

The Causal Effect of Public Service Motivation on Ethical Behaviour in the Public Sector: Evidence from a Large-Scale Survey Experiment

Jan-Hinrik Meyer-Sahling, University of Nottingham

Kim Sass Mikkelsen, Roskilde University

Christian Schuster, University College London

Abstract

Public Service Motivation (PSM) and ethical behaviour are central concerns in public administration. Yet, experimental evidence on the causes of ethical behaviour and the causal effects of PSM remains scarce, curtailing our understanding of both. This paper draws on a novel survey experimental design to improve this understanding. The design is based on a simple insight: asking about PSM can render salient PSM-oriented identities of respondents. By randomizing the order of PSM and outcome questions, PSM may be exogenously activated among survey respondents, and the causal effects of this activation assessed. Drawing on this design and a sample of over 5,000 Chilean central government employees – the largest experimental PSM survey sample to-date – we find that PSM activation enhances willingness to report ethical problems to management. This provides the first experimental evidence that PSM may promote ethical behavioural intent, and suggests that activating public employees' PSM can benefit public sector ethics.

Keywords: Public Service Motivation; Question Order Effects; Survey Experiment; Ethical

Reporting

JEL Code: D73

Introduction

Understanding how ethical decision making among public servants can be strengthened is a central concern in public administration (Menzel, 2015). Unethical behaviour by public servants is widespread in both developing and developed countries, according to survey evidence (e.g. Kolthoff et al. 2010; Meyer-Sahling, Schuster and Mikkelsen, 2018; OPM, 2012). It can undermine trust in government and foster corruption, among many other ills (Perry, 2015; Vigoda-Gadot, 2007).

How can public sector institutions encourage ethical decision making among public servants? Our paper looks at one potential lever: activating public servants' public service motivation (PSM) – that is their “orientation to delivering services to people with a purpose to do good for others and society” (Perry and Hondgehem, 2008, p. vii). Public employees often have greater PSM than their private sector counterparts (see, e.g., Pedersen, 2013). PSM may thus be a potentially potent lever to harness for public managers to encourage ethical decisions and behaviour. Understanding PSM more generally is also of scholarly weight. It is, as a public sector ethics, a core concern in public administration, with two recent meta studies counting over 400 public administration studies on the two topics (Menzel, 2015; Ritz, Brewer and Neumann, 2016).

Despite their centrality in public administration, however, experimental evidence on ethics and public service motivation remains scarce. In a recent review of 109 articles of ethics in public administration (Menzel, 2015), for instance, *none* employed experimental designs. While PSM has seen more experimental research (e.g. Bellé, 2014; Bellé and Cantarelli, 2015; Clerkin et al., 2009; Christensen et al., 2013; Esteve et al., 2015; Esteve, van Witteloostuijn and Boyne, 2015; Neuman, 2016; Tepe 2016), most of this research has *not* sought to experimentally manipulate – and thus estimate a causal effect of – PSM itself.¹ To our knowledge, only three experiments which manipulate PSM have been conducted so far (Bellé 2013; Pedersen 2015; Christensen and Wright, 2018). As detailed below, however, methodological concerns limit how much can be learnt from these three studies alone.

The lack of experimental evidence limits our understanding of both the causes of unethical conduct and the consequences of PSM. Experimental findings about unethical behaviour in other disciplines – such as economics and management – have been strikingly different from those in observational studies in public administration (Bellé and Cantarelli, 2017). The reliance on observational designs in public administration studies to-date thus implies, according to a recent meta analysis, that we lack a “solid understanding of the causal

¹ Instead, PSM serves as a moderator, dependent variable or non-experimentally-manipulated independent variable in most prior experimental studies.

mechanisms underlying unethical behaviour in public organizations.” (Bellé and Cantarelli, 2017, p. 2)

Assessments of the validity of observational studies of PSM are similarly sceptical. Much of the PSM research endeavour is motivated by positive statistical associations between PSM and a range of favourable attitudes and behaviour: performance, work motivation, job satisfaction and organizational commitment, among many (e.g. Alonso and Lewis 2001; Andersen and Serritzlew 2012; Brewer and Selden 2000; Kim 2005; Naff and Crum 1999). Yet, while these observational associations are suggestive, they are at risk of suffering from omitted variable biases, reverse causality, and other threats to validity (Wright and Grant, 2010; see also Antonakis, Bendahan, Jacquart and Lalive, 2010).^{2,3}

² This risk should not be taken lightly. As Wright and Grant (2010) had noted several years ago, omitted variable, reverse causality and common source bias all threaten the validity of inferences from partial correlations between PSM and other variables. To illustrate these potential biases with the example of PSM consequences: observational PSM studies infer effects from partially correlating a presumably favourable attitude or trait (such as PSM) with other presumably favourable attitudes, traits or behaviour (such as work motivation). Responses to both – and thus correlations between them – may be caused by social desirability and common source biases (cf. Kim and Kim 2016). Moreover, work motivation and other favourable attitudes and behaviour may cause PSM on their own. Perhaps most importantly, a range of other, omitted favourable attitudes and behaviour are likely to correlate with both PSM and other positive outcomes. In conjunction, these biases imply that partial correlations may not enable valid inferences about PSM (cf. Wright and Grant 2010).

³ Methods with stronger claims for causal identification in observational studies – in particular instrumental variables and regression discontinuity designs – of course exist (Antonakis, Bendahan, Jacquart and Lalive, 2010). To our knowledge, these, however, have not been employed in studies of PSM or public sector ethics. This may not surprise: finding suitable instruments or discontinuities in PSM and public sector ethics is challenging, thus putting a premium on a research agenda which experimentally manipulates PSM to assess its causal effects.

As a result, rigorous evidence about both the causes of ethical behaviour in the public sector and the causal implications of PSM remains highly circumscribed. This paper seeks to address this gap. By experimentally estimating the effect of PSM on ethical behavioural intent, it can further our understanding of these two core concepts in public administration.

To assess the causal effects of PSM, the paper draws on a novel survey experimental question order design. The design is based on a simple insight: asking about PSM can render salient PSM-oriented identities of respondents. By randomly assigning respondents to groups being asked PSM questions before and after outcome questions, PSM may be exogenously activated in respondents, and the causal effects of this activation assessed. Contrary to prior experimental research which has largely focused on students, we undertake this experiment with public servants and one of the largest original survey samples on PSM to-date: over 5,000 public employees in eleven central government institutions in Chile. Our findings provide the first experimental evidence for the importance of PSM in encouraging ethics among public servants and, to our knowledge, the first experimental evidence on the causes of ethical behavioural intent in the public sector. To derive these findings, our paper, initially, reviews the literature on, first, the relationship between PSM and public sector ethics and, second, the scarce existing experimental designs

to assess causal effects of PSM. Subsequently, we derive our hypotheses and develop our experimental design to test them. Thereafter, we delineate our survey sample and data. Lastly, we present our results, followed by a discussion and conclusion.

PSM, Ethical Behaviour and the Reporting of Ethical Problems

The past years have seen a marked increase in studies of ethics in public administration (see Menzel 2015). Ethical behaviour is commonly understood as behaviour “that is subject to or judged according to generally accepted moral norms of behaviour” (Reynolds and Cerani 2007, p. 1610) and that thus “reach[es] or exceed[s] some minimal moral standard . . ., such as being honest, obeying the law, and whistle-blowing” (Reynolds and Cerani 2007, p. 1610).

A subset of this literature has explored the relationship between PSM and ethical behaviour and decision making, underscoring its substantive importance in the field (e.g. Brewer and Selden 1998; Lim Choi 2004; Maesschalck, van der Wal and Huberts 2008; Vandenabeele and Kjeldsen 2011; Kwon 2014; Stazyk and Davis 2015; Caillier 2015; Wright, Hassan, and Park 2016). Yet, with the exception noted in the subsequent section, all of these studies are observational and, as detailed below for the case of ethical reporting, potentially vulnerable to biases.

This is, of course, not to say that prior studies do not provide helpful insights into the relationship between PSM and ethical behaviour. Not least, they provide several theoretical

rationales to link the four component dimensions of PSM – self-sacrifice, commitment to public values, attraction to public service and compassion⁴ – to ethical behaviour. These theoretical rationales also imply that we could plausibly expect PSM activation to shape ethical attitudes, decision making, and behaviour.

Most intuitively, self-sacrifice offers clear and direct linkages to ethics. Unethical behaviour is often driven by greed and furthering one’s self-interest (Wang and Murnighan 2011). A willingness to sacrifice one's own interest thus leads to an expectation of more ethical decisions and behaviour (Wright, Hassan, and Park 2016). For instance, corruption is frequently defined as the abuse of public office *for private gain*, leading to a clear expectation that employees willing to forego private gains for the common good are less prone to corruption (cf. Kwon 2014). Congruent with this line of reasoning, self-sacrifice is found to exert the strongest effect of all PSM dimensions on ethical conduct in at least some studies (Lim Choi 2004).

Self-sacrifice, however, is not the only PSM dimension that may be expected to affect ethical decision making favourably. “High road” ethical decision-making depends on values which are embedded in one’s internal moral compass and grounded in personal integrity, reflection and virtue (Stazyk and Davis 2015). A commitment to public values – a second important PSM dimension – is thus likely to lead to ethical behaviour: employees with

⁴ For the sake of congruence with our empirical test below, we discuss the dimensions developed in Kim et al. (2013) here. Our overarching empirical expectation that PSM encourages ethical behavior is not affected by this specific operationalization of PSM dimensions, and could have been equally derived with other operationalizations of PSM dimensions (e.g. Perry, 1996).

higher levels of PSM act more consistently with their own values when they behave ethically (Stazyk and Davis 2015).⁵

Next to this normative rationale, attraction to public service may foster ethical decision making. Individuals who value public service often view contributing to the common good, tackling social problems and helping one's community as primary work rewards (Rainey 1982; Crewson 1997). As such, those attracted to public service are more likely to engage in a range of prosocial and ethical behaviors as they find them more rewarding (e.g. Brewer and Selden 1998; Stazyk and Davis 2015; Wright, Hassan and Park 2016). In fact, experimental studies in other disciplines have shown that individuals who are less focused on their own self-interest tend to behave more honestly (e.g., Winterich et al., 2014).

Lastly, an affective motive could be at play: compassion could curb unethical behaviour. While, to our knowledge, prior public administration studies have not explicitly theorized an affective mechanism, individuals who feel sympathetic to the welfare of others may feel emotionally more compelled to behave ethically (or at least altruistically, see e.g. DeSteno 2015).

