Introduction

A growing stream of research in public administration is concerned with how burdensome rules and requirements negatively affect citizen satisfaction, the quality of public service delivery, and citizenship rights (e.g., Christensen et al., 2020; Kaufmann et al., 2019; Kaufmann & Tummers, 2017; Tummers et al., 2016). This type of research is timely, as many existing regulatory reform programs aim to alleviate red tape and administrative burdens for citizens (e.g., Kaufmann et al., 2018; Ntaliani & Costopoulou, 2018). However, there remains a knowledge gap when it comes to understanding how citizens perceive red tape and administrative burdens, especially because a certain amount of administrative burden is required for the effective functioning of bureaucracy (Kaufman, 1977; Moynihan & Herd, 2010). This is where we seek to make our contribution.

In this article, we draw on insights from the procedural fairness literature and hypothesize that rule consistency and outcome favorability both affect the level of perceived citizen red tape. Rule consistency is a fundamental part of effective citizen–state interactions. A lack of rule consistency implies that citizens may not be treated fairly, which can result in feelings of resentment toward government (Lens, 2009). Inconsistently applied rules fuel perceptions of ineffective government rules and—consequently—high levels of red tape (De Jong & van Witteloostuijn, 2015).

Another key part of the red tape and procedural fairness puzzle is the degree to which decision outcomes favorably (or unfavorably) affect individuals. Studies have shown that individuals evaluate social exchanges more positively if they receive a favorable outcome (e.g., Blader & Tyler, 2003; Brockner, 2002). As such, we hypothesize that individuals receiving an unfavorable outcome perceive a higher level of red tape compared with individuals receiving a favorable outcome. Indeed, earlier experimental work has already shown that outcome favorability has a statistically significant effect on perceived red tape (Kaufmann & Feeney, 2014).

We also hypothesize an interaction effect between consistency and outcome favorability. Studies have found that perceptions of procedural fairness, as evidenced by consistent rule application, are especially important if outcomes are unfavorable (e.g., Brockner et al., 2007; Y. Chen et al., 2003). At the same time, citizens may feel more deserving of a favorable outcome when rules are applied consistently (Blader & Tyler, 2003; Skitka, 2009) and consider decision outcomes as mere capricious bureaucratic behavior (Bozeman & Feeney, 2011) if rules are not. Either way, we expect an interaction effect of rule consistency and outcome favorability on perceived red tape.
We test our reasoning with a survey experiment. A pilot study is first conducted using a sample of first-year university students. Next, we administer our main experiment—which is based on an actual federal jury duty procedure in the United States—to a sample of U.S. citizens on TurkPrime. Participants were randomly assigned to one of the four experimental groups and presented with a vignette that included the jury duty procedure. The procedure described in the experimental vignettes was manipulated on the level of consistency and outcome favorability, and participants were asked to rate the level of red tape therein. Next, a set of factorial analyses of variance (ANOVA) was used to determine if consistency and outcome favorability affect perceived red tape.

The remainder of the article is structured as follows. We first discuss the concept of citizen red tape and introduce our hypotheses on how consistency and outcome favorability affect perceived red tape. We then present our methods and results, followed by a conclusion section.

**Theoretical Framework**

**Citizen Red Tape**

To date, most red tape research has focused on how unnecessarily burdensome organizational rules affect the functioning of public organizations and the well-being of public servants (Bozeman & Feeney, 2011; Bozeman & Scott, 1996; C. Chen, 2012; Kaufmann & Haans, 2020; van den Bekerom et al., 2017). Far less attention has been paid in the literature to how red tape affects citizens (Bozeman, 2000; Moynihan & Herd, 2010; Tummers et al., 2016). This dearth of research is surprising, as policymakers are increasingly concerned with the negative effects of red tape for citizens and clients (e.g., Kaufmann et al., 2018; Ntaliani & Costopoulou, 2018). In this light, Bozeman (1993, p. 290) refers to red tape created by public organizations that negatively affects citizens and clients as “ordinary” red tape “because it is the type of red tape that is most familiar and receives the most attention in both the literature and in popular discussion.” Our focus here is on this type of red tape.

Existing academic research on citizen red tape is limited to only a handful of studies. Tummers et al. (2016) show in an experiment that citizens are less satisfied with a procedure when the level of red tape, operationalized by asking participants to provide duplicative and unnecessary information in the context of a passport application procedure, is high. Moynihan and Herd (2010) link the study of red tape to a policy feedback framework and discuss how red tape can negatively affect citizen political and social rights. Yang and Pandey (2011) show that red tape constrains information flows from governments to citizens, thus acting as a barrier to citizen participation. Kaufmann et al. (2019) use a more indirect approach to study citizen red tape. Using survey data, these authors conclude that administrative delays negatively affect the organizations’ ability to serve clients.

