

Fairness at a Price: Support for Complex Benefit Systems

Magdalena Wasilewska*

October 30, 2026

[Click here for the most recent version](#)

Abstract. Directing benefits to those in need requires complex rules and procedures, which impose substantial transaction costs on claimants. This has sparked debates about simplifying benefit systems. I study how people trade off distributional fairness against the transaction costs of achieving it. In a spectator experiment, impartial participants choose between two bonus systems for workers, some of whom experienced a random earnings shock; the systems differ in targeting accuracy and effort costs workers incur to receive the bonus. In a Dutch survey experiment, I provide information about the costs of complexity and measure its effect on support for simplification. The results speak to the political economy of reform. First, most spectators are willing to impose substantial effort costs on workers for a modest improvement in accuracy. Second, this willingness decreases when effort costs are higher and when imposing them improves distributional fairness rather than achieving it perfectly, as real benefit systems rarely do. Third, information shifts Dutch respondents' beliefs but increases support for a simpler, less targeted system only among non-recipients of benefits, although recipients hold equally inaccurate priors and update as much. Recipients' open-ended answers instead point to their stake in the system and to uncertainty about their entitlements after reform.

JEL Classifications: D63; I38; H53; D72; D83; C91

Keywords: fairness, transaction costs, targeting, welfare simplification, spectator experiment, survey experiment

*University of Amsterdam, Tinbergen Institute (email: m.j.wasilewska@uva.nl). I thank Joël van der Weele and Thomas Douenne for their invaluable guidance and support. I am grateful to Alexander Cappelen, Peter Courtney, Cristina Figueroa-Sisniega, Oda Sund, Bertil Tungodden, and participants at seminars and conferences in Bergen, Amsterdam, and Helsinki for helpful comments and discussions. I gratefully acknowledge financial support from Amsterdam Center for Behavioral Change (ACBC), Amsterdam Centre for European Studies (ACES), and A Sustainable Future (ASF) initiative of the University of Amsterdam. Ethics approval number: EB-26693. The [spectator experiment](#) and the [survey experiment](#) were pre-registered on OSF.

1. Introduction

We are convinced that these ancient laws are scrupulously administered; nevertheless it is an extremely painful thing to be ruled by laws that one does not know.

— Franz Kafka, *The Problem of Our Laws* (1920)

Welfare benefit systems are often complex, and claiming and keeping benefits therefore involves substantial transaction costs.¹ Claimants spend time and effort filling in forms, reporting changes, and understanding the rules that apply to them. In the United Kingdom, the claim form for the Personal Independence Payment, a disability benefit, runs to 50 pages. In the Netherlands, only 43% of recipients know what changes in their situation they need to report (Kok et al. 2022). These costs reduce take-up of benefits (Castell et al. 2025; Kleven and Kopczuk 2011) and lead eligible recipients to lose their benefits when they fail to meet procedural requirements (Homonoff and Somerville 2021).

Simplifying benefit systems is on the policy agenda in many countries, but progress has been slow. In France, a reform consolidating several benefits into a single social allowance, first proposed in 2017 and revived by each of the last three prime ministers, has yet to be introduced (Jacquot 2026). The Dutch parliament unanimously adopted motions calling for a fundamental reform of the allowance system in 2024 and 2025 (Tweede Kamer 2024, 2025),² and two thirds of the 2025 party programmes proposed simplifying it.³ Yet beyond small adjustments, comprehensive reform has not been undertaken (van Dijk 2026).

If complexity is so costly, why has reform been slow? Complex rules and procedures can direct benefits more accurately to those in need, by sharpening the signal of need or by deterring the less needy from claiming (Kleven and Kopczuk 2011; Nichols and Zeckhauser 1982). The design of benefit systems therefore involves a trade-off between transaction costs and accurate targeting. Simplification proposals often focus on the costs of complexity for claimants, yet little is known about whether these costs matter for public support for simplification. Does the public object to these costs, or does it accept them as the “price of fairness”?

I study three aspects of this question. First, are people less willing to ensure distributional fairness, that is, that a transfer reaches its intended recipients, when doing so comes at higher effort costs to these recipients? Second, does this willingness depend on whether

¹In public administration research, these costs are known as administrative burden (Burden et al. 2012; Moynihan, Herd, and Harvey 2015).

²Complexity became nationally salient during the childcare allowance scandal, over which the government resigned in 2021: thousands of parents were accused of fraud and required to repay their allowances, mostly over small administrative errors (Kok et al. 2022).

³Author’s reading of the 2025 election programmes (NOS 2025).

these effort costs achieve perfect distributional fairness or only improve it, that is, whether some intended recipients still miss out? Both questions are motivated by two features of complex welfare systems: (i) claiming benefits is costly, and (ii) these systems fail to reach everyone who is eligible, which may make the costs of claiming harder to justify.⁴ Support for simplification may also be constrained by incorrect beliefs about the costs of complexity for claimants, rather than only by the weight people attach to distributional fairness. This motivates the third question: do people hold inaccurate beliefs about the costs of complexity for claimants, and do these beliefs have a causal effect on support for simplification?

I address the first two questions in an incentivised spectator experiment with 990 participants (Almås, Cappelen, and Tungodden 2020; Cappelen et al. 2013), where impartial participants (“spectators”) make choices that determine the earnings and effort of other participants (“workers”). The controlled setting allows me to cleanly isolate how people trade off distributional fairness against the effort costs of ensuring it, separately from considerations that confound inferences about this trade-off from support for real reforms, such as reforms’ budgetary cost and uncertainty about their distributional effects. Workers complete real-effort tasks, and some of them have their task earnings reduced by a random shock. Each spectator is matched with a group of 12 workers and chooses between two systems for distributing extra payments, added on top of task earnings. Both systems have the same budget and pay 8 of the 12 workers. One system, which I call Complex, directs these payments to the workers affected by the shock, but only after they complete a form reporting details of their working session, which takes extra, unpaid time. The other, which I call Simple, pays automatically based on a colour assigned to each worker, an imperfect proxy for experiencing the shock (an observable tag, Akerlof 1978), and therefore targets the payments less accurately. I vary two features of the Complex system: (i) how long the form takes (two vs. seven minutes), and (ii) whether it reaches all the workers affected by the shock.

Across treatments, I hold fixed the total budget for these extra payments and the Complex system’s targeting advantage over the Simple system: even when it fails to reach everyone, it reaches exactly one more worker affected by the shock. This gives the second question a sharp benchmark. I show that a utilitarian spectator’s choice should depend on the length of the form, but not on whether the Complex system fails to reach one

⁴Targeted systems may fail to reach eligible people for several reasons. Some do not claim (Castell et al. 2025; Kleven and Kopczuk 2011), others lose their benefits when they fail to meet procedural requirements (Homonoff and Somerville 2021), and some eligible applicants are wrongly rejected (Benitez-Silva, Buchinsky, and Rust 2004). Wrongful rejections are not necessarily caused by complexity, but they occur in these systems.

affected worker.⁵ A change in choices would instead indicate that spectators' evaluations are not separable across workers: they value properties of the allocation as a whole, such as whether it reaches every worker it is meant for.

I find that a majority of spectators choose the Complex system even when workers have to spend 56 minutes of unpaid time in total on the form.⁶ Spectators thus accept imposing a substantial effort cost on workers so that one more worker affected by the shock receives a payment. However, support for the Complex system falls by 19 percentage points when the form takes seven minutes instead of two, and by 15 percentage points when the system fails to reach a single affected worker, even though its targeting advantage over Simple is kept constant (both $p < 0.01$). I provide evidence that spectators' evaluations are not separable across workers: they depend not only on each worker's earnings and effort, but also on properties of the allocation as a whole, such as whether it delivers perfect distributional fairness. Left-wing spectators, who are more likely to choose the Complex system when it targets perfectly (75% against 59%), are also more sensitive to a single exclusion error: their support falls by 22 percentage points, against a statistically insignificant 5 percentage points among centre-right spectators.

Unlike spectators, when voters evaluate real reforms they weigh the costs of complexity against considerations held constant by design in the spectator experiment, such as budgetary costs, labour supply responses, their own interests and trust in government (Kuziemko et al. 2015; Stantcheva 2026). They also have to form their beliefs about these costs, and these beliefs may be inaccurate. I thus study the third question, whether people misperceive the costs of complexity for claimants and whether these beliefs have a causal effect on support for simplification, in a survey experiment ($N = 2,728$) in the Netherlands, where reform of the allowance system is high on the political agenda but has made little progress (van Dijk 2026). I elicit respondents' beliefs about three salient costs of complexity in the Dutch welfare system (non-take-up, how difficult the rules are to understand, and the risk that claimants' benefits are later corrected) and show the treatment group correct figures and explanations for all three, before measuring support for simplification reforms.

I find that respondents, on average, underestimate the costs of complexity and that information about these costs shifts their beliefs, but does not raise support for simplification. This average, however, hides heterogeneity, most notably by whether respondents have received benefits. While both recipients and non-recipients update their beliefs,

⁵More generally, this holds for any spectator whose evaluation is separable across workers, that is, depends only on each worker's own earnings and effort. This includes fairness views under which the spectator penalises each worker's deviation from a fair allocation (Section 3).

⁶The form takes seven minutes, and each of the eight workers paid by the Complex system completes it.

only non-recipients become more supportive of a general simplification of the system. The difference is not explained by prior beliefs, belief updating, baseline support for simplification, attention, or recipients' own experience of complexity, is unlikely to reflect experimenter demand, and survives allowing the treatment effect to vary with all other respondent characteristics. Recipients' open-ended justifications point to a personal stake in the system: they more often argue that a reform could cut their benefits and reason from their own situation.

My results highlight two explanations for why complex benefit systems remain in place. First, spectators accept imposing a substantial effort cost on claimants for a modest improvement in targeting accuracy, suggesting that even small decreases in accuracy may make reforms unpopular. Yet this support falls once the targeted system fails to reach a single worker it is meant for. Since real benefit systems rarely reach everyone who is eligible, exclusion errors may weaken the fairness argument that justifies complexity. Second, across both experiments, the support for complexity of people who judge the trade-off impartially is sensitive to the cost it generates for claimants. However, this is not the case for those with personal stakes in the system, i.e., Dutch benefit recipients. This suggests that for those who incur these costs, other considerations, such as uncertainty about how a reform would affect their entitlements, may be more salient. Reducing this uncertainty may therefore matter as much for the feasibility of reform as informing the public about the costs of complexity.

This paper contributes to three literatures. First, it adds to the experimental literature on fairness views (Almås, Cappelen, and Tungodden 2020; Andre 2025; Cappelen, Cappelen, and Tungodden 2023; Cappelen et al. 2013; Durante, Putterman, and Van der Weele 2014; Konow 2000). Previous experiments have priced fairness as lost efficiency (Almås, Cappelen, and Tungodden 2020; Durante, Putterman, and Van der Weele 2014) or as the decider's own forgone payoff (Piraino and Ryan 2024). I introduce a price of a different kind: an unpaid effort cost incurred by the beneficiaries themselves, mirroring the transaction costs in real welfare states. Methodologically, I extend the spectator design, in which spectators usually allocate earnings between individuals, to the choice of a rule for a group: like policymakers, the spectators in my experiment choose between systems that differ both in the errors they make and in the effort costs they impose on recipients. This allows me to show that fairness judgements of allocation rules depend not only on how many people a rule reaches but also on whether it reaches everyone it is meant for.

Second, the survey experiment adds to the literature on information experiments about redistribution (Alesina, Stantcheva, and Teso 2018; Cruces, Perez-Truglia, and Tetaz 2013; Fehr, Mollerstrom, and Perez-Truglia 2022; Karadj, Mollerstrom, and Seim 2017;

Kuziemko et al. 2015). These studies typically shift beliefs that influence how much people want to redistribute. I instead shift beliefs about how redistribution is delivered, i.e., beliefs about the costs that the complexity of the welfare system imposes on claimants. It also adds to survey experiments on administrative burden in political science (Baekgaard, Moynihan, and Thomsen 2021; Halling and Van de Walle 2026; Keiser and Miller 2020). The closest, Keiser and Miller (2020), shows that more administrative burden increases support for policies because respondents see it as a signal that recipients are deserving and the programme legitimate. I inform respondents not about the burden itself, but about its consequences for claimants. Among respondents who have never received benefits (and so have a less direct stake in the system), this information raises support for a simpler system.

Finally, the paper adds to the public economics literature on how to target transfers when need is unobserved (Akerlof 1978; Diamond and Sheshinski 1995; Nichols and Zeckhauser 1982). Conceptually closest is Kleven and Kopczuk (2011), who model complexity as a by-product of screening: it improves the accuracy of targeting but imposes transaction costs on claimants. Empirical work measures how these costs affect take-up and targeting (Alatas et al. 2016; Bhargava and Manoli 2015; Finkelstein and Notowidigdo 2019), and the literature on administrative burden documents them in citizens' interactions with the state (Burden et al. 2012; Christensen et al. 2020; Moynihan, Herd, and Harvey 2015). To my knowledge, I provide the first experimental evidence in economics on how the public weighs targeting accuracy against transaction costs, and show that the public accepts substantial transaction costs for accurate targeting, but that this support is fragile when the system fails to reach everyone.

The rest of the paper is organised as follows. Section 2 describes the design of the spectator experiment. Section 3 presents a conceptual framework and derives the hypotheses. Section 4 reports the results of the spectator experiment and examines potential mechanisms. Section 5 presents the survey experiment, including its institutional context, design, and results. Section 6 concludes.

2. Experimental Design

I develop an experimental set-up that allows me to study how individuals trade off the transaction costs of a compensation system against the accuracy with which it reaches its intended recipients. It builds on the spectator design used in experimental studies of fairness (Almås, Cappelen, and Tungodden 2020; Cappelen et al. 2013; Konow 2000), in which an impartial spectator makes a decision that affects other people's earnings. This

set-up differs from the standard spectator design in three ways. First, spectators choose a system that determines an allocation for a group of twelve workers, rather than deciding on an allocation for a pair. Second, the experimental manipulation concerns the properties of this system, rather than the source of the underlying inequality. Third, the spectator's choice determines not only what workers earn but also what they must do to receive it.

The experiment involves two types of participants. Workers complete a paid real-effort task, and for some of them part of their output is, at random, not registered, so they earn less. Spectators, who are the focus of the analysis, decide how a compensation is allocated among a group of twelve workers. They choose between two systems. Both systems have the same budget and compensate the same number of workers by the same amount; they differ only in who receives the compensation and what workers must do to receive it. Under the Complex system,⁷ compensation is accurately targeted at the workers whose output was not fully registered, but they must first complete a verification form that takes extra, unpaid time. Under the Simple system, compensation is paid automatically and at no transaction cost, but is targeted less accurately, since it is conditioned on the worker's colour (a tag correlated with lost output).⁸ I vary two features of this trade-off across treatments: (i) the length of the verification form, that is, the cost of complexity (Kleven and Kopczuk 2011), and (ii) the targeting accuracy of the Complex system. The primary outcome of interest is the spectators' choice between the two systems. Figure 1 shows, for the group of twelve workers each spectator decides for, who is compensated under each system, both in the treatments where the Complex system targets perfectly (T1 and T2) and in the treatment where it does not (T3); Section 2.2 describes the treatments in detail.

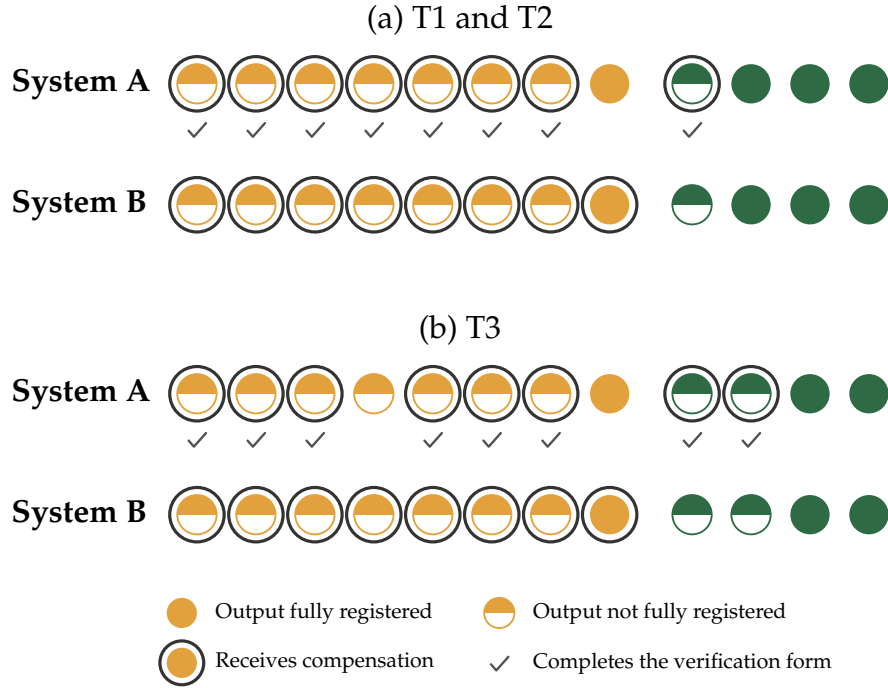
2.1. Workers

Worker characteristics. Each worker is randomly assigned a colour, Yellow or Green. They are also assigned to have their output fully registered or not. All workers work for the full fifteen minutes of the working session, but for some of them only the tasks completed in the first eight minutes count towards their performance bonus. They are therefore paid for fewer tasks than they actually complete, and earn less than otherwise identical workers. I refer to these workers as the workers affected by the shock. Colour is an informative but imperfect proxy for output not being fully registered: in the group

⁷I use the word "complex" in the sense of Kleven and Kopczuk (2011): complexity is the rigour of the screening a scheme uses to establish eligibility, which provides a better signal of who is truly entitled but imposes transaction costs on claimants. In the online experiment, this takes the form of a single requirement, the verification form, which imposes high or low transaction costs on the workers, depending on the treatment.

⁸In the experiment, spectators see the two systems under neutral labels, System A (Complex) and System B (Simple).

Figure 1. The two compensation systems, by treatment



Note: Each circle is one of the twelve workers a spectator decides for. In every treatment both systems compensate eight of the twelve workers and pay each of them the same amount. The verification form in System A takes seven minutes in T1 and T3 and two minutes in T2.

each spectator decides for, seven of the eight Yellow workers have output that is not fully registered, against one or two of the four Green workers, depending on the treatment (Section 2.2).

Workers' information. At the start of the experiment, workers are told that part of their completed output might not be registered, as the result of a random process outside their control. They are not told how likely that is, from which point in the session it might happen, whether it differs by colour, or which colour they have been assigned. This means that workers cannot condition their performance on whether their output is fully registered. Workers are also told that a decision by another participant may affect their final payment, but not which systems are possible, how that decision is made, or how the compensation is determined.

Tasks and earnings. In the first part of the experiment, workers complete a series of real-effort tasks for fifteen minutes. These tasks include counting how many times a given digit appears in a string of numbers, counting how many times a given letter appears in a sequence of letters, and encoding word sequences into numbers. They receive a fixed participation fee and a performance bonus that increases in the number of tasks

they complete correctly. On top of that, depending on the spectator's choice, they may receive compensation. The compensation is defined to spectators (but not to workers) as the difference between the average performance bonus of workers whose output is fully registered and that of workers whose output is not, so it equalises the two groups' bonuses in expectation. Workers were paid a fixed fee of £3.20 and a median performance bonus of £0.55 (£0.98 for workers whose output was fully registered, £0.50 for workers whose output was not). The compensation was equal to £0.47. Spectators were not told this amount, because it depends on workers' performance and was determined only after spectators had made their choices.⁹

The second part of the workers' experiment depends on the decision of the spectator. If the spectator chooses the Simple system, workers finish the experiment after the fifteen-minute working session. If, on the other hand, the spectator chooses the Complex system, those workers whose output was not fully registered complete a verification form. The form mirrors the verification requirements of real benefit programmes that ask claimants to report their circumstances, background, or activity, often in great detail.

Depending on the treatment, the form takes two or seven minutes to fill in.¹⁰ The form contains more items than can be completed in the allocated time, and is timed; workers cannot submit it before the allocated time ends and are moved on automatically once it ends. It contains several types of questions, about (i) administrative details of their working session (e.g., date, start and end times, minutes worked, device, operating system); (ii) estimates of their own performance in different types of tasks and the strategy they used; (iii) the environment in which they completed the tasks (e.g., location, noise, lighting); and (iv) their Prolific activity and habits. Workers who fill in the seven-minute form complete the same items and are then asked for a more granular account of their performance. They are shown non-copiable tables listing words or letters, some of which appeared in their tasks, each paired with an arbitrary six-character code, and must report the codes of the items they believe they worked on. They are also asked to provide further details of their working session and to evaluate the tasks they worked on.

⁹Not stating an amount also avoids inviting spectators to compare the minutes on the form with the small amounts paid in an online study, which would make their choice depend on how they convert time into money.

¹⁰Workers' responses in the form do not affect their payment; only completion matters, and every worker who completes the study completes the form. Tying payment to the *content* of the form would make the design more complicated; spectators would then have to form beliefs about workers' error rates, and the perfect targeting in some of the treatments (see Section 2.2) could no longer be guaranteed. This is an interesting avenue for future research. Workers are told that a rule of the study requires them to keep working on the form for its full allocated time. They appear to comply: the median last click on the form was at 97% of the allocated time in both form lengths.

2.2. Treatments and Compensation Systems

Spectators are randomly assigned to one of three between-subject treatments, T1, T2 and T3, summarised in Table 1. Each spectator is shown a group of twelve workers: eight assigned to Yellow and four to Green. Figure 1 shows this group, whose composition is fixed within each treatment, and who is compensated under each system. In T1 and T2, seven of the eight Yellow workers and one of the four Green workers have output that is not fully registered. In T3, seven of the eight Yellow and two of the four Green workers have output that is not fully registered.

The Complex system (System A). Compensation is conditioned on whether output was fully registered, and receiving it requires workers to complete the verification form, which imposes a transaction cost on them. Workers whose output was not fully registered complete the form and receive the compensation; workers whose output was fully registered do not. In T1 and T2 the system targets perfectly: all eight workers whose output was not fully registered complete the form and are compensated. In T3 the form is unavailable to one of the seven Yellow workers whose output was not fully registered, who therefore receives no compensation.

The Simple system (System B). Compensation is conditioned on colour. Every Yellow worker receives the compensation automatically, and Green workers receive nothing; no worker completes a form or spends any additional time. The system therefore pays one worker whose output was fully registered and misses the Green worker (T1 and T2) or workers (T3) whose output was not. It is the experimental analogue of a scheme that conditions on an observable tag correlated with need (Akerlof 1978).

In every treatment both systems compensate exactly eight of the twelve workers, and each compensated worker receives the same amount, so total expenditure is identical under both systems and across all treatments. In T1 and T2, the Complex system compensates every worker whose output was not fully registered; to receive the compensation, these workers must complete a verification form that takes seven minutes (T1) or two minutes (T2). In T3, the form again takes seven minutes, but it is unavailable to one Yellow worker whose output was not fully registered. This worker does not fill in the form and therefore does not receive the compensation, so the Complex system misses one such worker. Spectators are told that this is outside the worker's control: the worker neither causes nor can prevent it. Comparing T1 with T2 isolates the effect of the transaction cost, holding the resulting distribution fixed. Comparing T1 with T3 captures the effect of the Complex system losing its perfect targeting, holding fixed the form length, total expenditure, and the Complex system's advantage over the Simple system. To keep the budget fixed, the

T3 group has one more Green worker whose output was not fully registered than the T1 group.¹¹ Together, these changes mean that both systems miss one more worker whose output was not fully registered than in T1: the Complex system goes from zero to one, and the Simple system from one to two, so the Complex system retains its relative advantage of compensating one more such worker than the Simple system (see Section 3).

Table 1. Treatments

	Targeting of the Complex system	Form length	Isolates	<i>N</i>
T1 (reference)	perfect	7 min	—	313
T2	perfect	2 min	transaction cost (vs. T1)	320
T3	one worker excluded	7 min	imperfect targeting (vs. T1)	295

Notes: Treatments are between subjects. In every treatment both systems compensate eight of the twelve workers, each by the same amount, so total expenditure is identical across systems and treatments. In T3 the verification form is unavailable to one worker whose output was not fully registered, who is therefore not compensated; the group composition adjusts to preserve budget neutrality, and both systems miss one more worker whose output was not fully registered than in T1. The column “Isolates” reports the effect each treatment isolates relative to T1. *N* is the number of spectators in the main sample (928 in total), after excluding respondents who failed the attention checks or were flagged by Prolific’s bot check.

2.3. Spectators

The spectators’ experiment starts with a consent form, a pledge not to use AI tools, a pledge to read the instructions carefully, and standard attention checks. Next, spectators read the instructions. They learn that the workers are recruited to complete the task *individually* for payment, and they learn about the other properties of the worker set-up: that workers are assigned a colour, are paid for the tasks they complete, and that, for some of them, part of the output is not registered and therefore does not count towards their performance bonus.¹² Spectators then learn that they will choose how the compensation is allocated among the workers, and that there is more than one way to do so.

They are then shown the twelve workers they will decide for on a figure that visualises the group’s composition (see the full instructions in Appendix D.1). This is followed by

¹¹Without this change, excluding one worker whose output was not fully registered would leave the Complex system compensating only seven workers, while the Simple system would still compensate eight, so the two systems would no longer have the same budget.

¹²The instructions emphasise the individual nature of the task, avoid the word *group* throughout, and inform spectators that workers do not know what others receive. This reduces the risk that spectators judge the systems by imagining group dynamics among workers. For example, if workers were to work together on a group task, spectators might believe it is fair to grant compensation to everyone in one group (i.e., Yellow or Green) in order to minimise feelings of envy or regret, which is something the experiment does not aim to capture.

a description of the two systems: who is paid, who is not, and what is required. Under the Complex system, spectators are told that workers whose output is not fully registered “go through a required verification step: completing a form that reports details of their working session”, and that the time spent on it is unpaid; they are not told how the answers on the form are used. For each system a figure shows which workers receive the compensation and which complete the form, and a table summarises the time each type of worker spends and what they are paid. Spectators are reminded that the compensation amount is identical under both systems and that each compensates eight of the twelve workers. They then answer a comprehension quiz of eight true/false items, with a gated option to re-read the instructions beforehand; a wrong answer triggers a short clarification of the item concerned.

After the quiz, spectators choose which of the two systems to implement for the twelve workers they were shown. The allocation implied by each system is displayed alongside the choice, together with a reminder that both systems pay the same amount to each compensated worker and compensate eight of the twelve. Spectators are told that their decision is anonymous and is not observed by other participants. They are also told that workers know that their final payment may depend on a decision made by a participant in a related study, but not how that decision is made or which systems are possible. Because workers are unaware of the systems, spectators have no reason to believe their choice can meet or disappoint workers’ expectations. The choice is probabilistically incentivised: for one in thirty spectators, the chosen system is implemented for a group of twelve workers, which determines who is compensated and whether workers must fill in the form. To reduce the risk that spectators infer that I want a particular type of worker to be compensated, I tell them that whether they are selected does not depend on which system they choose and that there are no right or wrong answers.

After the choice, spectators are asked questions about their perceptions of the systems and their choice. They are first asked an open-ended question about the reasons for their choice. Next, they are asked closed-ended questions about the effort (the transaction cost) and fairness of the systems: how much more effort the Complex system requires of workers than the Simple system, from 0 (no extra effort) to 10 (far more effort), and which system’s distribution of the compensation is fairer, judged only on who is and is not compensated, from 0 (Complex much fairer) to 10 (Simple much fairer), with 5 labelled neither. In T3, where a worker whose output was not fully registered can miss out either because of the system’s rule (Simple) or because the form is unavailable (Complex), spectators are also asked whether they find one reason more unfair than the other, from 0 (missing out under

the Complex system is much more unfair) to 10 (missing out under the Simple system is much more unfair).

In the last part of the study, spectators answer a short battery of questions about their trust in and perceptions of the competence of government departments and institutions; their view of the trade-off between targeting and ease of claiming as it arises in real benefit systems, on a bipolar item from 0 (“benefits should be universal and easier to claim”) to 10 (“benefits should be targeted, even if harder to claim”); their support for two simplification policies (an automatic income-based payment in the style of a negative income tax, and a universal basic income), fairness attitudes, generalised trust, political attitudes, and demographics.

2.4. Design Choices

Budget neutrality. Because both systems compensate eight of the twelve workers by the same amount, a spectator’s choice cannot express a view about how much compensation should be paid, only about the form it should take: an accurate but costly system or a simpler but less accurate one. This is an advantage of the experimental setting. In reality, simplification reforms may also change how generous or costly the system is (e.g., some variants of a universal basic income; Goderis and Vlekke (2023)), so that support for reform can mix views about administrative design with views about the size of the welfare state.

Deciding for a group. The spectator chooses for a group of twelve workers rather than for the pair of the standard spectator design. This mirrors real welfare states, where transaction costs are borne by many claimants: the experiment tests whether spectators would impose these costs on many workers in exchange for a small increase in distributional fairness.

Fairness of the two systems. To create a sharp tension between distributional fairness and transaction costs, the Complex system is designed to distribute the compensation more fairly under the main fairness views. For egalitarians, it produces a more equal distribution of final earnings; for meritocrats, it pays for work done but not registered;¹³ and for luck egalitarians, it corrects brute luck unrelated to effort or ability (Cappelen et al. 2013; Mollerstrom, Reme, and Sørensen 2015). Finally, a libertarian could regard the low pay as the just outcome of a voluntary exchange, since workers agreed to work knowing their output might not be registered. Even so, of the two systems, the Complex

¹³All workers complete the full fifteen-minute session. Had some workers instead been prevented from working part-way through, their claim to compensation would be less clear, since people reward effort even when the opportunity to exert it is outside one’s control (Andre 2025; Bhattacharya and Mollerstrom 2025).

one better honours the entitlement to the fruits of one's labour (Nozick 1974), since it directs compensation to the workers who produced the output that was not registered.

Exclusion in T3. The exclusion of a worker whose output was not fully registered is framed neutrally, as the form being unavailable to that worker. The aim is to make targeting explicitly imperfect without narrowing the exclusion to one specific reason, such as errors of the screening technology or non-take-up, so that the results speak to exclusion in general rather than to one of its causes. The framing does not appear to make this exclusion seem particularly unfair: asked after their choice which reason for missing out is more unfair, spectators in T3 on average rate missing out because the form is unavailable as slightly less unfair than missing out because of one's colour (Section 4). Because the exclusion is outside the worker's control, however, the results may not extend to exclusion that people attribute to claimants' own decisions rather than to the system, such as non-take-up if it is seen as the claimant's choice.

Fixed take-up. Studying non-take-up would require letting workers decide whether to apply for the compensation. Allowing this in only one treatment, or with different probabilities across treatments, would give spectators an additional signal about the value of the compensation or the severity of the transaction costs, and would do so differently across the cost treatments. I therefore hold take-up fixed and require every worker whose output was not fully registered to complete the form. Endogenising take-up, so that it responds to the transaction cost and to the value of the compensation as in Kleven and Kopczuk (2011), is a natural avenue for future research.

2.5. Experimental Procedures

Spectators were recruited on Prolific, restricted to adult UK nationals whose first language was English, who resided in the UK, and who completed the study on a desktop, laptop, or tablet. 990 spectators completed the study and were approved on Prolific. As per the pre-registration, I exclude respondents who failed the attention checks or were flagged by Prolific's bot check, which leaves a final sample of 928. The pre-registered robustness sample additionally excludes the bottom 5% of time spent within each treatment and respondents with more than one incorrect answer on the comprehension quiz (709 spectators). Spectators were paid a fixed fee of £3.50 and a bonus of £0.25 if they answered all comprehension questions correctly. The median completion time was 14.4 minutes.

