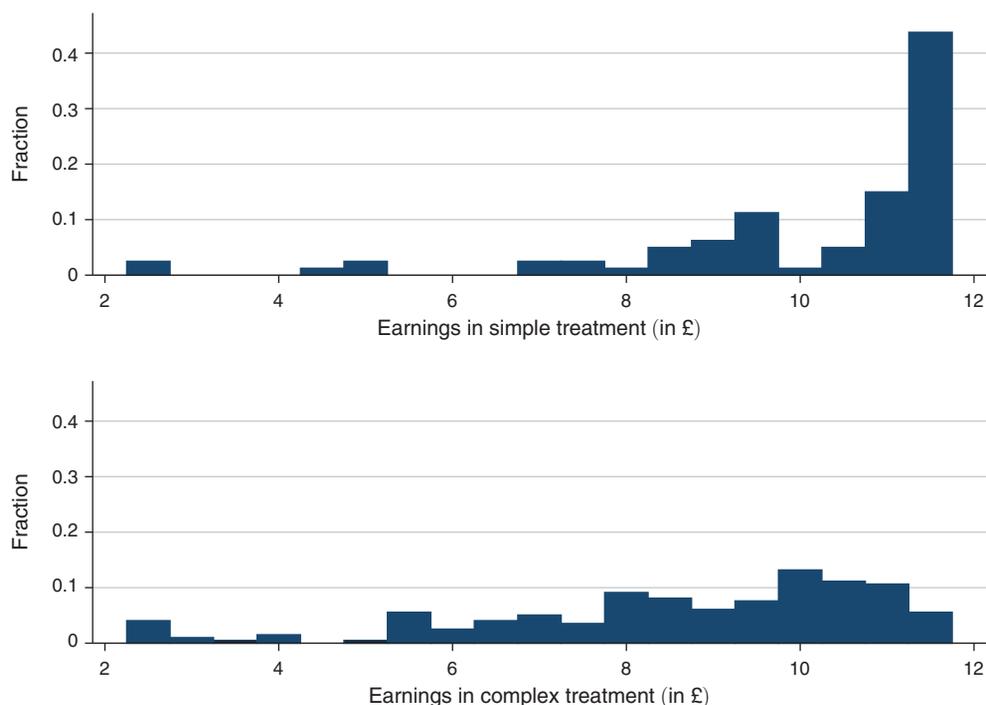


FIGURE 7. HISTOGRAM OF EARNINGS BY TREATMENT

optimum and earn less money.¹⁴ 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),¹⁵ 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 effort costs of moving the sliders are positive (if small) 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.

¹⁴In the framework of Bernheim and Rangel (2009), this means that subjects in CT underperform compared to both of our benchmarks: they realize only 65 percent of theoretically possible earnings and 77 percent of the average earnings in ST (see also Table 6 in the online Appendix). Differences in real-effort cost of moving the sliders do not add to these welfare differences since the total number of sliders moved across all three rounds is not significantly different between treatments (*t*-test, $p = 0.290$). In contrast, subjects in CT spend much longer thinking about their choice (see Result 5) and thus incur higher “mental” effort costs. We calculate a proxy for the time cost of complying with complexity in CT at around £1.16, quite considerable compared to the treatment difference in monetary earnings of £1.51 (see below).

¹⁵To derive these numbers, we take the actual choice in round $t - 1$ and check whether a subject changed their output level in the direction of the payoff-maximizing level in round t . If we alternatively assume that subjects chose or thought they had chosen the payoff-maximizing level in round $t - 1$ and check whether their direction of adjustment is in line with how the payoff maximum moved, the resulting shares are very similar: 68 percent in CT and 86 percent in ST adjust in the correct direction.