The negative effects of burdensome government rules and requirements are also being studied in the nascent literature on administrative burden. The administrative burden concept can be defined as “an individual’s experience of policy implementation as onerous” (Burden et al., 2012, p. 741). While the red tape literature is concerned with pathological (organizational) rules, administrative burden research assumes that citizens may be burdened in their interactions with the state more broadly as a result of learning, psychological, and compliance costs (Moynihan et al., 2015). Furthermore, while red tape is by definition “bad” (Bozeman, 1993; Bozeman & Feeney, 2011), administrative burden may very well be the result of functional rules that serve legitimate purposes but at the same time constrain citizens (Moynihan et al., 2015). But, while the concepts of citizen red tape and administrative burden are distinct, the mechanisms that link them to negative outcomes for citizens can in many cases be similar. Indeed, literature suggests that citizens’-state relationships can be harmed as a result of both citizen red tape (e.g., Tummers et al., 2016) and administrative burden (e.g., Fox et al., 2020; Hattke et al., 2020; Jilke et al., 2018).

Particularly relevant for the current study is existing research that focuses on how administrative burdens can prevent citizens from getting access to benefits that they are entitled to, or how they can exacerbate inequality in society (Heinrich, 2018; Jilke et al., 2018). In this light, Heinrich (2016) finds that a majority of child support recipients in South Africa experienced interruptions or disconnections in grant receipts as a result of administrative burden.

Similarly, Nisar (2018) uses the context of the Khawaja Sira—a genderqueer group in Pakistan—to show how marginalized groups can be disproportionately disadvantaged by administrative burden due to social factors, the role of third parties, and policy implementation. Using a data set with state-level data covering 2008 to 2017, Fox et al. (2020) find that a reduction in administrative burden was associated with increased enrollments for Medicaid.

These, and similar, studies on red tape and administrative burden point out that bureaucratic encounters are often particularly burdensome for poorer populations. First, effectively navigating red tape requires a certain understanding of bureaucratic processes, as well as an investment of scarce resources (in terms of time and money). These requirements will normally be more difficult to meet for poorer as opposed to richer citizens. Second, the stakes involved in procedures entailing high levels of red tape and administrative burden may be much higher for poorer citizens than richer citizens. For example, child support may well be a crucial income supplement for low-income families but play only a minor role for families that are financially better off. Yet, because poorer citizens are less likely to be able to effectively navigate red tape than richer citizens, the former run a greater risk of missing out on benefits than the latter. As such, improving our understanding of how citizen perceptions of
red tape are shaped can help address concerns about fairness and equality in citizen–state interactions.

In sum, existing research in public administration has shown that red tape and administrative burdens can negatively affect citizen satisfaction, well-being, and political and social rights. At the same time, we still know very little about citizens’ perceptions of red tape (or, administrative burden for that matter). This is where research on rule consistency and outcome favorability comes in, to which we turn next.

**Rule Consistency and Red Tape**

Rule consistency, which can be defined as “the reliability with which rules are applied to individuals and groups” (Borry et al., 2018, p. 370), is an important dimension in the procedural fairness literature (Brockner et al., 1997; Gilliland, 1993; Leventhal, 1980). Various studies have identified a positive relationship between rule consistency and perceptions of procedural fairness (e.g., Sheppard & Lewicki, 1987; van den Bos et al., 1996). Furthermore, rule consistency has been found to have positive effects on employee perceptions of rule fairness (DeHart-Davis, 2017) and rule following (Borry et al., 2018; Fleming, 2020), to mention just a few examples.

Consistent rule application also means that citizens are less likely to perceive particular rules and regulations as red tape. Generally speaking, citizens expect their government to act in an unbiased way (Box, 1999; Lens, 2009; Olsen, 2008). Indeed, “one of the basic principles of good administrative practice is the equal treatment of all citizens” (Grohs et al., 2016, p. 155). Inconsistent rule application implies that citizens are not treated equally, which can evoke feelings of frustration and anger (Lens, 2009), manifest itself in intentions to break rules (Fleming, 2020) or gaming the system (Taylor, 2020), and feed into perceptions of ineffective government rules and high levels of red tape (De Jong & van Witteloostuijn, 2015). Equality is not always fair, nor is inequality always unfair. For present purposes, we assume that rules consistently applied to citizens in similar circumstances (and vice versa) result in perceptions of both equality and fairness. As such, rule consistency results in less perceived red tape, which leads to our first hypothesis.

**Hypothesis 1:** Consistent rule application leads to less perceived red tape than inconsistent rule application.

**Outcome Favorability and Red Tape**

In a nutshell, individuals are more positive about interactions and relationships when they receive a positive rather than a negative outcome, which is known in the literature as outcome favorability (Brockner & Wiesenfeld, 1996; Skitka et al., 2003). For example, Marien and Kern (2018) find that political support goes up after a referendum, but only for individuals who voted for the winning option. Similarly, Esaìasson et al. (2019) use a combination of survey and field experiments to show that citizens’ decision acceptance is most strongly affected by outcome favorability.

Outcome favorability also plays a role in perceptions of red tape. An unfavorable procedural outcome likely means that an individual views the procedure in a negative light and will therefore perceive more red tape. By contrast, a favorable outcome has a positive effect on the individual’s procedural assessment and consequently leads to less perceived red tape. Using a survey experiment administered to a sample of U.S. graduate students, Kaufmann and Feeney (2014) indeed find that a negative procedural outcome leads to a higher level of perceived red tape. This leads to our second hypothesis:

**Hypothesis 2:** A favorable procedural outcome leads to less perceived red tape than an unfavorable procedural outcome.