Spectator data were collected first, because a spectator's choice determines what the workers do: whether they fill in the verification form, and for how long. The spectator instructions describe the workers' set-up in general terms and do not state whether workers

have already completed the task; every statement in them holds for the workers recruited afterwards. One in thirty of the 990 spectators who completed the study was then drawn at random, so that 33 spectators had their choice implemented. For each drawn spectator, I recorded the treatment they faced and the system they selected, and the system was implemented for a group of twelve workers with the composition the spectator was shown, so that 33 groups, or 396 workers, were run in total.¹⁴ The worker sessions were run in five separate Prolific studies, one per distinct (treatment, chosen system) worker experience. All workers completed the fifteen-minute working session; workers matched to a spectator who chose the Complex system, and whose own output was not fully registered, additionally completed the verification form, which lasted seven or two minutes according to the spectator's treatment (except, in T3, the single worker for whom the form was unavailable).¹⁵

Workers were recruited from the UK and the US. Their pay is described in Section 2.1.

3. Conceptual Framework and Hypotheses

This section presents a simple framework to derive hypotheses about how support for the Complex system changes across treatments. The main implication is that a utilitarian spectator should be unaffected by the exclusion in T3. Appendix B gives the full derivations.

3.1. A Utilitarian Benchmark

Each worker i earns y_i in the study: y_L if their output was not fully registered and $y_H > y_L$ if it was. A worker's consumption is $c_i = y_i + b_i$, where $b_i = b$ if the worker is compensated and $b_i = 0$ otherwise. All workers exert effort e_T on the tasks, and workers who fill in the verification form exert additional effort e_V . Workers value consumption and dislike effort:

$$U_i = u(c_i) - v(e_i), \quad u' > 0, u'' < 0, v' > 0. \quad (1)$$

¹⁴Of the 33 groups, 27 went through the Complex system (13 in T1, 12 in T2, and 2 in T3) and 6 through the Simple system (3 in T1 or T2, and 3 in T3).

¹⁵In practice, 8 of the 349 workers who consented to the Complex-system worker studies (2.3%) left the study when they reached the form, without answering any of its questions. As with other participants who dropped out at any point of the study, they were not included in the implemented groups, so every worker in an implemented group whose output was not fully registered completed the form.

I assume that the spectator is utilitarian and chooses the system that maximises social welfare over the group of twelve workers, defined as the sum of their utilities:¹⁶

$$W = \sum_{i=1}^{12} U_i. \quad (2)$$

The *targeting gain* N is the gain in social welfare from paying the compensation to a worker whose output was not fully registered rather than to one whose output was, and the *burden* B^t is the loss in utility of a worker who fills in the form in treatment t , in which the form requires effort e_V^t :

$$N \equiv [u(y_L + b) - u(y_L)] - [u(y_H + b) - u(y_H)] > 0, \quad B^t \equiv v(e_T + e_V^t) - v(e_T) > 0. \quad (3)$$

N is positive because u is concave: the compensation raises utility more for a worker who earns less.

In T1 and T2, both systems compensate the seven Yellow workers whose output was not fully registered, and differ only in who receives the eighth payment (Figure 1, panel (a)): the Simple system pays it to the Yellow worker whose output was fully registered, the Complex system to the Green worker whose output was not. Redirecting this payment is the Complex system's targeting advantage, worth N . In exchange, all eight workers it compensates fill in the form, at a total burden of $8B^t$. The difference in social welfare between the Simple and the Complex system is therefore $\Delta^t \equiv W^S - W^C = 8B^t - N$, and the spectator chooses the Complex system if $\Delta^t < 0$, that is, if the targeting gain outweighs the burden on the eight workers who fill in the form.

A shorter form lowers the burden, $B^{T2} < B^{T1}$, and leaves the targeting gain unchanged, so the Complex system becomes more attractive.

Prediction 1 (Transaction cost) *Any spectator who chooses the Complex system in T1 would also choose it in T2, so the share choosing the Complex system is weakly higher in T2.*

In T3, the form is unavailable to one Yellow worker whose output was not fully registered, and, to keep the budget fixed, a second Green worker's output is not fully registered (Figure 1, panel (b)). Both systems therefore leave one more such worker uncompensated than in T1, and eight workers still fill in the form under the Complex system. Its advantage over the Simple system is unchanged: it still compensates one more

¹⁶In W , u , v and b describe the spectator's beliefs about how workers value earnings and effort and about the size of the compensation. Spectators are not told the amounts in pounds, but they know how long the form takes and that the compensation closes the average earnings gap, $y_L + b = y_H$. Spectators may differ in these beliefs; I suppress the spectator index.

worker whose output was not fully registered than the Simple system does, at the cost of eight forms. Hence $\Delta^{T3} = 8B^{T1} - N = \Delta^{T1}$. Although the Complex system no longer reaches every worker it is meant to compensate, a utilitarian spectator makes exactly the same choice as in T1.

Prediction 2 (Exclusion) *Every spectator makes the same choice in T1 and T3, so the share choosing the Complex system is the same.*

For a utilitarian spectator, the exclusion of a single worker does not matter, because it takes nothing away from the Complex system's advantage over the Simple system, and thus support for Complex should be the same in T1 and T3. The prediction also holds for any generalised-utilitarian welfare function, $\sum_i \varphi(U_i)$ with φ increasing.¹⁷ More generally, the prediction does not require the spectator to be utilitarian - it holds whenever spectator's evaluation is separable across workers, i.e., whenever each worker's contribution to the evaluation depends only on their own consumption and effort. This includes fairness views under which the spectator penalises each worker's deviation from what they consider fair (Cappelen et al. 2007, 2013; Konow 2000) and, separately, dislikes the burden of the form.¹⁸ I pre-registered this benchmark (H2, with Prediction 1 as H1) together with the two-sided alternative that support for the Complex system changes in T3.

3.2. Alternative Hypotheses

Below, I discuss three cases in which support for the Complex system may nonetheless change in T3; these seem the most plausible.

Losing perfect targeting matters. A utilitarian spectator's evaluation is separable across workers. A spectator may instead value a system for compensating *every* worker whose output was not fully registered, as under a premium on perfect targeting (Isaac and Spangenberg 2021), or be less concerned with each additional exclusion. In T3, the exclusion is the Complex system's first, which ends its perfect targeting, but the Simple system's second, so support for the Complex system falls. A spectator may also judge each system

¹⁷Compared with T1, the T3 group has one more Green worker whose output was not fully registered. Under the Simple system, this worker is not compensated. Under the Complex system, this worker fills in the form and is compensated, but the Yellow worker to whom the form is unavailable is not. Under both systems, the group therefore ends up with one fewer worker whose output was fully registered and one more worker whose output was not fully registered and who is left uncompensated, while all other workers are as in T1. Social welfare under both systems changes by the same amount, whatever φ is, so the choice does not change.

¹⁸Let the spectator minimise $\sum_i f(|c_i - c_i^*|) + \beta n_V$, where c_i^* is worker i 's fair earnings, f is increasing, and β is the disutility of filling in one form. In T1, the Simple system overpays one worker and misses one, a loss of $2f(b)$, and the Complex system costs 8β . In T3, the Simple system loses $3f(b)$ and the Complex system $f(b) + 8\beta$. The difference between the two systems is $2f(b) - 8\beta$ in both treatments.

by its worst-off worker, as under a Rawlsian criterion, so that the Complex system loses its advantage as soon as it leaves one such worker uncompensated: in T3 both systems leave such a worker uncompensated and tie on their worst-off worker, and support falls as long as some spectators who become indifferent choose the Simple system.¹⁹

The reason for exclusion matters. A utilitarian spectator also judges the systems only by the welfare they give workers (welfarism), not by why a worker goes without compensation. Under the Simple system, workers miss out because of their colour; under the Complex system in T3, because the form is unavailable to them. A spectator who finds exclusion through an unavailable form more (less) unfair than exclusion by colour supports the Complex system less (more) in T3.

Design-related explanations. Finally, support for the Complex system could fall if spectators perceive T3 differently from what the design intends. In T3, unlike in T1, both systems miss some workers whose output was not fully registered, so the Complex system's fairness advantage is less obvious; spectators may then choose more noisily or miss the advantage, both of which lower support for the Complex system.

The effort costs associated with filling in the form may also become more salient once the fairness comparison is less obvious. Furthermore, learning that the form is unavailable to one worker may change spectators' beliefs about the setting in ways the design does not intend.

4. Results

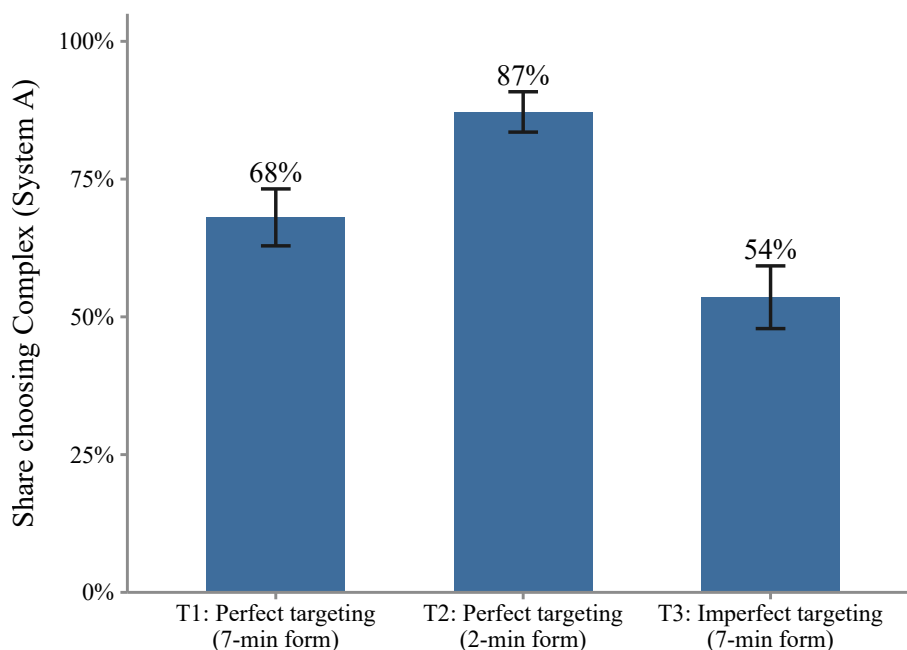
4.1. Main Results

Figure 2 shows the share of spectators who chose the Complex system (System A) in each treatment, and thus how their choices respond to the length of the form (T1 vs. T2) and to whether the Complex system reaches all the workers affected by the shock (T1 vs. T3). A majority chooses it in every treatment: 68% in T1, where the form takes seven minutes, almost half the working session, and the Complex system reaches all the workers affected by the shock; 87% in T2, where the form takes two minutes and the Complex system again reaches all of them; and 54% in T3, where it fails to reach one of them. In T1, choosing the Complex system means that eight workers spend 56 unpaid minutes in total on the form

¹⁹This requires that the compensation outweighs the burden of the form, $B^{T1} < u(y_L + b) - u(y_L)$; otherwise the workers who complete the form are the worst-off under the Complex system in both treatments, and a Rawlsian spectator chooses the Simple system in T1 and T3 alike. This condition need not hold for all spectators. If ties are broken by comparing the next worst-off workers (leximin), the Complex system still wins in T3. Appendix B.2 discusses rank-dependent welfare functions and the conditions under which they predict a fall in support for the Complex system in T3.

so that one extra payment, which spectators know makes up for the pay lost in the last seven minutes of the session (Section 2.1), goes to a worker affected by the shock instead of one who is not. The time cost of the Complex system is thus 8 times the value of the payment it redirects, when both are measured in minutes of work²⁰, yet two thirds of spectators choose it.

Figure 2. Spectators' choice of compensation system, by treatment



Notes: The figure reports the share of spectators who chose to implement the Complex system (System A) for the group of workers. Error bars indicate the 95% confidence interval. (N = 928)

To test whether these differences are statistically significant, Table 2 reports the corresponding regressions of system choice on treatment indicators; the reference category is T1 (perfect targeting, high transaction costs). The regression in Column (2) controls for respondents' characteristics including age, gender, education, income, and whether the respondent has ever received benefits.

These regressions show that the differences in Figure 2 are statistically significant and robust to controlling for spectators' characteristics and to using the pre-registered

²⁰This assumes that spectators regard (i) a minute on the form as costly as a minute of work, and (ii) the marginal cost of effort as constant. If spectators instead see the cost of effort as convex, each minute on the form, which comes after the fifteen-minute session, costs more than a minute of work, and the ratio underestimates the burden they accept.

Table 2. Choice of the Complex system

	Chose Complex (System A)	
	(1)	(2)
T2 (low cost)	0.191*** (0.032)	0.190*** (0.032)
T3 (imperfect targeting)	-0.145*** (0.039)	-0.145*** (0.039)
Constant	0.681*** (0.026)	0.710*** (0.063)
Controls		✓
Observations	928	928
Adjusted R^2	0.088	0.088

Notes: The table reports OLS regressions where the dependent variable equals 1 if the spectator chose to implement the Complex system (System A) and 0 if they chose the Simple system (System B). The independent variables are treatment indicators; T1 is omitted. Column (2) adds controls: age (in years), an indicator for being female, an indicator for having a Bachelor's degree or higher, an indicator for a yearly household income of at least £45,000, and an indicator for having ever received benefits. The sample (main sample) consists of all spectators who completed the study and passed both attention checks, excluding submissions returned on Prolific or flagged as likely bots. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

robustness sample (Appendix Table A1).²¹ Shortening the form (T2) raises the share choosing the Complex system by 19 percentage points ($p < 0.01$), in line with Prediction 1. When the Complex system fails to reach one worker affected by the shock (T3), the share falls by 15 percentage points ($p < 0.01$), contrary to Prediction 2. Spectators are thus less willing to accept the effort costs when these costs only improve distributional fairness than when they achieve it fully, even though the Complex system's advantage over the Simple system, measured in workers reached, is unchanged. The two effects are of comparable size, and the difference between their absolute sizes is not statistically significant ($p = 0.46$): the loss of support from failing to reach one worker is comparable to the gain from shortening the form by five minutes, or 40 unpaid minutes in total across the eight workers.

²¹Two of the eight quiz items differ across treatments: one has a different correct answer in T3, and another states the length of the form, which differs by treatment. When comprehension is measured only on the six items that are identical across treatments (see the comprehension quiz in Appendix D.1), both treatment effects are similar (Appendix Table A2).

4.2. Mechanisms

To understand what drives spectators' choices, I use perceptions and open-ended justifications elicited after the choice. Eliciting them afterwards avoids influencing the choice, but they may partly reflect it rather than the considerations behind it, so this evidence is only suggestive. Spectators perceive the Complex system as requiring significantly less effort in T2 than in T1 (Appendix Tables A3 and A4, Columns (1)–(2)). This is consistent with the lower burden of the Complex system driving its higher support in T2: controlling for perceived effort reduces the T2 effect from 19 to 6 percentage points, and from 17.5 to 3.1 percentage points in the robustness sample (Appendix Table A5). Perceived effort is also related to the choice within every treatment, in both samples (Appendix Table A6). To understand why support falls in T3, I examine the three cases discussed in Section 3.2 in turn.

The reason for exclusion matters One possibility is that spectators object to *how* the Complex system excludes a worker, by denying access to the form, more than to how the Simple system does, by conditioning on an imperfect proxy. This is a departure from welfarism (Section 3.2), because the way spectators evaluate the systems then depends on more than workers' welfare. To test this, in T3 I elicited which reason for exclusion spectators find more unfair, setting aside how many workers each affects. The distribution of responses is shown in Appendix Figure A1. If anything, spectators find missing out under the Simple system more unfair than missing out under the Complex system: the mean response of 5.41 lies significantly above the neutral midpoint of 5 ($t = 3.01$, $p = 0.003$),²² and more spectators regard exclusion under the Simple system as more unfair (36.9%) than under the Complex system (21.7%), with 41.4% seeing the two as equally unfair. Moreover, even spectators who find the two reasons for missing out equally unfair, and for whom the reason for exclusion therefore cannot explain the fall, choose the Complex system far less often than spectators in T1 (43% against 68%).²³ Procedural concerns are therefore unlikely to be the main explanation for the fall in support in T3.

Design-related explanations The effort costs are the same in T1 and T3 by design. If spectators nevertheless perceive the form as more burdensome in T3, support for the Complex system could fall. Perceived effort required to receive compensation indeed

²²Responses were elicited on a scale from 0 ("missing out under System A is much more unfair") to 10 ("missing out under System B is much more unfair"). The result holds on the robustness sample: the mean response is 5.42, significantly above the neutral midpoint of 5 ($t = 2.60$, $p = 0.010$, $N = 208$).

²³Furthermore, spectators who find missing out under the Complex system more unfair choose it about as often as those who find the two reasons equally unfair (39% against 43%; two-sample test of proportions, $\chi^2(1) = 0.18$, $p = 0.68$).

increases by 0.31 points on the 0–10 scale in T3 relative to T1 in the main sample (marginally significant) and by 0.57 points in the robustness sample ($p < 0.01$; Appendix Tables A3 and A4). This increase could reflect a general disapproval of the imperfect-targeting condition or, since the question is asked after the choice, post-choice rationalisation. The spectators were aware of the length of the form; in every treatment, almost all correctly answered a comprehension question about it (T1: 95.8%; T2: 93.1%; T3: 92.5%). Furthermore, controlling for perceived effort reduces the T3 effect only from 14.5 to 12.9 percentage points, and from 19.0 to 15.9 percentage points in the robustness sample (Appendix Table A5). It is thus unlikely that the difference in perceived effort explains the lower support in T3.

An alternative explanation is that the choice is more difficult in T3: unlike in T1, both systems miss some workers whose output was not fully registered, so the Complex system’s fairness advantage may be less obvious. If the difficulty of the choice explained the fall in support, it should be concentrated among spectators who understand the setting less well, since they may choose more noisily or miss the Complex system’s advantage. I find no evidence of this: the T3 effect does not differ by education or comprehension, the two available proxies for understanding (Table 3). Even among spectators who answered every comprehension question correctly, and who therefore understood the setting very well, support still falls by 15.6 percentage points in T3 relative to T1 ($p < 0.01$; Table 3, Column (2)).²⁴

Finally, when neither system reaches all the workers affected by the shock, the difference in fairness between the two systems may become less salient, shifting spectators’ attention to the burden of the form. To examine this, I use spectators’ open-ended justifications of their choice, coded with a large language model for whether each response mentions (i) the fairness of the resulting distribution and (ii) the effort costs.²⁵ Appendix Figures A2 and A3 report the share of spectators mentioning each consideration, by treatment and by the system they chose. The responses provide no support for this explanation: the form is not mentioned more often in T3 than in T1 (45% against 50%; $\chi^2(1) = 1.78$, $p = 0.18$). Among spectators who choose the Simple system, those in T3 do not mention the form more often than those in T1 (70% against 77%; $\chi^2(1) = 1.08$, $p = 0.30$). If spectators had turned to the Simple system in T3 because the burden of the form had become more salient, this share would have risen instead. This comparison is, however, only suggestive, since the treatment changes who chooses each system.

²⁴In the robustness sample, it falls by 16.7 percentage points (Appendix Table A7, Column (2)).

²⁵Responses were coded with GPT-5.6 (gpt-5.6-luna) via the GABRIEL package (Asirvatham, Mokski, and Shleifer 2026). A response is coded as mentioning a consideration whenever it raises it, even if it is not the decisive reason for the choice: a response such as “I disliked the form, but I went with what is fair” is coded as mentioning both fairness and the effort costs. Appendix C reports the prompts.

Table 3. Choice of the Complex system: heterogeneity

	Chose Complex (System A)				
	Interaction with moderator			Split sample	
	(1)	(2)	(3)	(4)	(5)
T2 (low cost)	0.244*** (0.058)	0.211*** (0.038)	0.265*** (0.050)	0.130*** (0.041)	0.267*** (0.051)
T3 (imperfect targeting)	−0.151** (0.070)	−0.156*** (0.052)	−0.052 (0.057)	−0.224*** (0.055)	−0.046 (0.058)
Moderator	0.039 (0.058)	−0.051 (0.056)	0.144*** (0.054)		
T2 × Moderator	−0.080 (0.070)	−0.057 (0.069)	−0.133** (0.065)		
T3 × Moderator	0.011 (0.085)	0.035 (0.081)	−0.172** (0.080)		
Constant	0.698*** (0.071)	0.723*** (0.066)	0.622*** (0.072)	0.760*** (0.082)	0.603*** (0.098)
Moderator	High education	Imperfect comprehension	Left-wing		
Sample	All	All	All	Left-wing	Centre-right
Controls	✓	✓	✓	✓	✓
Observations	928	928	928	466	462
Adjusted R^2	0.088	0.090	0.093	0.096	0.083

Notes: The table reports OLS regressions where the dependent variable is the choice of the Complex system. The independent variables are the treatment indicators, interacted in columns (1)–(3) with a moderator: having a Bachelor’s degree or higher (column 1), an indicator for at least one incorrect answer on the eight comprehension questions (column 2), and an indicator for left-wing self-placement, 0–4 on a scale from 0 (left) to 10 (right) (column 3). Columns (4) and (5) estimate the treatment effects separately for left-wing and centre-right (5–10) spectators. All columns include controls, except education in column (1), where it is the moderator. The sample is the main sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Losing perfect targeting matters Another explanation is that spectators’ evaluation of the systems is not separable across workers (Section 3). For example, they may value the Complex system for compensating *every* worker affected by the shock, beyond the number of such workers it reaches, as under a perfection premium, or judge each system by its worst-off worker, as under a Rawlsian criterion. In both cases, the Complex system loses its advantage as soon as it leaves one such worker uncompensated. Two pieces of evidence are consistent with this explanation.

First, the Complex system’s targeting advantage over the Simple system is the same in T1 and T3. If spectators’ evaluations of the systems are separable across workers, the perceived fairness advantage of the Complex system should be the same in both treatments

as well. If, instead, spectators give extra weight to a system reaching every worker affected by the shock, beyond the number of workers it reaches, or judge each system by the outcome of its worst-off worker, this advantage should shrink in T3 (Section 3.2). To test this, I use spectators' ratings of the relative fairness of the two distributions.²⁶

The perceived fairness advantage of the Complex system indeed shrinks in T3: spectators still rate the distribution of compensation under the Complex system as fairer than the Simple system's, but by less than in T1 (Appendix Table A3, Columns (3)–(4)). Among spectators who chose the Complex system, those in T3 also perceive its fairness advantage as smaller than those in T1 (Appendix Table A8), so the result is unlikely to reflect only post-choice rationalisation.²⁷

Second, if the fall in support is explained by how spectators evaluate who receives the compensation, rather than by considerations unrelated to the distribution (e.g., confusion or a general dislike of complicated systems), it should be concentrated among spectators who care most about distributional fairness. To test this, I split spectators by political self-placement, using left-wing affiliation (0–4 on a scale from 0, left, to 10, right) as a proxy for concern for distributional fairness and for the worst-off (Alesina and Giuliano 2011; Almås, Cappelen, and Tungodden 2020). The fall in support for the Complex system between T1 and T3 is indeed concentrated among left-wing spectators (Table 3, Columns (3)–(5)). In T1, 75% of left-wing spectators choose the Complex system, against 59% of centre-right spectators (5–10 on the scale); in T3, support among left-wing spectators falls to 52%, slightly below the 55% among centre-right spectators. With controls, this is a fall of 22 percentage points, against an insignificant 5 percentage points among centre-right spectators; the difference between the two groups is significant (interaction: -0.17 , $p < 0.05$, Column (3)), also in the robustness sample (Appendix Table A7).²⁸

²⁶Spectators were asked which system's distribution of the compensation is fairer, from 0 (System A's distribution is much fairer) to 10 (System B's distribution is much fairer), and were instructed to base their answer only on who receives the compensation and who does not, ignoring all other properties of the two systems.

²⁷The perceived fairness advantage of the Complex system is also larger in T2, although who receives the compensation is the same as in T1. This suggests that spectators judge a system's distribution as fairer when it requires less effort, even though they were asked to consider only who receives the compensation. Since spectators perceive that slightly more effort is required to receive the compensation in T3, this could also lower the Complex system's fairness advantage in T3. Controlling for perceived effort, however, the effect of T2 on the Complex system's perceived fairness advantage disappears, while the effect of T3 barely changes (Appendix Tables A3 and A4, Column (5)).

²⁸The results are robust to measuring political self-placement on the continuous 0–10 scale instead of splitting it at 4 (Appendix Table A10). Left-wing spectators in T3 are also not more likely than centre-right spectators to regard missing out under the Complex system as more unfair (19% against 24%; $\chi^2(1) = 0.79$, $p = 0.37$), so the larger fall in support among left-wing spectators is unlikely to reflect procedural concerns.

Together with the smaller fairness advantage of the Complex system in T3, the concentration of the fall among left-wing spectators suggests that support falls in T3 because spectators' evaluation of the Complex system depends on whether it leaves any worker affected by the shock uncompensated, not only on how many more of them it reaches than the Simple system.²⁹ The design cannot, however, distinguish between a premium on perfect targeting, diminishing sensitivity to targeting errors and a Rawlsian criterion, since all three are consistent with a fall in support in T3. Separating them is left to future research.

Cappelen, Cappelen, and Tungodden (2023) show that left-leaning spectators are more averse than right-wing spectators to false negatives, that is, to leaving deserving individuals without compensation. Between T1 and T3, the only change is one extra false negative under each system. Differences in how spectators view these two errors therefore cannot explain the fall in support. Extending Cappelen, Cappelen, and Tungodden (2023), my results thus suggest that spectators do not simply add up errors across workers, but also have preferences over properties of the allocation as a whole.

4.3. Choices and Policy Attitudes

To see whether choices in the experiment reflect spectators' views on real benefit systems, I regress four items collected after the choice on an indicator for implementing the Complex system, controlling for the treatment indicators, demographic characteristics, left-wing self-placement and generalised trust (Appendix Tables A11–A14). On a scale from 0 ("benefits should be universal and easier to claim") to 10 ("benefits should be targeted, even if harder to claim"), spectators who implemented the Complex system place themselves 0.57 points higher ($p < 0.01$; Appendix Table A11, Column (3)): those willing to impose transaction costs on workers for accurate targeting also tend to favour targeting in real benefit systems.³⁰ Support for the two simplification policies, a negative income tax and a universal basic income, is unrelated to the choice, possibly because it turns on considerations (e.g., financing and budget costs of a reform). Support for simplification reforms of an existing benefit system is the focus of the survey experiment (Section 5).

²⁹Left-wing spectators also respond less to the shorter form (Table 3, Column (3)), but this likely reflects a ceiling effect: their support is already 75% in T1, and support in T2 is close to 87% in both groups.

³⁰Confidence in government is unrelated to the choice, so the choice does not simply reflect a general disposition towards the state and its administration (Appendix Table A14).

5. Survey Experiment

The spectator experiment shows that the costs of complexity, both the effort it requires of claimants and its failure to reach everyone it is meant for, reduce support for complex systems. In a real welfare state, however, support for reform depends on many other factors, such as its fiscal cost, concerns about fraud, attitudes towards reform in general, and personal experience with the benefit system, and it is therefore an empirical question whether beliefs about these costs affect it. The survey experiment examines this question, focusing on three costs of complexity that are salient in the Dutch welfare state: non-take-up, the difficulty of understanding the rules, and the risk of having the amount of their benefits corrected at the end of the year, i.e., having to pay money back or receiving an additional payment.

The Dutch welfare state provides support through many targeted tax credits, allowances, and benefits,³¹ and experts broadly agree that the system has become too complex for many citizens to manage (Centraal Planbureau and Sociaal en Cultureel Planbureau 2020; IBO Vereenvoudiging Sociale Zekerheid 2023; Inspectie Belastingen, Toeslagen en Douane 2026; Kok et al. 2022).

5.1. Experimental Design

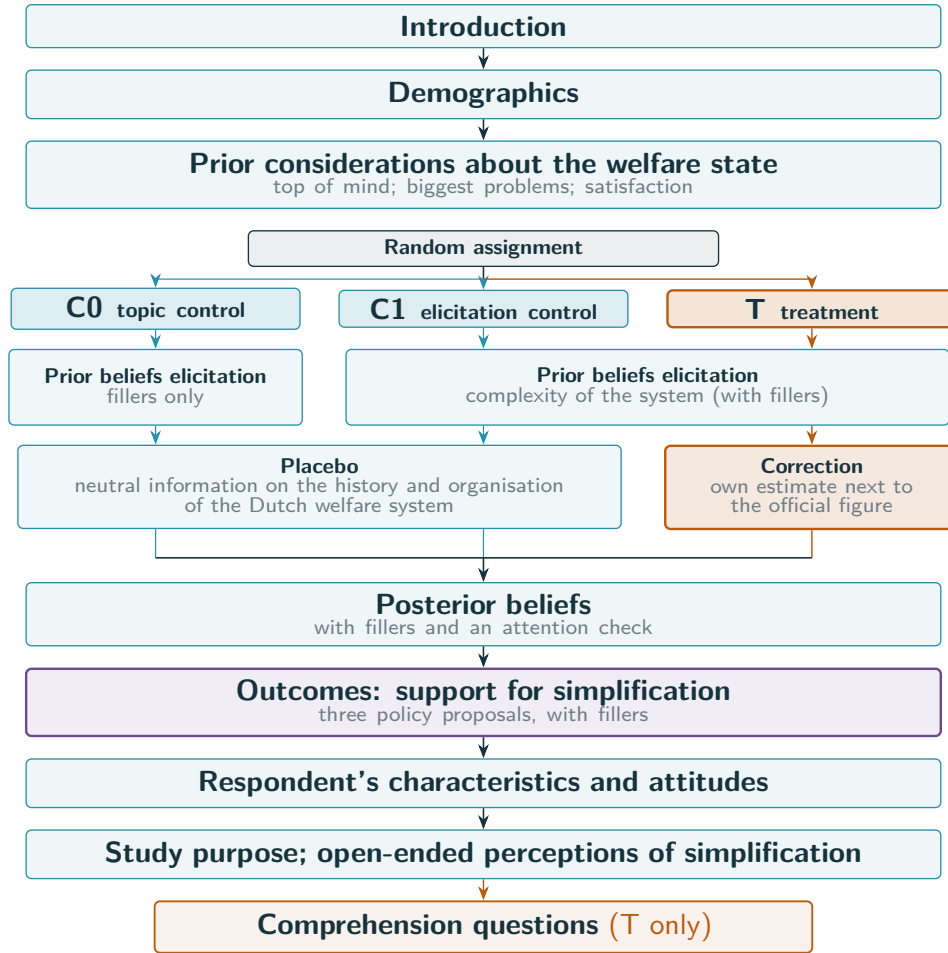
5.1.1. Structure of the Survey

Figure 3 shows the structure of the survey, and Appendix D.2 provides the full questionnaire. After reporting their baseline views, respondents are randomly assigned to one of three arms. In the treatment arm (T), I elicit their prior beliefs about three costs of complexity and then show them the correct figures, with additional information on these costs; in the elicitation-control arm (C1), I elicit the same beliefs without providing this information; in the topic-control arm (C0), respondents estimate only unrelated quantities about the welfare state and read placebo information. To conceal the purpose of the study, I include filler questions in several blocks of the survey.

Baseline considerations. After giving consent and reporting their age, gender, and education, respondents are told that the study concerns their opinions on and experiences with social security. They then describe in open text what works well and what works poorly, select its two most important problems from a list that includes fraud, over-generosity, in-

³¹Benefits (*uitkeringen*) include, for example, social assistance (*bijstand*) and unemployment benefit (WW). Allowances (*toeslagen*) are income-dependent contributions to specific costs, such as health insurance, rent, or childcare. Tax credits (*heffingskortingen*) reduce the income tax and national insurance contributions that individuals pay.

Figure 3. Outline of the survey experiment



adequacy, complicated rules, the number of schemes and agencies, and work disincentives (the fourth and fifth concern complexity), and rate their satisfaction with the welfare state.