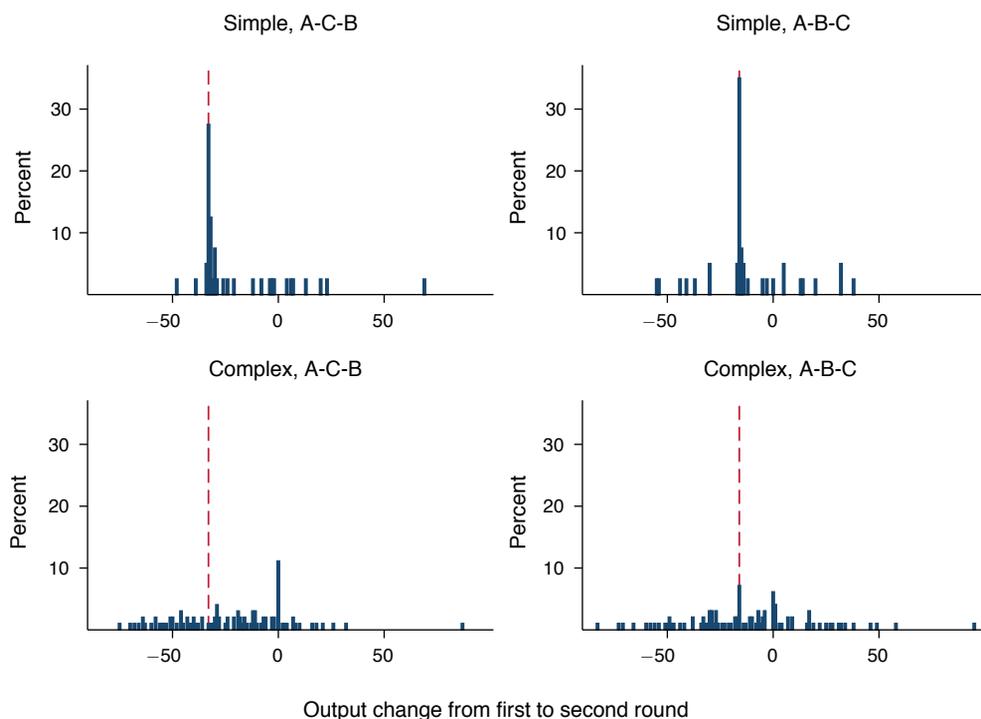


FIGURE 8. HISTOGRAM OF OUTPUT CHANGES FROM FIRST TO SECOND ROUND

Note: The dashed line marks the optimal change in output coming from a payoff-maximizing output in the previous round.

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.

Figures 8 and 9 depict histograms of the change of output from round to round. The histograms are split by the choice order (A-B-C or A-C-B) and by treatment (CT or ST). The dashed line marks the optimal change in output coming from a payoff-maximizing output in the previous round. One can clearly see that many subjects in ST are able to hit this optimum while subjects in CT are less able to do so and often stick to their previous choice.

Also overall, subjects in CT react less than subjects in ST. In Table 2, columns 1–3, we take the change in output 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 output changes comparable across treatments, the output change in the second decision in the choice order A-C-B is multiplied by -1 . It is thus always optimal to *reduce*

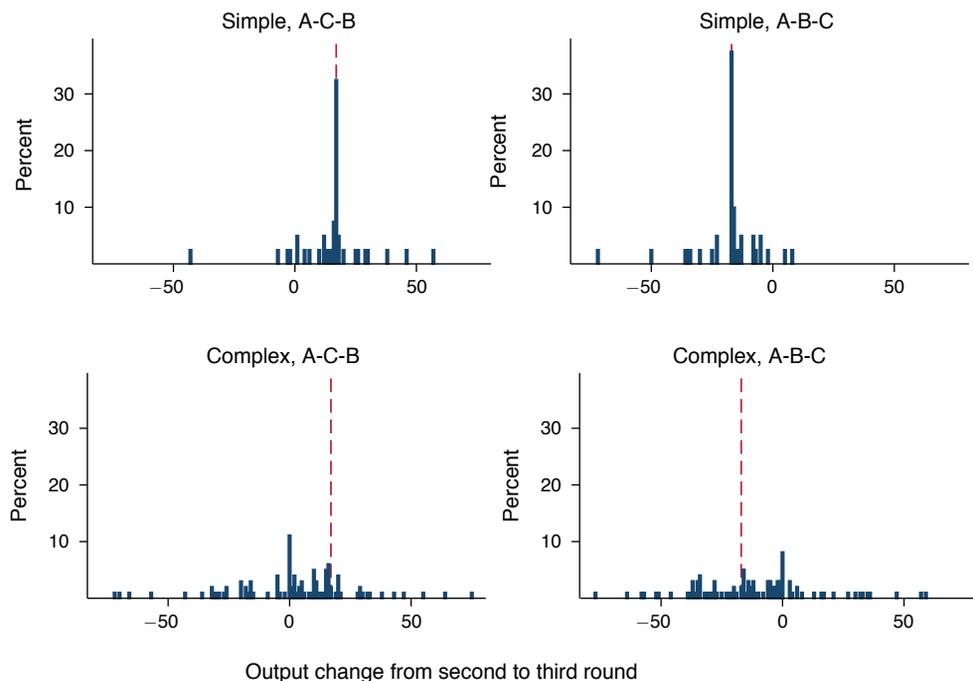


FIGURE 9. HISTOGRAM OF OUTPUT CHANGES FROM SECOND TO THIRD ROUND

Note: The dashed line marks the optimal change in output coming from a payoff-maximizing output in the previous round.

TABLE 2—CHANGE IN OUTPUT LEVEL FROM PREVIOUS ROUND

Dependent variable:	Change in output level from previous round			1 if subject chose same output as in previous round		
	(1)	(2)	(3)	(4)	(5)	(6)
1 if CT	5.00*** (1.53)	5.00*** (1.60)	5.34*** (1.72)	0.09*** (0.02)	0.09*** (0.02)	0.09*** (0.02)
1 if choice order A-B-C		1.00 (1.13)	0.91 (1.62)		-0.02 (0.02)	-0.02 (0.02)
Age			1.39* (0.73)			0.00 (0.01)
1 if female			2.07 (1.68)			-0.02 (0.02)
IQ measure			-0.30 (1.05)			-0.01 (0.01)
Constant	-17.00*** (0.45)	-17.00*** (0.98)	-45.51*** (15.04)	0.01 (0.01)	0.02 (0.01)	-0.04 (0.17)
Observations	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 output level from the previous round; the output 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 output level, an output decrease is always optimal. The dependent variable in columns 4–6 is a dummy equaling 1 if the subject chose the same output 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.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

the output 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 add in subsequent regressions controls for the choice order (A-B-C or A-C-B) and for age, gender, and IQ.¹⁶ Since output choices and also changes in output choices are spread out over the whole range of possible choices, we use median regressions to limit the influence of extreme outliers. We include both output 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 output by subjects in the simple treatment is -17 , which is actually the optimal reaction if they started from a payoff-maximizing choice. The median change in output 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.