**Rule Consistency, Outcome Favorability, and Red Tape**

Rule consistency is expected to also interact with outcome favorability to affect perceived red tape beyond either dimension alone. This expectation is in line with existing studies on procedural fairness that look for interaction effects between fairness and outcome, in addition to their respective main effects (e.g., Brockner, 2002; Y. Chen et al., 2003). Some studies have found that the positive effects of procedural fairness are more pronounced when outcomes are unfavorable (e.g., Brockner et al., 2007). To illustrate, Paternoster et al. (1997) find that suspects of domestic violence who were arrested, but treated in a procedurally fair way, show the same low subsequent assault rates as suspects who were given the more favorable outcome of only being warned by the police. Put differently, “the recidivism-inhibiting effect of perceived procedural justice completely offset the criminogenic main effect of arrest” (Paternoster et al., 1997, p. 192). Colquitt and Chertkoff (2002) similarly show that giving individuals an explanation about a decision heightens the positive effect on perceptions of procedural and distributive justice in case of unfavorable outcomes.

There are also theoretical reasons to expect an opposite dynamic, namely that the positive effect of procedural fairness on perceived red tape is strongest in case of favorable outcomes. First, procedural fairness conveys information to individuals that a favorable outcome is deserved (Blader & Tyler, 2003; Skitka, 2009), rather than arbitrary. For example, Krehbiel and Cropanzano (2000, p. 339) find that “[h]igher levels of guilt and anxiety were reported when an unfair procedure resulted in a favorable outcome.” Furthermore, green tape research suggests that the five attributes of effective organizational rules, namely, consistent rule application, written requirements, valid means–ends relationships, optimal control, and purposes understood (DeHart-Davis, 2009a,
for students. The negative binding study advice procedure is the perceived level of red tape due to the high stakes involved in consistency in how the procedure is applied to have an impact on assessments of the procedure. Furthermore, we expect uncertainty and delay for students, which likely affects receiving a negative binding study advice results in financial problems.

Problem definition

In line with this existing research, we expect that the effect of rule consistency on perceived red tape can be more pronounced in both a favorable and an unfavorable outcome setting. However, consistent rule application can ameliorate negative feelings caused by an unfavorable procedural outcome, whereas the absence of consistency may further strengthen beliefs that the procedural outcome is unjust. In this scenario, the positive effect of rule consistency on perceived red tape is strongest in case of an unfavorable outcome. However, a favorable outcome coupled with consistent rule application may give citizens the impression that a favorable outcome is deserved. Given the mixed record in prior research, it is difficult to predict a specific directional effect for the interaction. This leads to our third, nondirectional hypothesis.

Hypothesis 3: Rule consistency and outcome favorability will interact to affect perceived red tape more strongly than rule consistency or outcome favorability alone.

Data and Methods

We tested our reasoning with a survey experiment. First, we did a pilot study using a sample of first-year students to test our design. To this end, we created vignettes based on the actual procedure used at a Dutch research university to determine if first-year students are allowed to continue studying after their first year. The main experiment was administered to a sample of U.S. citizens through the online crowdsourcing platform TurkPrime. This experimental design was based on a jury service procedure in the United States.

Pilot Study

The pilot study is based on the so-called negative binding study advice of a Law School at a large Dutch research university. This procedure stipulates that all first-year students must obtain at least a prespecified number of course credits (in this case, 42 of 60) to be allowed to continue their degree program, unless certain personal circumstances (e.g., serious illness, or a death in the family) prevent students from obtaining the required number of course credits. The procedure was selected because of its salience; failure to obtain enough course credits means that students will need to switch to another study program, or stop studying altogether. As such, receiving a negative binding study advice results in financial uncertainty and delay for students, which likely affects assessments of the procedure. Furthermore, we expect consistency in how the procedure is applied to have an impact on the perceived level of red tape due to the high stakes involved for students. The negative binding study advice procedure is also in line with the university context from the experimental red tape study by Kaufmann and Feeney (2014).

The experiment was created with the survey software Qualtrics and administered to a sample of first-year public administration and law students as part of an undergraduate course on research methods in social sciences. The experiment was conducted in a classroom setting. Upon entering the classroom, participants were given a unique access code with which they could participate in the experiment. Next, participants were automatically assigned at random to one of the four experimental groups. Participants were informed that their school’s education and examination regulations stipulate that students must obtain at least 42 of 60 study credits (European Credit Transfer System [ECTS]) in the first academic year, or they will receive what is known as a negative binding study advice. Participants were also informed that the procedure itself consists of 17 paragraphs and that receiving a negative binding study advice means they will need to quit their degree program unless certain personal circumstances apply.

Participants were then told that they were unable to study for about 2 weeks halfway through the academic year due to a serious illness and that they only obtained 36 (instead of the required 42) course credits at the end of the academic year. We manipulated consistency of how the procedure was applied (consistently or inconsistently) and procedural outcome (favorable or unfavorable), which resulted in four treatments. In the consistent application condition, participants were informed that “You later learn that other cases similar to your own have received the same decision from the Examination Board.” In the inconsistent application condition, participants were told that “You later learn that other cases similar to your own have received a different decision from the Examination Board.” The favorable outcome condition reads as follows: “The Examination Board reviews your case, and you do not receive a negative binding study advice. This means you can continue your degree programme next academic year,” while the unfavorable procedural outcome condition stated that “The Examination Board reviews your case, and you receive a negative binding study advice. This means you need to quit your degree programme.”