Prior beliefs. Respondents in T and C1 then estimate, out of 100, three costs of complexity prominent in the Dutch system: how many people entitled to social assistance receive it (*take-up*; true figure 65, Inspectie SZW (2021)), how many allowance recipients correctly identified which assets they must report to the Benefits Agency (*understanding*; 26, Kok et al. (2022)), and how many rent-allowance households had to repay money or received an additional payment at the year-end reconciliation (*corrections*; 68, Centraal Planbureau (2019)).³² All arms also answer two filler items that are never corrected, on the age of

³²In the Netherlands, allowances are paid in advance on the basis of the household's estimated income and circumstances, and recalculated at the end of the year. Because rules are complicated and one's income may be hard to predict, many households receive too much or too little and must repay part of their allowance or receive an additional payment. Repayments often leads to debt (IBO Vereenvoudiging Sociale Zekerheid 2023; Kok et al. 2022).

allowance recipients and the share of adults receiving benefits; these are the only items shown to C0. Items appear in random order.

I phrase all questions neutrally and never describe the phenomenon as a problem.³³ Rather than beliefs about the complexity of the welfare state as a whole, I elicit a few quantities that are most likely to shift beliefs about its costs. I therefore interpret gaps between these estimates and the true figures as misperceptions of these three costs only, not as systematic misperceptions of the complexity of the welfare state; measuring such misperceptions is left to future work.

Information treatment. In Section 5, when you introduce the three costs: say that non-take-up mirrors the exclusion in T3.

Respondents then read short information screens with a minimum viewing time. In C0 and C1 these are placebo screens on the history and organisation of the Dutch welfare system. In T, each screen shows respondent their prior estimate next to the official figure and explains how the cost relates to complexity³⁴ Because respondents might otherwise attribute the costs to other causes, the treatment combines information with an interpretation of it, so its effect cannot be attributed to the information alone. I combine information on the three costs in a single treatment to maximise power.

Posterior beliefs. I do not re-elicite the corrected figures, which could capture recall and invite experimenter demand (Haaland, Roth, and Wohlfart 2023; Stantcheva 2023). Instead, all respondents rate whether having many separate schemes, each with its own rules, conditions, and application procedure, makes things better or worse for those entitled to support (0 = much worse, 10 = much better).³⁵ Below, I refer to this rating as whether complexity makes things better or worse for claimants (“Complexity helps claimants” in the tables). Respondents in C1 and T also answer the corrections question for the healthcare allowance, which the treatment does not mention, to measure whether respondents extend the information to another allowance (“Benefit corrections” in the tables); I do not ask this in C0, to avoid priming its respondents on the topic.

³³For example, I ask how many entitled people *receive* a benefit rather than how many do not, and I avoid leading questions such as how many people “find the rules complicated”.

³⁴For example, the screen on corrections says that the Benefits Agency sees complexity of the allowance system as an important cause of these repayments and that the number of households in debt because of this has increased. The screen on missing out on support adds that EUR 1 billion in allowances went unclaimed in 2021, that many people do not apply out of fear of having to pay money back or because the procedure is too complicated, and that half of the people on a Wajong benefit do not know they can get extra support.

³⁵The block opens with two filler items, so that this question does not directly follow the information screens.

Outcomes: support for simplification. The main outcomes are support for three proposals, each rated from 0 (keep the current system) to 10 (fully support the change), so that values above the midpoint indicate a preference for the change over the current system.³⁶ *Policy 1* consolidates the many benefits and allowances into fewer schemes, making the system easier to navigate at the cost of less accurately targeted support. *Policy 2*, from the government's review of alternatives to the allowance system, replaces most allowances with a single income-independent allowance, financed by abolishing tax credits and deductions (Dijksterhuis and van Baaren 2020; Ministerie van Financiën 2020). Like the choice in the spectator experiment, both trade simplicity against accurate targeting. *Policy 3* instead keeps all schemes in place and makes the consequences of mistakes milder, reducing the costs of complexity without sacrificing targeting. All proposals are presented as roughly budget-neutral and as not making it easier to obtain support without being entitled to it.³⁷

Respondent characteristics and attitudes. Finally, respondents report demographics and political attitudes, say what they think the study is about and whether any part of the survey steered them towards a particular answer, and give their main arguments for and against the proposals. Respondents in T also answer three comprehension questions.

5.1.2. *Experimental Procedures, Comprehension, Demand Effects*

I conducted the survey with a sample of Dutch respondents recruited through the online panel provider Bilendi, using quotas on age, gender, and education to match the Dutch adult population. Respondents received Bilendi's standard panel incentive. Of the 2,728 respondents who completed the survey, I exclude, as pre-registered, those who failed the instructed attention check or were among the fastest 5% of these respondents within their arm in total completion time, leaving an analysis sample of 2,212 (C0: 755; C1: 730; T: 727), with a median completion time of 9.5 minutes. The sample matches the Dutch adult population closely on gender and is somewhat older (51.8 against 49.1 years); as is common in online panels, respondents with a low level of education make up a smaller share than in the population (17.4% against 26.0%; Appendix Table A19).

Most treated respondents understood the information: 72.6% identified its main message, and 58.8% correctly recalled the rent-allowance figure. Only 8.5% of respondents

³⁶The full wording of all proposals is provided in Appendix D.2.

³⁷Policies 1 and 3 state that total spending on support and taxes stay about the same. Policy 2 instead states how it is financed, by abolishing tax credits, deductions and the current allowances, following the government's review of alternatives to the allowance system (Ministerie van Financiën 2020).

identified the purpose of the study,³⁸ a share that does not increase with the treatment (Appendix Table A21); 12.0% felt steered by some part of the questionnaire, 4.3 percentage points more in T than in C0 ($p < 0.05$).

5.2. Results

Beliefs about complexity. Respondents, on average, slightly overestimate non-take-up of social assistance (mean take-up estimate 59.8 against 65), although individual estimates vary widely in both directions.³⁹ They overestimate how many recipients understand the asset rules (47.5 against 26) and underestimate how many rent-allowance households face year-end corrections (49.7 against 68; Appendix Figures A5–A7 show the distributions of prior beliefs).⁴⁰

Impact of the intervention on beliefs and policy support. The intervention moves beliefs about the costs of complexity, but not support for simplification. Table 4 reports the effects.⁴¹ The intervention makes respondents see complexity as worse for claimants: it lowers their rating by 0.586 points on the 0–10 scale ($p < 0.01$). Column (2) shows that treated respondents raise their estimate of year-end corrections in the healthcare allowance by 6.19 points ($p < 0.01$). However, this belief change does not translate into policy preferences (columns 3–5). The effect on policy support also does not differ between respondents who underestimated these costs and those who did not (Appendix Tables

³⁸I code the open-ended answers with a large language model, using the GABRIEL package (Asirvatham, Molski, and Shleifer 2026) with three runs and a majority rule. An answer counts as identifying the purpose if it states that the study is about support for simplifying the benefits system. Appendix C reports the prompts.

³⁹Complexity is a common concern among respondents: 13% mention it spontaneously when asked what works well and what works poorly in the Dutch welfare state (answers coded with a large language model; Appendix C.2.1 reports the prompts), and 58% select a complexity-related item as one of the system's biggest problems (Appendix Figure A4). In the topic-control arm, 48% of respondents think that the many separate schemes make things worse for the people entitled to support, against 30% who think they make things better (22% see no difference).

⁴⁰I measure the gap between each prior and the official figure so that a positive gap means the respondent underestimates the cost of complexity. Averaged over the three items, the gap is 11.6 points ($t = 34.96$, $p < 0.001$). The gap is 21.5 points for understanding the rules ($t = 36.72$) and 18.3 points for year-end corrections ($t = 32.56$). For take-up of social assistance it is -5.2 points ($t = -8.24$): respondents slightly overestimate non-take-up, and although the difference is small, it is statistically significant (all $p < 0.001$).

⁴¹Throughout, I report the effect of the full intervention (T vs. C0), which combines eliciting beliefs with correcting them, as merely eliciting beliefs has no statistically significant effect on any outcome (C1 vs. C0). The T vs. C1 contrast is reported in Tables 4 and 5.

A24–A26).⁴² That beliefs move more easily than policy preferences is a common finding in information provision experiments (Haaland and Roth 2023; Haaland, Roth, and Wohlfart 2023; Kuziemko et al. 2015).

Heterogeneity in responses to the intervention. To examine whether the null average effect on policy support hides differences between groups, I estimate heterogeneous treatment effects by pre-registered respondent characteristics (Appendix Table A28).⁴³ For Policy 1, the treatment effect is smaller for benefit recipients (respondents who have ever received at least one of the listed benefits or allowances, excluding the state pension) and for the highly educated (respondents with a tertiary degree).⁴⁴ In case of Policy 2, it does not vary with any characteristic. For Policy 3, which would soften the consequences of mistakes without reducing targeting, the treatment increases support among respondents who see complexity as a problem, those less concerned with fraud, and those with higher incomes (the latter two differences are significant only at the 10% level).

Since Policy 1, unlike Policy 3, involves the trade-off between simplicity and accurate targeting that the spectator experiment studies, I examine its heterogeneous effects further, starting with benefit receipt. Benefit recipients make up 61% of the sample; this measure also captures likely future receipt (IBO Vereenvoudiging Sociale Zekerheid 2023). Most of them have received a means-tested scheme (85%), most often the healthcare or rent allowance, and the median recipient has received three different schemes (Appendix Figure A8).

At baseline, recipients' prior beliefs about the costs of complexity are as accurate as those of non-recipients, controlling for other characteristics (Appendix Table A32),⁴⁵ and the two groups do not differ in whether they think complexity makes things better or worse for claimants and have a similar level of support for a simpler system (mean 5.9

⁴²The effect of the treatment on numerical beliefs depended on the accuracy of prior beliefs: respondents who had underestimated year-end corrections in the rent allowance more raised their estimate for the healthcare allowance by more, whereas the shift in their rating of whether complexity makes things better or worse for claimants is similar for all respondents (Appendix Table A27). This is consistent with both parts of the treatment playing a role: the figures corrected numerical beliefs, while the interpretation may have changed how respondents evaluate complexity regardless of their priors.

⁴³The pre-analysis plan specified the comparison with the elicitation control (C1); Appendix A.3 reports this and the other deviations.

⁴⁴Appendix A.2.1 describes how these and other respondent characteristics relate to support for simplification in the topic-control arm.

⁴⁵The distributions of prior beliefs by benefit receipt are shown in Appendix Figure A9. Recipients are not significantly more accurate than non-recipients on any of the three items (Appendix Table A33, odd columns). The only difference is that recipients' estimates of take-up of social assistance are higher, by 2.6 points controlling for other characteristics (column 2), which brings their mean closer to the official figure (60.8 against 58.4; official figure 65).

Table 4. Treatment effects on beliefs and policy support

	Beliefs		Policy support (0–10)		
	Complexity helps claimants (1)	Benefit corrections (0–100) (2)	Policy 1: simpler system (3)	Policy 2: single allowance (4)	Policy 3: milder consequences (5)
Treatment (T)	–0.586*** (0.104)	6.189*** (1.038)	0.163 (0.116)	0.150 (0.137)	0.166 (0.126)
Elicitation (C1)	–0.086 (0.103)	—	0.031 (0.118)	0.165 (0.137)	0.126 (0.130)
Constant	5.042*** (0.214)	60.472*** (2.550)	5.824*** (0.238)	6.034*** (0.276)	5.367*** (0.260)
<i>p</i> -value: T = C1	<0.001	—	0.28	0.92	0.75
Controls	✓	✓	✓	✓	✓
Omitted category	C0	C1	C0	C0	C0
Observations	2,212	1,457	2,212	2,212	2,212
Adjusted R^2	0.165	0.073	0.031	0.051	0.031

Notes: The table reports OLS regressions where the dependent variables are: *Complexity helps claimants*, whether the many separate schemes, each with its own rules, conditions and application procedure, make things better or worse for people entitled to support, from 0 (much worse) through 5 (no difference) to 10 (much better), so a negative coefficient means complexity is seen as worse for claimants; *Benefit corrections*, the estimated number of every 100 zorgtoeslag (healthcare allowance) recipients who had to repay or receive extra at the end of the year; and support for Policy 1 (a simpler system with fewer, more general schemes), Policy 2 (a single basic allowance replacing all allowances except the childcare allowance) and Policy 3 (milder consequences of mistakes), each from 0 (keep the current system as it is) to 10 (fully support this change). The independent variables are indicators for the treatment arm (T; prior beliefs elicited and corrected) and the elicitation control arm (C1; prior beliefs elicited, no correction). Controls are age (in years), an indicator for being female, an indicator for high education (hbo or wo), an indicator for a net yearly household income of at least EUR 40,000, an indicator for left-wing self-placement (0–4 on a scale from 0, left, to 10, right), trust in the government and public institutions (0–10), and an indicator for benefit recipients. Appendix Table A44 reports the same regressions with the controls specified in the pre-analysis plan. The sample (analysis sample) consists of respondents who completed the survey, passed the attention check and were not among the fastest 5% within their arm. In columns (1) and (3)–(5), the topic control (C0) is omitted. Benefit corrections were elicited only in C1 and T, so column (2) uses these two arms with C1 omitted. The *p*-value is from a Wald test that the coefficients on T and C1 are equal. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

among recipients against 6.0 among non-recipients, on the 0–10 scale; Appendix Table A22). The treatment moves the beliefs of both recipients and non-recipients. After receiving the information, recipients see complexity as worse for the people entitled to support (by 0.71 points, against 0.39 points for non-recipients; Table 5), and both groups raise their estimate of year-end corrections in healthcare allowance by a similar amount (Appendix

Table A29).⁴⁶ Yet only non-recipients change their policy views. The treatment increases their support for a simpler system by 0.48 points,⁴⁷ while the effect among recipients is close to zero.⁴⁸

The pattern is unlikely to be driven by experimenter demand, as it persists when I exclude respondents who guessed the purpose of the study or felt steered (Appendix Table A34), and among non-recipients, the treatment effect does not vary significantly with whether respondents guessed the purpose or felt steered (Appendix Table A35). Inattention is also unlikely: recipients are about six percentage points more likely than non-recipients to pass the attention check and to be included in the analysis sample, and among treated respondents, they are more likely to identify the main message of the information and to recall the corrections figure (Appendix Table A36). Finally, the pattern is not explained by observable differences between recipients and non-recipients: when the treatment effect may vary with all respondent characteristics at once, the benefit receipt interaction remains negative and significant (Appendix Table A37).

Receiving benefits means having experienced the complexity of the system first-hand, which could explain the difference in two ways. First, recipients may already know more about the costs of complexity than the survey measures, so that the information adds little to their knowledge. If so, however, their prior beliefs should be more accurate and the information should move their beliefs less. Instead, their prior beliefs are as inaccurate as those of non-recipients, and the information moves them at least as much. Second, having dealt with complexity themselves, recipients may see it as less of a problem. However, in the topic-control arm, recipients do not differ from non-recipients in whether they think complexity makes things better or worse for claimants (Appendix Table A22). Finally, under either explanation, the treatment effect should be smaller among recipients who have experienced more complexity. To test this, I use a measure of *experienced complexity*,

⁴⁶In a pooled regression with an interaction between the treatment and benefit receipt, the difference between recipients and non-recipients is not statistically significant, either for whether complexity makes things better or worse for claimants ($-0.33, p = 0.12$) or for the estimate of corrections in the healthcare allowance ($-1.47, p = 0.49$; Appendix Table A30).

⁴⁷Just before the three simplification proposals, respondents also rate two other policy proposals: raising the social minimum and abolishing the health insurance deductible (*eigen risico*). I include them as filler questions, but they can also be used to check whether the information changes general attitudes towards policy reform or towards the generosity of the welfare state. It does not: the information changes support for neither proposal, among either recipients or non-recipients (Appendix Table A31).

⁴⁸These are effects of the full intervention, which combines eliciting beliefs with correcting them. The difference between the treatment (T) and the elicitation control (C1), that is, the effect of the information alone, is not statistically significant among either non-recipients or recipients (Table 5).

that is, how complex the schemes each recipient has received are.⁴⁹ Among recipients, neither the effect on beliefs nor the effect on support for a simpler system varies with this measure (Appendix Table A40).

As a possible explanation of these results, the open-ended answers point to a personal stake in the system and to uncertainty about how a reform would affect recipients (Appendix Table A38). Compared with non-recipients, recipients are more likely to say that a reform could cut their support, to mention the loss of targeted support, and to reason from their own situation.⁵⁰ This is consistent with what recipients name as the main problems of the welfare state before the treatment: compared with non-recipients, they are more likely to name inadequate benefits (Appendix Table A23), so they may be less willing to support a reform that could reduce their benefits.

The difference in policy responses between recipients and non-recipients is specific to a general simplification of the system: for Policies 2 and 3, the effect of the treatment does not differ between the two groups (Appendix Table A28). The two proposals differ from Policy 1 in different ways. Policy 2 replaces most allowances with a single allowance for all households, financed by abolishing tax credits and deductions, so respondents may find it harder to tell how it would affect them. Policy 3 keeps targeted support, so it does not involve the trade-off between simplicity and accurate targeting, and it leaves recipients' benefits unchanged.

The effect on Policy 1 also differs by education. Both the highly educated and respondents without higher education update their beliefs (Appendix Tables A41 and A42), but only respondents without higher education also change their policy views: the treatment raises their support for a simpler system by 0.37 points.⁵¹ A likely explanation is that the highly educated start from a different baseline: they are more likely to name complexity as one of the main problems of the system (Appendix Table A23), see complexity as worse for

⁴⁹Following the pre-analysis plan, experienced complexity is the sum, over all schemes in the benefit-receipt question that the respondent has ever received, of each scheme's mean experienced-complexity score (on a 0–10 scale) from IBO Vereenvoudiging Sociale Zekerheid (2023), Table 5.3. Kindgebonden budget, which the review does not score, is assigned the mean score of the comparable income-dependent allowances (about 3.8). The measure increases with both the number of schemes received and their complexity, and is zero for respondents who have never received any of them. Among recipients, it averages 10.9, and one scheme adds between 2.6 and 5.1 points.

⁵⁰The results replicate on the subsample of respondents who gave a substantive answer (Appendix Table A39).

⁵¹The pattern is unlikely to be driven by experimenter demand, as it holds when I exclude respondents who guessed the purpose of the study or felt steered (Appendix Table A43), or by inattention, as education is unrelated to passing the attention check or to inclusion in the analysis sample (Appendix Table A36). These are effects of the full intervention, which combines eliciting beliefs with correcting them. The difference between the treatment (T) and the elicitation control (C1), that is, the effect of the information alone, is not statistically significant among either respondents without or with higher education (Table A41).

claimants, and are more supportive of a simpler system even without the information, in the topic-control arm (Appendix Table A22).⁵² Their weaker response to the information is also consistent with evidence that more educated individuals hold more stable and constrained political attitudes (Converse 1964; Zaller 1992).

Taken together, correcting beliefs about the costs of complexity is not enough on its own to raise support for simplification. Respondents underestimate two of the three elicited costs and update their beliefs when informed, but only non-recipients become more supportive of a simpler system, similar to the impartial spectators in Section 4. Their own stake in the system, and uncertainty about how a reform would affect them, weighs more heavily than the costs of complexity.

Table 5. Treatment effects on beliefs and policy support, by benefit receipt

	Complexity helps claimants		Policy 1: simpler system	
	Non-recipients (1)	Recipients (2)	Non-recipients (3)	Recipients (4)
Treatment (T)	-0.386** (0.164)	-0.709*** (0.135)	0.480*** (0.174)	-0.039 (0.155)
Elicitation (C1)	-0.069 (0.160)	-0.093 (0.135)	0.204 (0.181)	-0.062 (0.155)
Constant	4.997*** (0.308)	4.877*** (0.280)	5.530*** (0.332)	5.712*** (0.317)
<i>p</i> -value: T = C1	0.065	<0.001	0.128	0.890
Controls	✓	✓	✓	✓
Adjusted <i>R</i> ²	0.138	0.178	0.021	0.029
Observations	865	1,347	865	1,347

Notes: The table reports OLS regressions where the dependent variables are Complexity helps claimants (columns 1 and 2) and support for Policy 1 (columns 3 and 4). The independent variables are the treatment and elicitation indicators; C0 is omitted. Controls are age, female, high education, higher income, an indicator for left-wing self-placement (0–4 on a scale from 0, left, to 10, right), and trust in the government and public institutions (0–10). The sample is the analysis sample, split into benefit recipients, who have ever received at least one of 14 listed benefit schemes (AOW excluded), and non-recipients. The *p*-value is from a Wald test that the coefficients on T and C1 are equal. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

⁵²Their prior beliefs about the costs of complexity are, on average, as inaccurate as those of respondents without higher education (Appendix Table A32); they are more accurate about understanding the rules, but less accurate about year-end corrections (Appendix Table A33). Their open-ended justifications of their policy view also suggest that they weigh costs of complexity against other considerations: they are more likely to mention costs and efficiency and the risk that a reform would increase fraud (Appendix Table A38), which may offset the case for simplification made by the information; they are, however, also more likely to mention complexity and bureaucracy (by 7.2 percentage points). The open-ended results replicate on the subsample of respondents who gave a substantive answer (Appendix Table A39).

6. Conclusion

Complexity of benefit systems improves the accuracy of targeting, but imposes transaction costs on claimants. This paper provides insight into why, despite political appetite for simplification in many countries, progress on reform has been slow.

First, the spectator experiment shows that people accept a high price for accurate targeting: two thirds of spectators choose a system in which eight workers spend 56 unpaid minutes in total so that one payment reaches a deserving worker. This means that proposed reforms that sacrifice targeting may be unpopular.

Second, the two experiments suggest different binding constraints on simplification. When people judge the trade-off impartially, as the spectators or benefit non-recipients do, their support for complexity depends on its costs. Informing people outside of the system about them might therefore be an effective way to boost their support for a simpler system. However, when people have a personal stake in the system, as benefit recipients do, the constraint appears to be different. Recipients update their beliefs about the costs of complexity at least as much as non-recipients, but do not change their support. For them, the constraint may instead be uncertainty about how a reform would affect their own benefits. Simplification may create uncertainty about who would gain and who would lose from the policy change, and such uncertainty can sustain support for the status quo (Fernandez and Rodrik 1991).

Support for complexity, however, depends on the complex system reaching everyone it is meant for. In the spectator experiment, a single exclusion error reduces support for the Complex system, even though its targeting advantage is kept constant. If the public evaluates real benefit systems in a similar way, exclusion in complex systems weakens the fairness argument that justifies complexity in the first place.⁵³ This may be particularly relevant for gaining the support of left-wing voters, who in both experiments are more supportive of complexity, but whose support in the spectator experiment falls to the level of other spectators once the complex system no longer reaches everyone. An interesting avenue for future research is to understand why this single exclusion error matters so much, since my design does not distinguish between a premium on perfect targeting, diminishing sensitivity to targeting errors and a Rawlsian concern for the worst-off.

⁵³The information in the survey experiment also covered non-take-up and did not raise average support for simplification. Respondents' beliefs about non-take-up were, however, already close to the official figure on average (59.8 against 65 out of every 100 people entitled to social assistance), but individual estimates were spread in both directions, so for some respondents this part of the information suggested that the system reaches more people than they thought.

Apart from resolving uncertainty about the distributional effects of simplification reforms, another possible way forward may be to keep accurate targeting but reduce the costs that it imposes on claimants. Complexity is not a problem the public overlooks. Even without any information, 48% of Dutch respondents in the control group think that the many separate schemes make things worse for people entitled to support, against 30% who think they make things better, and 58% of all respondents name complexity as one of the main problems of the welfare state. There may therefore be broad support for measures that keep targeting but reduce these costs. Such measures would shift the costs of complexity from claimants onto the state, for example through more staff and advisers who help claimants navigate the system, outreach to people who are entitled to support but do not claim it, and linking administrative data across agencies. Recent advances in AI may also make it possible for governments to develop cost-effective tools that explain the rules, check entitlements across schemes and help claimants complete their applications.

References

- Akerlof, George A. (1978). "The economics of "tagging" as applied to the optimal income tax, welfare programs, and manpower planning". *The American Economic Review* 68.1, 8–19.
- Alatas, Vivi, Ririn Purnamasari, Matthew Wai-Poi, Abhijit Banerjee, Benjamin A. Olken, and Rema Hanna (2016). "Self-targeting: Evidence from a field experiment in Indonesia". *Journal of Political Economy* 124.2, 371–427.
- Alesina, Alberto and Paola Giuliano (2011). "Preferences for Redistribution". In: *Handbook of Social Economics*. Ed. by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson. Vol. 1. Amsterdam: Elsevier, 93–131.
- Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso (2018). "Intergenerational Mobility and Preferences for Redistribution". *American Economic Review* 108.2, 521–554.
- Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden (2020). "Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians?" *Journal of Political Economy* 128.5, 1753–1788.
- Andre, Peter (2025). "Shallow Meritocracy". *The Review of Economic Studies* 92.2, 772–807.
- Asirvatham, Hemanth, Elliott Moksiki, and Andrei Shleifer (2026). "GPT as a Measurement Tool". Working Paper 34834. Cambridge, MA: National Bureau of Economic Research. URL: <https://doi.org/10.3386/w34834>.
- Baekgaard, Martin, Donald P. Moynihan, and Mette Kjærgaard Thomsen (2021). "Why do policy-makers support administrative burdens? The roles of deservingness, political ideology, and personal experience". *Journal of Public Administration Research and Theory* 31.1, 184–200.
- Benitez-Silva, Hugo, Moshe Buchinsky, and John Rust (2004). "How large are the classification errors in the social security disability award process?" NBER Working Paper 10219. National Bureau of Economic Research.
- Bhargava, Saurabh and Dayanand Manoli (2015). "Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment". *American Economic Review* 105.11, 3489–3529.
- Bhattacharya, Puja and Johanna Mollerstrom (2025). "Lucky to work". *The Review of Economics and Statistics*. Advance online publication.
- Burden, Barry C., David T. Canon, Kenneth R. Mayer, and Donald P. Moynihan (2012). "The effect of administrative burden on bureaucratic perception of policies: Evidence from election administration". *Public Administration Review* 72.5, 741–751.

- Cappelen, Alexander W, Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden (2007). "The Pluralism of Fairness Ideals: An Experimental Approach". *American Economic Review* 97.3, 818–827.
- Cappelen, Alexander W., Cornelius Cappelen, and Bertil Tungodden (2023). "Second-best fairness: The trade-off between false positives and false negatives". *American Economic Review* 113.9, 2458–2485.
- Cappelen, Alexander W., James Konow, Erik Ø. Sørensen, and Bertil Tungodden (2013). "Just Luck: An Experimental Study of Risk-Taking and Fairness". *American Economic Review* 103.4, 1398–1413.
- Castell, Laura, Marc Gurgand, Clément Imbert, and Todor Tochev (2025). "Take-up of social benefits: Experimental evidence from France". *American Economic Journal: Economic Policy* 17.4, 1–29.
- Centraal Planbureau (2019). "Gebruik (en niet-gebruik) van toeslagen in Nederland: Empirische analyse van huurtoeslag en kindgebonden budget". CPB Achtergronddocument. Den Haag: Centraal Planbureau.
- Centraal Planbureau and Sociaal en Cultureel Planbureau (2020). "Kansrijk armoedebeleid". Effective anti-poverty policy. Den Haag: Centraal Planbureau and Sociaal en Cultureel Planbureau.
- Christensen, Julian, Lene Aarøe, Martin Baekgaard, Pamela Herd, and Donald P. Moynihan (2020). "Human capital and administrative burden: The role of cognitive resources in citizen-state interactions". *Public Administration Review* 80.1, 127–136.
- Converse, Philip E. (1964). "The nature of belief systems in mass publics". In: *Ideology and Discontent*. Ed. by David E. Apter. Free Press, 206–261.
- Cruces, Guillermo, Ricardo Perez-Truglia, and Martin Tetaz (2013). "Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment". *Journal of Public Economics* 98, 100–112.
- Diamond, Peter and Eytan Sheshinski (1995). "Economic aspects of optimal disability benefits". *Journal of Public Economics* 57.1, 1–23.
- Dijksterhuis, Ap and Rick van Baaren (2020). "Toeslagen en doenvermogen: Alternatieven voor het toeslagstelsel getoetst vanuit burgerperspectief". Nijmegen: Ministerie van Financiën.
- Durante, Ruben, Louis Putterman, and Joël Van der Weele (2014). "Preferences for redistribution and perception of fairness: An experimental study". *Journal of the European Economic Association* 12.4, 1059–1086.
- Fehr, Dietmar, Johanna Mollerstrom, and Ricardo Perez-Truglia (2022). "Your Place in the World: Relative Income and Global Inequality". *American Economic Journal: Economic Policy* 14.4, 232–268.

- Fernandez, Raquel and Dani Rodrik (1991). "Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty". *American Economic Review* 81.5, 1146–1155.
- Finkelstein, Amy and Matthew J. Notowidigdo (2019). "Take-up and targeting: Experimental evidence from SNAP". *The Quarterly Journal of Economics* 134.3, 1505–1556.
- Goderis, Benedikt and Marente Vlekke (2023). "Tax and Benefit Policies to Reduce Poverty in the Netherlands: A Microsimulation Analysis". *IJM* 16.1, 108–133.
- Haaland, Ingar and Christopher Roth (2023). "Beliefs about racial discrimination and support for pro-Black policies". *The Review of Economics and Statistics* 105.1, 40–53.
- Haaland, Ingar, Christopher Roth, and Johannes Wohlfart (1, 2023). "Designing Information Provision Experiments". *Journal of Economic Literature* 61.1, 3–40.
- Halling, Aske and Steven Van de Walle (2026). "Do justifications affect tolerance for administrative burdens? Evidence from a survey experiment among policymakers". *Social Policy & Administration* 60.1, 12–26.
- Homonoff, Tatiana and Jason Somerville (2021). "Program Recertification Costs: Evidence from SNAP". *American Economic Journal: Economic Policy* 13.4, 271–298.
- IBO Vereenvoudiging Sociale Zekerheid (2023). "Moeilijk makkelijker maken". Interdepartementaal beleidsonderzoek. Den Haag: Rijksoverheid.
- Inspectie Belastingen, Toeslagen en Douane (2026). "Het ongemak van complexiteit". Tech. rep. Den Haag: Inspectie Belastingen, Toeslagen en Douane (IBTD). URL: <https://www.inspectiebt.nl/site/binaries/site-content/collections/documents/2026/04/30/het-ongemak-van-complexiteit/het-ongemak-van-complexiteit.pdf>.
- Inspectie SZW (2021). "Niet-gebruik van de algemene bijstand: Een onderzoek naar de omvang, kenmerken, langdurigheid en aanpak". Den Haag: Inspectie SZW.
- Isaac, Mathew S. and Katie Spangenberg (2021). "The perfection premium". *Social Psychological and Personality Science* 12.6, 930–937.
- Jacquot, Guillaume (26, 2026). *Allocation de solidarité unique : le projet de loi va être « retravaillé »*. Public Sénat. URL: <https://www.publicsenat.fr/actualites/societe/allocation-de-solidarite-unique-le-projet-de-loi-va-etre-retravaille>.
- Karadja, Mounir, Johanna Mollerstrom, and David Seim (2017). "Richer (and Holier) Than Thou? The Effect of Relative Income Improvements on Demand for Redistribution". *The Review of Economics and Statistics* 99.2, 201–212.
- Keiser, Lael R. and Susan M. Miller (2020). "Does administrative burden influence public support for government programs? Evidence from a survey experiment". *Public Administration Review* 80.1, 137–150.