A useful way of assessing the magnitude of the differences in reactions to new taxes is to calculate the ratio of reactions in CT and ST. This is analogous to Chetty, Looney, and Kroft (2009) who define a measure of underreaction θ as the ratio of (nonsalient) tax elasticities and (salient) price elasticities. Based on the results in column 1 of Table 2, we arrive at a ratio of reactions in CT and ST of $\theta = 0.706$, implying that the impact of new taxes is attenuated by about 30 percent in CT. One way to benchmark this attenuation against the magnitude of the increase in complexity is to consider the increase in the number of rules which increases tenfold from 2 to 22. A different quantification of the increase in complexity can be gained by comparing decision times, which we analyze in more detail below. The median subject in the first round of CT takes about 2.8 times as long to reach a decision as the median subject in ST, suggesting that complexity in CT is about 3 times as high as in ST.

We also explore whether complexity affects subjects' tendency to completely ignore the change in incentives across rounds. A spike at zero in the complex treatment is clearly visible in Figures 8 and 9. We can also show this effect in a regression analysis.

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 output choice unchanged.

In Table 2, columns 4–6, the dependent variable is a dummy that equals 1 if a subject chose the same output as in the previous round. The estimates show that subjects in ST almost never choose the same output as in the previous round. This happens

¹⁶We 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 et al. 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.

only once out of the 160 output changes in ST (0.6 percent of output changes). In contrast, 9.1 percent of output changes in CT equal zero. These treatment differences are highly significant and robust to the inclusion of the aforementioned control variables.

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 attenuated 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).

To shed light on these status quo effects, we regress the time subjects in CT needed for their decision making in a given round on a dummy variable indicating whether a subject chose the same output level as in the previous one. Subjects who did not adjust their output choices across rounds took on average more than one minute less to make their decision than subjects who changed their output 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.

Most of the underreaction is driven by the fact that more subjects in CT do not change their output at all. When we repeat the regressions of Table 2, columns 1–3, restricting the sample to subjects who changed output levels from round to round, the point estimates are much smaller at 2 (columns 1 and 2) and 3.67 (column 3).¹⁷ At around half of the effect size measured for the whole sample, these effects in the restricted sample are still sizable and positive but mostly not significant.

As noted above, if subjects change their output level, they most often change it (in both treatments) in the direction of the new payoff maximum. If it is cognitively simpler to know how to qualitatively respond to a change in incentives than to know the optimal response, this is what one would expect.

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 (1,087 points for ignoring the larger tax versus 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.

¹⁷The results of these regressions are reported in Table 8 in the online Appendix. Note that we condition on an outcome variable in these specifications.

TABLE 3—EFFECT OF THE MAGNITUDE OF INCENTIVE CHANGE

	(1)	(2)
<i>Dependent variable: 1 if subject chose same output in first and second round</i>		
1 if choice order A-C-B	0.05 (0.04)	0.05 (0.04)
Age		0.01 (0.01)
1 if female		−0.07* (0.04)
IQ measure		−0.02 (0.02)
Constant	0.06*** (0.02)	−0.09 (0.25)
Observations	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 output as in the previous round. Bootstrapped standard errors are in parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

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

In Table 3, we regress a dummy that equals 1 if a subject chose the same output 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 output 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. Overall, we interpret the results of this exercise as evidence not in line with rational inattention as the 95 percent confidence interval ranges from -0.025 to 0.125 and we can thus reject even quite small negative effects.

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 and third decision.

Figure 10 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.