Our dependent variable, red tape, was measured in two different ways. First, respondents were asked to indicate on a scale of 0 to 10: “If red tape is defined as burdensome administrative rules and procedures that have negative effects on the organization’s effectiveness, how would you assess the level of red tape in the negative binding study advice procedure?” This common general red tape (GRT) measure is adapted from Rainey et al. (1995). Second, red tape was measured using the three-item red tape (TIRT) scale introduced by Borry (2016). This three-item scale captures different dimensions of perceived red tape, namely, rule effectiveness, necessity, and burdensomeness. Responses to the three items on a seven-point Likert-type scale were averaged to arrive at an overall TIRT score for the final pilot.
study sample (the Cronbach’s α for the three items is .61, which is rather low). Participants were also asked to indicate their age, gender, and political orientation (on a seven-point Likert-type scale, ranging from very left wing to very right wing, which is appropriate for the Dutch context). Finally, we included attention and manipulation checks to ensure the reliability of the experiment.

In total, 129 valid responses were collected. Randomization appears to have been successful, as we do not find significant differences between the four experimental groups with regard to age, \(F(3, 125) = 1.48, p = .22\); gender, \(F(3, 125) = 0.74, p = .53\); or political orientation, \(F(3, 125) = 0.04, p = .99\). Eleven participants (8.5%) failed the attention check and were subsequently removed from the sample. Of the remaining 118 participants, 90.7% answered both manipulation checks for process and rationale transparency correctly. Participants failing one or both of the manipulation checks were also removed from the sample. This means our final pilot study sample consists of 107 participants.

As expected, the average red tape (GRT) scores of participants who were assigned the consistency condition (\(M = 4.75, SD = 2.38\)) are lower than the average red tape scores of participants who were assigned to the inconsistency condition (\(M = 5.49, SD = 2.17\)). Similarly, the average red tape scores of participants who were given a favorable procedural outcome (\(M = 4.84, SD = 2.34\)) are lower than those of participants with an unfavorable outcome (\(M = 5.39, SD = 2.23\)). Results for the TIRT scale are similar. The average red tape scores of participants who were told that the procedure was applied consistently (\(M = 3.63, SD = 1.15\)) are somewhat lower than the average red tape scores of participants who were told the opposite (\(M = 3.91, SD = 1.05\)). Similarly, the difference between the average red tape scores of participants who received a favorable outcome (\(M = 3.41, SD = 0.95\)) and scores of participants who received an unfavorable outcome (\(M = 4.11, SD = 0.95\)) is in line with expectations.

The results of a factorial ANOVA, shown in Appendix A, do not show a statistically significant effect of consistency or outcome favorability, or an interaction effect of these variables, on the general level of red tape (GRT), but we do find a statistically significant effect of outcome favorability on the TIRT. Yet, we are careful to draw conclusions from these pilot study results because of the limited sample size and the low Cronbach’s alpha score for the TIRT scale.

**Main Study**

For our main study, we adapted the context of our experimental design from a university negative binding study advice procedure to a jury service procedure. Specifically, we focus on the context of U.S. citizens asking to be excused from federal jury service. The vignettes are based on the information provided on the U.S. Courts website. Participants are informed that they have been summoned for federal jury service and need to fill out the Juror Qualifications Questionnaire that consists of 12 questions online. Next, participants are told that they can be excused from jury duty on the grounds of undue hardship or extreme inconvenience and that they are currently suffering from a serious medical condition that severely affects their mobility and day-to-day activities. Participants are then informed they have sent a letter to the clerk of court explaining their situation and asking to be excused from federal jury service.

Again, consistency of how the procedure was applied (consistently or inconsistently) and procedural outcome (favorable or unfavorable) are experimentally manipulated. In the consistent application condition, participants were informed that “You later learn that the decisions of the court are consistent, because other cases similar to your own have received the same decision.” In the inconsistent application condition, participants were told that “You later learn that the decisions of the court are not consistent, because other cases similar to your own have received a different decision.” The favorable outcome condition read as follows: “After four weeks, you receive a response from the court. Your request to be excused has been accepted, and you do not need to fill out the Juror Qualifications Questionnaire,” while the unfavorable procedural outcome condition stated that “After four weeks, you receive a response from the court. Your request to be excused has been denied, and you will need to fill out the Juror Qualifications Questionnaire within ten days.” We expect a strong effect of our outcome favorability manipulation on perceived red tape, as receiving an unfavorable outcome in and of itself may lead to a more critical assessment of the procedure. Furthermore, receiving the unfavorable outcome implies more paperwork, which is another potential cause of perceived red tape. The full text of the vignettes is shown in Appendix B.