- 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.
- Kok, Lucy, Lieke Heil, Joeri Athmer, Remco van Eijkel, Boukje Cuelenaere, and Millie Elsen (2022). “Evaluatie toeslagenstelsel: Evaluatie Awir, uitvoering en dienstverlening Toeslagen”. SEO-rapport 2022-123. Amsterdam: SEO Economisch Onderzoek.
- Konow, James (2000). “Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions”. *American Economic Review* 90.4, 1072–1092.
- Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva (1, 2015). “How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments”. *American Economic Review* 105.4, 1478–1508.
- Ministerie van Financiën (2020). “Eindrapportage Alternatieven voor het toeslagenstelsel: Uitwerking motie-Bruins-Van Weyenberg”. Den Haag: Ministerie van Financiën.
- Mollerstrom, Johanna, Bjørn-Atle Reme, and Erik Ø. Sørensen (2015). “Luck, choice and responsibility—An experimental study of fairness views”. *Journal of Public Economics* 131, 33–40.
- Moynihan, Donald, Pamela Herd, and Hope Harvey (2015). “Administrative burden: Learning, psychological, and compliance costs in citizen-state interactions”. *Journal of Public Administration Research and Theory* 25.1, 43–69.
- Nichols, Albert L. and Richard J. Zeckhauser (1982). “Targeting transfers through restrictions on recipients”. *The American Economic Review* 72.2, 372–377.
- NOS (2025). *Verkiezingsprogramma’s 2025*. URL: <https://nos.nl/collectie/13999-verkiezingsprogramma-s>.
- Nozick, Robert (1974). *Anarchy, State, and Utopia*. New York: Basic Books.
- Piraino, Patrizio and Joanna Ryan (2024). “The price of fairness: Experimental evidence on the limits to demand for redistribution”. *Journal of Economic Behavior & Organization* 227, 106737.
- Stantcheva, Stefanie (2023). “How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible”. *Annual Review of Economics* 15, 205–234.
- Stantcheva, Stefanie (2026). “Coase Lecture—Perceptions, mindsets and beliefs shaping policy views”. *Economica* 93.372.
- Tweede Kamer der Staten-Generaal (4, 2024). “Gewijzigde motie van de leden Bikker en Eerdmans ter vervanging van die gedrukt onder nr. 35”. Kamerstukken II 2023/24, 36 471, nr. 94. Den Haag.

Tweede Kamer der Staten-Generaal (12, 2025). “Motie van het lid Kouwenhoven c.s. over fundamentele beleidskeuzes voor de hervormingsagenda van het belasting- en toeslagenstelsel”. Kamerstukken II 2024/25, 36 708, nr. 23. Den Haag.

Van Dijk, Jasper J. (2026). “Dit is hét moment voor de hervorming van het toeslagenstelsel”. *ESB* 111.4853, 42–44.

Zaller, John R. (1992). *The Nature and Origins of Mass Opinion*. Cambridge University Press.

Appendices

A. Additional Tables and Figures

A.1. Spectator experiment

Table A1. Choice of the Complex system (robustness sample)

	Chose Complex (System A)	
	(1)	(2)
T2 (low cost)	0.179*** (0.035)	0.175*** (0.035)
T3 (imperfect targeting)	-0.187*** (0.045)	-0.190*** (0.045)
Constant	0.715*** (0.029)	0.750*** (0.068)
Controls		✓
Observations	709	709
Adjusted R^2	0.106	0.106

Notes: The table reports OLS regressions where the dependent variable is the choice of the Complex system. The independent variables are the treatment indicators; column (2) adds controls. The sample is the pre-registered robustness sample, which excludes from the main sample spectators in the bottom 5% of completion time within their treatment and spectators with more than one incorrect answer on the eight comprehension questions. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2. Choice of the Complex system (robustness sample, comprehension on items identical across treatments)

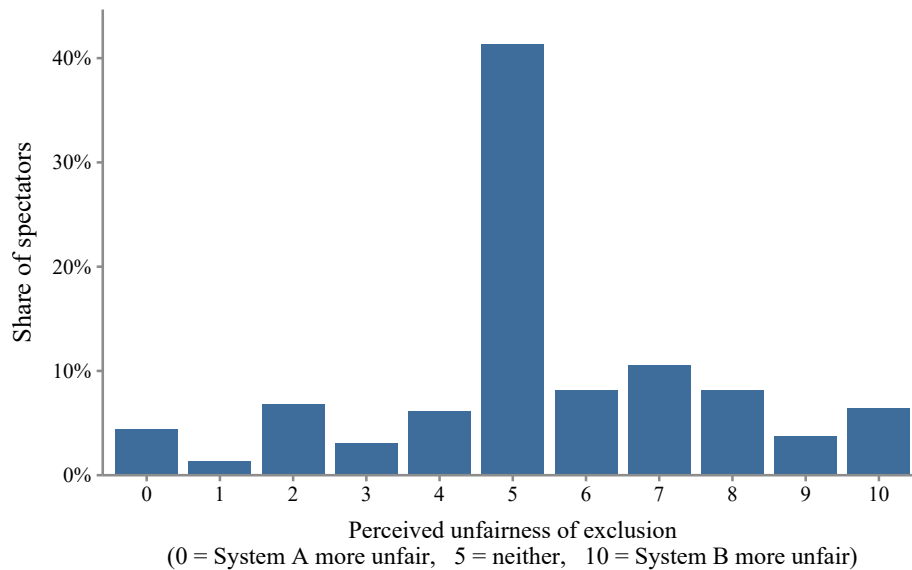
	Chose Complex (System A)	
	(1)	(2)
T2 (low cost)	0.187*** (0.034)	0.185*** (0.034)
T3 (imperfect targeting)	-0.161*** (0.043)	-0.161*** (0.043)
Constant	0.703*** (0.028)	0.749*** (0.067)
Controls		✓
Observations	769	769
Adjusted R^2	0.097	0.096

Notes: The table reports OLS regressions where the dependent variable is the choice of the Complex system. The independent variables are the treatment indicators; column (2) adds controls. The sample excludes from the main sample spectators in the bottom 5% of completion time within their treatment and spectators with more than one incorrect answer on the six comprehension questions that are identical across treatments (items 1, 2, 3, 5, 6 and 7 of the comprehension quiz in Appendix D.1); item 4 has a different correct answer in T3, and item 8 states the length of the form, which differs across treatments. The share of the main sample retained is 82.7% in T1, 85.0% in T2 and 80.7% in T3 ($\chi^2(2) = 2.02$, $p = 0.36$). Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3. Perceived effort and fairness

	Perceived effort		Perceived fairness (B vs. A)		
	(1)	(2)	(3)	(4)	(5)
T2 (low cost)	−2.342*** (0.169)	−2.380*** (0.169)	−0.678*** (0.205)	−0.632*** (0.204)	−0.071 (0.227)
T3 (imperfect targeting)	0.311* (0.176)	0.305* (0.176)	1.155*** (0.220)	1.152*** (0.220)	1.080*** (0.219)
Perceived effort					0.236*** (0.040)
Constant	6.649*** (0.120)	6.174*** (0.320)	2.588*** (0.154)	3.041*** (0.379)	1.586*** (0.445)
Controls		✓		✓	✓
Observations	928	928	928	928	928
Adjusted R ²	0.232	0.235	0.075	0.085	0.119

Notes: The table reports OLS regressions where the dependent variables are perceived effort (columns 1 and 2), how much more effort System A (verification form) requires than System B (automatic) for workers to receive the compensation, from 0 (no extra effort) to 10 (far more effort), and perceived fairness (columns 3 and 4), which system's distribution of the compensation is fairer, from 0 (System A's much fairer) to 10 (System B's much fairer), so higher values mean the Complex system is judged less fair. The independent variables are the treatment indicators; columns (2) and (4) add controls, and column (5) adds controls and perceived effort. Perceived effort is elicited after the choice, so its coefficient is a correlation. The sample is the main sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A1. Perceived unfairness of exclusion in T3

Notes. Spectators in T3 rated whether missing out on the compensation is more unfair under System A or System B, from 0 (System A much more unfair) to 10 (System B much more unfair). Bars show the share of spectators giving each response. $N = 295$.

Table A4. Perceived effort and fairness (robustness sample)

	Perceived effort		Perceived fairness (B vs. A)		
	(1)	(2)	(3)	(4)	(5)
T2 (low cost)	−2.540*** (0.188)	−2.575*** (0.188)	−0.646*** (0.209)	−0.618*** (0.212)	0.002 (0.235)
T3 (imperfect targeting)	0.569*** (0.208)	0.563*** (0.207)	1.412*** (0.243)	1.411*** (0.244)	1.275*** (0.240)
Perceived effort					0.241*** (0.042)
Constant	6.642*** (0.136)	6.267*** (0.356)	2.203*** (0.162)	2.659*** (0.398)	1.149** (0.461)
Controls		✓		✓	✓
Observations	709	709	709	709	709
Adjusted R ²	0.284	0.286	0.105	0.108	0.148

Notes: The table reports OLS regressions where the dependent variables are perceived effort (columns 1 and 2) and perceived fairness (columns 3 and 4). The independent variables are the treatment indicators; columns (2) and (4) add controls, and column (5) adds controls and perceived effort. Perceived effort is elicited after the choice, so its coefficient is a correlation. The sample is the robustness sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5. Choice of the Complex system, controlling for perceived effort

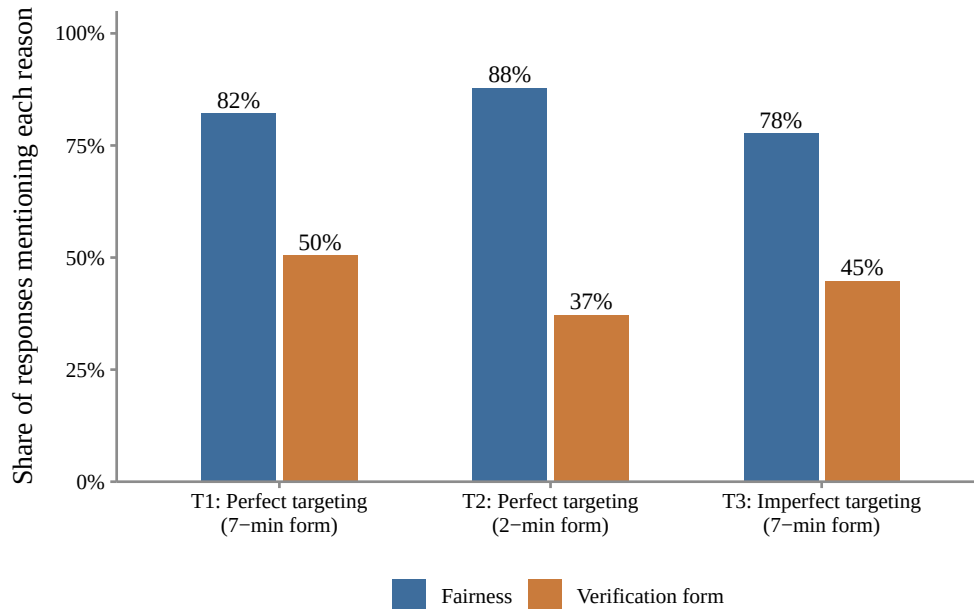
	Chose Complex (System A)			
	Main sample		Robustness sample	
	(1)	(2)	(3)	(4)
T2 (low cost)	0.190*** (0.032)	0.062* (0.035)	0.175*** (0.035)	0.031 (0.036)
T3 (imperfect targeting)	−0.145*** (0.039)	−0.129*** (0.038)	−0.190*** (0.045)	−0.159*** (0.043)
Perceived effort		−0.053*** (0.006)		−0.056*** (0.007)
Controls	✓	✓	✓	✓
Observations	928	928	709	709
Adjusted R ²	0.088	0.150	0.106	0.178

Notes: The table reports OLS regressions where the dependent variable is the choice of the Complex system. Columns (2) and (4) add perceived effort, spectators' rating of how much more effort the Complex system requires of workers than the Simple system, from 0 (no extra effort) to 10 (far more effort). Perceived effort is elicited after the choice, so its coefficient is a correlation. Controls are age, gender, education, income, and whether the respondent has ever received benefits. The reference category is T1. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6. Perceived effort and the choice of the Complex system, within each treatment

	Chose Complex (System A)					
	Main sample			Robustness sample		
	T1 (1)	T2 (2)	T3 (3)	T1 (4)	T2 (5)	T3 (6)
Perceived effort	−0.070*** (0.013)	−0.033*** (0.009)	−0.059*** (0.012)	−0.080*** (0.011)	−0.019** (0.009)	−0.065*** (0.015)
Controls	✓	✓	✓	✓	✓	✓
Observations	313	320	295	246	255	208
Adjusted R^2	0.120	0.035	0.055	0.161	0.017	0.081

Notes: The table reports OLS regressions, estimated separately for each treatment, where the dependent variable is the choice of the Complex system and the independent variable is perceived effort, spectators' rating of how much more effort the Complex system requires of workers than the Simple system, from 0 (no extra effort) to 10 (far more effort). Perceived effort is elicited after the choice, so its coefficient is a correlation. Controls are age, gender, education, income, and whether the respondent has ever received benefits. Columns (1)–(3) use the main sample, columns (4)–(6) the robustness sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A2. Reasons mentioned by spectators, by treatment

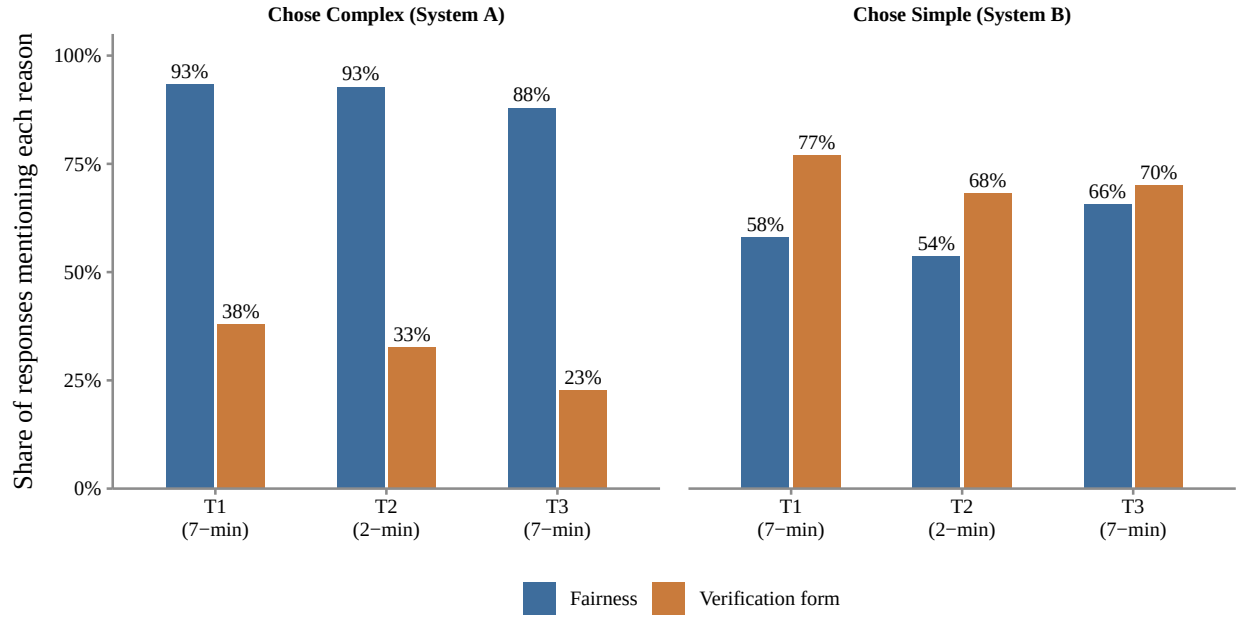
Notes. Share of spectators whose open-ended justification of their choice mentions the fairness of the distribution or the verification form, coded with a large language model. A response can mention both, so the shares within a treatment need not sum to 100%.

Table A7. Choice of the Complex system: heterogeneity (robustness sample)

	Chose Complex (System A)				
	Interaction with moderator			Split sample	
	(1)	(2)	(3)	(4)	(5)
T2 (low cost)	0.271*** (0.062)	0.212*** (0.039)	0.254*** (0.058)	0.124*** (0.042)	0.256*** (0.058)
T3 (imperfect targeting)	−0.179** (0.084)	−0.167*** (0.053)	−0.063 (0.069)	−0.290*** (0.061)	−0.061 (0.069)
Moderator	0.039 (0.066)	0.074 (0.067)	0.166*** (0.060)		
T2 × Moderator	−0.136* (0.075)	−0.159* (0.085)	−0.128* (0.071)		
T3 × Moderator	−0.013 (0.100)	−0.100 (0.101)	−0.227** (0.092)		
Constant	0.716*** (0.078)	0.729*** (0.070)	0.647*** (0.079)	0.777*** (0.088)	0.666*** (0.109)
Moderator	High education	Imperfect comprehension	Left-wing		
Sample	All	All	All	Left-wing	Centre-right
Controls	✓	✓	✓	✓	✓
Observations	709	709	709	375	334
Adjusted R^2	0.108	0.106	0.116	0.142	0.080

Notes: The table reports OLS regressions where the dependent variable is the choice of the Complex system. The independent variables are the treatment indicators, interacted in columns (1)–(3) with a moderator: having a Bachelor’s degree or higher (column 1), at least one incorrect comprehension answer (column 2), and left-wing self-placement (column 3). Columns (4) and (5) estimate the treatment effects separately for left-wing and centre-right spectators. All columns include controls, except education in column (1). The sample is the robustness sample; in column (2), the comparison is therefore between spectators with no and with one incorrect answer. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A3. Reasons mentioned by spectators, by chosen system and treatment



Notes. As in Figure A2, separately for spectators who chose the Complex system (System A, left) and the Simple system (System B, right). These comparisons condition on the choice, which the treatment affects.

Table A8. Perceived fairness, by chosen system

	Perceived fairness (B vs. A)			
	Chose Complex (System A)		Chose Simple (System B)	
	(1)	(2)	(3)	(4)
T2 (low cost)	−0.205 (0.172)	−0.198 (0.172)	0.604 (0.566)	0.801 (0.557)
T3 (imperfect targeting)	0.724*** (0.213)	0.692*** (0.216)	0.710* (0.368)	0.769** (0.378)
Constant	1.624*** (0.131)	1.803*** (0.341)	4.640*** (0.307)	6.121*** (0.751)
Controls		✓		✓
Observations	650	650	278	278
Adjusted R ²	0.032	0.040	0.007	0.024

Notes: The table reports OLS regressions where the dependent variable is perceived fairness. The independent variables are the treatment indicators; columns (2) and (4) add controls. The sample is the main sample, restricted to spectators who chose the Complex system in columns (1) and (2) and to those who chose the Simple system in columns (3) and (4). Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9. Perceived fairness, by chosen system (robustness sample)

	Perceived fairness (B vs. A)			
	Chose Complex (System A)		Chose Simple (System B)	
	(1)	(2)	(3)	(4)
T2 (low cost)	−0.276 (0.175)	−0.278 (0.177)	0.527 (0.694)	0.761 (0.713)
T3 (imperfect targeting)	0.632*** (0.215)	0.603*** (0.219)	1.278*** (0.448)	1.320*** (0.469)
Constant	1.477*** (0.136)	1.715*** (0.342)	4.029*** (0.376)	5.580*** (0.919)
Controls		✓		✓
Observations	514	514	195	195
Adjusted R^2	0.035	0.034	0.034	0.038

Notes: The table reports OLS regressions where the dependent variable is perceived fairness. The independent variables are the treatment indicators; columns (2) and (4) add controls. The sample is the robustness sample, restricted to spectators who chose the Complex system in columns (1) and (2) and to those who chose the Simple system in columns (3) and (4). Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A10. Choice of the Complex system: heterogeneity by political self-placement on the continuous scale

	Chose Complex (System A)	
	(1)	(2)
T2 (low cost)	0.402*** (0.097)	0.398*** (0.110)
T3 (imperfect targeting)	0.171 (0.108)	0.233* (0.128)
Left scale	0.035*** (0.012)	0.043*** (0.014)
T2 $\times$ Left scale	−0.036** (0.015)	−0.037** (0.017)
T3 $\times$ Left scale	−0.055*** (0.017)	−0.072*** (0.020)
Constant	0.496*** (0.100)	0.486*** (0.112)
Sample	Main	Robustness
Controls	✓	✓
Observations	928	709
Adjusted R^2	0.097	0.123

Notes: The table reports OLS regressions where the dependent variable is the choice of the Complex system. The treatment indicators are interacted with political self-placement on the continuous 0–10 scale, reversed so that 0 is right and 10 is left (Left scale). The coefficients on T2 and T3 are therefore the treatment effects for spectators at the right end of the scale. Controls are age, gender, education, income, and whether the respondent has ever received benefits. Column (1) uses the main sample, column (2) the robustness sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A11. Choice and preference for targeting

	Preference for targeting (0–10)		
	(1)	(2)	(3)
Chose Complex (System A)	0.461 ^{**} (0.210)	0.504 ^{**} (0.207)	0.574 ^{***} (0.199)
T2 (low cost)	0.131 (0.221)	0.107 (0.220)	0.015 (0.214)
T3 (imperfect targeting)	0.167 (0.229)	0.210 (0.226)	0.082 (0.220)
Constant	5.657 ^{***} (0.208)	4.874 ^{***} (0.415)	5.680 ^{***} (0.412)
Demographic controls		✓	✓
Attitudinal controls			✓
Observations	928	928	928
Adjusted R^2	0.003	0.043	0.100

Notes: The table reports OLS regressions where the dependent variable is the preference for targeting, from 0 (“benefits should be universal and easier to claim”) to 10 (“benefits should be targeted, even if harder to claim”). The independent variables are an indicator for having chosen the Complex system and the treatment indicators. Demographic controls (columns 2 and 3) are the controls used in the previous tables. Attitudinal controls (column 3) are an indicator for left-wing self-placement and an indicator for generalised trust of 6 or above on a scale from 0 (“you can’t be too careful”) to 10 (“most people can be trusted”). The sample is the main sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12. Choice and support for a negative income tax

	Support for NIT (0–10)		
	(1)	(2)	(3)
Chose Complex (System A)	0.044 (0.217)	0.023 (0.216)	−0.030 (0.212)
T2 (low cost)	−0.224 (0.232)	−0.180 (0.232)	−0.112 (0.230)
T3 (imperfect targeting)	−0.109 (0.234)	−0.121 (0.233)	−0.024 (0.230)
Constant	5.245*** (0.208)	6.376*** (0.439)	5.774*** (0.445)
Demographic controls		✓	✓
Attitudinal controls			✓
Observations	928	928	928
Adjusted R^2	−0.002	0.003	0.032

Notes: The table reports OLS regressions where the dependent variable is support for a negative income tax (NIT), described to respondents as a policy that pays money to people on low incomes based on income data the government already holds, with no application and nothing to report, and which does not adjust for circumstances such as rent or who they live with; it is measured from 0 (“keep the current system as it is”) to 10 (“fully support this change”). The independent variables are the indicator for having chosen the Complex system and the treatment indicators. The sample is the main sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A13. Choice and support for a universal basic income

	Support for UBI (0–10)		
	(1)	(2)	(3)
Chose Complex (System A)	–0.327 (0.233)	–0.330 (0.234)	–0.363 (0.232)
T2 (low cost)	–0.155 (0.239)	–0.123 (0.239)	–0.075 (0.238)
T3 (imperfect targeting)	–0.182 (0.247)	–0.144 (0.246)	–0.085 (0.248)
Constant	4.012*** (0.230)	4.388*** (0.453)	3.976*** (0.463)
Demographic controls		✓	✓
Attitudinal controls			✓
Observations	928	928	928
Adjusted R^2	–0.000	0.008	0.020

Notes: The table reports OLS regressions where the dependent variable is support for a universal basic income (UBI), described to respondents as a policy that pays every adult resident the same monthly amount, enough to cover basic living costs, paid automatically with no application and nothing to report, and which also goes to people who do not need it and is therefore more expensive; it is measured on the same 0–10 scale as support for NIT. The independent variables are the indicator for having chosen the Complex system and the treatment indicators. The sample is the main sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A14. Choice and confidence in government

	Confidence in government (0–10)		
	(1)	(2)	(3)
Chose Complex (System A)	0.110 (0.127)	0.122 (0.127)	0.117 (0.124)
T2 (low cost)	0.009 (0.135)	0.047 (0.135)	0.090 (0.130)
T3 (imperfect targeting)	0.261* (0.137)	0.297** (0.135)	0.303** (0.130)
Constant	4.296*** (0.120)	4.220*** (0.255)	3.874*** (0.244)
Demographic controls		✓	✓
Attitudinal controls			✓
Observations	928	928	928
Adjusted R^2	0.002	0.021	0.107

Notes: The table reports OLS regressions where the dependent variable is confidence in government, the mean of four items each measured on a 0–10 disagree–agree scale: trust in the civil service and government departments; whether government departments are generally able to run public services efficiently; whether the rules and procedures they follow usually serve a real purpose; and whether they waste a lot of money through errors and mismanagement (reverse-coded). The independent variables are the indicator for having chosen the Complex system and the treatment indicators. The sample is the main sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A15. Choice and preference for targeting (robustness sample)

	Preference for targeting (0–10)		
	(1)	(2)	(3)
Chose Complex (System A)	0.551** (0.246)	0.603** (0.244)	0.702*** (0.234)
T2 (low cost)	0.244 (0.248)	0.206 (0.248)	0.087 (0.243)
T3 (imperfect targeting)	0.304 (0.259)	0.335 (0.256)	0.231 (0.249)
Constant	5.577*** (0.244)	4.821*** (0.474)	5.655*** (0.478)
Demographic controls		✓	✓
Attitudinal controls			✓
Observations	709	709	709
Adjusted R^2	0.006	0.037	0.096

Notes: The table reports OLS regressions where the dependent variable is the preference for targeting. The independent variables are the indicator for having chosen the Complex system and the treatment indicators. The sample is the robustness sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A16. Choice and support for a negative income tax (robustness sample)

	Support for NIT (0–10)		
	(1)	(2)	(3)
Chose Complex (System A)	−0.157 (0.251)	−0.165 (0.250)	−0.235 (0.246)
T2 (low cost)	−0.254 (0.259)	−0.209 (0.258)	−0.124 (0.257)
T3 (imperfect targeting)	−0.103 (0.270)	−0.102 (0.267)	−0.033 (0.265)
Constant	5.446*** (0.240)	6.548*** (0.486)	5.956*** (0.496)
Demographic controls		✓	✓
Attitudinal controls			✓
Observations	709	709	709
Adjusted R^2	−0.002	0.003	0.030

Notes: The table reports OLS regressions where the dependent variable is support for NIT. The independent variables are the indicator for having chosen the Complex system and the treatment indicators. The sample is the robustness sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A17. Choice and support for a universal basic income (robustness sample)

	Support for UBI (0–10)		
	(1)	(2)	(3)
Chose Complex (System A)	–0.393 (0.275)	–0.374 (0.274)	–0.440 (0.271)
T2 (low cost)	–0.044 (0.267)	–0.029 (0.265)	0.051 (0.263)
T3 (imperfect targeting)	–0.221 (0.287)	–0.190 (0.282)	–0.123 (0.283)
Constant	3.948*** (0.273)	4.470*** (0.512)	3.913*** (0.517)
Demographic controls		✓	✓
Attitudinal controls			✓
Observations	709	709	709
Adjusted R^2	–0.001	0.015	0.036

Notes: The table reports OLS regressions where the dependent variable is support for UBI. The independent variables are the indicator for having chosen the Complex system and the treatment indicators. The sample is the robustness sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A18. Choice and confidence in government (robustness sample)

	Confidence in government (0–10)		
	(1)	(2)	(3)
Chose Complex (System A)	–0.007 (0.150)	0.034 (0.150)	–0.000 (0.147)
T2 (low cost)	–0.151 (0.152)	–0.113 (0.153)	–0.054 (0.147)
T3 (imperfect targeting)	0.212 (0.156)	0.251 (0.154)	0.232 (0.147)
Constant	4.450*** (0.143)	4.233*** (0.289)	3.884*** (0.278)
Demographic controls		✓	✓
Attitudinal controls			✓
Observations	709	709	709
Adjusted R^2	0.003	0.019	0.107

Notes: The table reports OLS regressions where the dependent variable is confidence in government. The independent variables are the indicator for having chosen the Complex system and the treatment indicators. The sample is the robustness sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.2. Survey experiment

Table A19. Sample composition and the Dutch adult population

	Sample	Dutch population
Age (mean, years)	51.8	49.1
Female	0.506	0.505
<i>Education</i>		
Low	0.174	0.260
Mid	0.430	0.370
High	0.396	0.364
Observations	2,212	

Notes. The sample is the analysis sample. Education follows the CBS classification — Low: primary education, VMBO, MBO-1; Mid: HAVO/VWO diploma, MBO-2/3/4; High: HBO, university (WO) — with shares computed over the 2,204 respondents with a known education level (8 respondents answering “other / don’t know” are excluded). Female is the share reporting “Vrouw”. Dutch-population figures are from CBS: mean age for those aged 16 and over and the female share for those aged 15 and over in 2026 (Stat-Line), and education shares from *The Netherlands in Numbers 2024*.

Table A20. Sample exclusion, by arm

Arm	Completers	In sample	Excluded	Excluded (%)
C0 (topic control)	916	755	161	17.6
C1 (elicitation control)	908	730	178	19.6
T (treatment)	904	727	177	19.6
Total	2,728	2,212	516	18.9

A.2.1. Correlates of support for simplification

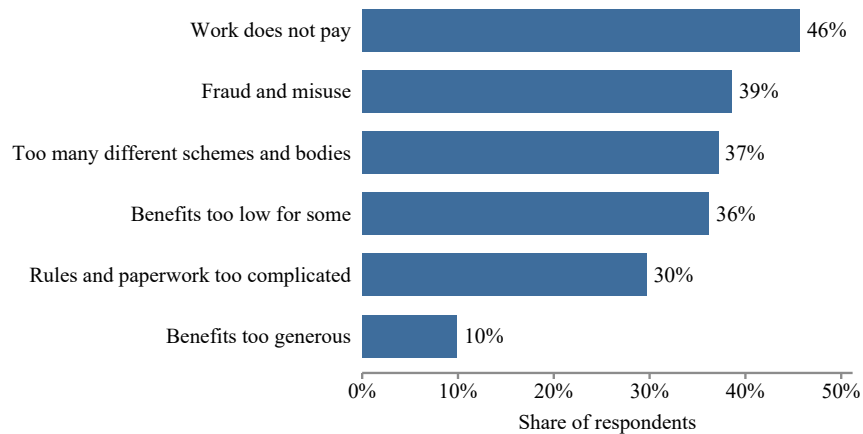
Table A22 reports correlates of support for simplification and of beliefs about whether complexity makes things better or worse for claimants for respondents in the topic control (C0) arm. These beliefs and support for reform do not always go together. Compared with other respondents, the highly educated (respondents with a tertiary degree, hbo or wo) see complexity as worse for claimants and are more supportive of a simpler system; left-wing respondents (those placing themselves at 0–4 on a scale from 0, left, to 10, right)

Table A21. Exclusion, guessing the purpose, and feeling steered, by arm

	Excluded (1)	Guessed purpose (2)	Felt steered (3)
Elicitation (C1)	0.020 (0.018)	-0.019 (0.014)	-0.003 (0.016)
Treatment (T)	0.020 (0.018)	-0.006 (0.015)	0.043** (0.017)
Constant	0.176*** (0.013)	0.093*** (0.011)	0.107*** (0.011)
Observations	2,728	2,212	2,212
Adjusted R^2	-0.000	-0.000	0.003

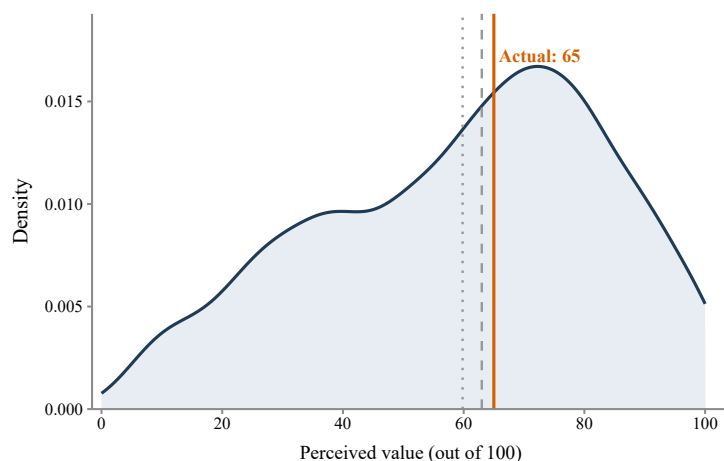
Notes: The table reports OLS regressions (linear probability models) where the dependent variables are indicators for: being excluded from the analysis sample (column 1); an open-ended answer that identifies the purpose of the study as support for simplifying the benefits system (column 2); and reporting that some part of the questionnaire steered the respondent toward an answer (column 3). Open-ended answers are LLM-coded; Appendix C reports the prompts. The independent variables are the treatment and elicitation indicators; C0 is omitted. The sample is all respondents who completed the survey in column (1) and the analysis sample in columns (2)–(3). Robust standard errors are reported in parentheses. $^{\dagger} p < 0.10$, $^{**} p < 0.05$, $^{***} p < 0.01$.