All four PSM dimensions – and PSM more generally – may thus be plausibly associated with ethical decision making and behaviour. Our empirical expectation – which is formalized into hypotheses further below – is thus that PSM activation affects ethical behavioural intent positively. Our paper thereby focuses on one form of ethical behavioural intent in

⁵ This holds not least as PSM and ethical behavior share many of the same underlying values (Maasschalck et al. 2008).

particular: the willingness of public servants to report ethical problems to management. The motivation for this choice is two-fold. First, as reflected in the recent increase in scholarly attention on the topic, ethical reporting plays an important role in the prevention of corruption and unethical behavior in public sectors (e.g. Kaptein et al. 2005, Caillier 2015, Hassan, Wright, and Yukl 2014, Brewer and Selden 1998). Huberts and de Graaf (2008, p. 645), for instance, found based on Dutch corruption cases that close colleagues often realise that corrupt officials are overstepping formal boundaries but, in the cases where corruption ensues, "decided not to report anything or speak to their superiors." Similarly, a national business survey in the U.S. found that less than 55 percent of employees who observed misconduct reported it to management (Ethics Resource Center, 2005). Assessing the causal effect of PSM on ethical reporting intent is thus important in its own right.

Second, observational studies on the effect of PSM on ethical reporting have yielded inconclusive findings to-date. Three studies found a significant positive effect (Caillier 2017; Vandenabeele and Kjeldsen 2011; Wright, Hassan, and Park 2016). A fourth study found no robust effect in a full model specification with a slightly different set of controls (Caillier 2015).

That PSM is not invariably associated with ethical reporting suggests, on the one hand, that the two are conceptually distinct. Their relationship thus cannot be taken for granted and is, hence, worth studying empirically. On the other hand, these inconsistent findings underscore the fragility of drawing inferences about the effects of PSM on ethical behaviour

and decision making from partial correlations. These might or might not become insignificant with additional controls (among other threats to validity). As such, prior inconsistent findings also underscore the utility of causally and experimentally identifying the effect of PSM on ethical behavioural intent.

With this in mind, the next section briefly reviews prior studies which have sought to experimentally manipulate PSM to contextualize our own experimental design.

Experimental Studies on the Causal Effects of PSM

Since its initial conceptualization by Perry and Wise (1990), research on PSM has grown exponentially to more than three hundred studies (Ritz et al. 2016). Existing quantitative studies are based overwhelmingly on partial correlations. This may not surprise. Assessing the *causal* effect of PSM requires random manipulation of PSM. Yet, PSM has been found to be in good part a function of antecedents which can usually not be randomly assigned, such as gender, age, education or parental socialization (Ritz et al. 2016). In other words, many of the antecedents of PSM are not dynamic, suggesting PSM might at least be in part a relatively stable trait, settled early in life (cf. Brewer 2003; Houston 2006; Karl and Peat 2004). At the same time, however, panel studies suggest that the PSM of individuals can change over time (e.g. Kjeldsen 2013). Public sector institutions in particular can play a role in these changes, socializing public servants into – or out of – PSM (see Perry, 2000; Kroll and Vogel, 2015; Vandenabeele, 2007). This suggests that at least part of an individual's PSM is amenable to change due to more proximate factors – and may thus be experimentally manipulated.

Three studies have sought to experimentally manipulate PSM to-date (Bellé 2013; Christensen and Wright 2018; and Pedersen 2015; see also Bellé 2014; Grant 2007, 2008). They offer two solutions to random assignment of PSM: PSM *cultivation* and PSM *activation* (Bellé 2013; Grant 2007, 2008; Pedersen 2015; Wright and Grant 2010). PSM *cultivation* refers to the fostering and increase of PSM. Experiments may, for instance, randomly assign employees to treatments that bring them into contact with program beneficiaries in a way that highlights meaningful impact and appreciation of the employees' work, or to self-persuasion interventions to commit them to public service (Grant 2007, 2008; Bellé 2013; Christensen and Wright 2018).⁶ By contrast, PSM *activation* refers to active engagement of existing levels of PSM. Activation renders salient "PSM as the motivational basis for action" (Pedersen 2015, p. 736).

Our own experimental design builds on the insight from these studies that PSM may be experimentally manipulated and activated through low-intensity survey treatments. At the same time, we depart from prior studies in regards to *how* PSM is experimentally manipulated, with a view to facilitating greater replicability and applicability.

Consider, first, the validity and applicability limitations of Pedersen's (2015) design. Pedersen (2015) shows that law students in a Danish university indicate a greater willingness to spend more time on a future university survey if the survey's purpose is

⁶ PSM cultivation and activation are, of course, not mutually exclusive. Treatments which cultivate PSM through, for instance, beneficiary contact arguably also concurrently activate PSM (cf. Christensen and Wright 2018).

public service-oriented, relative to a control group for which no survey purpose is stipulated. This effect is larger for respondents with greater PSM. Pedersen (2015) thus usefully underscores that low-intensity survey treatments may potentially activate PSM. Generalizing from and applying this design to other contexts and outcomes is thorny, however. The study focuses on the willingness of students to fill out future surveys for a university researcher, rather than attitudes and behaviours undertaken on-the-job by public servants in state institutions. In addition, the design does not attempt to check for the possibility of social desirability bias (SDB). Both higher PSM scores and higher willingness to fill out future surveys in response to public service appeals may be due to SDB, not least as recent studies suggest PSM scores are prone to SDB (cf. Kim and Kim 2016). As such, the design may have simply tested the effect of activating SDB rather than PSM.⁷

External validity and replicability limitations also extend to Bellé's (2013) otherwise remarkable study. Bellé (2013) shows in a field experiment that nurses whose PSM was exogenously manipulated through beneficiary contact or self-persuasion perform better. These inferences, however, are based on data from 90 nurses in a single hospital in Italy, who volunteered to help with a humanitarian emergency in a former war zone. Arguably, if PSM could ever be expected to affect performance, it is among staff in social sectors who volunteer for a humanitarian emergency. Generalizability to harder cases of more ordinary public sector work and to behaviour other than job performance is thus anything but

⁷ Though, as we argue in the conclusion, activating socially desirable behavior may in and of itself be desirable for public sector institutions.

certain. Moreover, Bellé's (2013) approach is not easily replicable, particularly when seeking a larger number of respondents. It requires the careful design of a field or lab experiment to randomize beneficiary contact, or a survey-based or field-based self-persuasion intervention, for which the risk of confounding – effects of self-persuasion on variables other than PSM – is likely to be high (cf. Dafoe, Zhang and Caughey 2016).

Lastly, in the study which is arguably closest to our own, Christensen and Wright (2018) assess, with a group of U.S. undergraduate students from a religious university, whether a series of pro-social primes and self-persuasion exercises inspired by Bellé (2013) and Ariely et al. (2014) enhance ethical behaviour. In three separate studies, they find no effect. As the authors themselves note, however, limitations in their sample, priming intervention, and outcome measure may explain their null findings. In two of three studies, their intervention was ineffective in priming the PSM of participants, plausibly as students at a religious university already receive frequent pro-social primes in their university environment, as the sample size was small relative to the present study⁸, and as the prime did not relate to specific work outcomes.

In the third study, only 18 students engaged in unethical behaviour across treatment and control groups, complicating the identification of treatment effects.

⁸ Although these samples were larger (with over 100 and in one case 200 in each treatment) than many other experiments conducted in public administration, they may still be underpowered. We return to the implications of our sample size and statistical power in experimental studies in public administration the conclusion.

In sum, existing works provide some initial evidence for a causal effect of PSM, and suggest that PSM may be experimentally manipulated. Also in light of their methodological and external validity limitations, however, our understanding of the causal implications of PSM remains limited relative to the otherwise sizable body of works on PSM. This also holds for the relationship between PSM and ethical behaviour. The next section thus develops a novel experimental design – and associated hypotheses – to manipulate PSM and assess its effect on ethical behavioural intent.

Hypotheses and Experimental Design

In our experimental design, we build on studies of PSM activation (Pedersen, 2015), social identity theory (Akerlof and Kranton, 2000; Brown, 1986; Tajfel and Turner, 1986) as well as insights from question order survey experiments in other disciplines (e.g. Cohn, Fehr, and Maréchal 2014).

Social identity theory suggests that public servants have multiple identities, based on social categories into which they classify themselves (cf. Brewer et al., 2002). Identities are tied to norms prescribing appropriate behavior. Identities shape behavior as individuals experience disutility if they deviate in their behavior from what their identities prescribe, and utility if they comply. Decision-making in this sense is identity fulfilment (cf. March and Olsen, 1989). The extent to which a given identity guides behaviour in a specific situation depends on the relative weight (salience) individuals attach to that identity in that moment – that is on the extent to which a given situation renders salient (i.e. activates) each identity

(Burke and Stets 2000).⁹ As a result, public service-oriented behaviour and behavioural intent by public employees may be encouraged by rendering more salient – that is activating – public service identities of public employees (Pedersen 2015; Perry and Vandenabeele 2008; Vandenabeele 2007).

In our experimental design, we render salient – that is *activate* – PSM through a novel design, which relies on question order randomization. Question order survey experiments have been made important contributions to other disciplines (e.g. Cohn, Fehr, and Maréchal 2014), and have been a staple in the survey methodology literature (see Oldendick 2008 for an overview). Yet, somewhat curiously, they have hardly been drawn on in public administration and mostly focused on citizen satisfaction surveys (Andersen and Hjortskov 2016; Van de Walle and Van Ryzin 2011).

The underlying rationale of question order survey experiments is simple: “preceding questions provide the context in which the respondent answers an item, and changing this context can make a large difference in survey results” (Oldendick 2008, p. 2). While survey responses can be shaped by a range of distinct question order effects (see, for instance, McFarland 1981; Strack 1992; Moore 2002), these effects typically rely on the common intuition that earlier questions prime respondents to think about the issues covered in earlier questions when answering subsequent questions.

⁹ We follow social identity theorists in equating activation and salience of identities. Identity theorists, by contrast, typically distinguish between the likelihood that an identity will be activated (salience) and an identity actually playing out in a situation (activation) (Burke and Stets 2000). For identity theorists, our treatment may thus activate PSM, but this activation is contingent upon the prior salience of PSM (see hypothesis 2).

Our particular question order design draws on this logic and a simple insight informed by social identity theory: asking about public service motivation can render salient – and thus activate – public service motivation. Its intuition is straightforward. A standard PSM battery requires respondents to answer 16 successive questions about public service motivation. This implies that respondents spend time answering questions concerning their commitment to public values, their willingness to make sacrifices, their compassion, and their attraction to public service.¹⁰ In other words, asking about PSM serves as a reminder to respondents that PSM-related values exist and - to the extent respondents actually hold these values - matter to respondents' identity at work (cf. e.g. Cohn, Fehr, and Maréchal 2014). This engagement with PSM may be expected to make more salient – that is activate – the PSM-founded aspects of the respondents' identities as public officials.¹¹

If asking about public service motivation activates public service motivation, however, then this activation may be randomly assigned – and the causal effect of this activation assessed – by randomizing whether respondents are asked PSM questions before or after outcome variables. To assess the causal effect of this activation, in the treatment condition, respondents are asked a battery of PSM questions *before* being asked about the outcome variable of interest – in our case their willingness to report ethical problems to management. By contrast, in the control condition, respondents are asked the PSM battery

¹⁰ Or other PSM-related dimensions, depending on which PSM scale is employed. Our example here uses the dimensions of the scale developed in Kim et al. (2013).