The experiment was again implemented in Qualtrics and administered to a sample of U.S. citizens using TurkPrime (Litman et al., 2017). TurkPrime is a research tool that integrates with Amazon Mechanical Turk (MTurk). In a nutshell, MTurk is a large online crowdsourcing platform that enables scholars to recruit participants for performing (short) tasks (e.g., Buhrmester et al., 2011; Paolacci & Chandler, 2014). Often, social scientists pay MTurk users a modest amount to complete a short survey, or survey-based experiment (e.g., Kaufmann et al., 2019). MTurk is becoming increasingly popular as a source of data collection in the social sciences because it allows researchers to quickly gather data in a relatively inexpensive way (e.g., Jilke et al., 2016; Stritch et al., 2017). This defining feature of MTurk also raises concerns about data quality. Most notably, MTurk users may quickly click through surveys, fill out the same survey more than once, or create bots to earn more money (e.g., Jacobs & Kaufmann, 2019; Necka et al., 2016).

Fortunately, a number of steps can be taken to ensure data quality. First, TurkPrime includes quality controls that, among others, preclude participants from taking the same
survey twice or using bots to complete surveys. Furthermore, TurkPrime uses IP addresses to verify that users are located in the country they have indicated. Second, including attention checks and manipulation checks in an experimental design can help researchers verify that participants are paying attention to the task at hand. Indeed, many studies have found that the quality of MTurk data is similar to, if not better than, other data collection platforms (e.g., Germin et al., 2012; Sheehan, 2018). For this particular study, we paid participants US$0.80 to complete our assignment. Participants were required to be U.S. based, have an average approval rating of at least 98%, and have experience of at least 1,000 MTurk surveys completed and approved. The mean completion time for our survey experiment is just less than 4 min.

In total, 999 valid responses were collected. Randomization appears to have been successful, as we do not find significant differences between the four experimental groups with regard to age, $F(3, 993) = 0.56, p = .64$; gender, $F(3, 992) = 0.98, p = .40$; or political orientation, $F(3, 993) = 0.36, p = .78$. Only seven participants (0.7%) failed the attention check and were subsequently removed from the sample. Of the remaining 992 participants, 83.8% answered both manipulation checks for consistency and procedural outcome correctly. Participants failing one or both of the manipulation checks were also removed from the sample. This means our final sample consists of 831 participants. The background characteristics of the final sample are shown in Table 1.

We first calculated the TIRT scale by averaging the three items on rule necessity, rule effectiveness, and rule burden-someness. The TIRT scale has a Cronbach’s alpha of .63. This rather low score implies that we need to be cautious in interpreting our results using the TIRT measure of red tape. As a robustness check, we also reran the main analyses with the three TIRT items on effectiveness, necessity, and burden-someness separately. These results are qualitatively similar to those reported below and available from the authors upon request.

As expected, the average red tape (GRT) scores of participants (on a scale of 1–10) who were told that the jury duty procedure was applied consistently ($M = 5.56, SD = 2.45$) are lower than the average red tape scores of participants who were told the procedure was applied inconsistently ($M = 6.28, SD = 1.97$). Similarly, the average red tape scores of participants who received a favorable outcome ($M = 5.00, SD = 2.19$) are lower than those of participants who received an unfavorable outcome ($M = 6.75, SD = 1.96$). Results for the TIRT scale are similar. The average red tape scores of participants in the scenario where the procedure was applied consistently ($M = 3.21, SD = 1.34$) are lower than the average red tape scores of participants in the opposite scenario ($M = 3.95, SD = 1.21$). Similarly, the difference between the average red tape scores of participants who received a favorable outcome ($M = 3.05, SD = 1.13$) and scores of participants who received an unfavorable outcome ($M = 4.07, SD = 1.31$) is in the expected direction.

The results of a factorial ANOVA, reported in Table 2, confirm these descriptive statistics.

| Variable          | N     | M     | SD  |
|-------------------|-------|-------|-----|
| **Age**           |       |       |     |
| Group 1: inconsistent, negative outcome | 220   | 39.50 | 11.59 |
| Group 2: inconsistent, positive outcome  | 199   | 41.55 | 12.23 |
| Group 3: consistent, negative outcome   | 215   | 40.32 | 12.92 |
| Group 4: consistent, positive outcome   | 195   | 40.52 | 11.03 |
| **Gender**        |       |       |     |
| Group 1: inconsistent, negative outcome | 220   | 0.53  | 0.50 |
| Group 2: inconsistent, positive outcome | 199   | 0.49  | 0.50 |
| Group 3: consistent, negative outcome  | 214   | 0.57  | 0.50 |
| Group 4: consistent, positive outcome   | 195   | 0.49  | 0.50 |
| **Political orientation**                |       |       |     |
| Group 1: inconsistent, negative outcome | 220   | 3.52  | 1.77 |
| Group 2: inconsistent, positive outcome | 199   | 3.58  | 1.88 |
| Group 3: consistent, negative outcome   | 216   | 3.67  | 1.83 |
| Group 4: consistent, positive outcome   | 196   | 3.55  | 1.80 |

As expected, the average red tape (GRT) scores of participants (on a scale of 1–10) who were told that the jury duty procedure was applied consistently ($M = 5.56, SD = 2.45$) are lower than the average red tape scores of participants who were told the procedure was applied inconsistently ($M = 6.28, SD = 1.97$). Similarly, the average red tape scores of participants who received a favorable outcome ($M = 5.00, SD = 2.19$) are lower than those of participants who received an unfavorable outcome ($M = 6.75, SD = 1.96$). Results for the TIRT scale are similar. The average red tape scores of participants in the scenario where the procedure was applied consistently ($M = 3.21, SD = 1.34$) are lower than the average red tape scores of participants in the opposite scenario ($M = 3.95, SD = 1.21$). Similarly, the difference between the average red tape scores of participants who received a favorable outcome ($M = 3.05, SD = 1.13$) and scores of participants who received an unfavorable outcome ($M = 4.07, SD = 1.31$) is in the expected direction.