Figure A4. Which problems respondents find most salient at baseline (“select exactly two”).



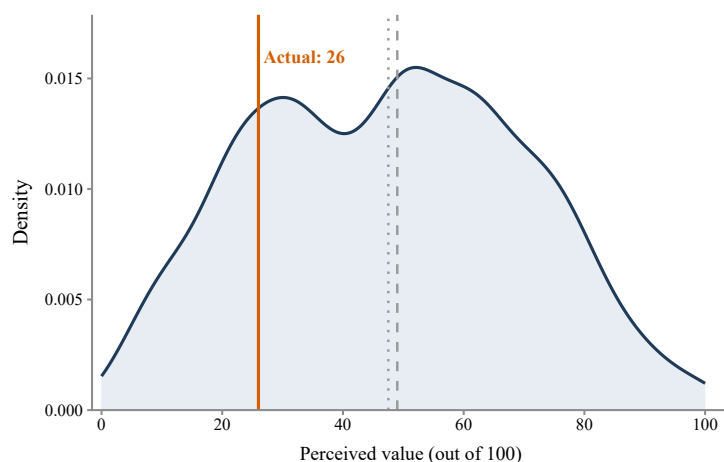
Notes. Share of respondents selecting each item when asked to choose exactly two of the biggest problems of the Dutch welfare state. The option “Other” (selected by 2.7% of respondents) is not shown. $N = 2,212$.

Figure A5. Non-take-up: perceived share of entitled who receive bijstand vs. the true value (65).



Notes. Distribution of respondents' prior beliefs in the arms where priors were elicited (C1 and T). The dotted line marks the mean, the dashed line the median, and the orange line the true value.

Figure A6. Understanding own rules: perceived share who answer the assets question correctly vs. the true value (26).

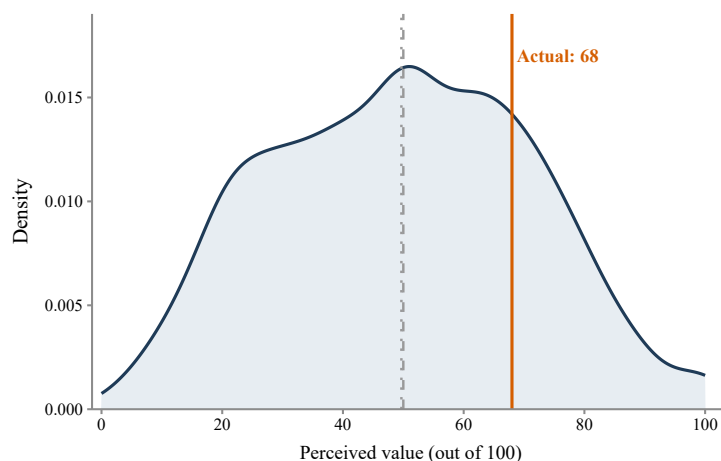


Notes. As in Figure A5.

see it as worse but are less supportive;⁵⁴ and respondents with higher trust in government see it as better but are more supportive, which plausibly reflects confidence that reform can be implemented well. Benefit recipients (respondents who have ever received at least one of the listed benefits or allowances, excluding the state pension) do not differ from non-recipients in whether they think complexity makes things better or worse for

⁵⁴Similarly, in the spectator experiment, left-wing spectators are more likely to choose the Complex system in T1 (Section 4).

Figure A7. Retroactive corrections: perceived share of huurtoeslag households corrected vs. the true value (68).



Notes. As in Figure A5.

claimants⁵⁵ or in their support for a simpler system, but are less supportive of replacing most allowances with a single basic allowance (Policy 2).⁵⁶ Left-wing respondents and recipients are also more likely to name inadequate benefits as a main problem of the welfare state, and recipients are less likely to name complexity (Table A23); left-wing respondents may thus care about the costs of complexity but be unwilling to give up targeted support.

⁵⁵One possible explanation is that recipients, who often receive support from several schemes at once, experience not only the costs of many separate schemes but also their benefits. By contrast, non-recipients know the system mainly from public debate, in which its costs feature prominently.

⁵⁶Among recipients, these views are also unrelated to *experienced complexity*, that is, how complex the schemes they have received are (Table A22, Panel B; see Section 5 for the definition).

Table A22. Correlates of policy support and posterior beliefs in the topic-control arm

	Policy 1: simpler system (1)	Policy 2: single allowance (2)	Policy 3: milder consequences (3)	Complexity helps claimants (4)
<i>Panel A: Topic-control arm (extensive margin)</i>				
Benefit recipient	−0.053 (0.169)	−0.493** (0.200)	0.224 (0.189)	−0.056 (0.146)
Left-wing	−0.446** (0.201)	−0.648*** (0.243)	0.221 (0.220)	−0.656*** (0.176)
Higher income	0.038 (0.174)	0.160 (0.211)	−0.085 (0.195)	0.064 (0.147)
High education	0.563*** (0.173)	−0.004 (0.217)	0.103 (0.204)	−0.518*** (0.153)
Trust in government	0.112*** (0.039)	0.123*** (0.045)	0.157*** (0.046)	0.279*** (0.033)
Age	0.000 (0.005)	−0.026*** (0.006)	−0.009 (0.006)	−0.024*** (0.005)
Female	−0.247 (0.167)	−0.363* (0.197)	−0.073 (0.186)	−0.113 (0.148)
Complexity salient	0.321* (0.168)	0.240 (0.201)	0.330* (0.198)	−0.171 (0.148)
Constant	5.256*** (0.399)	6.440*** (0.455)	5.332*** (0.441)	5.059*** (0.349)
Observations	755	755	755	755
Adjusted R ²	0.044	0.056	0.031	0.149
<i>Panel B: Benefit recipients in the topic-control arm (intensity margin)</i>				
Experienced complexity	0.004 (0.017)	−0.015 (0.019)	−0.009 (0.018)	0.002 (0.015)
Left-wing	−0.256 (0.242)	−0.765*** (0.295)	0.397 (0.267)	−0.702*** (0.216)
Higher income	0.000 (0.228)	0.120 (0.277)	−0.043 (0.252)	−0.039 (0.195)
High education	0.720*** (0.231)	0.391 (0.283)	0.339 (0.266)	−0.454** (0.202)
Trust in government	0.146*** (0.054)	0.117* (0.064)	0.128** (0.062)	0.313*** (0.045)
Age	−0.003 (0.007)	−0.037*** (0.008)	−0.015* (0.008)	−0.023*** (0.006)
Female	−0.233 (0.225)	−0.434* (0.262)	0.033 (0.241)	−0.258 (0.201)
Complexity salient	0.149 (0.216)	0.197 (0.260)	−0.065 (0.251)	−0.326* (0.196)
Constant	5.152*** (0.572)	6.677*** (0.636)	6.059*** (0.592)	5.004*** (0.538)
Observations	460	460	460	460
Adjusted R ²	0.047	0.079	0.027	0.160

Notes: The table reports OLS regressions where the dependent variables are support for Policies 1–3 (0 = keep the current system, 10 = fully support the change) and Complexity helps claimants (0 = many separate schemes make things much worse, 10 = much better for those entitled to support). The independent variables are the respondent characteristics shown. *Experienced complexity* is the sum, over all schemes the respondent has received, of each scheme’s mean experienced-complexity score from IBO (2023). *Complexity salient* is an indicator for naming complexity (rules and paperwork too complicated, or too many different schemes and bodies) as one of the two biggest problems of the social security system, asked before the treatment. The sample is the topic-control arm (C0) of the analysis sample in Panel A and the benefit recipients in it in Panel B; respondents in this arm neither estimated the costs of complexity nor received information about them, so no treatment-arm indicators are included. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A23. Correlates of problems named at baseline

	Complexity	Fraud	Too generous	Work does not pay	Too low
	(1)	(2)	(3)	(4)	(5)
Benefit recipient	−0.045** (0.021)	−0.020 (0.022)	−0.047*** (0.014)	−0.018 (0.022)	0.122*** (0.020)
Left-wing	0.110*** (0.024)	−0.105*** (0.024)	−0.069*** (0.012)	−0.047* (0.025)	0.069*** (0.024)
Higher income	0.005 (0.022)	0.057*** (0.022)	0.038*** (0.014)	0.073*** (0.022)	−0.155*** (0.021)
High education	0.074*** (0.022)	−0.035 (0.022)	−0.007 (0.014)	−0.025 (0.023)	−0.067*** (0.021)
Trust in government	0.022*** (0.004)	0.000 (0.004)	0.001 (0.003)	−0.024*** (0.004)	−0.004 (0.004)
Age	0.005*** (0.001)	0.003*** (0.001)	−0.002*** (0.000)	−0.005*** (0.001)	−0.003*** (0.001)
Female	0.002 (0.021)	−0.018 (0.021)	−0.029** (0.013)	0.031 (0.021)	0.037* (0.020)
Constant	0.185*** (0.047)	0.273*** (0.048)	0.218*** (0.029)	0.826*** (0.048)	0.537*** (0.047)
Observations	2,212	2,212	2,212	2,212	2,212
Adjusted R^2	0.056	0.021	0.031	0.052	0.076

Notes: The table reports OLS regressions (linear probability models) where the dependent variables are indicators for selecting a given item as one of the two biggest problems of the social security system: complexity (column 1, as in Complexity salient), fraud and misuse (column 2), benefits too generous (column 3), work does not pay (column 4), and benefits too low (column 5). The independent variables are the respondent characteristics shown. The dependent variables are measured before the treatment, so no treatment-arm indicators are included. The sample is the analysis sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A24. Treatment effects on policy support, by prior underestimation

	Policy 1: simpler system	Policy 2: single allowance	Policy 3: milder consequences
Treatment (T)	0.254 (0.198)	0.009 (0.237)	0.310 (0.223)
Underestimator	−0.303* (0.178)	−0.116 (0.205)	0.152 (0.197)
T × Underestimator	−0.168 (0.249)	−0.033 (0.292)	−0.407 (0.270)
Constant	6.227*** (0.285)	5.994*** (0.346)	5.369*** (0.322)
Controls	✓	✓	✓
Observations	1,457	1,457	1,457
Adjusted R^2	0.034	0.045	0.030

Notes: The table reports OLS regressions where the dependent variables are support for Policies 1–3. The independent variables are the treatment indicator, an indicator (*Underestimator*) for underestimating the severity of the problem by more than 5 points on at least two of the three prior-belief items (non-take-up, understanding of the rules, and year-end corrections), and their interaction, α_3 (H4 predicts $\alpha_3 > 0$). Controls are age, female, high education, higher income, left-wing, trust in government and benefit recipient. The sample is restricted to the C1 and T arms of the analysis sample, because prior beliefs were elicited only in these arms; C1 is omitted. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A25. Treatment effects on policy support, by prior underestimation: alternative measures

	Policy 1: simpler system	Policy 2: single allowance	Policy 3: milder consequences
<i>Panel A: Underestimator (all 3 items)</i>			
Treatment (T)	0.118 (0.130)	−0.094 (0.149)	0.073 (0.136)
Underestimator (all 3 items)	−0.264 (0.244)	−0.720*** (0.274)	0.061 (0.254)
T × Underestimator (all 3 items)	0.040 (0.342)	0.375 (0.380)	−0.201 (0.359)
Constant	6.118*** (0.279)	5.961*** (0.329)	5.459*** (0.303)
Adjusted R^2	0.029	0.051	0.029
<i>Panel B: Mean gap</i>			
Treatment (T)	0.261 (0.164)	0.058 (0.193)	0.213 (0.178)
Mean gap	−0.007 (0.007)	−0.011 (0.008)	0.005 (0.008)
T × Mean gap	−0.011 (0.010)	−0.006 (0.012)	−0.015 (0.011)
Constant	6.083*** (0.281)	5.938*** (0.333)	5.402*** (0.307)
Adjusted R^2	0.032	0.049	0.030
<i>Panel C: Under-size</i>			
Treatment (T)	0.165 (0.263)	0.082 (0.306)	0.409 (0.278)
Under-size	−0.006* (0.003)	−0.006 (0.004)	0.001 (0.004)
T × Under-size	−0.001 (0.005)	−0.002 (0.005)	−0.007 (0.005)
Constant	6.292*** (0.300)	6.104*** (0.360)	5.389*** (0.331)
Adjusted R^2	0.032	0.049	0.031
Controls	✓	✓	✓
Observations	1,457	1,457	1,457

Notes: The table reports OLS regressions where the dependent variables are support for Policies 1–3. The independent variables are the treatment indicator, a measure of prior underestimation, and their interaction, α_3 . The measure is: in Panel A, an indicator for underestimating the severity of the problem by more than 5 points on all three items; in Panel B, the mean signed gap between the prior and the true value across the three items, oriented so that positive values indicate underestimation; in Panel C, the sum of the positive (underestimating) gaps. Controls are age, female, high education, higher income, left-wing, trust in government and benefit recipient. The sample is the C1 and T arms of the analysis sample; C1 is only in the Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A26. Treatment effects on policy support, by prior underestimation and overestimation

	Policy 1: simpler system	Policy 2: single allowance	Policy 3: milder consequences
<i>Underestimators (N = 961)</i>			
Treatment (T)	0.089 (0.152)	−0.015 (0.170)	−0.100 (0.153)
Constant	5.792*** (0.363)	5.810*** (0.416)	5.465*** (0.378)
Adjusted R^2	0.027	0.052	0.043
<i>Neither (N = 321)</i>			
Treatment (T)	0.349 (0.231)	0.368 (0.287)	0.491* (0.257)
Constant	6.531*** (0.554)	6.001*** (0.727)	5.571*** (0.630)
Adjusted R^2	0.035	0.050	0.012
<i>Overestimators (N = 175)</i>			
Treatment (T)	0.033 (0.370)	−0.470 (0.411)	−0.005 (0.421)
Constant	6.764*** (0.807)	5.870*** (0.901)	5.215*** (0.941)
Adjusted R^2	0.020	0.003	0.012
Controls	✓	✓	✓

Notes: The table reports OLS regressions where the dependent variables are support for Policies 1–3, estimated separately within each group of prior beliefs. The independent variable is the treatment indicator. Controls are age, female, high education, higher income, left-wing, trust in government and benefit recipient. The sample is the C1 and T arms of the analysis sample; C1 is omitted. Overestimators overestimated the severity of the problem by more than 5 points on at least two of the three items, and the remaining respondents (Neither) are neither underestimators nor overestimators: for example, they may underestimate one item and overestimate another. The overestimator group is small, so those estimates are imprecise. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A27. Treatment effects on beliefs, by prior underestimation

Moderator (H):	Complexity helps claimants			Benefit corrections		
	Under- estimator (1)	Mean gap (2)	Gap, corrections (3)	Under- estimator (4)	Mean gap (5)	Gap, corrections (6)
Treatment (T)	−0.474*** (0.178)	−0.406*** (0.141)	−0.513*** (0.146)	3.246* (1.821)	3.921*** (1.468)	0.851 (1.313)
H	0.425*** (0.154)	0.024*** (0.006)	−0.001 (0.004)	−4.395*** (1.663)	−0.287*** (0.062)	−0.599*** (0.036)
T × H	−0.056 (0.222)	−0.009 (0.009)	0.001 (0.005)	4.562** (2.216)	0.206** (0.089)	0.303*** (0.049)
Constant	4.789*** (0.263)	4.902*** (0.260)	4.973*** (0.260)	62.710*** (2.678)	61.675*** (2.555)	62.895*** (2.153)
Controls	✓	✓	✓	✓	✓	✓
Observations	1,457	1,457	1,457	1,457	1,457	1,457
Adjusted R^2	0.173	0.178	0.166	0.077	0.089	0.299

Notes: The table reports OLS regressions where the dependent variables are Complexity helps claimants (columns 1–3; 0 = much worse, 10 = much better for those entitled to support) and Benefit corrections, the estimated number of healthcare-allowance recipients out of 100 who faced a year-end correction (columns 4–6), which the treatment does not mention. The independent variables are the treatment indicator, a moderator (H), and their interaction. The moderators are *Underestimator*, as in Table A24 (columns 1 and 4); the mean gap between the priors and the true values across the three items (columns 2 and 5); and the gap on the rent-allowance corrections item (columns 3 and 6). Gaps are positive when the respondent underestimates the cost of complexity. If respondents who underestimated more update more, $T \times H$ is negative in columns 1–3 and positive in columns 4–6. Controls are age, female, high education, higher income, left-wing, trust in government and benefit recipient. The sample is restricted to the C1 and T arms of the analysis sample, because prior beliefs were elicited only in these arms; C1 is omitted. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

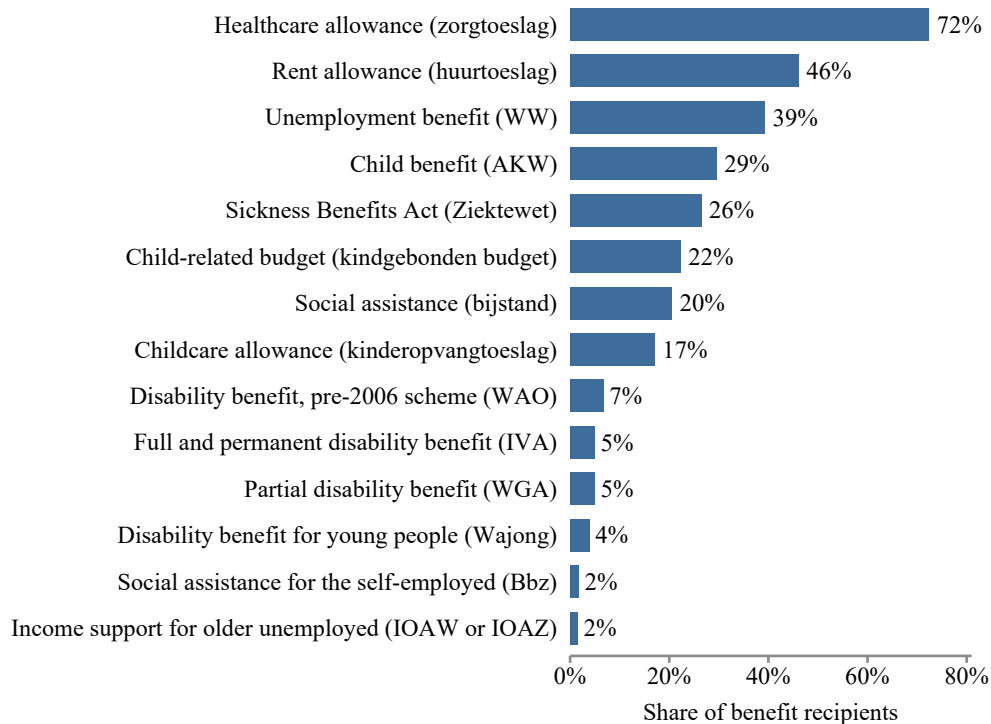
Table A28. Treatment effects on policy support, by respondent characteristic

Moderator (H):	(1) Left-wing	(2) Benefit recipient	(3) Higher income	(4) High education	(5) Fraud salient	(6) Complexity salient	(7) High experienced complexity
<i>Panel A. Policy 1: simpler system</i>							
Treatment (T)	0.103 (0.134)	0.483*** (0.174)	0.278 (0.170)	0.366** (0.153)	0.186 (0.147)	0.202 (0.186)	0.289** (0.135)
Elicitation (C1)	0.032 (0.132)	0.183 (0.180)	-0.054 (0.176)	0.169 (0.156)	-0.010 (0.148)	-0.018 (0.186)	0.141 (0.138)
H	-0.307 (0.198)	-0.084 (0.165)	0.091 (0.166)	0.567*** (0.164)	-0.366** (0.170)	0.364** (0.166)	0.084 (0.193)
T × H	0.252 (0.274)	-0.525** (0.232)	-0.238 (0.233)	-0.514** (0.234)	-0.017 (0.240)	-0.053 (0.238)	-0.376 (0.260)
C1 × H	-0.003 (0.292)	-0.250 (0.237)	0.175 (0.235)	-0.346 (0.238)	0.114 (0.243)	0.086 (0.239)	-0.326 (0.260)
Constant	5.836*** (0.237)	5.678*** (0.252)	5.819*** (0.250)	5.712*** (0.245)	5.920*** (0.244)	5.761*** (0.253)	5.767*** (0.243)
Adjusted R ²	0.030	0.032	0.031	0.032	0.035	0.036	0.031
<i>Panel B. Policy 2: single allowance</i>							
Treatment (T)	0.080 (0.155)	0.182 (0.196)	0.096 (0.204)	0.143 (0.178)	0.013 (0.177)	0.059 (0.214)	0.152 (0.160)
Elicitation (C1)	0.049 (0.155)	0.157 (0.203)	0.077 (0.202)	0.062 (0.181)	0.119 (0.172)	0.131 (0.217)	0.120 (0.162)
H	-0.639*** (0.237)	-0.503** (0.196)	0.163 (0.199)	0.001 (0.203)	-0.649*** (0.201)	0.175 (0.199)	-0.237 (0.228)
T × H	0.299 (0.331)	-0.052 (0.271)	0.105 (0.274)	0.015 (0.278)	0.400 (0.281)	0.172 (0.278)	0.001 (0.304)
C1 × H	0.489 (0.329)	0.012 (0.274)	0.174 (0.274)	0.259 (0.277)	0.131 (0.283)	0.059 (0.280)	0.142 (0.301)
Constant	6.097*** (0.277)	6.028*** (0.295)	6.078*** (0.288)	6.072*** (0.283)	6.214*** (0.282)	6.025*** (0.296)	6.074*** (0.283)
Adjusted R ²	0.051	0.050	0.050	0.050	0.057	0.051	0.050
<i>Panel C. Policy 3: milder consequences</i>							
Treatment (T)	0.178 (0.146)	0.181 (0.197)	-0.079 (0.190)	0.110 (0.164)	0.380** (0.152)	-0.116 (0.204)	0.146 (0.151)
Elicitation (C1)	0.168 (0.149)	0.189 (0.205)	0.021 (0.190)	0.177 (0.169)	0.186 (0.157)	0.136 (0.211)	0.094 (0.154)
H	0.282 (0.215)	0.266 (0.187)	-0.080 (0.188)	0.113 (0.191)	-0.531*** (0.196)	0.337* (0.192)	-0.040 (0.216)
T × H	-0.051 (0.294)	-0.023 (0.257)	0.492* (0.253)	0.143 (0.257)	-0.462* (0.264)	0.520** (0.258)	0.063 (0.275)
C1 × H	-0.178 (0.302)	-0.104 (0.265)	0.201 (0.260)	-0.130 (0.263)	-0.152 (0.273)	-0.014 (0.266)	0.096 (0.282)
Constant	5.347*** (0.263)	5.338*** (0.280)	5.470*** (0.272)	5.367*** (0.266)	5.487*** (0.265)	5.348*** (0.279)	5.383*** (0.268)
Adjusted R ²	0.030	0.030	0.032	0.030	0.051	0.042	0.030
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	2,212	2,212	2,212	2,212	2,212	2,212	2,212

Notes: The table reports OLS regressions where the dependent variables are support for Policies 1–3. The independent variables are the treatment and elicitation indicators, a moderator (H), and the interactions of both indicators with H; C0 is omitted. *Fraud salient* is an indicator for naming fraud and misuse as one of the two biggest problems of the social security system, asked before the treatment. *High experienced complexity* is an indicator for the top tercile of the sum, over all schemes the respondent has received, of each scheme's mean experienced-complexity score from IBO (2023); the score is zero for respondents who have never received benefits, so the tercile is computed on the full analysis sample. Controls are age, female, high education, higher income, left-wing, trust in government and benefit recipient, excluding the moderator of the column. The sample is the analysis sample. Robust standard errors are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A8. Benefit schemes ever received by benefit recipients.



Notes. Share of benefit recipients (respondents who have ever received at least one of the 14 listed schemes, AOW excluded) who report having ever received each scheme. Respondents could select several schemes, so the shares sum to more than 100%. Means-tested schemes are the four allowances (zorgtoeslag, huurtoeslag, kinderopvangtoeslag, kindgebonden budget) and the social assistance schemes (bijstand, IOAW or IOAZ, Bbz). Of all recipients, 85.1% have ever received a means-tested scheme, and 64.4% one other than zorgtoeslag. $N = 1,347$.

Table A29. Treatment effect on the quantitative posterior belief, by benefit receipt

	Benefit corrections (0–100)	
	Non-recipients (1)	Recipients (2)
Treatment (T)	6.727*** (1.656)	5.559*** (1.352)
Constant	59.362*** (3.725)	61.269*** (3.280)
Controls	✓	✓
Adjusted R^2	0.077	0.072
Observations	570	887

Notes: The table reports OLS regressions where the dependent variable is Benefit corrections. Zorgtoeslag corrections were not part of the information treatment, so this measure tests whether the information generalises. The independent variable is the treatment indicator. Controls are age, female, high education, higher income, left-wing and trust in government. The sample is the C1 and T arms of the analysis sample, split by benefit receipt; C1 is omitted. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A30. Treatment effects on beliefs: difference between recipients and non-recipients

	Complexity helps claimants (1)	Benefit corrections (0–100) (2)
Treatment (T)	-0.386** (0.162)	7.082*** (1.645)
Elicitation (C1)	-0.076 (0.160)	—
Benefit recipient	-0.093 (0.144)	0.761 (1.606)
T × Benefit recipient	-0.328 (0.212)	-1.465 (2.134)
C1 × Benefit recipient	-0.015 (0.209)	—
Constant	4.984*** (0.222)	60.117*** (2.633)
Controls	✓	✓
Omitted category	C0	C1
Adjusted R^2	0.165	0.073
Observations	2,212	1,457

Notes: The table reports OLS regressions where the dependent variables are Complexity helps claimants (column 1; 0 = many separate schemes make things much worse, 10 = much better for those entitled to support) and the estimated number of every 100 healthcare allowance recipients who had to repay or received extra money at the end of the year (column 2). The independent variables are the treatment and elicitation indicators, an indicator for benefit recipients, and their interactions. The coefficient on T is the effect for non-recipients, and T × Benefit recipient is the difference between the effects for recipients and non-recipients. Benefit corrections were elicited only in C1 and T, so column (2) uses these two arms with C1 omitted. Controls are age, female, high education, higher income, an indicator for left-wing self-placement (0–4 on a scale from 0, left, to 10, right), and trust in the government and public institutions (0–10). The sample is the analysis sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A31. Treatment effects on support for other policy proposals, by benefit receipt

	Raise the social minimum		Abolish the deductible	
	Non-recipients (1)	Recipients (2)	Non-recipients (3)	Recipients (4)
Treatment (T)	-0.177 (0.201)	-0.043 (0.148)	0.216 (0.255)	-0.110 (0.199)
Elicitation (C1)	0.084 (0.209)	-0.197 (0.154)	-0.001 (0.264)	-0.103 (0.203)
Constant	5.693*** (0.404)	7.871*** (0.304)	7.723*** (0.476)	7.338*** (0.386)
<i>p</i> -value: T = C1	0.221	0.319	0.410	0.973
Controls	✓	✓	✓	✓
Adjusted <i>R</i> ²	0.038	0.075	0.045	0.053
Observations	865	1,347	865	1,347

Notes: The table reports OLS regressions where the dependent variables are support for raising the social minimum (columns 1 and 2) and for abolishing the health insurance deductible (*eigen risico*; columns 3 and 4), each from 0 (completely against) to 10 (completely in favour). Both questions were asked as filler items just before the three simplification proposals. The independent variables are the treatment and elicitation indicators; C0 is omitted. Controls are age, female, high education, higher income, an indicator for left-wing self-placement (0–4 on a scale from 0, left, to 10, right), and trust in the government and public institutions (0–10). The sample is the analysis sample, split into benefit recipients, who have ever received at least one of 14 listed benefit schemes (AOW excluded), and non-recipients. The *p*-value is from a Wald test that the coefficients on T and C1 are equal. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

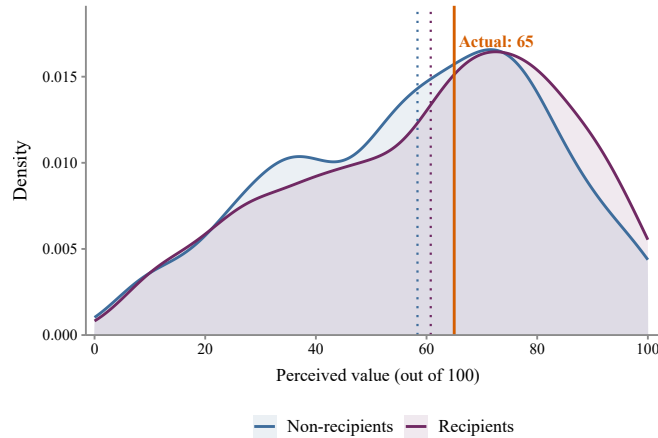
Table A32. Correlates of prior underestimation

	Underestimator (all 3 items) (1)	Absolute gap (accuracy) (2)	Mean gap (signed) (3)
Complexity salient	−0.033 (0.020)	−1.278** (0.513)	−0.782 (0.689)
Left-wing	0.034 (0.024)	0.546 (0.631)	0.606 (0.793)
Higher income	0.010 (0.021)	0.348 (0.535)	−0.236 (0.703)
High education	0.026 (0.022)	0.038 (0.553)	0.644 (0.724)
Benefit recipient	0.020 (0.020)	−0.288 (0.513)	0.528 (0.682)
Trust in government	0.012*** (0.004)	0.054 (0.104)	0.907*** (0.136)
Age	0.002*** (0.001)	0.077*** (0.016)	0.105*** (0.020)
Female	0.028 (0.021)	−0.529 (0.492)	0.369 (0.656)
Constant	−0.028 (0.045)	19.318*** (1.165)	1.572 (1.515)
Observations	1,457	1,457	1,457
Adjusted R^2	0.013	0.016	0.044

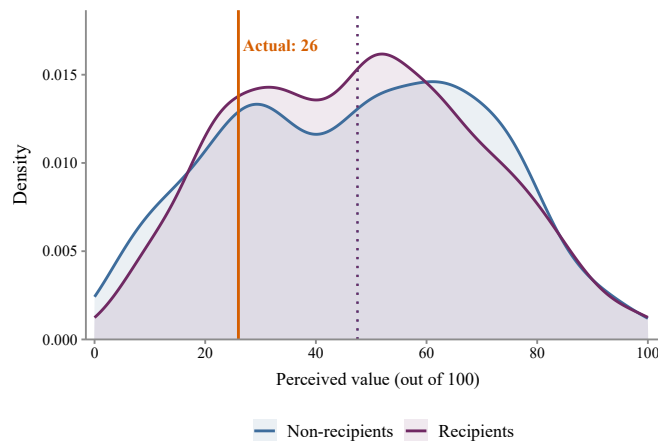
Notes: The table reports OLS regressions where the dependent variables are Underestimator (all 3 items) (column 1, a linear probability model), the mean absolute gap between the priors and the true values across the three items (column 2), and the mean gap (column 3). The independent variables are the respondent characteristics shown. The sample is the C1 and T arms of the analysis sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A9. Prior beliefs, by benefit receipt.