¹¹ Our treatment is unlikely to cultivate PSM. Answering a battery of questions is unlikely to change people's work-related identities and core attitudes.

only *after* the outcome variable of interest. PSM is thus only activated in the treatment group – not the control group.¹²

With this experimental activation design, our core expectation – PSM has a positive effect on the willingness to report ethical problems to management – can be translated into two testable hypotheses:

Hypothesis 1 (H1): Activating PSM will make respondents more willing to report ethical problems to management.

Hypothesis 2 (H2): Activating PSM will have a larger effect on ethical reporting for respondents with higher levels of PSM.

The rationale for adding H2 is straightforward. Activating PSM is predicated on the assumption that respondents, in fact, count on PSM – or, in social identity terms, that respondents have a public service-oriented identity that can be rendered salient (cf. Burke and Stets 2000). Where respondents count on little or no PSM, there is little PSM to activate in the treatment group. The effect of activating PSM on the outcome variable of interest should thus be larger among respondents with high levels of PSM.¹³

¹² We randomized the order of PSM dimensions within the question order experiment to not give undue weight to any one dimension in particular in the PSM activation.

¹³ H2 might appear to contradict Linos' (2018) argument that PSM-oriented messages are less effective at attracting applicants to public sector institutions as potential applicants are already aware of the public service-oriented nature of public sector work. In Linos' (2018) treatment, however, PSM messages shape behavior of potential *future* public sector workers by *informing* them of salient characteristics of the

The observable implications of these hypotheses are intuitive. For H1 to be true, there should be a significant difference in the outcome variable between respondents in the treatment and control group. For H2 to be true, the difference in the effect on the outcome variable between treatment and control groups should be larger for high-PSM respondents than for low-PSM respondents.

Asking about PSM may, of course, not only activate PSM, but also social desirability bias (SDB). In other words, being asked a series of PSM questions might make respondents respond to subsequent questions in a more socially desirable way, not least as prior list experimental research underscores the risk of SDB in the measurement of PSM (Kim and Kim 2016). To at least partially address this risk, we add a placebo question in randomized order to the outcome question – that is a question that would be affected by social desirability bias but could, contrary to the outcome variable of interest, not be affected by a greater salience of public service-oriented identities. Inspired by Paulhus’ (1984) SDB scale, our SDB check asked respondents whether they had usually accepted constructive criticism at work *in the past*. Since activating PSM should not affect past experience, we should not see an effect of our treatment on this variable. However, since accepting constructive criticism (as opposed to arguing or being vengeful with colleagues) is socially desirable in most settings, we should expect our treatment to have an effect on the SDB

organization that seeks to attract them. By contrast, our treatment does not provide new information to public employees about the public service-oriented nature of the organization they currently work for. Rather, it renders salient the PSM-founded aspects of their work-related identity. Having noted this, our treatment relates to Linos (2018) in that asking about PSM might plausibly also remind respondents of the public service-oriented nature of the organization they work for. This might strengthen perceived person-organization fit for high PSM individuals in particular, which in turn might encourage more ethical behavior (cf. Kim 2012).

question if increased social desirability bias is at play. As with other SDB checks, this check, of course, cannot fully rule out that our findings are driven by SDB. Instead, it only disentangles, to some extent, whether asking about PSM activates a general tendency of respondents for socially desirable answers.¹⁴

Survey and Data

Our survey was conducted in the Chilean central government, with support and authorization from the Chilean Civil Service Agency (*Dirección Nacional del Servicio Civil*).

Chile's central government is a propitious environment for inferring about the causes of ethical behaviour in OECD contexts. Like other OECD countries, Chile has very limited public sector corruption, ranking 24 out of 176 in Transparency International's Corruption Perception Index (Transparency International 2017). Findings might thus travel to other OECD country settings. Moreover, our sample of central government employees offers important external validity advantages over prior experimental ethics research, which has overwhelmingly draw on students (Belle and Cantarelli 2017; Christensen and Wright, 2018).

¹⁴ SDB may not only be driven by the respondent's general tendency for socially desirable responses, but also by what particular forms of behavior are socially desirable in a given situation. A PSM treatment may indicate to respondents that ethical behavior itself is socially desirable, and our SDB check cannot rule out that this alternative SDB mechanism accounts for our findings. Moreover, SDB may be more pronounced for future (ethical) intent than past (accepting criticism) behavior. In addition, SDB checks more generally are widely used but nonetheless contested in the literature (Uziel, 2010). We thus cannot rule out conclusively SDB as the underlying mechanism with our SDB check. We return to this limitation of our study in the conclusion.

Our survey was conducted online on Qualtrics between November 2016 and May 2017. The survey frame comprised all employees in 11 central government institutions: the Treasury, Economic Development Agency (CORFO), Civil Service Agency (DNSC), Attorney General (MP), Social Security Administration (IPS), Planning Directorate in the Ministry of Public Works (MOP), Solidarity and Social Investment Fund (FOSIS), Directorate for Libraries, Archives and Museums (DIBAM), Legal Medical Service (SML), National Fishery Service (SERNAPESCA) and National Health Fund (FONASA).^{15,16}

Of the 15,706 employees in our survey frame, 5,742 employees completed the survey, yielding a response rate of 37 percent. Of these 5,742 respondents, 974 choose not to reply to either our reporting measure or at least one of our sixteen PSM questions and were excluded from the subsequent analysis. Excluding these respondents yields 4,763 survey responses and a response rate of 30 percent.^{17,18} Our respondents are roughly representative of public employees in the eleven surveyed central government institutions in terms of gender, albeit slightly younger and with a greater share of graduates of

¹⁵ The survey experiment in this paper was embedded in a larger survey on civil service management and bureaucratic attitudes and behavior in Chile. The eleven institutions which participated in the survey were invited to this end by Chile's Civil Service Agency.

¹⁶ The results we report below do not take account of the resulting nested data structure. Given our experimental setup, which included simple randomization of our treatment for all respondents, our findings are not at risk of bias from differences in the institutional setting of our respondents. Consequently, we opt for simplicity in our models. Having noted this, we also ran our models using a fixed effects specification, which sustained our conclusions (see appendix 6).

¹⁷ Since H1 can be tested without relying on responses to the PSM battery, we are able to test this hypothesis using a larger sample (comprising 5,050 responses). The estimated ATE for this sample is near identical to the estimate from the 4,763 respondents we report in the main text (see model 2 in table A4. in appendix 4).

¹⁸ We tested whether assignment to the treatment or control group affects non-response to our outcome measure. A logit model finds no evidence for this proposition (est = 0.001, p = 0.888).

vocational (rather than university) degrees. They are also roughly representative of employees in Chile's central government as a whole in terms of gender, albeit slightly younger (Appendix 1). As detailed below in the robustness checks, our results remain significant when excluding individual subgroups, giving us no reason to believe that a fully representative sample of survey respondents would have yielded different results.

Ahead of survey authorization, the survey was presented in-person by one of the authors to the heads (or their representatives) of the participating institutions in September 2016. The leadership of all participating institutions endorsed the survey and encouraged participation from their employees.

To ensure measurement validity, the survey was extensively pre-tested prior to its implementation, including through revisions of the survey items with high-level and technical staff of Chile's Civil Service Agency and ten face-to-face cognitive interviews with public employees in a range of institutions and levels of hierarchy to ensure the meaning of survey questions was well-understood. As the PSM battery was developed in English, all survey questions were translated and subsequently back-translated between Spanish and English to avoid translation issues and ensure congruence between the meaning of translated questions in Spanish and the existing literature in English. These duties of care enhance confidence that respondents understood the items – including the PSM battery – in the intended fashion (see Appendix 2 for the Spanish translation of the PSM survey items).

To measure our dependent variable – willingness to report ethical problems – we duplicate the single item used in Wright, Hassan, and Park (2016): "I feel comfortable reporting ethical problems to upper management." Answer options to this and all PSM battery items were on a five point Likert scale (0-4), ranging from "strongly disagree" to "strongly agree." Following Wright, Hassan and Park (2016, p. 652), we thus only measure the *willingness* to report. This focus on behavioural intent rather than (past) behaviour is deliberate: activating PSM during a survey response cannot affect the frequency of prior ethical reporting.

In addition to enabling meaningful experimentation in our setting, relying on behavioural intent measures of ethical reporting follows common practice in the literature (e.g. Callier 2012; Wright et al. 2016; Vandelabeele and Kjeldsen 2011, exceptions are e.g. Brewer and Selden 1998; Callier 2017). Nonetheless, of course, it is a limitation of our study: ethical behavioural intent and ethical behaviour need not stand in a one-to-one relation. In meta-analyses, however, they are closely related, thus suggesting ethical behavioural intent measures can provide insights into ethical behavior (Ajzen and Fishbein 1980; Armitage and Connor 2001; Hertz and Krettenauer 2016). Moreover, in our application, the measure varies greatly across respondents (mean = 2.329; SD = 1.312), suggesting responses are not merely driven by social desirability biases.

To measure PSM, we make recourse to the international PSM measurement scale developed in Kim et al. (2013). While PSM measurement is subject to an on-going discussion (see Perry and Vandelabeele 2015), Kim et al.'s dimensions are considered as

the “current authority” in at least some works (Prebble 2016, p. 2). We thus replicate Kim et al.’s 4 dimensions and 16 items: attraction to public service (APS), commitment to public values (CPV), compassion (COM), and self-sacrifice (SS) (table 1).¹⁹

We used block order randomization on Qualtrics to randomly assign respondents to a treatment group in which the PSM battery preceded the outcome question, and a control group in which the outcome question preceded the PSM battery. Balance tests suggest that randomization in our experiment was successful in relation to observable characteristics: treatment and control groups are not significantly different in age, gender, education and years of service in the public sector (Appendix 3).

¹⁹ Note that Kim et al.'s (2013) APS scale differs from Perry's (1996) attraction to policy making, as well as other scales in the literature, in order to strengthen discriminant validity vis-a-vis CPV.

