PUBLIC MANAGEMENT REVIEW
https://doi.org/10.1080/14719037.2024.2322159

The management of bias and noise in public sector decision-making: experimental evidence from healthcare

Nicola Belle[a], Paola Cantarelli[a] and Sophie Y. Wang[b]

[a]Management and Healthcare Laboratory, Institute of Management, Scuola Superiore Sant'Anna, Pisa, Italy; [b]Health Division, Directorate for Employment, Labour, and Social Affairs, Organization for Economic Cooperation and Development, Paris, France

ABSTRACT

In six randomized online experiments with 2,647 medical doctors we test whether – depending on the choice architecture – physicians engaged in prescribing decisions in public organizations fall prey to systematic error (i.e. bias) and make significantly different choices when faced with the same clinical case (i.e. level noise). Results show that experts tend to make irrational choices that are influenced by outgroup bias, social comparison, past behaviour, confirmation bias, loss aversion, equivalence framing, and asymmetric dominance. We also find evidence of significant level noise, that is, between-prescriber variability, with the distribution of responses differing remarkably across experimental arms.

ARTICLE HISTORY Received 4 July 2023; Accepted 19 February 2024

KEYWORDS Decision-making; bias; noise; nudge theory; public employees

> Our topic is human error. Bias and noise – systematic deviation and random scatter – are different components of error. (Kahneman, Sibony, and Sunstein 2021, 9)

Introduction

When doctors prescribe diagnostic tests and drugs in public healthcare organizations, does the decision-making context – that is, the architecture of prescription choices – influence the likelihood that they will fall prey to systematic bias or that different physicians will opt for significantly different solutions in identical clinical cases? Indeed, a prescribing decision is quintessentially a matter of judgement, namely, 'one with some uncertainty about the answer and where we allow for the possibility that reasonable and competent people might disagree' (Kahneman, Sibony, and Sunstein 2021, 44). A few decades of behavioural science research have revealed that experts, like all human beings, are prone to judgement error, in the form of systematic deviation from rationality (i.e. bias) or unwanted variability (i.e. noise). The investigation of these phenomena embodies the essence of Behavioural Public Administration

CONTACT Paola Cantarelli paola.cantarelli@santannapisa.it

(Grimmelikhuijsen et al. 2017), a burgeoning movement within our discipline in recent years. Research in this realm – rooted in seminal works on bounded rationality (Simon 1947), heuristics and biases (Kahneman and Tversky 1979; Tversky and Kahneman 1974), as well as nudging and choice architecture (Thaler and Sunstein 2008) – has spanned diverse decision domains, units of analysis, and countries (Battaglio et al. 2019). Its significance extends beyond public management, as evidenced by contemporary research on the instability and heterogeneity of beliefs spanning across decision domains (Bordalo et al. 2023).

Studying bias and noise in physician prescription decisions significantly impacts crucial dimensions essential for effective public service management. Prioritizing the adept handling of undesired variation is vital for upholding impartiality and equity in public service accessibility, core tenets of public bureaucracies (Cepiku and Mastrodascio 2021; Frederickson 1971, 2005; Rhys, Beynon, and McDermott 2019; Rivera and Connolly Knox 2023). Addressing and mitigating biases and inconsistencies in prescription practices is imperative to prevent the misallocation of public resources funded by taxation. In 2022, health expenditure by government schemes in Italy accounted for 7.1% of the gross domestic product (Organization for Economic Cooperation and Development. 2022). Our research demonstrates that decisions on drug prescriptions, like those under study, contribute significantly to public pharmaceutical expenditure, comprising approximately 19% of the *Fabbisogno Sanitario Nazionale (FSN)*–the national healthcare system budget (The Medicines Utilisation Monitoring Centre 2023). Moreover, the presence of biases and inconsistencies in prescription practices can have detrimental effects on public health, particularly through the emergence of antibiotic resistance – a substantial threat to humanity. According to the European Centre for Disease Control, 'more than 35,000 people die from antimicrobial-resistant infections in the EU/EEA each year. The health impact of antimicrobial resistance (AMR) is comparable to that of influenza, tuberculosis and HIV/AIDS combined' (European Centre for Disease Control 2022).

Behavioural research has long neglected noise and focused mostly on one or a few cognitive biases at a time, for a number of applications in healthcare (Blumenthal-Barby and Krieger 2015; Freeman, Robinson, and Scholtes 2021; Ganju et al. 2020; Levy Andrea and Hershey 2006; Meeker et al. 2016) and public management (Dinhof et al. 2023; Lee and Kim 2023; Liu, Qin, and Zhang 2022; Marvel 2015; Nagtegaal et al. 2019, 2020; Roberts and Wernstedt 2019; Sheeling et al. 2023; Walter and Barton Cunningham 2023; Weißmüller 2022). Kahneman et al. (2021) have recently urged to adopt a broader view of decision-making by considering noise alongside bias, because both contribute to the overall error. Recognizing that 'wherever the person making a judgment is randomly selected from a pool of equally qualified individuals [. . .] noise is a problem' (Kahneman, Sibony, and Sunstein 2021, 29), investigating judgement heterogeneity across public sector settings is imperative. In most cases, there is randomness in the allocation of judicial cases to judges, pupils to teachers, patients accessing emergency rooms to clinicians and nurses, and offenders to police officers. In this perspective, we study the causal impact that specific features of the context in which public sector doctors make prescribing decisions – i.e. the architecture of prescribing choices – have on systematic error and unwanted between-subject variability. Our work explicitly adopts a micro-level perspective with attention to several psychological aspects of experts' judgements. This choice aligns nicely with the foundations of behavioural public administration (Battaglio et al. 2019; Grimmelikhuijsen et al. 2017), and the latest developments in the study of unwanted variability within and between decisions made by professionals, street level bureaucrats included (Kahneman, Sibony, and Sunstein 2021). Although falling outside the scope of the current project due to a different research focus and low feasibility, estimating the impact of different political and institutional forces on decision-making holds value and requires a dedicated endeavour (A. M. Bertelli and Riccucci 2022; A. Bertelli et al. 2022). We selected experimental contexts and interventions jointly with our partner government institution to maximize relevance for management implications, timeliness in tackling real-world challenges, and ecological validity.