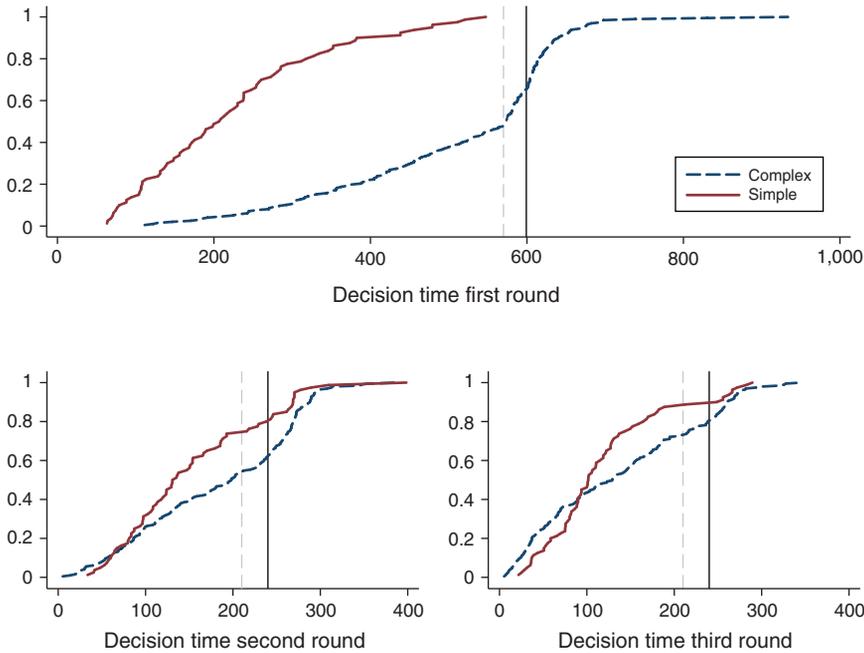


FIGURE 10. DECISION TIMES IN THE THREE ROUNDS (IN SECONDS)

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 10, subjects in CT take, on average, more than twice as long as subjects in ST (511 seconds versus 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. In contrast, average decision times in the second and third round are more similar between treatments. 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).

The differences in decision times between ST and CT allow for a computation of a measure of complexity based on compliance costs (Slemrod 1984). Taking the median difference of 366 seconds of decision time between ST and CT in the first round as an indicator of complexity of the tax schedule that subjects initially face in CT, one can calculate a money metric measure of decision time by valuing the additional time that subjects in CT spend with the payoff per unit of time that subjects earn in the experiment. Based on this calculation, the compliance costs of CT relative to ST would be estimated to be in the order of £1.16 (or USD 1.92) in the

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	285.5*** (17.0)	285.5*** (17.0)	277.8*** (17.9)	30.3*** (10.8)	30.2*** (10.9)	28.4** (11.2)	17.4* (10.0)	17.3* (10.0)	13.7 (9.9)
1 if choice order A-B-C		-7.2 (16.7)	-6.7 (16.4)		-9.9 (10.3)	-10.8 (10.4)		-9.7 (10.2)	-10.1 (10.2)
Age			1.1 (4.4)			-2.7 (3.3)			0.3 (3.5)
1 if female			-18.9 (16.5)			-17.6 (11.2)			-23.0** (10.6)
IQ measure			23.1*** (8.5)			2.6 (5.5)			7.7 (5.1)
Constant	225.2*** (13.7)	228.8*** (15.1)	221.5** (91.2)	153.1*** (8.6)	158.0*** (10.6)	221.3*** (67.9)	117.8*** (7.7)	122.6*** (9.9)	130.6* (69.9)
Observations	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 are in parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

first round.¹⁸ To put this number in perspective, overall earnings are £1.51 lower in CT due to the reduction in decision quality. In our experiment, the decision quality costs—as measured by earnings differences between treatments—are thus in the same order of magnitude as a measure of compliance costs based on decision time. Applied to real tax systems, this suggests that compliance cost measures will understate the overall costs of tax complexity.

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

Result 6: Subjects with a lower IQ measure react less to incentive changes in CT. Subjects in CT who choose output levels further away from the payoff optimum in the first round are more likely to ignore tax changes in the following rounds.

We focus on cognitive ability as one potential source of heterogeneity in subjects' reaction to complex tax incentives. In columns 1–3 of Table 9 in the online Appendix, we take up the specifications in Table 2 and additionally include an interaction of IQ with CT. In all specifications, we find that subjects with a lower IQ underreact significantly more in the Complex Treatment, indicated by the negative interaction effect. IQ does not have a significant effect on the change in output in ST. The magnitude of the effect of IQ in CT is large: given that the baseline effect of IQ is close to 0 in ST, a 2 standard deviation increase in the IQ measure in CT

¹⁸The median subjects earned £9.50 and stayed for about 50 minutes in the lab. This leads to a time cost of $366/(50 \times 60) \times £9.50 = £1.16$ for 366 seconds. As described above, decision times in the second and third round did not differ very much and thus do not add considerably to the treatment difference in compliance cost.

has an effect of the same order of magnitude as CT itself. In columns 4 to 6 of the same table, we document the effects of IQ on the likelihood of not reacting to a new tax rule at all. Here, IQ does not have a significant impact in either treatment; the point estimates in all specifications indicate, however, that low-IQ subjects in CT are slightly more likely to entirely ignore a new tax rule.¹⁹

Note that our IQ measure is a combination of math grade and answers to the Cognitive Reflection Test and to a numeracy test (see footnote 16). It might, thus, not fully capture the aspects of cognitive ability that are most important in our experiment, and measurement error plausibly biases the coefficients of IQ and its interactions downward. An alternative measure that is more closely related to how subjects react to new taxes in our experiment is the distance of a subject's output in the first round to the optimal output level, i.e., a measure of their within-round decision quality. As can be seen from column 6 in Table 7 in the online Appendix, IQ and distance to the optimal output level are highly correlated. Since this measure might be mechanically linked to overall round-to-round reaction, we focus on whether subjects change their output at all. In columns 1 and 2 of Table 5, we regress a dummy variable indicating whether a subject chose the same output level as in the previous round on the distance to the payoff-maximizing choice in the first round (as almost all subjects in ST change their choice of output from round to round we confine our analysis to subjects in CT). The regressions reveal that subjects who choose output levels further away from the payoff optimum in the first round are more likely to ignore tax changes in the following rounds. A 1 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 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.