Table 1. Survey Items in the PSM construct

Item ID	Survey item
<i>Attraction to public service</i>	
APS1	I admire people who initiate or are involved in activities to aid my community
APS2	It is important to contribute to activities that tackle social problems
APS3	Meaningful public service is very important to me
APS4	It is important for me to contribute to the common good
<i>Commitment to public values</i>	
CPV1	I think equal opportunities for citizens are very important
CPV2	It is important that citizens can rely on the continuous provision of public services
CPV3	It is fundamental that the interests of future generations are taken into account when developing public policies
CPV4	To act ethically is essential for public servants
<i>Compassion</i>	
COM1	I feel sympathetic to the plight of the underprivileged
COM2	I empathize with other people who face difficulties
COM3	I get very upset when I see other people being treated unfairly
COM4	Considering the welfare of others is very important
<i>Self-sacrifice</i>	
SS1	I am prepared to make sacrifices for the good of society
SS2	I believe in putting civic duty before self
SS3	I am willing to risk personal loss to help society
SS4	I would agree to a good plan to make a better life for the poor, even if it costs me money

To further enhance our confidence in the assumption that PSM is a meaningful construct in the Chilean setting, we developed a measurement model for our PSM construct.²⁰ This model will also enable us further below to test hypothesis 2.²¹ Table 2 shows the path

²⁰ This is necessary not least as applying the PSM construct outside North America (where it was developed) has occasionally proven to be difficult (e.g. Kim et al. 2013).

²¹ All analyses for the measurement model were conducted using the *lavaan* package for R (Rosseel 2012). Because our variables are, strictly speaking, ordinal, and because some variables show signs of skew, we estimated our measurement model using robust diagonally weighted least squares. Using ordinary maximum likelihood estimation does not qualitatively change our results.

coefficients and means in the measurement model.²² The model fits the data reasonably well ($\chi^2 = 1172.510$ [df = 100, $p < 0.001$] CFI = 0.975, RMSEA = 0.044), giving us some reassurance that PSM *is* a meaningful construct in Chile. To further examine the scale properties of the four PSM dimensions, we tested the scale reliability of the individual dimensions. Cronbach's Alpha for each dimension was appreciably above standard benchmarks (0.87 for APS, 0.83 for CPV, 0.80 for COM, and 0.84 for SS). To test the internal discriminant validity, we compared the four-dimensional model to a one-dimensional alternative. The former performs significantly better in terms of fit ($\Delta\chi^2 = 607.44$ [df = 4, $p < 0.000$], Δ CFI = 0.310, Δ RMSEA = 0.114). Furthermore, following Kim et al. (2015, 91), we tested the correlation between the four dimensions. These fall between 0.264 (between CPV and SS) and 0.670 (between APS and COM). In no case does the 95 pct. confidence interval include 1.000.

Thus, the scale properties of the PSM construct show acceptable validity and reliability. This also implies that we need not make any changes to the scale developed by Kim et al. (2013). With our PSM scale validated, we proceed to estimating our results.

²² We give each dimension scale by fixing one path coefficient to one. All path coefficients are significant at a 0.001 level. The model is fitted to 4.768 of our 5.742 respondents using listwise deletion.

Table 2. PSM Measurement Model

Item	Path coefficient	Mean	Item	Path coefficient	Mean
	<i>Attraction to public service</i>	0		<i>Commitment to public values</i>	0
APS1	1	3.778	CPV1	1	3.878
APS2	1.151	3.749	CPV2	0.954	3.853
APS3	0.965	3.858	CPV3	1.039	3.795
APS4	1.056	3.832	CPV4	0.785	3.932
	<i>Compassion</i>	0		<i>Self-sacrifice</i>	0
COM1	1	3.457	SS1	1	3.015
COM2	0.888	3.675	SS2	0.892	2.969
COM3	0.692	3.806	SS3	1.188	2.514
COM4	0.970	3.631	SS4	1.026	3.046
	<i>PSM</i>	0			
APS	1				
CPV	1.289				
COM	1.615				
SS	1.481				

Results

Congruent with prior studies, PSM and willingness to report ethical problems are significantly correlated in our survey at the 0.001 level ($r=0.12$). To assess whether this association is causal, we turn to our survey experiment.

With our randomized treatment, we can assess our two core hypotheses. Our first hypothesis suggests that activating PSM will make respondents more willing to report ethical problems to management. If this were true, we should observe an average treatment effect (ATE) of our question order experiment. Figure 1 shows that this is, in fact,

the case. The figure shows estimates from an ordinary least squares model regressing our PSM battery treatment on willingness to report (est = 0.200, $p_{\text{two-sided}} < 0.001$, see model 1 in table A4.1 in appendix 4 for further details). Respondents whose PSM is activated prior to answering how willing they would be to report ethical problems to management have a significantly higher willingness to report. In other words, just being reminded of PSM values through a PSM battery increases our respondents' average willingness to report ethical problems to management. This effect of activating PSM is also substantively relevant: the willingness to report ethical problems shifts upwards by 0.200 points on a 0 to 4 scale.²³

Figure 1. Average Treatment Effect: PSM Activation and Ethical Reporting

<Figure 1 about here>

Our second hypothesis posited that activating PSM will have a larger effect on ethical reporting for respondents with higher levels of PSM. The intuition for this was straightforward: if PSM activation is causally related to willingness to report we would expect respondents with *higher* levels of PSM to be more affected by our treatment. By contrast, we would expect a smaller, or perhaps no, treatment effect on respondents with *lower* levels of PSM, since the activation of identities that are only weakly developed in respondents should have weaker implications for behavioural intent.²⁴

²³ A Welch two-sample t-test gives virtually identical results (est = 0.200, $t = 5.320$, $p_{\text{two-sided}} < 0.001$; treatment group = 2.438; control group = 2.238). Similarly, a model including gender, age, and university education as control variables (see model 3 in table A4.1 in appendix 4) sustains our conclusions (est = 0.205, $p_{\text{two-sided}} < 0.001$)

²⁴ Theoretically, asking about PSM may also remind respondents without *any* PSM that they do not hold *any* PSM-related values – in other words, it might activate their lack-of-public-service-oriented identity – which in

To test this hypothesis, we estimated non-linear treatment effects at varying levels of PSM. The result is shown in figure 2 (using three natural splines, see model 4 in table A4.1 in appendix 4). We find that the treatment *only* has a positive significant effect at high levels of PSM – yet not a significant effect at low levels of PSM. Our analysis thus suggests that our treatment only affects respondents who have some level of PSM for our treatment to activate.²⁵

Figure 2. Treatment effect by PSM level (using three natural splines)

<Figure 2 about here>

Note: The figure plots the predicted scores for respondents in the PSM construct on the first axis (mean = 0.00, SD = 0.247) against the predicted mean difference for willingness to report between treatment and control group on the second axis. Five pct. of observations fall outside of the figure's frame indicating skew. We observe no significant effects for any region not shown. Re-estimating the model using robust regression to account for these outliers gives qualitatively similar results.

Our results hold throughout a range of robustness checks, which address several important internal and external validity threats: social desirability bias, consistency bias, satisficing, attrition, an effect of asking about ethical reporting on PSM, and sensitivity of our findings to the exclusion of specific subgroups or of specific Chilean state institutions.

First, social desirability bias (SDB) could be at play. As noted in the discussion of our experimental design, asking about PSM might simply make our respondents answer

turn might curb their willingness to report ethical problems. We do not find empirical support for this theoretical possibility in our data, however.

²⁵ Since our PSM measure shows some right skew, we re-estimated this model using robust regression. The estimates from this model are qualitatively similar to the results shown in figure 2 (see model 5 in table A4.1 in appendix 4).

questions in a more desirable way. Previous survey experimental research on the causal effects of PSM had not addressed this threat to validity. As noted above, to assess whether our findings are affected by SDB, we asked our respondents whether they had usually accepted constructive criticism at work *in the past*. While not ruling out altogether that our findings are caused by SDB, this allows us to assess one SDB channel – the respondent’s general tendency for socially desirable answers – which may drive our findings. We tested whether our treatment affects this variable (appendix 5). The effect of our treatment on the SDB question is insignificant.²⁶

For respondents at specific PSM-levels, SDB could, of course, still be at cause. In particular, SDB could be higher for respondents with greater PSM, who might be more prone to responding in a socially desirable manner (cf. Kim and Kim 2016). This would threaten the validity of our inference about H2. To address this threat to validity, we estimated the effect of our treatment on the SDB question for varying levels of PSM (using three natural spines to capture non-linear effects). As illustrated in Appendix 5, there is no significant positive treatment effect on the SDB question at *any* level of PSM. Our SDB check thus does not give us any indication that either H1 or H2 are affected by SDB.

We can also rule out that our results stem from satisficing (respondents’ shortcutting

²⁶ Paulhus’ (1984) index is measured on a seven-point frequency scale and subsequently commonly rescored such that the two highest scores (*in casu* those who claim to have been the most willing to accept criticism at work) are assigned a value of 1; the remaining answers are assigned a value of 0. Following this practice, we estimated a logit model predicting the rescored variable using only our order treatment. The resulting estimate is similarly insignificant (est. = -0.06, $p_{\text{two-sided}} = 0.392$). Estimating an OLS model on the original variable gives similar insignificant results (est. = -0.06, $p_{\text{two-sided}} = 0.170$). We re-estimated both models using robust GLM methods and found, again, insignificant effects of the treatment on the SDB check.

cognitive response processes by selecting the first response option they encounter) or consistency bias (respondents seeking to provide answers that are consistent with earlier responses). If either of these two biases were at play, high-PSM respondents in treatment group should select high-ethical reporting, while low-PSM respondents in the treatment group should select low-ethical reporting. As the control group did not receive the 16-item battery prior to the ethical reporting question, by contrast, satisficing and consistency biases could be expected to be less pronounced in the control group (if they were at play). Yet, as figure 2 underscores, the treatment *only* has a positive significant effect at high levels of PSM – yet not a significant negative effect at low levels of PSM, which satisficing or consistency biases would presuppose. Neither satisficing nor consistency biases are thus compatible with our results.

Further, we rule out attrition as a rival explanation. Respondents in our treatment group had to answer sixteen PSM questions before the ethical reporting question. Non-random attrition in the treatment group due to these additional pre-outcome survey questions might bias our inference. Less motivated respondents in the treatment group might be more inclined to drop out of the survey before the ethical reporting question due to having to answer sixteen additional questions before the ethical reporting question. Less motivated respondents might also be less willing to report ethical problems. To address this concern, we compared non-response rates to the ethical reporting question in the

treatment and control groups. There are no statistically significant differences, giving us no reason to believe that attrition is biasing our findings.²⁷

As a final threat to internal validity, we also rule out that our results stem from an effect of asking about ethical reporting on PSM – rather than asking about PSM on ethical reporting. This might be of concern as our control group receives, contrary to the treatment group, the ethical reporting question prior to the PSM battery. In this instance, the moderator variable in our analysis (figure 2) – that is PSM – could be affected by the treatment. To address this concern, we estimated a system of equations which incorporates this potential effect.²⁸ This model, similarly, supports H1 and H2 (see appendix 6).

A range of rival explanations and threats to the internal validity of our results can thus be addressed. Beyond internal validity, we assess, as a last duty of care, whether our data provides suggestive evidence for generalizability. We do so by assessing whether our findings remain robust to the exclusion of specific groups of respondents – in particular specific institutions or specific age and education groups. Assessing sensitivity for the exclusion of age and education groups also speaks to the aforementioned concern with survey representativeness. As noted, our online survey respondents are, on average, slightly younger and more prone to having vocational degrees than the survey population.