The results of a factorial ANOVA, reported in Table 2, confirm these descriptive statistics.

For the GRT scale, both consistency, $F(1, 827) = 27.28, p < .001$, partial $\eta^2 = .03$, and outcome favorability, $F(1, 827) = 153.44, p < .001$, partial $\eta^2 = .16$, have a statistically significant effect on the level of perceived red tape. There is also a statistically significant interaction effect for consistency and outcome favorability, $F(1, 827) = 6.89, p < .01$, albeit with a very small effect size (partial $\eta^2 < .01$). This interaction effect is shown graphically in Figure 3(A) and shows that the positive effect of rule consistency is more pronounced in favorable rather than unfavorable outcomes.

The results for the TIRT scale are similar to those reported for the GRT scale. Again, both consistency, $F(1, 827) = $
Table 2. Factorial ANOVAs for the Effects of Consistency and Positive Procedural Outcome on Red Tape.

| Dependent variable | Main/interaction effect | Effect size (partial $\eta^2$) |
|--------------------|-------------------------|-------------------------------|
| General red tape scale |                         |                               |
| Consistency (C)    | $F(1, 827) = 27.28, p < .001$ | .03                           |
| Procedural outcome (PO) | $F(1, 827) = 153.44, p < .001$ | .16                           |
| C × PO             | $F(1, 827) = 6.89, p < .01$ | <.01                          |
| Three-item red tape scale |                     |                               |
| Consistency (C)    | $F(1, 827) = 87.77, p < .001$ | .10                           |
| Procedural outcome (PO) | $F(1, 827) = 160.52, p < .001$ | .16                           |
| C × PO             | $F(1, 827) = 7.73, p < .01$ | <.01                          |

Note. ANOVAs = analyses of variance.

Figure 1. Interaction effect of rule consistency and outcome favorability on (A) general red tape (GRT) scale and (B) three-item red tape (TIRT) scale.
The current study also has a number of limitations that offer opportunities for future research. First, we focus on a specific procedure used in the United States for federal jury duty selection. This context applies to most U.S. citizens and is likely salient for our participants. At the same time, our design does not enable us to disentangle the U.S. court system at state and federal levels of government. Indeed, citizens’ level of perceived red tape may be affected in part by the level of government or government agency involved, aside from other contextual factors. It could be worthwhile to experimentally study if perceived red tape is affected by the government actors involved. Future research can also replicate and extend our study in other procedural and cultural contexts. For example, our experimental design can be replicated in other areas of citizen–state interactions, such as applying for benefits. Finally, future studies can test if outcome favorability has the same effect across different cultures.

Second, more can be done to integrate knowledge from the red tape and administrative burden literatures to experimentally study different dimensions of burdensome citizen–state interactions. In so doing, future research can open up the black box of red tape as perceived by citizens. Furthermore, existing red tape conceptualizations and measures can be refined and improved upon by taking into account work on citizen–state interactions from outside the red tape literature. Ultimately, we believe that joint research on citizen red tape and administrative burden can generate relevant new insights into how governments can serve their citizens in the most effective ways.

Third, the importance of outcome favorability as a cause of perceived red tape opens up possibilities for research on different types of outcomes. There was a clear positive and negative outcome in our experimental design, but in many cases, outcomes may be far less black and white. A given outcome that falls somewhere between favorable and unfavorable, or that entails both negative and positive elements, may affect perceived red tape differently compared with a situation in which the outcome is clearly defined. In this light, policymakers may be able to reduce perceived red tape by framing procedural outcomes (e.g., Brockner et al., 1995) in a more positive light.

Fourth, our results also have interesting implications for research on street-level bureaucracy. More specifically, a rich literature exists on how street-level bureaucrats cope with bureaucratic control, implement policies, and use discretion when interacting with clients (e.g., Lipsky, 1980; Maynard-Moody & Musheno, 2003; Tommiers & Bekkers, 2014). While our focus has been on citizens, our approach can be extended to contexts that also include the role of street-level bureaucrats in shaping citizen perceptions of red tape. For example, citizen perceptions of the trustworthiness of street-level bureaucrats may moderate the relationship between rule consistency and citizen red tape. We leave such extensions of our study for future research.
Appendix A

Table A1. Factorial ANOVAs for the Effects of Consistency and Positive Procedural Outcome on Red Tape, Pilot Study.

| Dependent variable | Main/interaction effect | Effect size (partial $\eta^2$) |
|--------------------|-------------------------|-------------------------------|
| General red tape scale | $F(1, 103) = 2.81, p = .10$ | .03 |
| Consistency (C) | | |
| Procedural outcome (PO) | $F(1, 103) = 1.44, p = .23$ | .01 |
| C $\times$ PO | $F(1, 103) = 0.98, p = .33$ | <.01 |
| Three-item red tape scale | $F(1, 103) = 1.52, p = .22$ | .01 |
| Consistency (C) | | |
| Procedural outcome (PO) | $F(1, 103) = 11.52, p < .01$ | .10 |
| C $\times$ PO | $F(1, 103) = 0.02, p = .90$ | <.01 |

Note. ANOVAs = analyses of variance.