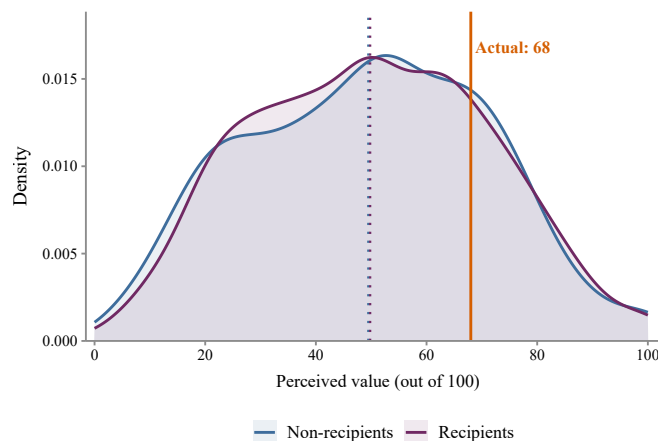
(a) Non-take-up: perceived share of entitled who receive bijstand vs. the true value (65).



(b) Understanding own rules: perceived share who answer the assets question correctly vs. the true value (26).



(c) Retroactive corrections: perceived share of huurtoeslag households corrected vs. the true value (68).



Notes. Distribution of respondents' prior beliefs in the arms where priors were elicited (C1 and T), separately for benefit recipients ($N = 887$) and non-recipients ($N = 570$). The dotted lines mark the group means, and the orange line the true value. The share of respondents within 5 points of the true value is 19.1% among recipients and 19.1% among non-recipients for non-take-up, 15.7% and 15.4% for understanding, and 15.1% and 17.0% for corrections.

Table A33. Correlates of prior beliefs, by item

	Non-take-up		Understanding		Corrections	
	Absolute gap (accuracy) (1)	Gap (signed) (2)	Absolute gap (accuracy) (3)	Gap (signed) (4)	Absolute gap (accuracy) (5)	Gap (signed) (6)
Complexity salient	−2.327*** (0.789)	−1.108 (1.326)	−0.697 (0.983)	−0.445 (1.203)	−0.810 (0.868)	−0.793 (1.143)
Left-wing	−0.200 (0.920)	−0.241 (1.477)	0.084 (1.179)	0.275 (1.419)	1.755* (1.059)	1.784 (1.357)
Higher income	0.656 (0.813)	1.235 (1.357)	−0.259 (1.028)	−1.518 (1.277)	0.647 (0.909)	−0.427 (1.191)
High education	0.553 (0.840)	1.478 (1.401)	−2.902*** (1.024)	−3.915*** (1.283)	2.462*** (0.928)	4.369*** (1.179)
Benefit recipient	0.484 (0.780)	2.601** (1.297)	−1.318 (1.000)	−0.775 (1.239)	−0.031 (0.895)	−0.241 (1.178)
Trust in government	−0.517*** (0.160)	1.311*** (0.264)	0.870*** (0.203)	1.460*** (0.252)	−0.191 (0.176)	−0.049 (0.236)
Age	0.064*** (0.024)	0.064 (0.039)	−0.054* (0.030)	−0.080** (0.036)	0.223*** (0.026)	0.333*** (0.034)
Female	−0.917 (0.751)	1.413 (1.278)	0.838 (0.951)	2.160* (1.180)	−1.509* (0.849)	−2.465** (1.115)
Constant	19.967*** (1.786)	−17.287*** (2.866)	26.124*** (2.224)	20.830*** (2.740)	11.864*** (1.935)	1.172 (2.599)
Observations	1,457	1,457	1,457	1,457	1,457	1,457
Adjusted R^2	0.016	0.019	0.016	0.030	0.054	0.071

Notes: The table reports OLS regressions where the dependent variables are the absolute gap between the prior and the true value (odd columns) and the signed gap (even columns), separately for non-take-up (columns 1–2, true value 65), understanding (columns 3–4, true value 26) and corrections (columns 5–6, true value 68). The gap is positive when the respondent underestimates the cost of complexity, as for the mean gap in Table A32. The independent variables are the respondent characteristics shown. The sample is the C1 and T arms of the analysis sample. Robust standard errors are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A34. Treatment effects on support for a simpler system, by benefit receipt: excluding possible demand effects

	Full sample		Excl. guessed purpose		Excl. felt steered	
	Non- recipients (1)	Recipients (2)	Non- recipients (3)	Recipients (4)	Non- recipients (5)	Recipients (6)
Treatment (T)	0.480*** (0.174)	-0.039 (0.155)	0.515*** (0.183)	-0.005 (0.161)	0.409** (0.189)	0.027 (0.164)
Elicitation (C1)	0.204 (0.181)	-0.062 (0.155)	0.243 (0.188)	-0.009 (0.163)	0.174 (0.191)	-0.064 (0.165)
Constant	5.530*** (0.332)	5.712*** (0.317)	5.576*** (0.352)	5.650*** (0.330)	5.271*** (0.354)	5.677*** (0.335)
Controls	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.021	0.029	0.022	0.029	0.024	0.034
Observations	865	1,347	794	1,231	756	1,190

Notes: The table reports OLS regressions where the dependent variable is support for Policy 1. The independent variables are the treatment and elicitation indicators; C0 is omitted. Controls are age, female, high education, higher income, left-wing and trust in government. The sample is the analysis sample, split by benefit receipt; columns (3) and (4) exclude respondents who guessed the purpose of the study (strict coding), and columns (5) and (6) those who felt steered. Both variables are measured after the treatment, so columns (3)–(6) are a robustness check only. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A35. Treatment effects on support for a simpler system, by awareness of the study purpose and benefit receipt

	Non-recipients		Recipients	
	(1)	(2)	(3)	(4)
Treatment (T)	0.511*** (0.183)	0.416** (0.189)	-0.007 (0.162)	0.020 (0.164)
Elicitation (C1)	0.250 (0.188)	0.171 (0.192)	-0.007 (0.163)	-0.068 (0.165)
Guessed purpose	1.055** (0.435)		0.099 (0.332)	
Treatment × Guessed purpose	-0.609 (0.574)		-0.394 (0.580)	
Elicitation × Guessed purpose	-0.576 (0.684)		-0.663 (0.534)	
Felt steered		0.039 (0.401)		0.032 (0.377)
Treatment × Felt steered		0.409 (0.500)		-0.412 (0.511)
Elicitation × Felt steered		0.287 (0.577)		0.069 (0.490)
Constant	5.458*** (0.333)	5.500*** (0.333)	5.713*** (0.317)	5.723*** (0.319)
Controls	✓	✓	✓	✓
Adjusted R^2	0.026	0.020	0.028	0.027
Observations	865	865	1,347	1,347

Notes: The table reports OLS regressions where the dependent variable is support for Policy 1. The independent variables are the treatment and elicitation indicators, an indicator for respondents who guessed the purpose of the study (strict coding; columns 1 and 3) or felt steered by some part of the questionnaire (columns 2 and 4), and the interactions of both indicators with it; C0 is omitted. The coefficient on T is the treatment effect among respondents who did not guess the purpose (did not feel steered). Controls are age, female, high education, higher income, left-wing and trust in government. The sample is the analysis sample, split by benefit receipt. Both awareness variables are measured after the treatment, so the table is a robustness check only. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A36. Correlates of data quality

	Passed attention check (1)	Included in analysis sample (2)	Main message correct (3)	Figure recalled correctly (4)
Benefit recipient	0.061*** (0.015)	0.067*** (0.015)	0.137*** (0.036)	0.101** (0.039)
Age	0.007*** (0.000)	0.008*** (0.000)	0.000 (0.001)	0.001 (0.001)
Female	0.064*** (0.014)	0.070*** (0.014)	0.066** (0.034)	-0.021 (0.038)
High education	0.021 (0.015)	0.009 (0.015)	0.044 (0.036)	0.120*** (0.040)
Higher income	0.039*** (0.015)	0.048*** (0.015)	0.001 (0.036)	0.046 (0.041)
Left-wing	0.011 (0.016)	0.022 (0.017)	0.101*** (0.036)	0.068 (0.042)
Trust in government	-0.017*** (0.003)	-0.020*** (0.003)	0.011 (0.007)	-0.012 (0.008)
Treatment (T)	-0.015 (0.016)	-0.018 (0.017)		
Elicitation (C1)	-0.007 (0.017)	-0.010 (0.017)		
Constant	0.465*** (0.035)	0.410*** (0.035)	0.538*** (0.081)	0.450*** (0.090)
Adjusted R^2	0.132	0.161	0.041	0.022
Observations	2,728	2,728	727	727

Notes: The table reports OLS regressions (linear probability models) where the dependent variables are indicators for passing the instructed attention check (column 1), for being included in the analysis sample (column 2), for identifying the main message of the information (column 3: “for many people, the system is so complex that they make mistakes or miss out on support”), and for correctly recalling the corrections figure (column 4: more than 50 of every 100 rent-allowance households). The independent variables are the respondent characteristics and, in columns 1–2, the arm indicators shown; C0 is omitted. The sample is all respondents who completed the survey in columns 1–2, and the treatment arm of the analysis sample in columns 3–4, since only treated respondents answered the comprehension questions. Answering “I don’t remember” counts as incorrect. Of treated respondents, 41% found the system worse than they had thought, 30% as they had thought and 29% better; these answers are related to the prior gap on year-end corrections ($r = -0.20$) but hardly to the gaps on take-up and understanding ($r = 0.09$ and 0.05) or to the mean gap ($r = -0.03$). Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A37. Treatment effects on support for a simpler system: all characteristics jointly

	Main effect (1)	× T (2)	× C1 (3)
Benefit recipient	-0.063 (0.169)	-0.611** (0.240)	-0.222 (0.244)
Age	0.002 (0.005)	-0.003 (0.007)	-0.005 (0.007)
Female	-0.245 (0.168)	-0.075 (0.243)	-0.224 (0.243)
High education	0.604*** (0.173)	-0.504** (0.252)	-0.474* (0.257)
Higher income	0.038 (0.175)	-0.201 (0.249)	0.240 (0.253)
Left-wing	-0.422** (0.201)	0.442 (0.282)	0.141 (0.297)
Trust in government	0.120*** (0.039)	-0.058 (0.057)	-0.006 (0.056)
Arm (T or C1)		1.184** (0.573)	0.599 (0.540)
Constant	5.309*** (0.400)		
Adjusted R^2		0.033	
Observations		2,212	

Notes: The table reports a single OLS regression where the dependent variable is support for Policy 1. The independent variables are the treatment and elicitation indicators, the respondent characteristics shown, and the interactions of both indicators with all characteristics; C0 is omitted. Column (1) reports the main effects of the characteristics, and columns (2) and (3) their interactions with T and C1. The arm coefficients are the effects for a respondent with all characteristics equal to zero and are not meaningful on their own. The sample is the analysis sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A38. Correlates of arguments in open-ended answers

	Benefit recipient	Age	Female	High education	Higher income	Left- wing	Trust in government	Constant	Treatment-arm controls	Adjusted R^2
<i>Own stake and risks of reform</i>										
Reform would cut support	0.034*** (0.009)	0.000 (0.000)	0.019** (0.009)	0.006 (0.010)	-0.000 (0.011)	0.019* (0.012)	-0.003* (0.002)	0.005 (0.021)	✓	0.009
Loss of targeting	0.030*** (0.011)	0.001* (0.000)	0.037*** (0.011)	-0.006 (0.012)	0.011 (0.012)	0.014 (0.014)	0.006*** (0.002)	-0.026 (0.021)	✓	0.009
Risk of change	0.016 (0.010)	0.001* (0.000)	-0.000 (0.010)	0.006 (0.011)	-0.015 (0.011)	-0.000 (0.012)	-0.002 (0.002)	0.019 (0.023)	✓	0.003
Own situation	0.045*** (0.010)	-0.000 (0.000)	0.022** (0.011)	-0.016 (0.012)	-0.017 (0.012)	0.011 (0.013)	-0.004* (0.002)	0.086*** (0.025)	✓	0.018
<i>Complexity and access</i>										
Complexity and bureaucracy	0.055*** (0.019)	0.003*** (0.001)	-0.040** (0.019)	0.072*** (0.021)	0.046** (0.020)	0.043* (0.023)	-0.001 (0.004)	0.011 (0.044)	✓	0.031
Errors and corrections	0.029*** (0.008)	-0.000 (0.000)	0.019** (0.008)	-0.005 (0.009)	0.011 (0.009)	-0.002 (0.009)	-0.001 (0.002)	0.017 (0.016)	✓	0.007
Access and take-up	0.007 (0.012)	-0.000 (0.000)	0.010 (0.013)	0.017 (0.014)	0.013 (0.013)	0.038** (0.016)	0.008*** (0.002)	0.007 (0.029)	✓	0.016
<i>Generosity and deservingness</i>										
Fraud and misuse	-0.045*** (0.013)	0.001*** (0.000)	0.004 (0.012)	0.002 (0.012)	0.036*** (0.013)	-0.020 (0.013)	-0.010*** (0.002)	0.090*** (0.028)	✓	0.022
Fraud from reform	-0.002 (0.006)	0.000*** (0.000)	0.007 (0.006)	0.016** (0.006)	0.002 (0.006)	-0.009 (0.006)	-0.001 (0.001)	-0.004 (0.013)	✓	0.003
Too generous	-0.025*** (0.010)	0.000 (0.000)	0.013 (0.009)	-0.004 (0.009)	0.022** (0.010)	-0.019** (0.008)	-0.010*** (0.002)	0.084*** (0.020)	✓	0.023
Work should pay	-0.019 (0.012)	-0.000 (0.000)	0.012 (0.011)	-0.002 (0.012)	0.046*** (0.012)	-0.021* (0.011)	-0.007*** (0.002)	0.105*** (0.026)	✓	0.014
Support too low	0.020** (0.009)	-0.000 (0.000)	0.030*** (0.010)	0.003 (0.010)	-0.012 (0.010)	0.011 (0.012)	-0.008*** (0.002)	0.083*** (0.022)	✓	0.017
<i>Other</i>										
Cost and efficiency	0.003 (0.008)	0.001*** (0.000)	-0.028*** (0.008)	0.027*** (0.009)	0.012 (0.008)	-0.006 (0.009)	-0.003* (0.001)	0.009 (0.021)	✓	0.016
Distrust of government	0.012** (0.006)	0.000* (0.000)	-0.011* (0.007)	0.007 (0.007)	-0.004 (0.006)	0.008 (0.008)	-0.008*** (0.002)	0.043*** (0.016)	✓	0.021
Immigration	-0.001 (0.005)	0.000 (0.000)	-0.000 (0.005)	0.000 (0.004)	0.002 (0.005)	-0.011*** (0.002)	-0.005*** (0.001)	0.038*** (0.012)	✓	0.016
No substantive answer	-0.105*** (0.021)	-0.004*** (0.001)	-0.036* (0.020)	-0.064*** (0.021)	-0.082*** (0.021)	-0.014 (0.023)	0.021*** (0.004)	0.572*** (0.048)	✓	0.048

Notes: The table reports OLS regressions (linear probability models), one per row, where the dependent variable equals 1 if the respondent's open-ended answer on the reasons for supporting or opposing simplification makes the given argument (LLM-coded, see Appendix C for the prompts; an answer can make several arguments). The independent variables are the respondent characteristics shown and the treatment-arm controls (coefficients not reported); C0 is omitted. The sample is the analysis sample, excluding 59 very short answers that the LLM did not classify (2,153 observations). Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A39. Correlates of arguments in open-ended answers: respondents with a substantive answer

	Benefit recipient	Age	Female	High education	Higher income	Left- wing	Trust in government	Constant	Treatment-arm controls	Adjusted R ²
<i>Own stake and risks of reform</i>										
Reform would cut support	0.039*** (0.013)	0.000 (0.000)	0.024* (0.013)	0.003 (0.015)	-0.006 (0.015)	0.024 (0.016)	-0.003 (0.002)	0.028 (0.031)	✓	0.007
Loss of targeting	0.029* (0.015)	0.000 (0.000)	0.045*** (0.015)	-0.018 (0.017)	0.006 (0.017)	0.015 (0.019)	0.011*** (0.003)	0.003 (0.032)	✓	0.011
Risk of change	0.010 (0.015)	0.000 (0.000)	-0.005 (0.014)	0.001 (0.015)	-0.030** (0.015)	-0.001 (0.017)	0.000 (0.003)	0.062* (0.034)	✓	0.003
Own situation	0.048*** (0.015)	-0.001 (0.000)	0.028* (0.015)	-0.033** (0.016)	-0.033** (0.017)	0.012 (0.019)	-0.003 (0.003)	0.160*** (0.038)	✓	0.025
<i>Complexity and access</i>										
Complexity and bureaucracy	0.024 (0.026)	0.002*** (0.001)	-0.074*** (0.025)	0.065** (0.027)	0.025 (0.027)	0.047 (0.030)	0.010** (0.005)	0.146** (0.062)	✓	0.029
Errors and corrections	0.035*** (0.011)	-0.001* (0.000)	0.024** (0.011)	-0.011 (0.013)	0.010 (0.013)	-0.005 (0.013)	0.000 (0.002)	0.042* (0.025)	✓	0.008
Access and take-up	-0.014 (0.018)	-0.001** (0.001)	0.004 (0.017)	0.009 (0.019)	0.003 (0.019)	0.053** (0.022)	0.015*** (0.003)	0.078* (0.043)	✓	0.030
<i>Generosity and deservingness</i>										
Fraud and misuse	-0.082*** (0.020)	0.001 (0.001)	0.001 (0.017)	-0.007 (0.017)	0.035* (0.018)	-0.024 (0.018)	-0.011*** (0.003)	0.170*** (0.042)	✓	0.027
Fraud from reform	-0.004 (0.008)	0.000** (0.000)	0.008 (0.008)	0.021** (0.009)	0.001 (0.009)	-0.014* (0.008)	-0.001 (0.002)	0.001 (0.019)	✓	0.003
Too generous	-0.044*** (0.015)	0.000 (0.000)	0.018 (0.012)	-0.010 (0.013)	0.023* (0.013)	-0.022** (0.011)	-0.012*** (0.003)	0.140*** (0.030)	✓	0.030
Work should pay	-0.040** (0.017)	-0.001** (0.000)	0.013 (0.015)	-0.010 (0.016)	0.052*** (0.016)	-0.028* (0.015)	-0.007** (0.003)	0.186*** (0.039)	✓	0.020
Support too low	0.015 (0.014)	-0.001** (0.000)	0.041*** (0.014)	-0.001 (0.014)	-0.026* (0.015)	0.016 (0.016)	-0.010*** (0.003)	0.149*** (0.034)	✓	0.022
<i>Other</i>										
Cost and efficiency	0.000 (0.012)	0.001** (0.000)	-0.041*** (0.011)	0.035*** (0.013)	0.012 (0.012)	-0.011 (0.013)	-0.002 (0.002)	0.025 (0.032)	✓	0.019
Distrust of government	0.013 (0.009)	0.000 (0.000)	-0.015 (0.009)	0.008 (0.010)	-0.009 (0.009)	0.010 (0.011)	-0.011*** (0.002)	0.065*** (0.023)	✓	0.025
Immigration	-0.004 (0.007)	0.000 (0.000)	0.000 (0.006)	0.000 (0.006)	0.001 (0.007)	-0.014*** (0.004)	-0.007*** (0.002)	0.057*** (0.018)	✓	0.019

Notes: The table repeats the regressions in Table A38 on the respondents whose open-ended answer was coded as substantive (1,521 observations), that is, excluding answers coded as *No substantive answer*. Each row is an OLS regression (linear probability model) where the dependent variable equals 1 if the respondent's open-ended answer makes the given argument (LLM-coded, see Appendix C for the prompts). The independent variables are the respondent characteristics shown and the treatment-arm controls (coefficients not reported); C0 is omitted. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A40. Treatment effects on beliefs and policy support, by experienced complexity

	Complexity helps claimants (1)	Policy 1: simpler system (2)
Treatment (T)	-0.709*** (0.135)	-0.038 (0.155)
Elicitation (C1)	-0.093 (0.135)	-0.062 (0.155)
Experienced complexity	-0.000 (0.015)	0.000 (0.016)
T × Experienced complexity	-0.008 (0.020)	-0.016 (0.024)
C1 × Experienced complexity	-0.019 (0.021)	-0.006 (0.024)
Constant	4.915*** (0.283)	5.734*** (0.319)
Controls	✓	✓
Adjusted R ²	0.177	0.027
Observations	1,347	1,347

Notes: The table reports OLS regressions where the dependent variables are Complexity helps claimants (column 1) and support for Policy 1 (column 2). The independent variables are the treatment and elicitation indicators, experienced complexity, and their interactions; C0 is omitted. Experienced complexity is centred at the recipients' mean (10.9), so the coefficients on T and C1 are the effects for a recipient with average experienced complexity. Controls are age, female, high education, higher income, left-wing and trust in government. The sample is the benefit recipients in the analysis sample. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A41. Treatment effects on beliefs and policy support, by education

	Complexity helps claimants		Policy 1: simpler system	
	No higher education (1)	Higher education (2)	No higher education (3)	Higher education (4)
Treatment (T)	-0.677*** (0.133)	-0.428*** (0.166)	0.369** (0.153)	-0.152 (0.177)
Elicitation (C1)	-0.143 (0.133)	-0.011 (0.164)	0.173 (0.156)	-0.169 (0.179)
Constant	4.763*** (0.275)	4.938*** (0.335)	6.330*** (0.305)	5.303*** (0.387)
<i>p</i> -value: T = C1	<0.001	0.014	0.208	0.930
Controls	✓	✓	✓	✓
Adjusted R^2	0.145	0.201	0.037	0.019
Observations	1,339	873	1,339	873

Notes: The table reports OLS regressions where the dependent variables are Complexity helps claimants (columns 1 and 2) and support for Policy 1 (columns 3 and 4). The independent variables are the treatment and elicitation indicators; C0 is omitted. Controls are age, female, higher income, left-wing, trust in government and benefit recipient. The sample is the analysis sample, split by education. The *p*-value is from a Wald test that the coefficients on T and C1 are equal. Robust standard errors are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A42. Treatment effect on the quantitative posterior belief, by education

	Benefit corrections (0–100)	
	No higher education (1)	Higher education (2)
Treatment (T)	4.682*** (1.345)	8.301*** (1.637)
Constant	65.234*** (3.154)	50.621*** (4.366)
Controls	✓	✓
Adjusted R^2	0.069	0.089
Observations	886	571

Notes: The table reports OLS regressions where the dependent variable is Benefit corrections. The independent variable is the treatment indicator. Controls are age, female, higher income, left-wing, trust in government and benefit recipient. The sample is the C1 and T arms of the analysis sample, split by education; C1 is omitted. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A43. Treatment effects on support for a simpler system, by education: excluding possible demand effects

	Full sample		Excl. guessed purpose		Excl. felt steered	
	No higher education	Higher education	No higher education	Higher education	No higher education	Higher education
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment (T)	0.369** (0.153)	-0.152 (0.177)	0.402** (0.160)	-0.123 (0.187)	0.336** (0.163)	-0.064 (0.188)
Elicitation (C1)	0.173 (0.156)	-0.169 (0.179)	0.229 (0.163)	-0.130 (0.185)	0.142 (0.166)	-0.167 (0.188)
Constant	6.330*** (0.305)	5.303*** (0.387)	6.299*** (0.319)	5.253*** (0.409)	6.303*** (0.326)	5.039*** (0.405)
Controls	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.037	0.019	0.032	0.018	0.039	0.023
Observations	1,339	873	1,224	801	1,166	780

Notes: The table reports OLS regressions where the dependent variable is support for Policy 1. The independent variables are the treatment and elicitation indicators; C0 is omitted. Controls are age, female, higher income, left-wing, trust in government and benefit recipient. The sample is the analysis sample, split by education; columns (3) and (4) exclude respondents who guessed the purpose of the study (strict coding), and columns (5) and (6) those who felt steered. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.3. Deviations from the pre-analysis plan

A.3.1. Survey experiment

Deviations from pre-registered analyses.

1. The pre-analysis plan specified the elicitation control (C1) as the reference for H2 and H3; I report effects relative to the topic control (C0), that is, of the whole intervention. I report the test of $T = C1$ in every table, and the conclusions are the same under both contrasts, so this concerns only the presentation.
2. I report how the effect of the full intervention (T vs. C0) varies with respondent characteristics, consistent with how the main results are presented.
3. **Controls.** The pre-analysis plan specified four controls for H2–H4 but a more extensive set for the rest of analysis. To reduce noise and keep the specifications consistent, I use the same full set of controls throughout; with the four pre-registered controls, the main results are unchanged (Table [A44](#)).
4. **Correlates.** The pre-analysis plan controls for the treatment indicators; I estimate the correlates of support on the topic-control arm only, without arm indicators. I use benefit receipt instead of experienced complexity as the respondent characteristic, because the survey results focus on recipients; experienced complexity, which is zero for non-recipients, is analysed among recipients separately.
5. **Demand effects.** Interactions of the treatment with guessing the study purpose and feeling steered were reported within the benefit-receipt and education analyses, rather than for the main analyses.

Analyses not in the pre-analysis plan.

- Belief effects by prior beliefs.
- Heterogeneity by benefit receipt and education: the interactions were pre-registered (as above), but to examine whether the experiment moved each group, I also estimate split-sample regressions.

A.3.2. Spectator experiment

Analyses not in the pre-analysis plan. Heterogeneity by left-wing self-placement, education and comprehension, and part of the mechanisms section: controlling for perceived effort, fairness split by the system chosen, and choice by the rating of exclusion fairness.

Table A44. Treatment effects on beliefs and policy support: pre-registered controls

	Beliefs		Policy support (0–10)		
	Complexity helps claimants (1)	Benefit corrections (0–100) (2)	Policy 1: simpler system (3)	Policy 2: single allowance (4)	Policy 3: milder consequences (5)
Treatment (T)	–0.599*** (0.110)	6.182*** (1.036)	0.160 (0.117)	0.146 (0.139)	0.160 (0.128)
Elicitation (C1)	–0.050 (0.109)	—	0.045 (0.119)	0.183 (0.139)	0.141 (0.131)
Constant	6.208*** (0.194)	60.713*** (2.120)	6.013*** (0.199)	6.207*** (0.234)	6.219*** (0.216)
<i>p</i> -value: T = C1	<0.001	—	0.35	0.79	0.88
Controls	✓	✓	✓	✓	✓
Omitted category	C0	C1	C0	C0	C0
Observations	2,212	1,457	2,212	2,212	2,212
Adjusted R^2	0.057	0.075	0.016	0.029	0.008

Notes: The table reports OLS regressions where the dependent variables are: *Complexity helps claimants*, whether the many separate schemes, each with its own rules, conditions and application procedure, make things better or worse for people entitled to support, from 0 (much worse) through 5 (no difference) to 10 (much better), so a negative coefficient means complexity is seen as worse for claimants; *Benefit corrections*, the estimated number of every 100 zorgtoeslag (healthcare allowance) recipients who had to repay or receive extra at the end of the year; and support for Policy 1 (a simpler system with fewer, more general schemes), Policy 2 (a single basic allowance replacing all allowances except the childcare allowance) and Policy 3 (milder consequences of mistakes), each from 0 (keep the current system as it is) to 10 (fully support this change). The independent variables are indicators for the treatment arm (T; prior beliefs elicited and corrected) and the elicitation control arm (C1; prior beliefs elicited, no correction). Controls are age (in years), an indicator for being female, an indicator for high education (hbo or wo), and an indicator for a net yearly household income of at least EUR 40,000. These are the controls specified in the pre-analysis plan; Table 4 reports the main specification, which adds left-wing self-placement, trust in government and benefit receipt. The sample (analysis sample) consists of respondents who completed the survey, passed the attention check and were not among the fastest 5% within their arm. In columns (1) and (3)–(5), the topic control (C0) is omitted. Benefit corrections were elicited only in C1 and T, so column (2) uses these two arms with C1 omitted. The *p*-value is from a Wald test that the coefficients on T and C1 are equal. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B. Derivations for the Conceptual Framework

B.1. Utilitarian benchmark

This section derives the welfare difference between the two systems, Δ^t , stated in Section 3.1. The notation is as in Section 3: a worker earns y_L if their output was not fully registered and $y_H > y_L$ if it was, and consumption is $c_i = y_i + b_i$.

Group composition. Let $d_i = 1$ if worker i 's output was not fully registered and $d_i = 0$ otherwise, and let $k_i \in \{Y, G\}$ be the worker's colour. Let $n_{k,d}$ be the number of workers of colour k with $d_i = d$, and $n_d \equiv n_{Y,d} + n_{G,d}$. In the experiment,

	$n_{Y,1}$	$n_{Y,0}$	$n_{G,1}$	$n_{G,0}$	n_1	n_0
T1, T2	7	1	1	3	8	4
T3	7	1	2	2	9	3

Welfare under each system. The Simple system pays b to every Yellow worker and nothing to Green workers, and no worker fills in the form:

$$W^S = n_{Y,1}[u(y_L + b) - v(e_T)] + n_{Y,0}[u(y_H + b) - v(e_T)] + n_{G,1}[u(y_L) - v(e_T)] + n_{G,0}[u(y_H) - v(e_T)]. \quad (4)$$

The Complex system pays b only to workers whose output was not fully registered, who bear the additional effort e_V . Of the n_1 such workers, n_V complete the form and $n_X = n_1 - n_V$ cannot; they are not compensated and exert only e_T :

$$W^C = n_V[u(y_L + b) - v(e_T + e_V)] + n_X[u(y_L) - v(e_T)] + n_0[u(y_H) - v(e_T)]. \quad (5)$$

In T1 and T2, $n_X = 0$; in T3, $n_X = 1$. Because both systems compensate eight workers, $n_V = 8$ in every treatment.

T1 and T2. Using $n_V = n_{Y,1} + n_{G,1}$ and $n_0 = n_{Y,0} + n_{G,0}$, the consumption terms of the Yellow workers whose output was not fully registered and all terms of the Green workers whose output was fully registered cancel, and

$$\begin{aligned} \Delta^t &= n_{Y,0}[u(y_H + b) - u(y_H)] - n_{G,1}[u(y_L + b) - u(y_L)] + n_V[v(e_T + e_V^t) - v(e_T)] \\ &= 8B^t - N, \quad t \in \{T1, T2\}, \end{aligned} \quad (6)$$

since $n_{Y,0} = n_{G,1} = 1$. The targeting term is the same in both treatments, so

$$\Delta^{T1} - \Delta^{T2} = 8[v(e_T + e_V^{T1}) - v(e_T + e_V^{T2})] > 0, \quad (7)$$

because v is increasing and $e_V^{T1} > e_V^{T2}$. This gives Prediction 1.

T3. In T3, $n_{G,1} = 2$ and $n_X = 1$, so $n_V = 8$ as in T1:

$$\begin{aligned} \Delta^{T3} &= n_{Y,0}[u(y_H + b) - u(y_H)] - (n_{G,1} - n_X)[u(y_L + b) - u(y_L)] + n_V[v(e_T + e_V^{T1}) - v(e_T)] \\ &= 8B^{T1} - N = \Delta^{T1}. \end{aligned} \quad (8)$$

The excluded worker offsets the second Green worker whose output was not fully registered, which gives Prediction 2.

B.2. Rank-dependent welfare functions

This section derives how support for the Complex system changes from T1 to T3 for a spectator with a rank-dependent welfare function,

$$W = \sum_{r=1}^{12} a_r U_{(r)}, \quad a_1 \geq a_2 \geq \dots \geq a_{12} > 0, \quad (9)$$

where $U_{(r)}$ is the r -th lowest utility in the group of twelve workers. The utilitarian welfare function is the case in which all weights are equal, and the Rawlsian welfare function is the limit in which all weight is on a_1 . Sen's welfare function and the generalised Gini welfare functions lie in between.