²⁷ Specifically, a logit model predicting non-response to our reporting item using our order treatment returns a tiny and insignificant estimate (est = 0.001, p_{two-sided} = 0.887).

²⁸ We estimated a model with two equations and three natural splines, using the systemfit package for the R environment (Henningsen and Hamann 2007). One equation replicates the model discussed in the main text, the other equation models a treatment effect of ethical reporting on PSM. We do indeed find a substantively weak but statistically significant effect of our treatment on PSM (est = 0.017, p_{two-sided} = 0.015). Our findings remain robust to modelling this effect.

To rule out that our findings are driven by specific groups, we re-estimate average treatment effects (ATEs), excluding in each iteration one specific subgroup. As detailed in appendix 6, ATEs remain significant when excluding individual institutions from the sample; when using a fixed (institutional) effects specification; and when excluding specific age or education groups from the sample (see appendix 7). This suggests that our results do not stem from individual institutions or demographic subgroups, but rather have relevance across institutions and across demographic groups.

In sum, our validity checks enhance confidence in our findings: activating PSM in our respondents does make them more willing to report ethical problems to management, and this effect is larger the higher the respondents' level of PSM – that is the more respondents count on PSM to activate. Moreover, this effect is not limited to individual groups of respondents or institutions, but holds across multiple of them; and we find no evidence that rival explanations explain these findings (satisficing, social desirability biases, consistency biases, attrition, and an effect of asking about ethical reporting on PSM). Our experimental evidence thus suggests that activating PSM has a causal effect on ethical behavioural intent in the public sector.

Discussion and Conclusion

Our findings have important implications for the scholarly understanding of PSM and ethical behaviour in the public sector, and for survey research on PSM.

Substantively, our findings provide the first experimental evidence for the importance of PSM in encouraging ethical behavioural intent among public servants. Experimental evidence thus suggests that PSM may not only benefit performance, as prior studies had suggested (Bellé 2013; see also Bellé 2014); but also integrity in public sectors. With these findings, our paper also validates and triangulates prior observational studies on the relationship between PSM and ethical behaviour, most of which had identified positive statistical associations (e.g. Brewer and Selden 1998; Wright, Hassan, and Park 2016; Kwon, 2014; Lim Choi 2004; Stazyk and Davis 2015; Caillier, 2015; Vandenabeele and Kjeldsen 2011). PSM thus appears to hold genuine promise for better functioning public sectors.

For practitioners, our findings suggest in particular that activating PSM among public employees is both feasible through low-intensity treatments (cf. Bellé 2013; Pedersen 2015) and beneficial for public sector ethics.²⁹ Public organizations might wish to take this lesson to heart. It provides suggestive evidence that public service motivation reminders to employees may enhance public sector integrity.

²⁹ The relatively small effect size of PSM activation (0.2 on a 0-4 scale of ethical reporting) does imply, though, that PSM (activation) is far from the only determinant of ethical behavior which practitioners and scholars should pay attention to.

Next to contributing to the scarce experimental literature on PSM, our paper also contributes to addressing the dearth of experimental evidence on ethics in public sectors (see Menzel 2015). Reliance on observational studies has led scholars to conclude that we lack a “solid understanding” of the causes of ethical behaviour in public organizations (Bellé and Cantarelli 2017, p. 2). This paper provides experimental evidence to enhance our understanding of one such cause: public service motivation.

Beyond contributing to our understanding of PSM and ethical behaviour, our findings also have important implications for survey research on PSM and experimental research in public administration more generally. Our results suggest that whether PSM questions precede or follow outcome questions shapes the size of correlations between PSM and outcome variables. Inconsistent findings in meta analyses (e.g. Ritz et al., 2016) about the consequences of PSM for specific outcomes may thus be partially due to survey design effects: whether PSM questions preceded or succeeded outcome questions in surveys may affect whether and how strongly PSM correlates with other variables. To avoid survey design effects, scholars designing PSM surveys should thus either randomize the order of PSM and outcome questions, or ensure outcome questions consistently precede PSM survey questions to disentangle correlation and activation.

For experimental research in public administration more generally, our findings also suggest the importance of large sample sizes to avoid type II errors for effects which are theoretically important but of limited size. In our experiment, Cohen’s d is 0.157. It

required at least 2,098 respondents for a power of 0.95 (see Appendix 8 for a distribution of statistical power across sample sizes). Most prior experimental studies in public administration would have been too underpowered to detect this effect (see James, Jilke and Van Ryzin, 2017).

While our paper thus contributes importantly to the literatures on PSM and ethical behaviour in the public sector, and advances experimental evidence and approaches in public administration more generally (see, e.g. Bækgaard et al. 2015; James, Jilke and Van Ryzin 2017), several limitations remain and point to important avenues for future research.

First, we only assess the effect of PSM on one form of ethical behavioural intent by public servants – willingness to report ethical problems. In doing so, we follow observational studies of PSM and ethical reporting, which had all similarly assessed intent, not behaviour (e.g. Callier 2012; Wright et al. 2016; Vandelabeele and Kjeldsen 2011). Intentions, of course, often predict behaviour (e.g. Ajzen and Fishbein 1980), and meta analyses suggest that ethical behavioural intent is closely associated with ethical behaviour (Armitage and Connor 2001; Hertz and Krettenauer 2016). Whether these behavioural intentions translate into actual behaviour of public servants in our case – and to other ethical behaviour beyond ethical reporting – however remains for future research to assess. Our design would lend itself to assessing this: PSM questions could be exogenously asked before or after ethical behavioural tasks, such as honesty games (cf. Gächter and Schulz 2016; Cohn, Fehr, and Maréchal 2014).

Second, while our SDB check gives us no reason to believe that SDB accounts for our findings, we cannot – as aforementioned – conclusively rule out this possibility. Even if SDB did account for our effects, however, this need not invalidate the appeal of PSM activation. It would implicate that public sector organizations can draw on PSM messages to signal socially desirable behaviour to public employees.

Third, we cannot disentangle with our survey design which PSM dimension contributed (most) to the effect on ethical behavioural intent. We randomized the order of the 16-item battery and the outcome question (next to randomizing the order of the four PSM dimensions within the 16-item battery to not give undue weight to any one dimension). Future research could adapt our survey design to assess the effects of individual PSM dimensions, by randomizing whether individual PSM dimensions precede or succeed outcome questions.

Substantively, our survey experimental design also cannot provide any evidence on how long-lived the effect of PSM activation on ethical behaviour is. As our control group receives the PSM battery after the outcome variable, PSM is also activated in our control group during the survey. A second-wave survey or behavioural game at a later point in time thus would not be able to identify any effect of PSM activation. This points to the utility of field experiments or iterative surveys, with high-intensity PSM treatments at the outset, to complement the evidence presented in this paper and assess how long-lived the effects of PSM activation on ethical behaviour in public sectors are.

Moreover, as prior survey experimental research on PSM activation (Pedersen, 2015), our survey experimental test drew inferences in part based on an assumption of convergent validity. We showcased that asking about PSM enhances the willingness of public servants to report ethical problems, and provided a panoply of additional pieces of evidence which are compatible with the assertion that this effect is due to PSM activation – and not social desirability bias, consistency bias, satisficing or other threats to validity. Our paper, however, does not directly test that PSM activation mediates the effect of asking about PSM on ethical reporting. We thus encourage scholars who embed question order randomization in future PSM research to include PSM activation checks with outcome questions.

Lastly, as any study, our study comes with external validity limitations. We do have some confidence that our findings are generalizable across OECD governments: we survey central government employees rather than students; draw on the, to our knowledge, largest survey experimental PSM sample to-date, with around 5,000 respondents; survey Chile, an OECD member with limited public sector corruption; and find significant effects across institutions and demographic groups within the central government. Whether our results do, in fact, travel to other countries – and other outcome variables beyond ethical reporting – remains an empirical question.

Our paper thus leaves fertile ground for future research to experimentally study the causes of ethical behaviour and causal effects of PSM. More broadly, it also points to a research

agenda on identity activation of public servants. In our study, a low-intensity survey treatment activated the PSM-founded aspects of the respondents' identities as public officials. Future studies could equally use low-intensity treatments – such as question order experiments – to activate other identities of public servants. From bureaucratic rule-following identities to professional (e.g. lawyers, policemen) identities in public service, among many, research opportunities to study which identities can be activated and what behavioural consequences they have are manifold.

References

Ajzen, Icek, and Martin Fishbein. 1980. *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, NJ: Prentice-Hall

Akerlof, George, and Rachel Kranton. 2000. "Economics and identity". *The Quarterly Journal of Economics*, 115 (3): 715-753.

Alonso, Pablo, and Gregory B. Lewis. 2001. Public Service Motivation and Job Performance: Evidence from the Federal Sector. *American Review of Public Administration*, 31(4): 363-80.

Andersen, Lotte Bøgh, and Søren Serritzlew. 2012. Does Public Service Motivation Affect the Behavior of Professionals? *International Journal of Public Administration*, 35(1): 19-29.

Andersen, Simon Calmar and Morten Hjortskov. 2016. "Cognitive Biases in Performance Evaluations." *Journal of Public Administration Research and Theory*, 26(4): 647-662.

Antonakis, John and Samuel Bendahan, Philippe Jacquart and Rafael Lalive. 2010. "On making causal claims: A review and recommendations." *The Leadership Quarterly*, 21, 1086-1120.

Arieli, Sharon, Grant, Adam and Lilach Sagiv. 2014. "Convincing yourself to care about others: An intervention for enhancing benevolence values." *Journal of Personality*, 82(1), 15-24.

Armitage, Christopher J. and Conner, Mark. 2001. "Efficacy of the theory of planned behaviour: A meta-analytic review." *British journal of social psychology*, 40(4), 471-499.

Bækgaard, Martin, Caroline Baethge, Jens Blom-Hansen, Claire A. Dunlop, Marc Esteve, and Morten Jakobsen. 2015. Conducting Experiments in Public Management Research: A Practical Guide. *International Public Management Journal* 18(2): 323-342.

Bellé, Nicola. 2013. "Experimental Evidence on the Relationship Between Public Service Motivation and Job Performance." *Public Administration Review* 73(1): 143-53.

Bellé, Nicola. 2014. "Leading to Make a Difference: A Field Experiment on the Performance Effects of Transformational Leadership, Perceived Social Impact, and Public Service Motivation." *Journal of Public Administration Research and Theory*, 24(1): 109-136.