Overall, our two measures, IQ and distance to the optimum in round 1, provide converging evidence that cognitive ability is an important dimension of heterogeneity in the response to tax complexity in our experiment.²⁰

¹⁹Higher IQ subjects also obtain higher earnings in the experiment, but the significance level and magnitude of this result depends on the exact specification. Splitting the sample at the median of the IQ measure shows a strong positive effect of IQ in both ST and CT (see Table 6 in the online Appendix). The difference is statistically significant in both cases ($p = 0.0049$ and $p = 0.0320$, respectively, in Wilcoxon rank-sum tests). If we regress overall earnings directly on IQ and a number of control variables, we similarly find positive point estimates, but the effect is at most weakly significant (see Table 7 in the online Appendix, column 9). When we consider earnings in the first period as outcome variable (Table 10 in the online Appendix), we find a positive baseline effect of IQ but actually a negative interaction term of IQ and CT. This interaction term is, however, not precisely estimated and the point estimate for the overall effect of IQ on earnings in the first round of CT, i.e., the sum of the baseline and the interaction effect, is small but positive.

²⁰While we focus on cognitive ability, other dimensions of unobserved heterogeneity might also be important for the response to tax complexity. Heckman, Stixrud, and Urzua (2006), for example, have documented the

TABLE 5—WITHIN-SUBJECT CORRELATION OF BEHAVIOR IN THE COMPLEX TREATMENT

Dependent variable:	1 if subject chose same output as in previous round		1 if subject chose same output in rounds 2 and 3	
	(1)	(2)	(3)	(4)
Distance to payoff-maximizing choice (round 1)/100	0.25** (0.12)	0.25** (0.13)		
1 if subject chose same output in rounds 1 and 2			0.35*** (0.12)	0.36*** (0.12)
Constant	0.04 (0.03)	0.03 (0.20)	0.07*** (0.02)	0.16 (0.18)
Control variables	No	Yes	No	Yes
Observations	394	394	197	197

Notes: OLS estimates. The sample is confined to individuals in the complex treatment. Control variables are choice order, age, gender, and IQ. Each subject enters the regression twice for columns 1 and 2; standard errors computed by block bootstrap are in parentheses and are clustered by subject in columns 1 and 2.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

III. 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 fix 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 remaining $N - k$ rules are completely ignored. This sharp drop in attention at k is similar to models of consideration sets (Eliaz and Spiegel 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.²¹ 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.

importance of noncognitive skills. With regard to tax complexity, a personality trait such as tenacity or grit (e.g., Duckworth et al. 2007) would be a plausible candidate for predicting how thoroughly an individual strives to understand his or her tax incentives.

²¹This 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 (2013). The ordering of rules could also be randomly chosen, but would then be fixed.

If we assume that $4 \leq k_i < 22$ 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. For $k_i \geq 4$, subjects in ST will choose the payoff-maximizing output level as the tax schedules in ST contain a maximum of four rules. The median subject in ST indeed fares quite well at finding the payoff-maximizing output level.²² 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 output level.²³ 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. Given that the new rules in our experiment move the optimal choice more than almost any of the initial rules, our framework predicts that if output is changed, it will mostly be changed in the direction of the new payoff maximum—which is in line with our data.

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 to include the new rule in the third round. 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 nonreaction to small and large taxes is similar points to the importance of this type of criteria. At the same time, this finding contrasts 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.²⁴ The aim of this paper is,

²²Taking the model literally, we can calculate the possible output levels that subjects with a given k would choose and can then assign each subjects' actual output level to the k which generates the closest predicted output level. According to this procedure and depending on which choice order we consider, 53 to 59 percent of subjects in ST have $k \geq 4$.

²³We 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.

²⁴For example, one could assume that subjects have a stock of mental resources that depletes when mental effort is exerted (see, e.g., Baumeister et al. 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 nonresponse to new rules that we observe frequently.

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.

IV. Conclusion

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

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 *nonsalient* tax rules (as in Chetty, Looney, and Kroft 2009) but will also underreact to *salient* rules if the existing system is too complex. As we document that tax complexity lowers the responsiveness to tax changes, a social planner may want to use complexity as a tool to influence elasticities; depending on the respective tax, obfuscation or simplification could be optimal. As some individuals are more strongly affected by complexity than other, complexity could also be used as a screening device.

Clearly, our results—taken at face value—cannot be quantitatively 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 and subjects did not receive feedback about their performance between rounds. 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),²⁵ understanding the role of learning and advice would be crucial for identifying ways to improve decision making.