Utility levels. Under the two systems, each worker has one of five utility levels:

	Output, compensation	Utility
L	not fully registered, not compensated	$u(y_L) - v(e_T)$
H	fully registered, not compensated	$u(y_H) - v(e_T)$
P	not fully registered, compensated, no form (Simple)	$u(y_L + b) - v(e_T)$
R	fully registered, compensated, no form (Simple)	$u(y_H + b) - v(e_T)$
F	not fully registered, compensated after the form (Complex)	$u(y_L + b) - v(e_T + e_V)$

Let $G \equiv u(y_L + b) - u(y_L)$ be the utility gain from the compensation for a worker whose output was not fully registered, and $B \equiv B^{T1}$ the burden of the form. Suppose the spectator believes, as the instructions state, that the compensation offsets the lost earnings,

$y_L + b = y_H$, and that the compensation outweighs the burden of the form, $B < G$. Then $P = H$, $H - L = G$, $H - F = B$ and $F - L = G - B > 0$, so that

$$L < F < H = P < R. \quad (10)$$

Ranks. Sorting each group from the worst-off to the best-off worker gives:

Ranks, from worst-off to best-off			
Simple, T1	L: 1	H or P: 2–11	R: 12
Simple, T3	L: 1–2	H or P: 3–11	R: 12
Complex, T1	F: 1–8	H: 9–12	
Complex, T3	L: 1	F: 2–9	H: 10–12

Collecting the weights of the ranks that each utility level occupies,

$$\begin{aligned}
W_{T1}^S &= a_1 L + (a_2 + \dots + a_{11}) H + a_{12} R, \\
W_{T3}^S &= (a_1 + a_2) L + (a_3 + \dots + a_{11}) H + a_{12} R, \\
W_{T1}^C &= (a_1 + \dots + a_8) F + (a_9 + \dots + a_{12}) H, \\
W_{T3}^C &= a_1 L + (a_2 + \dots + a_9) F + (a_{10} + \dots + a_{12}) H.
\end{aligned} \quad (11)$$

Change from T1 to T3. Under the Simple system, the additional uncompensated worker becomes the second worst-off, and all other workers keep their rank and utility:

$$W_{T3}^S - W_{T1}^S = -a_2(H - L) = -a_2 G. \quad (12)$$

Under the Complex system, the excluded worker becomes the worst-off, and each of the eight workers who complete the form moves up one rank, so that one of them replaces a worker whose output was fully registered at rank 9:

$$W_{T3}^C - W_{T1}^C = a_1 L + (a_9 - a_1) F - a_9 H = -a_1[G - B] - a_9 B. \quad (13)$$

Writing $G = [G - B] + B$ in the first expression and taking the difference,

$$\Delta^{T3} - \Delta^{T1} = (a_1 - a_2)[G - B] - (a_2 - a_9) B. \quad (14)$$

The first term is non-negative and works against the Complex system: the excluded worker is the worst-off under the Complex system but only the second worst-off under the Simple system. The second term is non-positive and works in its favour: the burden of the form

now falls on a worker at rank 9 rather than on one at rank 2. Support for the Complex system falls if the first term is larger, that is, if the weights decline steeply at the bottom of the distribution relative to the burden of the form, and rises if the second term is larger.

The utilitarian case, with all weights equal, gives $\Delta^{T3} - \Delta^{T1} = 0$, as in Prediction 2. The Rawlsian case, with all weight on a_1 , gives $a_1[G - B] > 0$. Sen’s welfare function, average welfare multiplied by one minus the Gini coefficient, gives the weights $a_r = 2(12 - r) + 1$, that is, 23, 21, 19, $\dots$, 3, 1 (up to a positive factor, which does not affect the choice). Consecutive weights differ by 2, so $a_1 - a_2 = 2$, $a_2 - a_9 = 14$, and

$$\Delta^{T3} - \Delta^{T1} = 2G - 16B, \quad (15)$$

which is positive if and only if $G > 8B$.

If the burden outweighs the compensation. If $B > G$, then $F < L$: the workers who complete the form are the eight worst-off under the Complex system in both treatments, and the excluded worker enters at rank 9. Then $W_{T3}^C - W_{T1}^C = -a_9 G$, while the change under the Simple system is unchanged, so

$$\Delta^{T3} - \Delta^{T1} = -(a_2 - a_9)G \leq 0, \quad (16)$$

and support for the Complex system weakly rises.

C. LLM Coding of Open-Ended Answers

This appendix reports the prompts used to code the open-ended answers in both experiments with a large language model, using the GABRIEL package (Asirvatham, Mokski, and Shleifer 2026).

C.1. Spectator experiment: reasons for the choice

Spectators’ answers to the question “Can you briefly explain the reasons behind your choice?” were classified with GABRIEL’s `classify` function, using the model `gpt-5.6-luna`. Each answer was classified in three independent runs, and a label was assigned if it was returned in at least half of the runs. An answer can receive both labels or none.

Instructions.

You will be given open-ended survey responses. They come from an experiment in which participants acted as impartial spectators and decided on a compensation system for a group of 12 workers. After their decision, we asked

spectators: “Can you briefly explain the reasons behind your choice?”. Your task is to classify these responses into two categories: Fairness and Verification form.

In the experiment, spectator’s task was to choose a compensation system for workers who had worked on a series of tasks. Spectators made these choices individually. For some workers, part of their output was not registered. Each worker was randomly assigned a colour (Yellow or Green) that was correlated with, but did not perfectly indicate, whether part of their output was unregistered. It was much more likely that a worker’s output was not fully registered if they were assigned to Yellow. Spectators could choose between two systems:

System A pays compensation to workers whose output was not fully registered, but they are required to fill in a verification form that takes time. It is possible that the form is unavailable to some workers whose output was not fully registered, in which case they do not fill it in and do not receive the compensation.

System B pays compensation automatically by colour group (paying Yellow workers), with no form, but since colour is an imperfect proxy, it ends up paying Yellow workers whose output was fully registered and does not pay Green workers whose output was not fully registered.

Therefore, the responses you will code will justify whether the respondent chose System A or System B.

Classify the response to all categories that apply, respondents often talk about several components of their choice. If a response fits neither category, return no label.

Labels. **Fairness.**

This label should be applied when the response mentions fairness of the distribution of the compensation, who deserves the compensation, and what is a fair or correct way to distribute it. For example, it should be applied when the response discusses whether different types of workers (i.e., those whose output was fully registered and those whose output was not fully registered) were paid compensation or not. If the response mentions fairness with relation to the verification form and time or effort required to fill it in, i.e., that it is unfair for the workers to be required to fill in the form, do not classify here; this is an argument related to the verification form. If the response mentions

fairness of the distribution only to set it aside as a determinant of the choice (e.g. “Although the other system distributes the compensation more fairly, I chose this one so the workers wouldn’t have to complete the form.”), still assign this label.

Verification form.

This label should be applied when the response mentions the verification form the workers are required to fill in under one of the systems. For example, it should be applied when the response discusses the time or effort it takes to fill in the form. It should also be applied when the response is mentioning that in one of the systems the workers are not required to fill in the form, or discusses this system’s simplicity or automatic nature. If the response mentions fairness with relation to who gets the compensation, do not classify here; this is a fairness argument. If the response mentions the verification form and time or effort it requires only to set it aside as a determinant of the choice (e.g. “Although I would prefer workers not to be required to fill in the form, I chose this system because it paid compensation to the right people.”), still assign this label.

C.2. Survey experiment

Respondents answered the open-ended questions in Dutch. The answers were classified with the same settings as in the spectator experiment: GABRIEL’s `classify` function, the model `gpt-5.6-luna`, and three independent runs, with a label assigned if it was returned in at least half of the runs. The prompts below were written in English and quote the Dutch survey questions.

C.2.1. Complexity mentioned at baseline

The answer to the first open-ended question, on what goes well and less well in social security, asked before randomisation, was coded for whether it mentions complexity (`Complexity_mention`).

Instructions.

You will be given open-ended responses (in DUTCH, often with typos/informal spelling) to: “Wat gaat er volgens u goed in de sociale zekerheid, en wat gaat er minder goed? Geef antwoord in een of twee zinnen.” — respondents say what goes well and less well about the Dutch social-security system (uitkeringen en toeslagen).

This is a BINARY judgement: apply complexity_mention = 1 if the answer refers anywhere to the administrative burden / complexity of the system or to difficulty of accessing support, and 0 otherwise. Judge ONLY the content of the text; a keyword counts only if it is actually about the system being complex or hard to access. Do not infer from the respondent's presumed politics/education.

Label. complexity_mention.

Apply this label (1) when the respondent's answer refers, anywhere, to the ADMINISTRATIVE BURDEN or COMPLEXITY of the system, or to DIFFICULTY OF ACCESS — e.g. that the system is complicated, bureaucratic, unclear or confusing, involves too much paperwork or too many rules/schemes/agencies, is hard to understand, or that it is difficult to apply for or to actually get the support one is entitled to. Leave it off (0) otherwise. Do NOT code 1 for unrelated points (benefits too high/low, fraud, immigration, work incentives, general (dis)satisfaction) unless the answer ALSO mentions complexity/administrative burden/difficulty of access.

C.2.2. Study purpose

The answer to the question on what respondents think the study is about was coded for whether it identifies the purpose of the study (Guessed_purpose).

Instructions.

You will be given open-ended responses (in DUTCH, often with typos/informal spelling) to the question: "Wat denkt u dat de onderzoeksvraag van dit onderzoek is? Wat willen de onderzoekers weten?" — i.e. what respondents think the study is about.

CONTEXT to judge correctness: the study actually examines how giving people information about the COSTS OF COMPLEXITY of the Dutch social-security system affects their SUPPORT FOR SIMPLIFYING that system.

This is a BINARY judgement: for each response decide whether it correctly identifies that aim (apply guessed_purpose = 1) or not (0). Judge ONLY the content of the text; do not infer from the respondent's presumed politics/education. A keyword alone is not enough — the respondent must actually state that aim as what the researchers want to know.

Label. guessed_purpose.

Apply this label (1) when the respondent CORRECTLY identifies what the study is about; leave it off (0) otherwise. The study's true aim is to examine how giving people information/figures about the costs of COMPLEXITY of the Dutch social-security system affects their SUPPORT FOR SIMPLIFYING that system. Count as correct if the respondent says the study is about opinions on, or support for, SIMPLIFYING the benefits-and-allowances system.

C.2.3. Feeling steered

The answer to the question on whether any part of the questionnaire steered respondents towards a particular answer was coded for whether the respondent reports feeling steered (Felt_steered).

Instructions.

You will be given open-ended responses (in DUTCH, often with typos) to: "Had u het gevoel dat een onderdeel van de vragenlijst u naar een bepaald antwoord stuurde? Zo ja, welk onderdeel?" — i.e. did any part steer them toward a particular answer, and which.

This is a BINARY judgement: for each response decide whether the respondent reports feeling steered toward an answer (felt_steered = 1) or not (0). Treat the instructed 'kies 7' control question as NOT substantive steering (an attention check), so code it 0 unless the respondent ALSO reports substantive steering. Judge only the text; a named survey part counts as steering only if the respondent frames it as having steered/led them.

Label. felt_steered.

Apply this label (1) when the respondent reports feeling STEERED, and leave it off (0) otherwise. Code 1 when the respondent affirmatively says some part of the questionnaire pushed/steered/nudged them toward a particular answer or view, OR describes the survey as leading, suggestive, or manipulative, OR names a SUBSTANTIVE part as having steered them, OR complains the answer options forced a direction. Leave 0 when the respondent says no / not really / that the questions were neutral, objective, open, or that they were free to choose and gave their own opinion; when the answer is blank, gibberish, off-topic, or only names a topic with no steering valence; and — importantly — when the ONLY thing identified is the instructed control/test question that told them to pick 7 (de vraag zeg 7, kies 7, de controlevraag, de testvraag), since that is an attention check, not substantive steering. If they mention the kies-7 question AND also substantive steering, code 1.

C.2.4. Arguments for and against the proposals

The answer to the final open-ended question, on respondents' main arguments for or against the simplification proposals, was coded into the arguments below (Tables A38 and A39). An answer can receive several labels.

Instructions.

You will be given open-ended responses (in DUTCH, often with typos/informal spelling) to: “U heeft net een paar voorstellen gezien om het stelsel van uitkeringen en toeslagen eenvoudiger te maken. Wat zijn voor u de belangrijkste argumenten om daar voor of tegen te zijn? Waarom bent u er zelf voor of tegen?”

CONTEXT (what respondents had just seen — use it to interpret references, not to guess anyone's position): This is a survey about the Dutch social-security system of benefits and allowances (uitkeringen en toeslagen). Just before this question, respondents rated their support for three proposals to SIMPLIFY that system: (1) making the system of benefits and allowances simpler and clearer in general (fewer, more general schemes with clearer rules); (2) replacing all toeslagen except the childcare allowance (kinderopvangtoeslag) with a single fixed monthly amount per household (a ‘basistoeslag’); (3) keeping the schemes but making the CONSEQUENCES OF MISTAKES milder (a mistake leads to less harsh repayments or penalties). Each proposal was rated on a 0–10 scale where 0 meant “keep the current system as it is” and 10 meant “fully support this change” — so being AGAINST means preferring the current system (status quo) and being FOR means supporting the reform. Their arguments may be about simplification in general or about any of these specific proposals.

Read the WHOLE response and apply ALL labels that fit — respondents often give several arguments and may argue both for and against.

IMPORTANT — distinguish the OBJECT of each argument. Some arguments are about the CURRENT system (why it is good or bad: it is complex now, there is fraud now, benefits are too low/high now, the corrections cause harm). Others are about the PROPOSED change itself (why to support or oppose the reform: the new system would be unclear, would increase fraud, would hurt vulnerable people, would be costly to implement). Use the current-system labels (complexity_burden, fraud_misuse, error_corrections, too_generous, inadequate_support, work_incentive, cost_efficiency) ONLY for points about the

EXISTING system. Route proposal-directed worries to change_risk_status_quo — EXCEPT ‘the reform would increase fraud’, which is fraud_from_reform. Judge only the content of the text; do not infer an argument from the respondent’s presumed politics/education/income. A keyword alone is not enough; the respondent must actually make that argument.

Labels. **self_interest.**

Apply ONLY when the respondent reasons from their own FINANCIAL stake — how the system affects their personal money or entitlements: they (used to) receive or financially depend on benefits/toeslagen, they would personally gain or lose money under the proposals, their own eligibility, income, or household finances are at issue, or they note that it does affect them financially. Must be about material impact on the respondent themselves — NOT merely having experience with the system, and NOT concern for other people’s finances or ‘people who need it’ unless tied to the respondent’s own case. (Can be pro OR anti; direction is not part of this label.)

complexity_burden.

Apply when the argument is that the CURRENT system is too complicated, bureaucratic, unclear, or hard to navigate/apply for, or has too many schemes/agencies — i.e. complexity of the EXISTING system, typically as a reason to simplify. Do NOT apply if the point is that the PROPOSAL / new system would be unclear, poorly specified, or add complexity — that goes to change_risk_status_quo.

error_corrections.

Apply when the argument is about harm from the CURRENT system’s year-end corrections, repayments, clawbacks, or penalties for honest mistakes (including the toeslagenaffaire), as a problem with the existing system.

access_takeup.

Apply when the argument is that people who are ENTITLED should be able to get support (more) easily, or about non-take-up — people missing out on support they are entitled to. This is about reaching the entitled, distinct from complexity_burden (which is about the system being complicated as such).

fraud_misuse.

Apply when the argument concerns fraud, abuse, or misuse in the CURRENT system: people receiving support who don’t deserve it, or the need for more

control/enforcement now. Do NOT apply for the worry that the PROPOSAL would create new fraud opportunities — that is `fraud_from_reform`.

`fraud_from_reform`.

Apply when the argument is that the PROPOSAL / simplification would INCREASE fraud, abuse, or misuse, or make cheating easier (e.g. simpler rules or fewer checks lead to more fraud). This is a reason to OPPOSE the change.

`work_incentive`.

Apply when the argument is that work should pay, that the CURRENT benefits/toeslagen discourage work, or that the gap between working and not working is too small.

`too_generous`.

Apply when the argument is that CURRENT benefits/toeslagen are too high, too easy to get, or that people get too much.

`inadequate_support`.

Apply when the argument is that CURRENT benefits/toeslagen are too low, that people cannot make ends meet, or that support should be higher — a critique of the EXISTING level. Do NOT apply for the worry that the PROPOSAL would cut support or leave people worse off — that is `reform_cuts_support`.

`reform_cuts_support`.

Apply when the argument is that the PROPOSAL/simplification would REDUCE support, cut benefits, cause individual income loss, or leave (vulnerable) people worse off or create losers. This is a distributional reason to OPPOSE the change, focused on people being made worse off. If the concern is specifically that support would no longer be DIRECTED to those who need it (loss of means-testing or tailoring), use `loss_of_targeting` instead.

`loss_of_targeting`.

Apply when the argument is that simplification, or a flatter/more uniform benefit, would stop directing support to those who need it most: support would no longer be means-tested or tailored to individual circumstances (maatwerk), so help is no longer concentrated on the needy. This includes worries that people who do NOT need support would then receive it, that everyone getting the same regardless of need is unfair or wasteful, or that differentiation by income/need/situation would be lost. It is about the PRINCIPLE of need-based targeting. Distinct from `reform_cuts_support` (an outright cut or people made worse off), from `too_generous` (the CURRENT system being too generous), and

from `cost_efficiency` (saving money). (Can be pro OR anti; direction is not part of this label.)

`change_risk_status_quo.`

Apply when the respondent opposes or is cautious about the PROPOSAL because the change itself is risky, uncertain, unclear or poorly specified, would ADD complexity, or would be costly /disruptive to implement — or simply prefers keeping the current system as it is. Distributional harms go to `reform_cuts_support`; higher-fraud worries go to `fraud_from_reform`.

`cost_efficiency.`

Apply when the argument is that simplification would SAVE money or make administration cheaper/leaner (fewer agencies or civil servants), or that the CURRENT system is wasteful or expensive to run. Do NOT apply for the worry that implementing the PROPOSAL would itself be costly or disruptive — that goes to `change_risk_status_quo`.

`distrust_government.`

Apply when the argument expresses distrust of government/politics, cynicism, or suspicion that reform serves the elite or has a hidden agenda, or that ordinary people will lose out because the government cannot be trusted. This is orientation toward the actors, not a specific design feature of the proposal.

`immigration.`

Apply when the argument links the issue to immigrants, asylum seekers, refugees.

`no_substantive_answer.`

Apply when there is no codable argument: blank, gibberish, don't-know, a bare meta answer (mijn mening, om mijn mening te geven), or off-topic. When this applies, no other label should apply.

D. Experimental Instructions

This appendix reproduces the instructions and questions of the two experiments as respondents saw them. Each block headed in small capitals is one screen. Text in italics and square brackets was not shown to respondents: it describes the survey flow, the answer format, or differences between treatments. Timers, embedded data, and other technical elements of the Qualtrics files are omitted.

D.1. Spectator experiment

The spectator experiment was programmed in Qualtrics and fielded on Prolific, in English. After the general instructions, each spectator was randomly assigned to one of three treatments (T1, T2, T3). Below I show the T1 version in full. Where T2 or T3 differ, I give the alternative text in square brackets, and I show the treatment-specific figures separately. In the instructions the Complex system is called *System A* and the Simple system *System B*. System A was always shown first.

	Form under System A	Who is compensated under System A
T1	7 minutes	every worker whose output is not fully registered (7 Yellow, 1 Green)
T2	2 minutes	every worker whose output is not fully registered (7 Yellow, 1 Green)
T3	7 minutes	all but one: the form is unavailable to 1 Yellow worker (6 Yellow, 2 Green)

In all treatments, System B compensates the 8 Yellow workers automatically, and both systems compensate 8 of the 12 workers with the same amount.

Introduction and consent

INTRODUCTION

Thank you for participating in this study.

This is a study conducted by researchers at the University of Amsterdam. All data collected in this study are for research purposes only.

The study should last approximately 16 minutes.

For your participation, you will be paid a participation fee of £3.50. You might also earn an additional bonus of £0.25. This bonus will be explained to you later. Due to organisational issues, we need some time to transfer the bonus. You will receive this payment within the next 4 weeks. We will try to be as fast as possible.

This study must be completed on a desktop computer, laptop, or tablet, as it includes figures and tables that need a larger screen to read properly. The type of device you use is recorded, so please make sure you are using one of these before you continue.

CONSENT

Confidentiality

Note that your identity will remain anonymous throughout the study. We will only use your Prolific ID to provide payment. The data may be published in aggregate

form in scientific articles or academic presentations. Your personal identity will not be revealed.

Your participation is voluntary. You may withdraw at any time during the study. However, if you withdraw, you will not receive any money.

Please indicate, in the box below, that you are at least 18 years old and have read and understood this consent form, and that you agree to participate in this study.

- ☐ I agree to participate in this study, I am at least 18 years of age and have read the consent form.
- ☐ I do not want to participate in this study. *[ends the survey]*

NO USE OF AI TOOLS

One more important point before we begin. Your honest and thoughtful answers are essential to the success of our research. We study how real people think and feel, and the value of this work depends entirely on the authenticity of your answers.

Therefore, please do not use AI tools to generate or support your answers. This undermines the integrity of our research.

All submissions are carefully reviewed to ensure quality. If we detect the use of AI tools, we consider this a violation of Prolific's policy, and we may report this and cancel participation.

We rely on your cooperation and thank you for your contribution to this important research! Please confirm below that you understand this.

- ☐ I will not use AI tools.
- ☐ I do not want to participate in this study. *[ends the survey]*

PLEDGE TO READ

In what follows, we will ask you to make decisions that can affect real people.

Because these decisions have real consequences, it is important that you understand the situation before you make them. Please read all the instructions that follow carefully. Afterwards, a short quiz will check your understanding.

- ☐ I understand that I need to read the instructions carefully in order to make an informed decision, and I will do so.

[Next screen: Prolific ID, filled in automatically.]

ATTENTION CHECK

This study should take approximately 16 minutes to complete. It is important that you take the time to read all instructions and that you read questions carefully before

you answer them. Previous research has found that some people do not take the time to read everything that is displayed. Below, there are two questions. To show that you read our questions carefully, please choose both 'Monday' and 'Tuesday' as your answer in the first question and type 'darts' into the 'Other' field of the second question.

What are your preferred days to do sports? (Click all that apply)

- ☐ Monday
- ☐ Tuesday
- ☐ Wednesday
- ☐ Thursday
- ☐ Friday
- ☐ Saturday
- ☐ Sunday

What is your favourite sport?

- ☐ Football
- ☐ Tennis
- ☐ Basketball
- ☐ Volleyball
- ☐ Golf
- ☐ Wrestling
- ☐ Other: _____

General instructions (all treatments)

INSTRUCTIONS 1 OF 6

We will now move to the main part of the study.

It is very important that you read the information that will follow, as in this study you will make a choice that can affect real people.

These people are recruited by us to complete tasks individually, in exchange for a payment. From now on, we will call these people "workers".

On the next pages, you will learn about the workers' situation and the choice you will be asked to make.

INSTRUCTIONS 2 OF 6

Workers are randomly assigned to be Yellow or Green.

Most workers are assigned to be Yellow.

Both types work on exactly the same set of tasks for 15 minutes.

INSTRUCTIONS 3 OF 6

Workers are paid for their work: the more tasks they complete, the more they earn. However, for some workers, part of their output is not registered; it does not count towards their pay.

They work for the full duration of 15 minutes, but after minute 8 their completed tasks are no longer registered. They are only paid for the tasks registered in the first 8 minutes.

Because of that, workers whose output is not fully registered are, on average, paid less than workers whose output is fully registered.

INSTRUCTIONS 4 OF 6

Workers assigned to Yellow are much more likely to have their output not fully registered than workers assigned to Green. Note however, as said before, workers have no control over the colour they are assigned.

Aside from being more likely for workers assigned to Yellow, whether part of a worker's output goes unregistered is random and unrelated to their effort or ability.

Workers are told that their completed output might not be fully registered, due to a random process outside their control. They are not told how likely this is, whether it differs between colours, or which colour they themselves are. As a result, during the session a worker cannot know whether their own output is fully registered.

INSTRUCTIONS 5 OF 6

Compensation

We have set aside funds to give additional compensation to some workers.

This compensation is added on top of their other earnings.

The compensation is equal to the difference between the average earnings of workers whose output is fully registered and those whose output is not fully registered.

There is more than one way to decide which workers actually receive this compensation.

We call each of these a compensation system.

INSTRUCTIONS 6 OF 6

Your task is to decide on a compensation system for a subset of 12 workers.

1 in 30 respondents of this study will be selected at random and their choice will be implemented for 12 real workers.

Whether you are selected does not depend on which system you choose, so please choose the system you genuinely think is best. There is no right and wrong answer.

Your decision is anonymous and not observed by other respondents. Workers know that a decision of another person may affect their payment, but they are not told the details, such as which systems are possible, how the decision is made, or what other workers receive. Each worker receives their payments without being told exactly how the amount was determined.

On the next page, you will learn more about the workers you are deciding for.

Treatment-specific instructions

[Random assignment to T1, T2, or T3.]

THE 12 WORKERS

You will be deciding which compensation system to implement for 12 workers. To help you picture them, we represent each worker below with a figure. Keep in mind that these figures stand for real people.

Each figure's colour and legs tell you about that worker:

- Colour shows the worker's assigned colour: Yellow or Green.
- Grey legs mean worker's output is not fully registered; full colour means their output is fully registered.

As mentioned, workers whose output is not fully registered (shown with grey legs) earn, on average, less than other workers.



[T1 and T2. T3 is shown in Figure D3.]

As you can see, among these 12 workers, there are 8 workers assigned to Yellow and 4 workers assigned to Green.

7 out of 8 workers assigned to Yellow have part of their output unregistered.

1 out of 4 workers assigned to Green has part of their output unregistered. *[T3: 2 out of 4 workers assigned to Green have part of their output unregistered.]*

TWO COMPENSATION SYSTEMS

You can choose between two different compensation systems for these 12 workers.

These systems differ in who is paid and what is required from the worker to access the compensation.

However, both systems pay the same amount to each compensated worker. Both systems also compensate 8 of the 12 workers. That means they have the same total budget.

On the next pages, you will learn more about these two systems.

DESCRIPTION OF THE SYSTEMS

System A	System B
<p>Who is paid the compensation</p> <p>Every worker whose output is not fully registered. <i>[T3: ..., except for one.]</i></p> <p>Who is not paid the compensation</p> <p>Workers whose output is fully registered. <i>[T3: ... and one worker whose output is not fully registered.]</i></p> <p>What is required</p> <p>Workers whose output is not fully registered go through a required verification step: completing a form that reports details of their working session.</p> <p>They complete it after the end of their working session. It takes about 7 minutes <i>[T2: 2 minutes]</i>. The time spent on the form is not paid; it is simply required in order to receive the compensation.</p> <p><i>[T3 only: However, the form is unavailable to 1 worker assigned to Yellow whose output is not fully registered. This worker does not fill in the form and so does not receive the compensation. This is outside the worker's control: they do not cause it and cannot prevent it.]</i></p> <p>All of them <i>[T3: All other workers whose output is not fully registered]</i> complete the form and receive the compensation.</p> <p>Remember: the compensation amount is identical under both systems, and each compensates 8 of the 12 workers!</p>	<p>Who is paid the compensation</p> <p>Every worker assigned to Yellow.</p> <p>Who is not paid the compensation</p> <p>Workers assigned to Green.</p> <p>What is required</p> <p>Nothing. Every worker assigned to Yellow receives the compensation automatically.</p> <p>There is no form to complete and nothing to report. Workers who receive the compensation spend no extra time after their working session.</p>

WHAT EACH SYSTEM MEANS FOR THE 12 WORKERS

We will show you what each system would mean for these 12 workers.

For each system, a figure shows which workers receive the compensation and which complete the form, and a table summarises the total time each type of worker spends and what they are paid.



beneath a worker means that worker completes the verification form, which takes **7 minutes**.



beneath a worker means that worker receives the compensation.

[T2: "...which takes 2 *minutes*."]

ILLUSTRATION OF THE TWO SYSTEMS

Remember:

- Colour shows the worker's assigned colour: Yellow or Green.
- Grey legs mean worker's output is not fully registered; full colour means their output is fully registered.
- Workers whose output is not fully registered end up with lower task earnings than workers whose output is fully registered.

[Followed by the figures and tables in Figure D1 (T1), Figure D2 (T2), or Figure D3 (T3).]

Comprehension quiz

BEFORE THE QUIZ

Before you make your decisions, we ask you to take a short quiz to check your understanding of the situation. For each of several statements, you will indicate whether it is true or false. If you answer all of them correctly, you will receive a bonus payment of £0.25.

Do you want to read the instructions again before you start the quiz?

- ☐ Yes
☐ No

[If "Yes": one screen that repeats all treatment-specific instructions and figures above, introduced by "Here are the instructions again. Take a moment to review, then continue to the quiz."]

QUIZ

Please indicate whether each of the following statements is true or false. If you answer any of them incorrectly, you will be shown a short clarification afterwards. As a reminder, the two systems are shown below.

[The System A and System B figures are shown here (as on the choice screen, Figure D4). Each statement is answered True or False. Correct answers are given in the table below and were not shown.]

Statement	T1, T2	T3
1 Whether part of a worker's output is (not) fully registered depends on how hard they work.	False	False
2 All workers work for the full 15 minutes, whether or not part of their output is unregistered.	True	True
3 During the session, workers can tell whether part of their own output is not being registered.	False	False
4 Under System A (verification form required), every worker whose output is not fully registered receives the compensation.	True	False
5 Under System A (verification form required), workers whose output is fully registered do not receive the compensation.	True	True
6 Under System B (no form, automatic), the compensation goes to all workers assigned to Yellow.	True	True
7 Under System B (no form, automatic), all workers whose output is not fully registered receive the compensation.	False	False
8 Under System A, all workers whose output is not fully registered fill in the verification form, which takes about 7 minutes [T2: 2 minutes]. [T3 adds: "... , except one worker to whom the form is unavailable."]	True	True

CLARIFICATIONS

[Shown only for statements answered incorrectly. Each clarification repeats the statement, followed by:]

1. This is false. Whether part of a worker's output is registered does not depend on how hard they work, or on any other behaviour. It is not something the worker can influence.
2. This is true. Every worker works the full 15 minutes, regardless of whether part of their output ends up unregistered.
3. This is false. Workers do not know during the session that part of their output is not being registered. They are told it might happen, but not whether it is actually happening to them.
4. [T1, T2] This is true. Under System A, every worker whose output is not fully registered fills in the verification form and receives the compensation.
[T3] This is false. Under System A, the verification form is unavailable to one worker whose output is not fully registered, so they do not receive the compensation. All the other such workers fill in the form and receive the compensation.

5. This is true. Under System A, the compensation goes *[T3: can go]* only to workers whose output is not fully registered. Workers whose output is fully registered do not receive it.
6. This is true. Under System B, the compensation goes automatically to all workers assigned to Yellow, with no form required.
7. This is false. Under System B, the compensation is based on colour. A worker whose output is not fully registered but who is assigned to Green does not receive it.
8. This is true. Under System A, all workers whose output is not fully registered fill in the verification form, which takes about 7 minutes *[T2: 2 minutes; T3 adds: “, except one worker to whom the form is unavailable”]*.

Choice

CHOICE

We now ask you to choose which compensation system to implement for these 12 workers.

Your decision is anonymous and not observed by other respondents. As mentioned, workers know that another person’s decision may affect their payment, but they are not told the details, such as which systems are possible, or how the decision is made, or what other workers receive. They simply do what is asked of them (complete their tasks and, where required, fill in a form) and receive a final payment.

Your decision may be selected to determine which system is implemented for these workers.