Bellé, Nicola and Paola Cantarelli. 2015. "Monetary Incentives, Motivation, and Job Effort in the Public Sector: An Experimental Study With Italian Government Executives." *Review of Public Personnel Administration*, 35(2), 99 - 123

Bellé, Nicola and Paola Cantarelli. 2017. "What Causes Unethical Behavior? A Meta-Analysis to Set an Agenda for Public Administration Research." *Public Administration Review*, early view

Brewer, Gene A. 2003. Building Social Capital: Civic Attitudes and Behavior of Public Servants. *Journal of Public Administration Research and Theory*, 13(1): 5-26

Brewer, Gene A, and Sally Coleman Selden. 1998. "Whistle Blowers in the Federal Civil Service: New Evidence of the Public Service Ethic." *Journal of Public Administration Research and Theory* 8(3): 413-40.

Brewer, Gene A., and Sally Coleman Selden. 2000. Why Elephants Gallop: Assessing and Predicting Organizational Performance in Federal Agencies. *Journal of Public Administration Research and Theory* 10(4): 685-711.

Brewer, Gene A., Sally Coleman Selden and Rex Facer. 2002. Individual Conceptions of Public Service Motivation. *Public Administration Review* 60(3): 254-264

Brown, Roger. 1986. *Social Psychology: The Second Edition*. New York: The Free Press

Caillier, James Gerard. 2012. 'Agency Retaliation against Whistle-blowers', *Public Integrity*, 15(1): 29–49.

Caillier, James Gerard. 2015. "Transformational Leadership and Whistle-Blowing Attitudes Is This Relationship Mediated by Organizational Commitment and Public Service Motivation?" *The American Review of Public Administration* 45(4): 458–75.

Caillier, James Gerard. 2017. "Public service motivation and decisions to report wrongdoing in US Federal Agencies: Is this relationship mediated by the seriousness of the wrongdoing". *American Review of Public Administration*, 47(7), 810-825.

Christensen, Robert, Tobin, Im, Eunju Rho, Justin M. Stritch and Jungho Park. 2013. "Public Service Motivation, Task, and Non-task Behavior: A Performance Appraisal Experiment with Korean MPA and MBA Students." *International Public Management Journal*, 16(1): 28-52.

Christensen, Robert and Bradley Wright. 2018. "Public service motivation and ethical behavior: Evidence from three experiments." *Journal of Behavioral Public Administration*, 1(1), 1-8

Clerkin, Richard., Sharon Paynter and Jami Taylor. 2009. "Public Service Motivation in Undergraduate Giving and Volunteering Decisions." *American Review of Public Administration*, 39(6): 675–698.

Cohn, Alain, Ernst Fehr, and Michel André Maréchal. 2014. "Business Culture and Dishonesty in the Banking Industry." *Nature*, 516: 86–89.

Crewson, Philip. 1997. "Public-Service Motivation: Building Empirical Evidence of Incidence and Effect", *Journal of Public Administration Research and Theory*, 7(4): 499–518.

Dafoe, Allan, Zhang, Baobao and Devin Caughey (2016) "Confounding in Survey Experiments: Diagnostics and Solutions." Available at: <http://www.allandafoe.com/confounding>

DeSteno, David. 2015. "Compassion and altruism: how our minds determine who is worthy of help". *Current opinion in behavioral Sciences*, 3, 80-83.

Esteve, Marc, Arjen van Witteloostuijn, and George Boyne. 2015. "The Effects of Public Service Motivation on Collaborative Behavior: Evidence from Three Experimental Games." *International Public Management Journal*, 18(2): 171-89.

Esteve, Marc, Diemo Urbig, Arjen van Witteloostuijn, and George Boyne. 2015. "Prosocial Behavior and Public Service Motivation." *Public Administration Review*, 76(1): 177-187

Ethics Resource Center. 2005. *National Business Ethics Survey*. Washington, DC: Ethics Resource Center

Gächter, Simon and Jonathan F. Schulz (2016) "Intrinsic honesty and the prevalence of rule violations across societies." *Nature*, 531: 496-499

Grant, Adam. 2007. "Relational Job Design and Motivation to Make a Difference." *Academy of Management Review*, 32(2): 393-417.

Grant, Adam. 2008. "Does Intrinsic Motivation Fuel the Prosocial Fire? Motivational Synergy in Predicting Persistence, Performance and Productivity." *Journal of Applied Psychology*, 93(1): 48-58.

Hassan, Shahidul, Bradley E Wright, and Gary Yukl. 2014. "Does Ethical Leadership Matter in Government? Effects on Organizational Commitment, Absenteeism, and Willingness to Report Ethical Problems." *Public Administration Review*, 74(3): 333-43.

Henningsen, Arne and Jeff D. Hamann. 2007. "systemfit: A Package for Estimating Systems of Simultaneous Equations in R." *Journal of Statistical Software*, 23(4).

Hertz, Steven G., and Krettenauer, Tobias. 2016. "Does moral identity effectively predict moral behavior?": A meta-analysis. *Review of General Psychology*, 20(2), 129.

Houston, David J. 2006. "Walking the Walk of Public Service Motivation: Public Employees and Charitable Gifts of Time, Blood, and Money." *Journal of Public Administration Research and Theory*, 16(1): 67–86.

Huberts, Leo, and Gjalt De Graaf. 2008. "Portraying the Nature of Corruption Using an Explorative Case Study Design." *Public Administration Review*, 68(4): 640–53.

James, Oliver, Sebastian Gilke and Gregg Van Ryzin. 2017. *Experiments in Public Management Research: Challenges and Contributions*. Cambridge: Cambridge University Press

Kaptein, Muel, Leo Huberts, Scott Avelino, and Karin Lasthuizen. 2005. "Demonstrating Ethical Leadership by Measuring Ethics: A Survey of US Public Servants." *Public Integrity*, 7(4): 299–311.

Karl, Katherine A., and Barbara Peat. 2004. "A Match Made in Heaven or a Square Peg in a Round Hole? How Public Service Educators Can Help Students Assess Person–Environment Fit." *Journal of Public Affairs Education*, 10(4): 265–78.

Kim, Sangmook. 2005. "Individual-Level Factors and Organizational Performance in Government Organizations." *Journal of Public Administration Research and Theory* 15(2): 245–62.

Kim, Sangmook. 2012. "Does Person-Organization Fit Matter in the Public-Sector? Testing the Mediating Effect of Person-Organization Fit in the Relationship between Public Service Motivation and Work Attitudes." *Public Administration Review* 72(6): 830-840

Kim, Seung and Sangmook Kim. 2016. "Social Desirability Bias in Measuring Public Service Motivation." *International Public Management Journal*, 19(3), 293-319.

Kim, Sangmook, Wouter Vandenberghe, Bradley E Wright, Lotte Bøgh Andersen, Francesco Paolo Cerase, Robert K Christensen, Céline Desmarais, et al. 2013. "Investigating the Structure and Meaning of Public Service Motivation Across Populations: Developing an International Instrument and Addressing Issues of Measurement Invariance." *Journal of Public Administration Research and Theory*, 23(1): 79-102.

Kjeldsen, Anne. 2013. "Dynamics of Public Service Motivation: Attraction-Selection and Socialization in the Production and Regulation of Social Services." *Public Administration Review*, 74(1): 101-112.

Kolthoff, E., R. Erakovich and K. Lasthuizen. 2010. 'Comparative Analysis of Ethical Leadership and Ethical Culture in Local Government: The USA, The Netherlands, Montenegro and Serbia', *International Journal of Public Sector Management*, 23(7): 596-612.

Kwon, Illoong. 2014. "Motivation, Discretion, and Corruption." *Journal of Public Administration Research and Theory*, 24(3): 765-94.

Lim Choi, Do. 2004. "Public Service Motivation and Ethical Conduct." *International Review of Public Administration*, 8(2): 99-106.

Linos, Elizabeth. 2018. "More Than Public Service: A Field Experiment on Job Advertisements and Diversity in the Police." *Journal of Public Administration Research and Theory*, 28(1), 67-85

Maesschalck, Jeroen, Zeger van der Wal and Leo Huberts. 2008. "Public Service Motivation and Ethical Conduct." In James Perry and Annie Hondeghem (eds), *Motivation in Public Management: The Call of Public Service*. Oxford: Oxford University Press, 157–76

March, James G., and Johan P. Olsen. 1989. *Rediscovering Institutions: The Organizational Basis of Politics*. New York: Free Press.

McFarland, Sam. 1981. "Effects of Question Order on Survey Responses." *Public Opinion Quarterly*, 45(2): 208-215.

Menzel, Donald C. 2015. "Research on Ethics and Integrity in Public Administration: Moving Forward, Looking Back." *Public Integrity*, 17(4): 343–70.

Meyer-Sahling, Jan, Christian Schuster and Kim Mikkelsen. 2018. *Civil Service Management in Developing Countries: What Works? Evidence from a Survey of 23.000 Public Servants in Africa, Asia, Latin America and Eastern Europe*. Report prepared for the UK-Department for International Development (DFID), February 2018

Moore, David. 2002. "Measuring New Types of Question-Order Effects: Additive and Subtractive." *Public Opinion Quarterly*, 66(1): 80-91.

Naff , Katherine, and John Crum. 1999. "Working for America: Does Public Service Motivation Make a Difference?" *Review of Public Personnel Administration*, 19(4): 5–16.

Neumann, Oliver. 2016. "Does misfit loom larger than fit? Experimental evidence on motivational person-job fit, public service motivation, and prospect theory." *International Journal of Manpower*, 37(5): 822-839.

Oldendick, Robert. 2008. "Question Order Effects." In Lavrakas, Paul (ed.), *Encyclopedia of Survey Research Methods*. New York: Sage, 1-4

OPM (US Office of Personnel Management). 2012. *Federal Employee Viewpoint Survey*. Washington, DC: OPM. Retrieved from: <http://www.fedview.opm.gov/>.

Paulhus, Delroy (1984) "Two-component models of socially desirable responding." *Journal of Personality and Social Psychology*, 15: 383-388.

Pedersen, Mogens Jin. 2015. "Activating the Forces of Public Service Motivation: Evidence from a Low-Intensity Randomized Survey Experiment." *Public Administration Review*, 75(5): 734-46.

Perry, James L. 1996. "Measuring Public Service Motivation: An Assessment of Construct Reliability and Validity." *Journal of Public Administration Research and Theory* 6 (1): 5-22.

Perry, James L. 1997. "Antecedents of Public Service Motivation." *Journal of Public Administration Research and Theory*, 7(2): 181-197.

Perry, James L. 2000. "Bringing Society In: Toward a Theory of Public-Service Motivation." *Journal of Public Administration Research and Theory*, 10(2), 471-488

Perry, James L. 2015. "Revisiting the Core of Our Good Government Ethos." *Public Administration Review*, 75(2): 186-187.

Perry, James L. and Lois Wise. 1990. "The Motivational Bases of Public Service." *Public Administration Review*, 50(3): 367-373.