APPENDIX

A. Real-Effort Costs

The idea behind having a real-effort task linked to the output 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. Two 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 allotted 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 since this cost is lower than the marginal incentive around the optima. Moreover, the behavior of subjects in CT and ST does not differ in this phase ($p = 0.208$) 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." ($p = 0.303$). We conclude that the real-effort cost is too small to have influenced the effects of the treatment. Welfare consequences for subjects could be influenced since the real-effort costs are probably positive for most subjects. The total number of sliders moved across all three rounds is, however, not significantly different between treatments. In contrast, subjects in CT spend much more time thinking about their decision. This "mental" effort cost would come on top of the treatment effects we report in Table 7 in the online Appendix.

REFERENCES

- Abeler, Johannes, and Simon Jäger. 2015. "Complex Tax Incentives: Dataset." *American Economic Journal: Economic Policy*. <http://dx.doi.org/10.1257/pol.20130137>.
- Baumeister, Roy F., Ellen Bratslavsky, Mark Muraven, and Dianne M. Tice. 1998. "Ego depletion: Is the active self a limited resource?" *Journal of Personality and Social Psychology* 74 (5): 1252–65.

²⁵The 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.

- Bernheim, B. Douglas, and Antonio Rangel.** 2009. "Beyond revealed preference: Choice theoretic foundations for behavioral welfare economics." *Quarterly Journal of Economics* 124 (1): 51–104.
- Bhargava, Saurabh, and Dayanand Manoli.** 2013. "Why Are Benefits Left on the Table? Assessing the Role of Information, Complexity, and Stigma on Take-up with an IRS Field Experiment." Unpublished.
- Blumenthal, Marsha, and Joel Slemrod.** 1992. "The Compliance Cost of the U.S. Individual Income Tax System: A Second Look after Tax Reform." *National Tax Journal* 45 (2): 185–202.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer.** 2012. "Salience Theory of Choice Under Risk." *Quarterly Journal of Economics* 127 (3): 1243–85.
- Brown, Jeffrey R., Arie Kapteyn, Erzo F. P. Luttmer, and Olivia S. Mitchell.** 2013. "Cognitive Constraints on Valuing Annuities." National Bureau of Economic Research (NBER) Working Paper 19168.
- Brown, Jennifer, Tanjim Hossain, and John Morgan.** 2010. "Shrouded Attributes and Information Suppression: Evidence from the Field." *Quarterly Journal of Economics* 125 (2): 859–76.
- Cabral, Marika, and Caroline Hoxby.** 2012. "The Hated Property Tax: Salience, Tax Rates, and Tax Revolts." National Bureau of Economic Research (NBER) Working Paper 18514.
- Caplin, Andrew, and Mark Dean.** 2013. "Behavioral Implications of Rational Inattention with Shannon Entropy." National Bureau of Economic Research (NBER) Working Paper 19318.
- Carlin, Bruce Ian, Shimon Kogan, and Richard Lowery.** 2013. "Trading Complex Assets." *Journal of Finance* 68 (5): 1937–60.
- Cheremukhin, Anton, Anna Popova, and Antonella Tutino.** 2011. "Experimental Evidence on Rational Inattention." Federal Reserve Bank of Dallas Working Paper 1112.
- Chetty, Raj.** 2012. "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica* 80 (3): 969–1018.
- Chetty, Raj, John N. Friedman, and Emmanuel Saez.** 2013. "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review* 103 (7): 2683–2721.
- Chetty, Raj, Adam Looney, and Kory Kroft.** 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99 (4): 1145–77.
- Chetty, Raj, and Emmanuel Saez.** 2013. "Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients." *American Economic Journal: Applied Economics* 5 (1): 1–31.
- Congdon, William, Jeffrey R. Kling, and Sendhil Mullainathan.** 2011. *Policy and Choice: Public Finance through the Lens of Behavioral Economics*. Washington, DC: Brookings Institution.
- Crosetto, Paolo, and Alexia Gaudeul.** 2012. "Do consumers prefer offers that are easy to compare? An experimental investigation." http://mpr.ub.uni-muenchen.de/41462/1/MPRA_paper_41462.pdf.
- Dahremöller, Carsten, and Markus Fels.** 2012. "Product Lines, Product Design and Limited Attention." Social Science Research Network (SSRN) Working Paper 1993128.
- Deary, Ian J., Steve Strand, Pauline Smith, and Cres Fernandes.** 2007. "Intelligence and educational achievement." *Intelligence* 35 (1): 13–21.
- de Bartolome, Charles A. M.** 1995. "Which tax rate do people use: Average or marginal?" *Journal of Public Economics* 56 (1): 79–96.
- DellaVigna, Stefano, and Joshua M. Pollet.** 2009. "Investor Inattention and Friday Earnings Announcements." *Journal of Finance* 64 (2): 709–49.
- Duckworth, Angela L., Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly.** 2007. "Grit: Perseverance and Passion for Long-Term Goals." *Journal of Personality and Social Psychology* 92 (6): 1087.
- Eliaz, Kfir, and Ran Spiegler.** 2011. "Consideration Sets and Competitive Marketing." *Review of Economic Studies* 78 (1): 235–62.
- Fehr, Ernst, and Jean-Robert Tyran.** 2001. "Does Money Illusion Matter?" *American Economic Review* 91 (5): 1239–62.
- Feldman, Naomi E., and Peter Katuscak.** 2009. "Effects of Predictable Tax Liability Variation on Household Labor Income." Unpublished.
- Feldman, Naomi E., and Bradley J. Ruffle.** 2015. "The Impact of Including, Adding, and Subtracting a Tax on Demand." *American Economic Journal: Economic Policy* 7 (1): 95–118.
- Finkelstein, Amy.** 2009. "E-ZTax: Tax Salience and Tax Rates." *Quarterly Journal of Economics* 124 (3): 969–1010.
- Fischbacher, Urs.** 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10 (2): 171–78.
- Fleming, Stephen M., Charlotte L. Thomas, and Raymond J. Dolan.** 2010. "Overcoming status quo bias in the human brain." *Proceedings of the National Academy of Sciences* 107 (13): 6005–09.