Appendix B

Table A2. Full Text of Jury Duty Procedure Vignettes: Introduction (Identical Across Treatments).

You receive an official court mailing summoning you for federal jury service. The mailing informs you that you are an eligible citizen, and that you should fill out a Juror Qualifications Questionnaire online. The Questionnaire consists of 12 questions. The court mailing also informs you that courts can excuse a juror from service at the time he or she is summoned on the grounds of “undue hardship or extreme inconvenience.” In this case, the juror should write a letter to the clerk of court requesting an excuse with an explanation of hardship.

You are currently suffering from a serious medical condition that severely impacts on your mobility and day-to-day activities. As a result, you decide to send a letter to the clerk of court explaining your situation and asking to be excused from federal jury service.

Vignette 1. Inconsistent Rule Application, Unfavorable Outcome.

After 4 weeks, you receive a response from the court. Your request to be excused has been denied, and you will need to fill out the Juror Qualifications Questionnaire within 10 days. You later learn that the decisions of the court are not consistent, because other cases similar to your own have received a different decision.

Vignette 2. Inconsistent Rule Application, Favorable Outcome.

After 4 weeks, you receive a response from the court. Your request to be excused has been accepted, and you do not need to fill out the Juror Qualifications Questionnaire. You later learn that the decisions of the court are not consistent, because other cases similar to your own have received a different decision.

Vignette 3. Consistent Rule Application, Unfavorable Outcome.

After 4 weeks, you receive a response from the court. Your request to be excused has been denied, and you will need to fill out the Juror Qualifications Questionnaire within 10 days. You later learn that the decisions of the court are consistent, because other cases similar to your own have received the same decision.

Vignette 4. Consistent Rule Application, Favorable Outcome.

After 4 weeks, you receive a response from the court. Your request to be excused has been accepted, and you do not need to fill out the Juror Qualifications Questionnaire. You later learn that the decisions of the court are consistent, because other cases similar to your own have received the same decision.

Acknowledgments

The authors would like to thank The American Review of Public Administration (ARPA) editors and the three ARPA anonymous reviewers for their review of this article and excellent suggestions put forward to improve it.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The study was financially supported by Tilburg Law School.

ORCID iD

Wesley Kaufmann https://orcid.org/0000-0002-4434-7175
Note
1. See https://www.uscourts.gov/services-forms/jury-service/juror-qualifications and https://www.uscourts.gov/services-forms/jury-service (retrieved last on April 1, 2020).

References
Blader, S. L., & Tyler, T. R. (2003). A four-component model of procedural justice: Defining the meaning of a “fair” process. *Personality and Social Psychology Bulletin, 29*(6), 747–758.

Bohrnstedt, M., Kwang, T., & Gosling, S. D. (2011). *Amazon’s Mechanical Turk: A new source of inexpensive, yet high-quality, data*. *Perspectives on Psychological Science, 6*(1), 3–5.

Burden, B. C., Canon, D. T., Mayer, K. R., & Moynihan, D. P. (2012). The effect of administrative burden on bureaucratic perception of policies: Evidence from election administration. *Public Administration Review, 72*(5), 741–751.

Chen, C. (2012). Sector imprinting: Exploring its impacts on managers’ perceived formalized personnel rules, perceived red tape, and current job tenure. *The American Review of Public Administration, 42*(3), 320–340.

Chen, Y., Brockner, J., & Greenberg, J. (2003). When is it “a pleasure to do business with you?” The effects of relative status, outcome favorability, and procedural fairness. *Organizational Behavior and Human Decision Processes, 92*(1–2), 1–21.

Christensen, J., Aroe, L., Baekgaard, M., Herd, P., & Moynihan, D. P. (2020). Human capital and administrative burden: The role of cognitive resources in citizen-state interactions. *Public Administration Review, 80*(1), 127–136.

Colquitt, J. A., & Chertkoff, J. M. (2002). Explaining injustice: The interactive effect of explanation and outcome on fairness perceptions and task motivation. *Journal of Management, 28*(5), 591–610.

DeHart-Davis, L. (2009a). Green tape: A theory of effective organizational rules. *Journal of Public Administration Research and Theory, 19*(2), 361–384.

DeHart-Davis, L. (2009b). Green tape and public employee rule abidance: Why organizational rule attributes matter. *Public Administration Review, 69*(5), 901–910.

DeHart-Davis, L. (2017). Creating effective rules in public sector organizations. Georgetown University Press.

DeHart-Davis, L., Chen, J., & Little, T. D. (2013). Written versus unwritten rules: The role of rule formalization in green tape. *International Public Management Journal, 16*(3), 331–356.

De Jong, G., & van Witteloostuijn, A. (2015). Regulatory red tape and private firm performance. *Public Administration, 93*(1), 34–51.