Therefore, please take this decision seriously - it might matter a lot for real people.

In this study there are other people who make the same decision as you. 1 in 30 respondents is selected at random to determine final payments for each set of 12 workers. If you are selected, your decision will be implemented for these 12 workers. Whether you are selected does not depend on your choice.

Please make your choice now.

[System A and System B figures, see Figure D4.]

Remember: both systems pay the same amount to each compensated worker and compensate 8 of the 12 workers.

- ☐ System A
- ☐ System B

Post-choice questions

REASONS

Can you briefly explain the reasons behind your choice?

Answer: open text box.

PERCEIVED BURDEN

Compared with System B (no form, automatic), how much more effort does System A (verification form of 7 minutes [T2: 2 minutes] required) require from workers to receive the compensation? Please answer on the following scale, where 0 means “No extra effort” and 10 means “Far more effort”.

Answer scale: 0 (No extra effort) ... 10 (Far more effort)

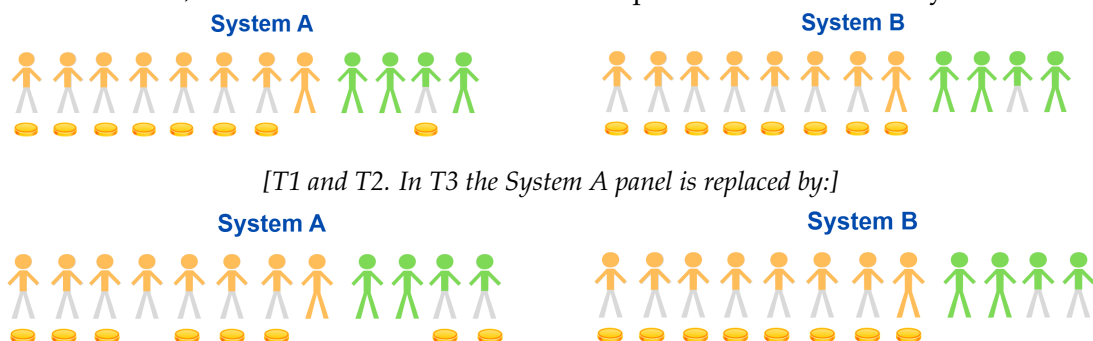
FAIRNESS OF THE DISTRIBUTION

Which system’s distribution of the compensation is fairer?

Base your answer only on who receives the compensation and who does not, and ignore all other properties of the two systems.

Please answer on the following scale, where 0 means “System A’s distribution is much fairer” and 10 means “System B’s distribution is much fairer”.

As a reminder, here is the distribution of the compensation under each system.



Answer scale: 0 (System A’s distribution is much fairer) ... 10 (System B’s distribution is much fairer); 5 (Neither is fairer)

FAIRNESS OF THE REASON FOR EXCLUSION (T3 ONLY)

Under both systems, a worker whose output is not fully registered can end up without the compensation:

- Under System A, because the form is unavailable to them.

- Under System B, because they are assigned to Green, and compensation goes only to workers assigned to Yellow.

In both cases the worker fills in no form and receives no compensation.

Ignoring how many workers this happens to under each system, is one reason for missing out more unfair than the other? Please answer on the following scale, where 0 means “missing out under System A is much more unfair” and 10 means “missing out under System B is much more unfair”.

Answer scale: 0 (missing out under System A is much more unfair) ... 10 (missing out under System B is much more unfair); 5 (Neither is more unfair)

Final questionnaire (all treatments)

INTRODUCTION

We will ask you a few final questions about your views and background.

ATTITUDES

Below, there are 6 statements about various issues. Please indicate how much you agree or disagree with each of the following statements on a scale from 0 (strongly disagree) to 10 (strongly agree).

1. Government departments are generally able to run public services efficiently.
2. The rules and procedures that government departments follow usually serve a real purpose.
3. Government departments waste a lot of money through errors and mismanagement.
4. I have trust in the civil service and government departments.
5. Large redistribution hurts the economy.
6. The government should take measures to reduce differences in income levels.

TARGETING VERSUS SIMPLICITY

Benefits like Universal Credit, Child Benefit and disability benefits each have their own conditions and their own application process: people fill in forms and report information about themselves.

This is not accidental - more conditions and more checks mean support can be targeted more closely on those who need it most. But it also means more time spent filling in forms, more to report, and more that can go wrong: some people who are entitled to support never claim it because the process is difficult.

It is possible to make benefit systems simpler by replacing targeted schemes with more universal ones. This makes benefits easier to claim and receive, but the trade-off is that support no longer fits everyone's situation as closely, so some people get more than they need and others less.

Overall, where do you stand on this trade-off? Please answer on the following scale, where 0 means "benefits should be universal and easier to claim" and 10 means "benefits should be targeted, even if harder to claim."

Answer scale: 0 (benefits should be universal and easier to claim) ... 10 (benefits should be targeted, even if harder to claim)

SIMPLIFICATION POLICIES

Below are two examples of policies that could be introduced in place of current benefits in order to simplify the system.

Policy 1. This policy pays money to people on low incomes, based on income data the government already holds. There is no application and nothing to report. Because it is based on income alone, it does not adjust for things such as rent or who you live with.

Policy 2. This policy pays every adult resident the same monthly amount, enough to cover basic living costs. It is paid automatically, with no application and nothing to report. Because everyone receives it, it also goes to people who do not need it, which makes it more expensive.

How much would you support introducing Policy 1? How much would you support introducing Policy 2? Please answer on the following scale, where 0 means "keep the current system as it is" and 10 means "fully support this change".

Answer scale: 0 (keep the current system as it is) ... 10 (fully support this change)

FAIRNESS VIEWS, IDEOLOGY, AND TRUST

We will now ask you about what you think of a hypothetical scenario.

Mark and Ben did the same task. Mark earned more than Ben for two reasons: he chose to work longer, and he also got a random bonus that Ben didn't. If you could adjust their pay, what would be fairest?

- ☐ Make their final earnings equal.
- ☐ Remove the part of the gap caused by the random bonus, but let Mark keep what he earned by working longer.
- ☐ Leave their earnings as they are. Mark is entitled to what he ended up with.

In political matters, people talk of "the left" and "the right". How would you place your views on this scale, generally speaking? Please answer on the following scale, where 0 means "left", and 10 means "right".

Answer scale: 0 (Left) ... 10 (Right)

Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?

Answer scale: 0 (You can't be too careful) ... 10 (Most people can be trusted)

BACKGROUND

What is your age? [text box]

What is your gender? Male; Female; Non-binary; Prefer not to say; Prefer to self-describe.

What is your country of birth? UK; Other.

What is the highest educational level that you have completed? Secondary school or lower; Post-secondary / vocational (e.g. BTEC, NVQ); A-levels or equivalent; Some university, no degree; Bachelor's degree; Postgraduate degree (Master's or equivalent); Doctoral degree (PhD or equivalent).

What is your current employment status? Full-time employee; Part-time employee; Self-employed or small business owner; Unemployed; Student; Not in the labour force (for example: retired, or full-time parent).

What was your yearly household income in 2025 pounds before taxes and deductions? (Note: household income = total amount of money earned by every member of your household) Under £15,000; £15,000–£29,999; £30,000–£44,999; £45,000–£59,999; £60,000–£79,999; £80,000–£99,999; £100,000 or more.

Which of the following benefits have you ever received? (select all that apply) Universal Credit or legacy means-tested benefits (Housing Benefit, Income Support, Tax Credits, Jobseeker's Allowance); Disability or health-related benefits; Child Benefit or Child Tax Credit; Carer's Allowance; Council Tax Reduction; Other; I have never received any welfare benefits or government financial support.

END

Thank you for taking part in this study. Please click the button below to be redirected back to Prolific and register your submission.

Figure D1. Illustration of the two systems, T1

System A



Workers	Total Time	Paid
output not fully registered	22 min (15 task + 7 form)	task earnings + compensation
output fully registered	15 min	task earnings

System B



Workers	Total Time	Paid
Yellow workers	15min	task earnings + compensation
Green workers	15 min	task earnings

Notes. Screen shown to spectators in T1, below the “Remember” bullets. A coin beneath a worker means the worker receives the compensation; a sheet means the worker completes the verification form (legend on the previous screen).

Figure D2. Illustration of the two systems, T2

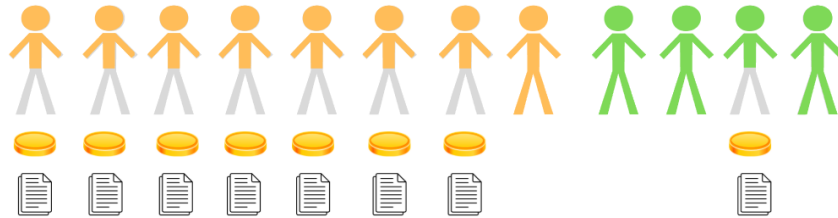


beneath a worker means that worker completes the verification form, which takes **2 minutes**.



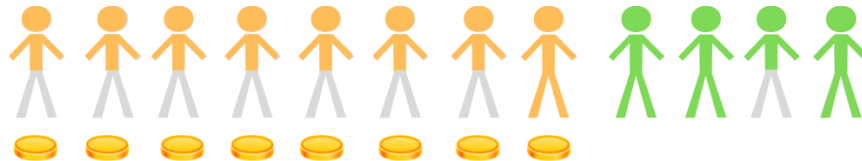
beneath a worker means that worker receives the compensation.

System A



Workers	Total Time	Paid
output not fully registered	17 min (15 task + 2 form)	task earnings + compensation
output fully registered	15 min	task earnings

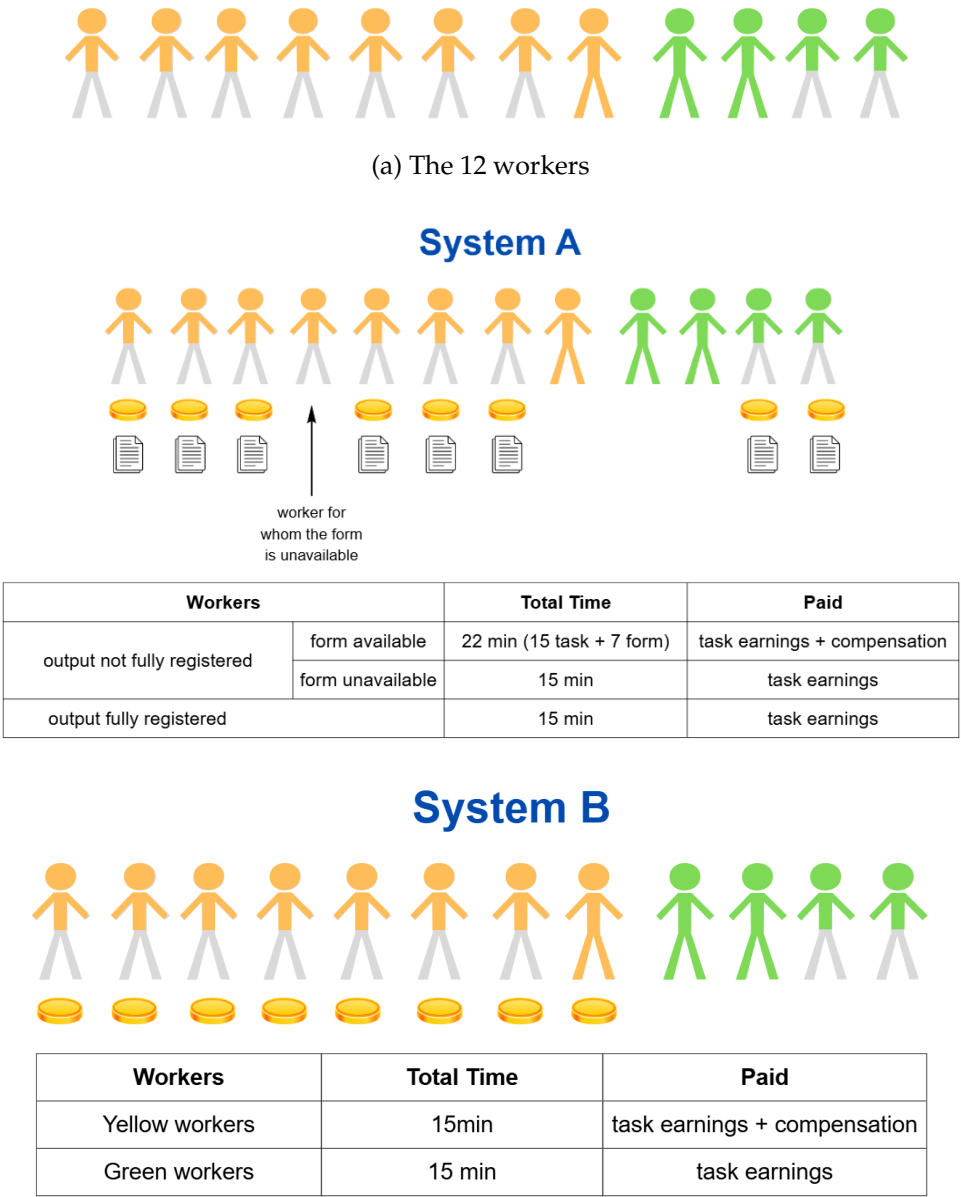
System B



Workers	Total Time	Paid
Yellow workers	15min	task earnings + compensation
Green workers	15 min	task earnings

Notes. T2 differs from T1 only in the length of the form (2 instead of 7 minutes). The top panel is the legend shown on the previous screen in T2; the other panels are the illustration screen.

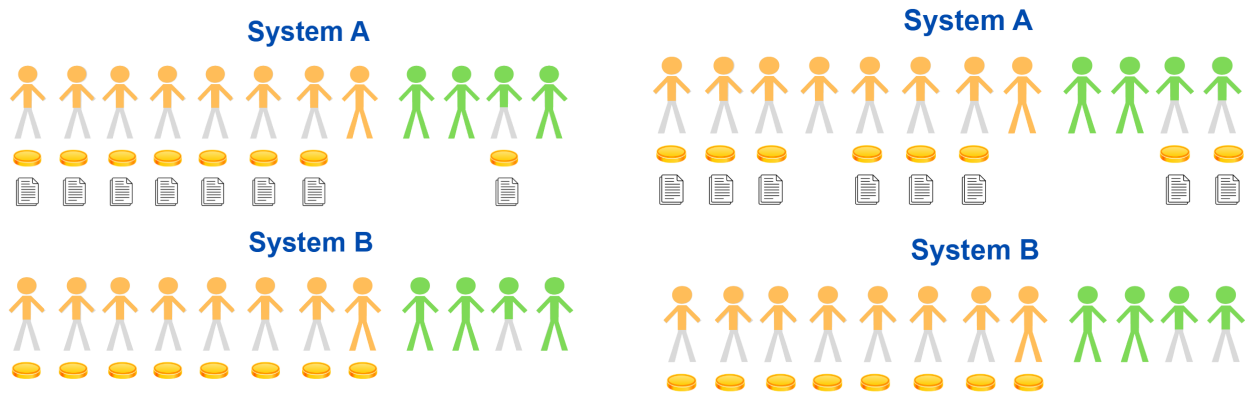
Figure D3. The 12 workers and illustration of the two systems, T3



(b) Illustration of the two systems

Notes. In T3 two of the four Green workers have output that is not fully registered, and the form is unavailable to one Yellow worker whose output is not fully registered (marked by the arrow). System A therefore still compensates 8 workers (6 Yellow, 2 Green). The legend is the same as in T1 (7 minutes).

Figure D4. Figures on the quiz and choice screens



(a) T1 and T2

(b) T3

Notes. The figures were shown one above the other, System A first.

D.2. Survey experiment

The survey experiment was programmed in Qualtrics and fielded in Dutch. The text below is my English translation; Dutch names of schemes are given in parentheses where useful. Respondents were randomly assigned to the topic control (C0), the elicitation control (C1), or the treatment (T). All questions on a 0–10 scale were displayed as eleven radio buttons with labelled endpoints. Several screens could not be skipped for a few seconds: the continue button appeared only after the time indicated in brackets.

Consent and background

WELCOME AND CONSENT

Welcome! Thank you for taking part in this study. This study is carried out by researchers at the University of Amsterdam. We use all the data we collect for research purposes only. Completing the survey takes about 15 minutes.

Your answers remain anonymous. The data may be published in aggregate form in scientific articles or presentations. Participation is voluntary. You can stop at any time. We study how real people think and feel. The value of this study therefore depends on the answers truly being your own. Please do not use AI tools to write your answers or to help write them.

Please indicate below that you have read and understood this consent form, and that you take part in this study.

- ☐ I take part in this study and I have read the consent form.
- ☐ I do not want to take part in this study. *[ends the survey]*

BACKGROUND

What is your age? [text box]

What is your gender? Man; Woman; Non-binary; I prefer not to say; Other, namely:

What is the highest level of education you have completed with a diploma? Primary education; Vmbo or mbo level 1; Havo or vwo; Mbo level 2, 3, or 4; Hbo; Wo; Other / I don't know.

INTRODUCTION TO THE STUDY

In this study we are interested in your opinion on and experiences with social security: the system of benefits and allowances (*uitkeringen en toeslagen*).

We ask what you think of this system and how satisfied you are with it. Please read each question carefully and give your honest opinion.

The results of this study may be shared with policymakers.

Views on the welfare state (all arms, before randomisation)

TOP OF MIND

In your view, what goes well in social security, and what goes less well? Please answer in one or two sentences.

Answer: open text box.

BIGGEST PROBLEMS

In your view, what are the biggest problems with social security in the Netherlands?

Choose exactly 2.

- ☐ The rules and paperwork are too complicated.
- ☐ Work does not pay, because people lose benefits and allowances when their income rises.
- ☐ There are too many different benefits and allowances, and too many agencies that administer them.
- ☐ Benefits and allowances are too high.
- ☐ Fraud and abuse.
- ☐ Benefits and allowances are too low for some people who depend on them.
- ☐ Other: _____

[The first six options were shown in random order; "Other" was always last.]

SATISFACTION

All in all, how satisfied are you with how social security (benefits and allowances) works? Please answer on the following scale, where 0 means "very dissatisfied" and 10 means "very satisfied".

Answer scale: 0 (very dissatisfied) ... 10 (very satisfied)

Prior beliefs

[Random assignment to C0, C1, or T. C1 and T answer the five questions below, one per screen, in random order. C0 answers only the two filler questions (the last two), also in random order.]

INTRODUCTION

In this part we ask you to estimate a few facts about social security in the Netherlands. All these figures come from official studies of the past few years. For each question, you give your answer by moving the slider. Please give your best estimate. Do not use other sources when answering the questions.

NON-TAKE-UP (C1, T)

Social assistance (*bijstand*) is a benefit for people who have too little income or wealth to make ends meet on their own. Someone is entitled to social assistance if they meet all the conditions, such as a low income and little wealth.

Out of every 100 people who are entitled to social assistance (who therefore meet all the conditions), how many do you think actually receive social assistance?

Answer: slider from 0 to 100 (steps of 1; current value shown).

UNDERSTANDING THE RULES (C1, T)

In a recent study, allowance recipients were asked which assets (savings and possessions) they must report for their allowance.

Out of every 100 people with an allowance, how many do you think answered the question correctly? This question is about people with healthcare allowance, rent allowance, or child-related budget (*zorgtoeslag, huurtoeslag, kindgebonden budget*).

Answer: slider from 0 to 100 (steps of 1; current value shown).

CORRECTIONS (C1, T)

Allowances are paid out during the year based on an estimate of income. If income turns out to be higher or lower than the estimate, people have to pay money back or receive extra money.

Out of every 100 households with rent allowance, how many do you think had to pay money back or received extra money at the end of the year, in a recent year?

Answer: slider from 0 to 100 (steps of 1; current value shown).

FILLER: AGE OF RECIPIENTS (ALL ARMS)

The healthcare allowance is a contribution to the cost of health insurance.

Out of every 100 people who receive healthcare allowance, how many do you think are aged 65 or older?

Answer: slider from 0 to 100 (steps of 1; current value shown).

Out of every 100 adults in the Netherlands, how many do you think receive some form of benefit or allowance, including pensions and child benefit?

Answer: slider from 0 to 100 (steps of 1; current value shown).

Information: placebo (C0 and C1)

INTRODUCTION

We will now show you a few facts about the history and organisation of social security in the Netherlands. For each topic you will see a year or a number, with a short explanation.

Each screen stays up for a few seconds so that you have time to read it calmly; after that, the button to continue appears.

PLACEBO 1 OF 4 [8 seconds]

The AOW

The year in which the General Old Age Pensions Act (AOW) was introduced: **1957**

The AOW is the state basic pension: everyone who reaches the AOW age receives the AOW. The AOW replaced the Emergency Act on Old Age Provision of 1947, a temporary scheme for older people. Under that emergency act, the number of benefits rose from 250 thousand in 1947 to 369 thousand in 1956.

Since 1980, the level of the AOW has been linked to the minimum wage: a single person receives 70 percent of the net minimum wage per month, married (and cohabiting) people 50 percent.

PLACEBO 2 OF 4 [8 seconds]

The SVB

The year in which the SVB, which pays out the AOW, was founded: **1901**

The Social Insurance Bank (SVB) pays out the AOW. When it was founded in 1901, the organisation was called the Rijksverzekeringsbank; at the time it was the first agency administering social insurance in the Netherlands. The name has changed a few times since then. Today the SVB administers several schemes, including the AOW and child benefit.

PLACEBO 3 OF 4 [8 seconds]

Unemployment insurance

The year in which unemployment insurance (WW) was introduced: **1952**

The Unemployment Insurance Act (WW) was passed in 1949 and came into force in 1952. It gives employees who lose their job a temporary income while they look for new work. The WW is administered by the UWV, which assesses the application and pays out the benefit.

PLACEBO 4 OF 4 [8 seconds]

Child benefit

The year since which child benefit has been granted: **1941**

Child benefit is a contribution to the cost of raising children and has been granted since 1941. In 1962 the scheme was brought under the General Child Benefit Act (AKW). The Social Insurance Bank (SVB) pays out child benefit.

Information: correction (T only)

INTRODUCTION

We will now show you a few of your earlier estimates, next to the figures from official research. This way you can see whether your estimate was correct.

Each screen stays up for a few seconds so that you have time to read it calmly; after that, the button to continue appears.

[Three correction screens follow, in the order below. “[your estimate]” was filled in with the respondent’s own slider answer. The number in the “In reality” line was shown large and in red. Below it, a pictogram of 100 people showed the true number in red, and the explanatory text was set off by a red vertical bar.]

CORRECTION 1 OF 3 [10 seconds]

Unclear rules

Your answer

Out of every 100 people with an allowance, **[your estimate]** could correctly answer a question about which assets they must report for their allowance.

In reality

Out of every 100 people with an allowance, **26** could correctly answer a question about which assets they must report for their allowance. (SEO Amsterdam Economics)



The rules for allowances and benefits are often complicated. Every year, **about 50,000 people pay a commercial agency to fill in their application**. Many of those applications contain errors. This is not only the case for allowances: 40 out of 100 people on social assistance find the rules that apply to them **hard to understand**.

CORRECTION 2 OF 3 [10 seconds]

Unexpected repayments

Your answer

Out of every 100 households with rent allowance, [**your estimate**] have to pay money back or receive extra money at the end of the year.

In reality

Out of every 100 households with rent allowance, **68** have to pay money back or receive extra money at the end of the year. (CPB Netherlands Bureau for Economic Policy Analysis)



The Benefits Agency (*Dienst Toeslagen*) names the complexity of the allowance system as an important cause of this. For example, many people find it hard to estimate their annual income in advance. The number of households that get into debt because they have to pay back allowances has risen sharply in recent years.

CORRECTION 3 OF 3 [12 seconds]

Missing out on support

Your answer

Out of every 100 people who are entitled to social assistance, [your estimate] actually receive social assistance.

In reality

Out of every 100 people who are entitled to social assistance, **65** actually receive social assistance. (Inspectorate of the Ministry of Social Affairs and Employment, Inspectie SZW)



So about 35 out of 100 people who are entitled to social assistance do not receive it. This is not only the case for social assistance: in 2021, €1 billion in allowances went unclaimed. Many people do not apply for support because they are afraid that they will have to pay money back later, or because they find the procedure too complicated. Because there are many different benefits, allowances, and agencies, some people do not always know what they are entitled to. Research shows that half of the people with a Wajong benefit do not know that they can get extra support if their income is too low to make ends meet.

Posterior beliefs (all arms)

IMPORTANCE

How important do you think the role of social security is in Dutch society? Please answer on the following scale, where 0 means “not important at all” and 10 means “very important”.

Answer scale: 0 (not important at all) ... 10 (very important)

ATTENTION TO THE TOPIC

How much attention do you usually pay to news or discussions about benefits and allowances? Please answer on the following scale, where 0 means “no attention at all” and 10 means “a great deal of attention”.

Answer scale: 0 (no attention at all) ... 10 (a great deal of attention)

COMPLEXITY: BETTER OR WORSE FOR CLAIMANTS

Social security in the Netherlands consists of many separate schemes, each with its own rules, its own conditions, and its own application procedure.

All in all, do you think these features of the system make things better or worse for the people who are entitled to support? Please answer on the following scale, where 0 means “much worse”, 5 “no difference”, and 10 means “much better”.

Answer scale: 0 (much worse) ... 10 (much better); 5 (no difference)

ATTENTION CHECK

People differ in how attentively they fill in a questionnaire. This is a check question: we want to know whether you read the questions attentively. Choose 7.

Please answer on the following scale, where 0 means “not at all” and 10 means “very much”.

Answer scale: 0 (not at all) ... 10 (very much)

NUMERIC POSTERIOR (C1 AND T ONLY)

Out of every 100 people who receive healthcare allowance, how many do you think had to pay money back or received extra money at the end of the year?

Answer: slider from 0 to 100 (steps of 1; current value shown).

Policy outcomes (all arms)

INTRODUCTION

We will now ask what you think of various proposals that are currently being discussed. Before we ask your opinion on a proposal, we first give some background information about it. Each screen stays up for a few seconds so that you have time to read it calmly; after that, the button to continue appears.

There are no right or wrong answers. We are interested in your own opinion!

SOCIAL MINIMUM [3 seconds]

The social minimum is the lowest income that the government guarantees to people who cannot support themselves. For a single adult with no other income, this is social assistance: about 70% of the net minimum wage.

To what extent are you for or against raising the social minimum? Please answer on the following scale, where 0 means “completely against” and 10 means “completely in favour”.

Answer scale: 0 (completely against) ... 10 (completely in favour)

HEALTH INSURANCE DEDUCTIBLE [3 seconds]

In Dutch health insurance, people pay the first €385 of healthcare costs per year themselves: the deductible (*eigen risico*).

To what extent are you for or against abolishing the €385 deductible? Please answer on the following scale, where 0 means “completely against” and 10 means “completely in favour”.

Answer scale: 0 (completely against) ... 10 (completely in favour)

POLICY 1: A SIMPLER SYSTEM [5 seconds]

Proposals have been made to make the system of benefits and allowances simpler. For example, the many separate benefits and allowances could be replaced by a smaller number of more general schemes. There are also other proposals.

As a result, the rules and conditions are clearer and simpler, so that people have more grip on the system. It is also easier to apply for support.

But support fits each person’s situation less well. Some people get a bit more than they need and others a bit less.

Total spending on support stays about the same, and so do taxes. It remains just as hard as now to get support without being entitled to it.

Would you, in general, support such a change? Please answer on the following scale, where 0 means “keep the current system as it is” and 10 means “fully in favour of this change”.

Answer scale: 0 (keep the current system as it is) ... 10 (fully in favour of this change)

POLICY 2: A SINGLE BASIC ALLOWANCE [7 seconds]

A proposal

All allowances disappear, except the childcare allowance. Instead, every household in the Netherlands receives a single basic allowance (*basistoelag*), paid out every month. Other benefits, such as social assistance, continue to exist.

How much a household receives depends on:

- the number of adults (one or two);
- the number of children;

- the ages of everyone in the household.

People can more easily get the support they are entitled to. Applying is simple and needs to be done only once: after that, people only have to check a few details.

The basic allowance stays the same every month and is not based on an estimate of income, so no money has to be paid back at the end of the year.

However, the basic allowance does not take into account things such as rent or health-care costs, and is therefore less well tailored to individual situations.

This proposal is financed by abolishing tax credits and deductions, and the current allowances. It remains just as hard as now to get support without being entitled to it.

Would you, in general, support such a change? Please answer on the following scale, where 0 means “keep the current system as it is” and 10 means “fully in favour of this change”.

Answer scale: 0 (keep the current system as it is) ... 10 (fully in favour of this change)

POLICY 3: Milder consequences of mistakes [5 seconds]

A proposal

At present, a mistake (for example, not reporting a change on time) can lead to a recovery of payments or a fine. In this proposal all schemes continue to exist, but the consequences of a mistake become milder.

Total spending on support stays about the same, and so do taxes. It remains just as hard as now to get support without being entitled to it.

Would you, in general, support such a change? Please answer on the following scale, where 0 means “keep the current system as it is” and 10 means “fully in favour of this change”.

Answer scale: 0 (keep the current system as it is) ... 10 (fully in favour of this change)

Final questions (all arms)

BACKGROUND

What is your current employment situation? Employee, full-time; Employee, part-time; Self-employed (freelancer or business owner); Unemployed or looking for work; Unable to work (because of illness or a disability); Student or pupil; Retired; Caring for the family or household; Other, namely: _____

Have you ever received a benefit, allowance, or other financial support from the government?

Do not count the AOW or your employer pension. Yes; No; I don't know.

Which of the schemes below have you ever received? Several answers possible. [only if “Yes”]

Healthcare allowance (zorgtoeslag); rent allowance (huurtoeslag); childcare allowance

(*kinderopvangtoeslag*); child-related budget (*kindgebonden budget*); social assistance (*bijs-tand*); IOAW or IOAZ; social assistance for the self-employed (Bbz); WW (unemploy-ment benefit); WGA; IVA; WAO; Wajong; Sickness Benefits Act (*Ziektewet*); child benefit (AKW); something else, namely: _____

What was the net annual income of your household in 2025? Count all income of all household members: wages, benefits, allowances, pension, and other income, after taxes and deductions. Less than €15,000; €15,000–24,999; €25,000–39,999; €40,000–54,999; €55,000–74,999; €75,000 or more; I don't know or prefer not to say.

How much trust do you personally have in the government and public institutions in the Netherlands? Please answer on the following scale, where 0 means “no trust at all” and 10 means “complete trust”.

In politics, people often talk about “left” and “right”. Where would you place yourself on the scale below, where 0 means left and 10 means right?

PURPOSE OF THE STUDY

What do you think is the research question of this study? What do the researchers want to know?

Answer: open text box.

FEELING STEERED

Did you feel that any part of the questionnaire steered you towards a particular answer? If so, which part?

Answer: open text box.

REASONS

You have just seen a few proposals to make the system of benefits and allowances simpler. What are, for you, the most important arguments for or against them? What is, for you personally, the most important reason for your opinion? Why are you yourself for or against?

Answer: open text box.

Comprehension questions (T only)

COMPREHENSION

Earlier in this study you had to estimate a few figures. After that, you saw the real figures. According to those real figures, was the system better or worse than you had thought? (0 = much worse than I thought, 5 = as I thought, 10 = much better than I thought)

Answer scale: 0 (much worse than I thought) ... 10 (much better than I thought); 5 (as I thought)

Which statement best fits the information you saw earlier in this study?

- ☐ Most people get the support they are entitled to without much effort.
- ☐ For many people the system is so complicated that they make mistakes or miss out on support.
- ☐ Fraud and abuse are the biggest problems in the system.
- ☐ I don't remember.

Earlier in this study you saw how many out of 100 households with rent allowance had to pay money back or received extra money at the end of the year. Do you remember whether that was more or less than 50 out of 100?

- ☐ More than 50 out of 100
- ☐ Less than 50 out of 100
- ☐ I don't remember