- Fochmann, Martin, and Joachim Weimann.** 2011. "The Effects of Tax Salience and Tax Experience on Individual Work Efforts in a Framed Field Experiment." Institute for the Study of Labor (IZA) Discussion Paper 6049.
- Frederick, Shane.** 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives* 19 (4): 25–42.
- Fujii, Edwin T., and Clifford B. Hawley.** 1988. "On the Accuracy of Tax Perceptions." *Review of Economics and Statistics* 70 (2): 344–47.
- Gabaix, Xavier.** 2011. "A Sparsity-Based Model of Bounded Rationality." National Bureau of Economic Research (NBER) Working Paper 16911.
- Gabaix, Xavier, and David Laibson.** 2006. "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets." *Quarterly Journal of Economics* 121 (2): 505–40.
- Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg.** 2006. "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model." *American Economic Review* 96 (4): 1043–68.
- Gerardi, Kristopher, Lorenz Goette, and Stephan Meier.** 2010. "Financial Literacy and Subprime Mortgage Delinquency: Evidence from a Survey Matched to Administrative Data." Federal Reserve Bank of Atlanta Working Paper 2010-10.
- Gideon, Michael S.** 2013. "Survey Measurement of Tax Rates: Estimation and Behavioral Implications." Unpublished.
- Gill, David, and Victoria Prowse.** 2011. "A Novel Computerized Real Effort Task Based on Sliders." Institute for the Study of Labor (IZA) Discussion Paper 5801.
- Gill, David, and Victoria Prowse.** 2012. "A Structural Analysis of Disappointment Aversion in a Real Effort Competition." *American Economic Review* 102 (1): 469–503.
- Goldin, Jacob.** 2012. "Optimal Tax Salience." Princeton University Working Paper 571.
- Goldin, Jacob, and Tatiana Homonoff.** 2013. "Smoke Gets in Your Eyes: Cigarette Tax Salience and Regressivity." *American Economic Journal: Economic Policy* 5 (1): 302–36.
- Golosov, Mikhail, Aleh Tsyvinski, and Iván Werning.** 2007. "New Dynamic Public Finance: A User's Guide." In *NBER Macroeconomics Annual 2006*, Vol. 21, edited by Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, 317–88. Cambridge: MIT Press.
- Gravelle, Jane G., and Thomas L. Hungerford.** 2012. *The Challenge of Individual Income Tax Reform: An Economic Analysis of Tax Base Broadening*. Congressional Research Service. Washington, DC, March.
- Greiner, Ben.** 2004. "An Online Recruitment System for Economic Experiments." In *Forschung und wissenschaftliches Rechnen*, edited by Kurt Kremer and Volker Macho, 79–93. Göttingen: Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG).
- Heckman, James J., Jora Stixrud, and Sergio Urzua.** 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24 (3): 411–82.
- Hirshleifer, David, Sonya Seongyeon Lim, and Siew Hong Teoh.** 2009. "Driven to Distraction: Extraneous Events and Underreaction to Earnings News." *Journal of Finance* 64 (5): 2289–2325.
- Huck, Steffen, and Georg Weizsäcker.** 1999. "Risk, complexity, and deviations from expected-value maximization: Results of a lottery choice experiment." *Journal of Economic Psychology* 20 (6): 699–715.
- Ito, Koichiro.** 2014. "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing." *American Economic Review* 104 (2): 537–63.
- Kahneman, David, Jack L. Knetsch, and Richard H. Thaler.** 1991. "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives* 5 (1): 193–206.
- Kalayci, Kenan, and Marta Serra-Garcia.** Forthcoming. "Complexity and Biases: An Experimental Study." *Experimental Economics*.
- Kleven, Henrik Jacobsen, and Wojciech Kopczuk.** 2011. "Transfer Program Complexity and the Take-Up of Social Benefits." *American Economic Journal: Economic Policy* 3 (1): 54–90.
- Kopczuk, Wojciech.** 2005. "Tax bases, tax rates and the elasticity of reported income." *Journal of Public Economics* 89 (11–12): 2093–2119.
- Kőszegi, Botond, and Adam Szeidl.** 2013. "A Model of Focusing in Economic Choice." *Quarterly Journal of Economics* 128 (1): 53–107.
- Liebman, Jeffrey B., and Richard J. Zeckhauser.** 2004. "Schmeduling." <http://www.hks.harvard.edu/jeffreyliebman/schmeduling.pdf>.
- Lipsey, R. G., and Kelvin Lancaster.** 1956. "The General Theory of Second Best." *Review of Economic Studies* 24 (1): 11–32.