Perry, James L, and Annie Hondeghem. 2008. *Motivation in public management: The Call of Public Service*. Oxford, UK: Oxford University Press.

Perry, James L, and Wouter Vandenabeele. 2015. "Public Service Motivation Research: Achievements, Challenges, and Future Directions." *Public Administration Review* 75 (5): 692–99.

Prebble, Mark. 2016. "Has the Study of Public Service Motivation Addressed the Issues That Motivated the Study?" *The American Review of Public Administration*, 46(3): 267-291.

Rainey, Hal. 1982. "Reward Preferences of Public and Private Managers: In Search of the Service Ethic." *American Review of Public Administration*, 16(4): 288–302.

Reynolds, Scott, and Tara Cerani. 2007. "The Effects of Moral Judgment and Moral Identity on Moral Behavior: An Empirical Examination of the Moral Individual." *Journal of Applied Psychology* 92: 1610–1624.

Ritz, Adrian, Gene A. Brewer and Oliver Neumann. 2016. "Public Service Motivation: A Systematic Literature Review and Outlook." *Public Administration Review*, 76(3): 414-426.

Rosseel, Yves. 2012. "lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software*, 48(2): 1–36.

Schott, Carina, Oliver Neumann, Muriel Baertschi and Adrian Ritz. 2017. "Public service motivation, prosocial motivation altruism and prosocial behavior: Towards disentanglement and conceptual clarity." *Conference paper for the Annual Conference of the International Research Society for Public Management (IRSPM)*, Budapest, Hungary,

Stazyk, Edmund C, and Randall S Davis. 2015. "Taking the 'High Road': Does Public Service Motivation Alter Ethical Decision Making Processes?" *Public Administration*, 93(3): 627–45.

Strack, Fritz. 1992. "Order Effects in Survey Research: Activation and Information Functions of Preceding Questions." In Schwarz, Norbert and Seymour Sudman (Eds.) *Context Effects in Social and Psychological Research*. New York: Springer, 23-34

Stets, Jan and Peter J. Burke. 2000. Identity Theory and Social Identity Theory. *Social Psychology Theory*, 63(3), 224-237

Tajfel, Henri, and John Turner. 2004. "The social identity theory of intergroup behavior". In S. Worchel and W. Austin (Eds.) *Psychology of Intergroup Relations*. Chicago: Nelson-Hall

Tepe, Markus. 2016. "In Public Servants We Trust?: A behavioural experiment on public service motivation and trust among students of public administration, business sciences and law." *Public Management Review*, 18(4), 508-538

Thaler, Julia and Bernd Helmig. 2016. "Do Codes of Conduct and Ethical Leadership influence Public Employees' Attitudes and Behaviours? An Experimental Analysis" *Public Management Review*, 18(9), 1365–1399.

Transparency International. 2017. "Corruption Perceptions Index." Available at https://www.transparency.org/news/feature/corruption_perceptions_index_2016

Uziel, Liad. 2000. Rethinking Social Desirability Scales: From Impression Management to Interpersonally Oriented Self-Control. *Perspectives on Psychological Science*, 5(3): 243-262.

Van de Walle, Steven and Van Ryzin, Gregg. 2011. "The Order of Questions in a Survey on Citizen Satisfaction with Public Services: Lessons from a Split-Ballot Experiment" *Public Administration*, 89(4), 1436–1450.

Vandenabeele, Wouter. 2007. "Toward a Public Administration Theory of Public Service Motivation." *Public Management Review*, 9(4), 545-556

Vandenabeele, Wouter and James Perry. 2008. "Behavioral Dynamics: Institutions, Identities and Self-Regulation." In: Perry, J. and Hondeghem A. (Eds). *Motivation in public management : the call of public service*. pp. 56-79. Oxford: Oxford University Press

Vandenabeele, Wouter, and Anne Kjeldsen. 2011. "The Relationship between Public Service Motivation and Whistle-blowing intentions: Interplay of Individual and Structural Elements." *Conference Paper for the European Group for Public Administration (EGPA) Conference*, Bucharest, Romania

Vigoda-Gadot, Eran. 2007. "Citizens' Perceptions of Politics and Ethics in Public Administration: A Five-Year National Study of their Relationship to Satisfaction with Services, Trust in Governance, and Voice Orientations." *Journal of Public Administration Research and Theory*, 17(2), 285–305.

Vogel, Dominik and Alexander Kroll. 2016. "The Stability and Change of PSM-Related Values Across Time: Testing Theoretical Expectations Against Panel Data." *International Public Management Journal*, 19:1, 53-77

Wang, Long and Keith Murnighan. 2011. "On Greed." *Academy of Management Annals*, 5(1): 279–316.

Winterich, Karen Page, Vikas Mittal, and Andrea C. Morales. 2014. "Protect Thyself: How Affective Self-Protection Increases Self-Interested, Unethical Behavior." *Organizational Behavior and Human Decision Processes*, 125(2): 151 – 61

Wright, Bradley and Adam Grant. 2010. "Unanswered Questions about Public Service Motivation: Designing Research to Address Key Issues of Emergence and Effects." *Public Administration Review*, 70(5): 691-700

Wright, Bradley, Shahidul Hassan, and Jongsoo Park. 2016. "Does a Public Service Ethic Encourage Ethical Behaviour? Public Service Motivation, Ethical Leadership and the Willingness to Report Ethical Problems." *Public Administration*, 94(3): 647-663.

Appendix

Appendix 1: Survey Representativeness

	All public employees (in central government)	All public employees (in sampled institutions)	Respondents (in sampled institutions)
Sex			
Female	58%	55%	57%
Male	42%	45%	43%
Age			
29 or less	18%	7%	21%
30-39	30%	30%	36%
40-49	27%	36%	24%
50-59	16%	18%	14%
60 and more	10%	8%	6%
Education			
High School	n/a	23%	21%
Vocational degree	n/a	9%	29%
University degree	n/a	67%	50%

Sources: Dirección de Presupuestos del Ministerio de Hacienda (2017); institutional-level data shared by institutions with authors

Table A1.1: Distribution of sampling frame and sample on demographic variables

Appendix 2: Spanish Translation of the PSM items

Item ID	Survey item
<i>Attraction to public service</i>	
APS1	Admiro a la gente que emprende o está involucrada en actividades que ayudan a mi comunidad
APS2	Es importante contribuir a actividades que enfrentan a problemas sociales
APS3	Un trabajo público con sentido es muy importante para mí
APS4	Es importante para mí contribuir al bien público
<i>Commitment to public values</i>	
CPV1	Pienso que la igualdad de oportunidades para ciudadanos es muy importante
CPV2	Es importante que los ciudadanos puedan confiar en la provisión continua de servicios públicos
CPV3	Es fundamental que los intereses de generaciones futuras sean tomados en consideración, al momento de formular políticas públicas
CPV4	Actuar éticamente es esencial para los funcionarios públicos
<i>Compassion</i>	
COM1	Siento compasión por las dificultades de los desfavorecidos
COM2	Empatizo con las personas que enfrentan dificultades
COM3	Me molesta mucho cuando veo que otra gente es tratada injustamente
COM4	Considerar el bienestar de otros es muy importante para mí
<i>Self-sacrifice</i>	
SS1	Estoy preparado para hacer sacrificios por el bien de la sociedad
SS2	Creo en priorizar el deber cívico por sobre mí mismo
SS3	Estoy dispuesto a correr el riesgo de pérdidas personales para ayudar a la sociedad
SS4	Estaría de acuerdo con un buen plan para mejorar el estándar de vida de los pobres, incluso si eso me cuesta dinero

Table A2.1: Spanish Translation of the PSM items

Appendix 3. Balance Test: Treatment vs. Control Group

We report a balance test of our treatment assignment in figure A3.1 below. The figure shows estimates from a logit model, regressing years of service, age, education, and gender on treatment assignment. No coefficient is statistically significant despite substantial power, indicating that our random assignment was successful.

<Appendix Figure 3 about here>

Figure A3.1: Balance test on observed covariates

Table A3.1 and A3.2 below shows that none of the demographic variables are affected by the treatment. On the other hand, PSM – our moderator – is affected by the treatment (albeit with a substantively minimal difference). We address this potential problem in a system of equations discussed in the paper (see in particular, footnote 31, see also Appendix 6).

	Control group Mean (standard deviation)	Treatment group Mean (standard deviation)	F-test (p value)
PSM	-0.008 (0.256)	0.009 (0.215)	5.979 (0.014)
Willingness to report	2.240 (1.330)	2.440 (1.260)	28.23 (0.000)

Table A3.1: Means and standard deviations for quantitative variables by treatment assignment

	Control group	Treatment group	Chi squared test (p value)
Gender	Female: 1.297 Male: 1.116	Female: 1.243 Male: 1.057	0.030 (0.863, df = 1)
Age	29 or less: 454 30-39: 834 40-49: 629 50-59: 356 60 or above: 156	29 or less: 454 30-39: 824 40-49: 575 50-59: 334 60 or above: 133	2.514 (0.642, df = 4)

Table A3.2: Distributions for qualitative variables by treatment assignment

Appendix 4. Regression Results

Table A4.1 below reports the results for (i) the models we report graphically in the paper to facilitate future meta-analyses; and (ii) the additional models mentioned in the paper.

Model 1 shows the estimated ATE for our treatment in a regression framework. This model uses only respondents with complete responses to the outcome question and the PSM battery. Model 2 shows the estimated ATE for all respondents responding to the outcome question. Model 3 expands Model 1 by including demographic variables in a control function.

Model 4 shows the estimates from the spline model reported in figure 2 of the main paper. Note the spline coefficients can be more meaningfully interpreted graphically. As the PSM construct is right skewed, Model 5 replicates Model 4 using robust regression to avoid outliers influencing results.

	Model 1 (Figure 1)	Model 2 (Full sample)	Model 3 (Including controls)	Model 4 (Figure 2)	Model 5 (Robust regression)
Intercept	2.238 (0.026)	2.228 (0.026)	2.158 (0.056)	2.571 (0.678)	2.744 (0.457)
Treatment	0.200 (0.037)	0.206 (0.037)	0.202 (0.038)	-0.064 (0.053)	-0.084 (0.055)
Gender: Male			0.219 (0.040)		
University education			-0.100 (0.040)		
Years of service			0.006 (0.002)		
Age: 29 or less			(reference)		
Age: 30-39			-0.128 (0.055)		
Age: 40-49			-0.009 (0.063)		
Age: 50-59			-0.071 (0.079)		
Age: 60 or above			-0.011 (0.109)		
PSM				0.697 (0.101)	0.781 (0.121)
Spline Coefficient (1)				-0.386 (0.363)	-0.480 (0.288)
Spline Coefficient (2)				0.015 (1.371)	-0.094 (0.890)
Spline Coefficient (3)				0.222 (0.361)	0.197 (0.286)
N	4.763	5.050	4.687	4.763	4.763

Table A4.1: Regression estimates

Appendix 5. Treatment Effect on Socially Desirable Question by Level of PSM

To test whether our treatment affects our respondents' general tendency to provide socially desirable answers, we estimate an ATE on a placebo item in the main text. Figure A5.1 below plots the predicted scores for respondents on the PSM construct (mean centred, SD = 0.237) against the predicted treatment effect of our question order treatment (using 3 natural splines to account for non-linear patterns). The result is that our treatment does not affect the placebo item for any level of PSM.