Esaissasson, P., Persson, M., Gilljam, M., & Lindholm, T. (2019). Reconsidering the role of procedures for decision acceptance. *British Journal of Political Science, 49*(1), 291–314.

Fleming, C. J. (2020). Prosocial rule breaking at the street level: The roles of leaders, peers, and bureaucracy. *Public Management Review, 22*(8), 1191–1216.

Fox, A. M., Feng, W., & Stazyk, E. C. (2020). Administrative easing: Rule reduction and Medicaid enrollment. *Public Administration Review, 80*(1), 104–117.

Germaine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review, 19*(5), 847–857.

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *The Academy of Management Review, 18*(4), 694–734.

Grohs, S., Adam, C., & Knill, C. (2016). Are some citizens more equal than others? Evidence from a field experiment. *Public Administration Review, 76*(1), 155–164.

Hattke, F., Hensel, D., & Kalucza, J. (2020). Emotional responses to bureaucratic red tape. *Public Administration Review, 80*(1), 53–63.

Heinrich, C. J. (2016). The bite of administrative burden: A theoretical and empirical investigation. *Journal of Public Administration Research and Theory, 26*(3), 403–420.

Heinrich, C. J. (2018). Presidential address: “A thousand petty fortresses”: Administrative burden in U.S. immigration policies and its consequences. *Journal of Policy Analysis and Management, 37*(2), 211–239.

Jacobs, D., & Kaufmann, W. (2019). The right kind of participation? The effect of a deliberative mini-public on the perceived legitimacy of public decision-making. *Public Management Review, 21*(6), 805–828.
Necska, E. A., Cacioppo, S., Norman, G. J., & Cacioppo, J. T. (2010). Red tape and democracy: How administrative burden and social equity in citizen–state interactions. Journal of Public Administration Research and Theory, 20(1), 104–119.

Ntaliani, M., & Costopoulou, C. (2018). E-Government for lowering administrative burden: An empirical research on European rural businesses. International Journal of Public Administration, 41(9), 700–711.

Olsen, J. P. (2008). The ups and downs of bureaucratic organization. Annual Review of Political Science, 11, 13–37.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. Current Directions in Psychological Science, 23(3), 184–188.

Paternoster, R., Brame, R., Bachman, R., & Sherman, L. (1997). Do fair procedures matter? The effect of procedural justice on spouse assault. Law and Society Review, 31(1), 163–204.

Rainey, H. G., Pandey, S. K., & Bozeman, B. (1995). Research note: Public and private managers’ perceptions of red tape. Public Administration Review, 55(6), 567–574.

Sheehan, K. B. (2018). Crowdsourcing research: Data collection with Amazon’s Mechanical Turk. Communication Monographs, 85(1), 140–156.

Sheppard, B. H., & Lewicki, R. J. (1987). Toward general principles of managerial fairness. Social Justice Research, 1(2), 161–176.

Skitka, L. J. (2009). Exploring the “lost and found” of justice theory and research. Social Justice Research, 22, 98–116.

Skitka, L. J., Winquist, J., & Hutchinson, S. (2003). Are outcome fairness and outcome favorability distinguishable psychological constructs? A meta-analytic review. Social Justice Research, 16(4), 309–341.

Stritch, J. M., Pedersen, M. J., & Taggart, G. (2017). The opportunities and limitations of using Mechanical Turk (MTurk) in public administration and management scholarship. International Public Management Journal, 20(3), 489–511.

Taylor, J. (2020). Public officials’ gaming of performance measures and targets: The nexus between motivation and opportunity. Public Performance & Management Review. DOI: 10.1080/15309576.2020.1744454

Tummers, L., & Bekkers, V. (2014). Policy implementation, street-level bureaucracy, and the importance of discretion. Public Management Review, 16(4), 527–547.

Tummers, L., Weske, U., Bouwman, R., & Grimmelikhuijsen, S. (2016). The impact of red tape on citizen satisfaction: An experimental study. International Public Management Journal, 19(3), 320–341.

Van den Bekerom, P., Torenvlied, R., & Akkermans, A. (2017). Constrained by red tape: How managerial networking moderates the effects of red tape on public service performance. The American Review of Public Administration, 47(3), 300–322.

Van den Bos, K., Vermunt, R., & Wilke, H. A. M. (1996). The consistency rule and the voice effect: The influence of expectations on procedural fairness judgements and performance. European Journal of Social Psychology, 26(3), 411–428.

Yang, K., & Pandey, S. K. (2011). Further dissecting the black box of citizen participation: When does citizen involvement lead to good outcomes? Public Administration Review, 71(6), 880–892.
Author Biographies

Wesley Kaufmann is an associate professor of public governance at Tilburg University, the Netherlands, and a research fellow at Arizona State University. His research focuses on good governance, rules and red tape, transparency, and citizen participation.

Alex Ingrams is assistant professor in the Institute of Public Administration, Leiden University, the Netherlands. He mainly teaches and researches on the topics of public management, governmental transparency, and the influence of digital technologies on public sector information management.

Daan Jacobs is a researcher and lecturer at the Department of Public Law and Governance, Tilburg University, the Netherlands. His work focuses mostly on the topics of legitimacy and citizen participation in public decision-making.