- Mill, John Stuart. 1848. *Principles of Political Economy*. London: John W. Parker.
- Mirrlees, J. A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38 (2): 175–208.
- Office of Tax Simplification (OTS). 2013. *Tax Complexity Project*. Her Majesty's Treasury. London, February.
- Office of the Taxpayer Advocate (OTA). 2010. *2010 Annual Report to Congress*. Internal Revenue Service (IRS). Washington, DC, January.
- Organisation for Economic Co-operation and Development (OECD). 2007. *OECD Economic Surveys: United Kingdom*. OECD, September.
- Ortoleva, Pietro. 2013. "The Price of Flexibility: Towards a Theory of Thinking Aversion." *Journal of Economic Theory* 148 (3): 903–34.
- Persson, Petra. 2012. "Attention Manipulation and Information Overload." Unpublished.
- Pommerehne, Werner W., and Friedrich Schneider. 1978. "Fiscal Illusion, Political Institutions, and Local Public Spending." *Kyklos* 31 (3): 381–408.
- PricewaterhouseCoopers (PwC). 2014. *Paying Taxes 2014: The Global Picture*. World Bank. <http://www.pwc.com/gx/en/paying-taxes/assets/pwc-paying-taxes-2014.pdf>.
- Reis, Ricardo. 2006. "Inattentive consumers." *Journal of Monetary Economics* 53 (8): 1761–1800.
- Rohaly, Jeffrey, and William G. Gale. 2004. "Effects of Tax Simplification Options on Equity, Efficiency, and Simplicity: A Quantitative Analysis." In *The Crisis in Tax Administration*, edited by Henry J. Aaron and Joel Slemrod. Washington, DC: Brookings Institution Press.
- Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (3): 180–212.
- Samuelson, William, and Richard Zeckhauser. 1988. "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty* 1 (1): 7–59.
- Sausgruber, Rupert, and Jean-Robert Tyran. 2005. "Testing the Mill hypothesis of fiscal illusion." *Public Choice* 122 (1–2): 39–68.
- Sims, Christopher A. 2003. "Implications of rational inattention." *Journal of Monetary Economics* 50 (3): 665–90.
- Sitzia, Stefania, Jiwei Zheng, and Daniel John Zizzo. 2014. "Inattentive Consumers in Markets for Services." *Theory and Decision* (2014): 1–26.
- Slemrod, Joel. 1984. "Optimal Tax Simplification: Toward a Framework for Analysis." In *Proceedings of the 76th Annual Conference of the National Tax Association*, 158–67. Columbus: National Tax Association.
- Slemrod, Joel. 1992. "Did the Tax Reform Act of 1986 simplify tax matters?" *Journal of Economic Perspectives* 6 (1): 45–57.
- Slemrod, Joel, and Wojciech Kopczuk. 2002. "The optimal elasticity of taxable income." *Journal of Public Economics* 84 (1): 91–112.
- Slemrod, Joel, and Nikki Sorum. 1985. "The Compliance Cost of the U.S. Individual Income Tax System." National Bureau of Economic Research (NBER) Working Paper 1401.
- Sunstein, Cass R. 2011. "Empirically Informed Regulation." *University of Chicago Law Review* 78 (4): 1349–1430.
- Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (4481): 453–58.
- Tversky, Amos, and Daniel Kahneman. 1986. "Rational choice and the framing of decisions." *Journal of Business* 59 (4): S251–78.
- United States General Accounting Office (GAO). 2002. *Income Ranges of Taxpayers Who May Have Overpaid Federal Taxes by Not Itemizing*. Report to House of Representatives. Washington, DC, April.
- Wagner, Richard E. 1976. "Revenue Structure, Fiscal Illusion, and Budgetary Choice." *Public Choice* 25 (1): 45–61.
- Wilcox, Nathaniel T. 1993. "Lottery Choice: Incentives, Complexity and Decision Time." *Economic Journal* 103 (421): 1397–1417.
- Wuppermann, Amelie C., Sebastian Bauhoff, and Markus M. Grabka. 2014. "The Price Sensitivity of Health Plan Choice: Evidence from Retirees in the German Social Health Insurance." Munich Discussion Paper 2014-34.
- You, Haifeng, and Xiao-jun Zhang. 2009. "Financial reporting complexity and investor underreaction to 10-K information." *Review of Accounting Studies* 14 (4): 559–86.