<Appendix Figure 5 About Here>

Figure A5.1: Treatment Effect on Socially Desirable Question by Level of PSM

Appendix 6. Estimate of Treatment Effect (Activating PSM) by Level of PSM, Accounting for Effect of Asking about Ethical Reporting on Level of PSM

As noted in the paper (in particular footnote 31, see also appendix 3), we estimate an effect of our treatment on our moderator. In other words, asking about ethical reporting influences PSM, though the effect is of marginal size (the OLS estimate in a system of equation is 0.017, $p_{\text{two-sided}} = 0.015$).

To assess whether our treatment effect on ethical reporting is affected by this potential threat to validity, figure A6.1 below shows the result of a system of equations accounting for this effect. The model replicates the model reported in figure 2 of the paper but adds an effect of the treatment on PSM. Hence, the treatment is allowed to affect the moderator, as well as affect the outcome variable (with moderation). Figure A6.1 replicates figure 2 using this model – plotting predicted PSM scores (mean centred, $SD = 0.237$) against the predicted effect of our treatment. Our core findings about the effect of PSM activation remain robust (though the estimate from the system of equations is less efficient).

<Appendix figure 6 about here>

Figure A6.1: Estimate of Treatment Effect by Level of PSM, Accounting for the Effect of Asking about Ethical Reporting on PSM

Appendix 7. Sensitivity of Results to Exclusion of Specific Institutions or Demographic Groups

Our analyses in the paper do not account for the hierarchical structure of our data. As our treatment is exogenous, this is not strictly necessary. However, we did implement our model using a fixed effects specification (using standard errors clustered by institution). The ATE in this model is 0.198 ($p_{\text{two-sided}} < 0.001$). A Tukey honestly significant differences test reveals that no two institutions are significantly different on our outcome variable regardless of our treatment.

To assess the validity of our finding across our sample of 11 institutions, we re-estimated an ATE which excluding institutions one at a time. The result is shown in figure A7.1. Each estimate in the figure represents the ATE for a model excluding the institution shown on the first axis. Figure A7.2 shows the result from a similar analysis excluding respondents by age group. Figure A7.3 shows the results from a similar analysis excluding respondents by education. In all instances, the ATE remains positive and significant, each analysis revealing only small differences to the original full-sample ATE.

<Appendix Figure 7A About Here>

Figure A7.1: Sensitivity of results to exclusion of individual institutions

<Appendix Figure 7B About Here>

Figure A7.2: Sensitivity of results to exclusion of respondents by age group

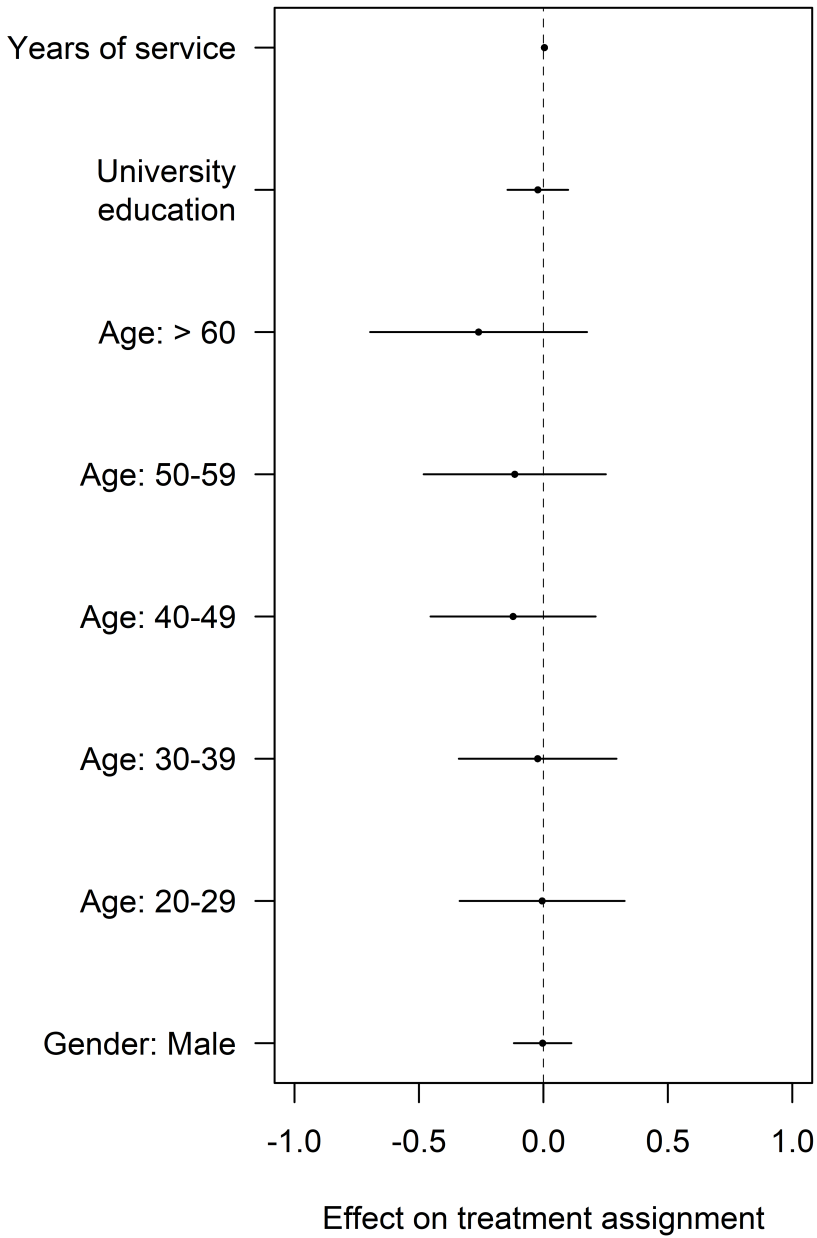
<Appendix Figure 7C About Here>

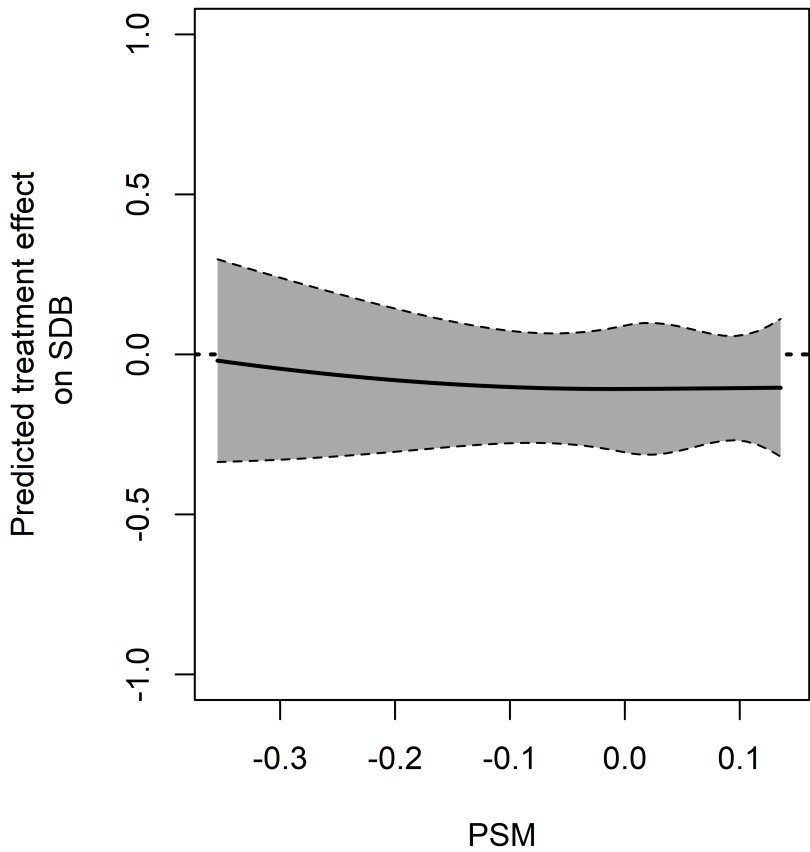
Figure A7.3: Sensitivity of results to exclusion of respondents by education

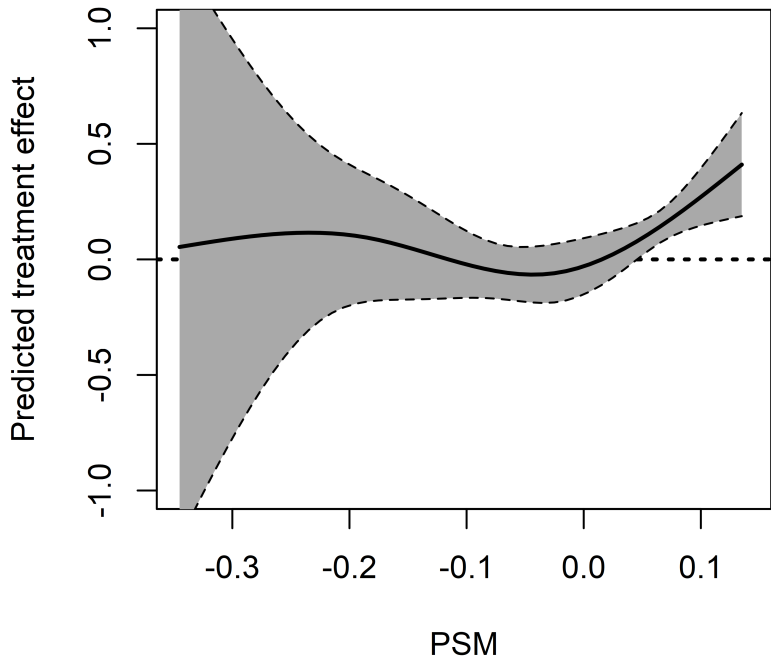
Appendix 8. Power Analysis

<Appendix Figure 8 About Here>

As discussed in the paper, we are able to detect our effect in part due to the size of our sample. The power of our analysis is substantially higher than in most current experimental public administration scholarship. Figure A8.1 shows power analyses for the effect of PSM activation on willingness to report. It underscores that studies with smaller samples might not have detected the effect. For instance, a study with 1,000 respondents would commit a Type II error in about 35 percent of samples. To reach a power level of 0.95, a study would have to include 2,098 respondents. This also implies that any replication of our study would require another large sample of public servants.







Treatment effect
omitting institutions