Theoretical background

Prescribing decisions are a prime example of mental operations called judgements (Blumenthal-Barby and Krieger 2015; Freeman, Robinson, and Scholtes 2021; Kahneman, Sibony, and Sunstein 2021). Kahneman et al. (2021) argue that 'some judgments are biased; they are systematically off target. Other judgments are noisy, as people who are expected to agree end up at very different points around the target' (Kahneman, Sibony, and Sunstein 2021, 9). Indeed, studies based on behavioural economics demonstrate that experts, like all human beings, systematically deviate from rational decisions and show patterns of unwanted variability (Kahneman, Sibony, and Sunstein 2021; Thaler and Sunstein 2008). This perspective explains the empirically observable violations of predictions based on rational choice models, which were the gold standard before the advent of behavioural economics. The decisions of politicians who are elected to serve the public interests, public managers called to implement government policies, and physicians working in public organizations are no exception to this phenomenon (Baekgaard et al. 2019; Belle and Cantarelli 2021; Belle et al. 2022; Javier, Van Ryzin, and Leth Olsen 2021).

Whereas research on heuristics and associated biases has focused on systematic deviations from rationality, more recent work by Kahneman, Sibony, and Sunstein (2021) has explored unwanted variability in professional judgement, the so-called noise. According to their reasoning, bias and noise are two independent components of overall error. Specifically, bias is the average of errors and noise is the variability of errors. The authors distinguish between level noise, which is the focus of our study, and pattern noise. Level noise is defined as the between-person variability that we observe, for instance, when different prescribers make significantly different decisions regarding the same clinical case. Examples of such variability can be found in different recommendations for back surgery to the same patient where some orthopaedists are more aggressive than the average orthopaedist, or in ordering painkillers for the same clinical case where some doctors are more or less generous than the average doctor. Pattern noise, instead, refers to within-person variability in decisions made by the same physician across cases. Pattern noise can be due to work overload, which has recently received attention in public management studies (Grima, Georgescu, and Prud'homme 2020).

Interestingly, bias and noise can sometimes be reduced through changes in the decision-making context. Choice architecture interventions are native to nudge theory (Thaler and Sunstein 2008), which advances the work of Daniel Kahneman and Amos Tversky by illuminating how we can predictably alter people's behaviour through changes in the decision environment that do not involve prohibitions or powerful incentives and that may even exploit systematic biases to lead to better decisions. Insights from nudge theory have now been applied across decision domains (Barnes et al. 2021; Belle and Cantarelli 2021; Belle et al. 2022; Beshears and Kosowsky 2020). Nudge theory portrays individuals – health professionals included – as *Humans* rather than *Econs*. The design of the context in which *Humans* make decisions is not neutral because, unlike *Econs*, *Humans* do not always have unbiased expectations, are not optimizers, lack self-control, and are not exclusively moved by self-interest.

Drawing on these insights, we test whether and to what extent in-group favouritism, past behaviour, confirmation bias, loss aversion, equivalence framing, and asymmetric dominance can cause bias and noise in public physicians' prescribing in the context of public organizations. Starting with in-group favouritism, social identity theory (Turner and Reynolds 2001) posits that individuals tend to conform to the behaviours of others with whom they identify (in-group members) and differentiate from those with whom they do not identify (out-group members). Thus, physicians may follow a prescription suggestion if made by in-group actors but not follow the same recommendation put forth by individuals they consider to be out-group.

The status quo bias represents the propensity to disproportionately stick to the current situation rather than moving away from it, even when alternative options are superior (Arad 2013; Jilke, Van Ryzin, and Van de Walle 2016; Samuelson and Zeckhauser 1988). As diverting from the status quo seems to be considered risky, how doctors prescribed in the past serves as an anchor for future prescriptions. The impact of past events on decisions in the present has received attention in public management (Hong, Hyoung Kim, and Son 2020; Olsen 2017), although in domains different from the one we are studying. With analogous logic, existing impressions trigger the confirmation bias, a phenomenon whereby decision makers selectively search for or interpret information that confirms their prior held beliefs while simultaneously ignoring any disconfirmatory evidence (Baekgaard and Serritzlew 2016; Nickerson 1998). This may in turn nurture overconfidence, which has been observed in many professional fields, including medicine (Baumann, Deber, and Thompson 1991; Christensen-Szalanski and Bushyhead 1981).

The popular 'Asian disease' problem developed by Tversky and Kahneman (1981) exemplifies reference dependence and loss aversion, which – along with diminishing sensitivity and probability weighting – are the core components of prospect theory (Barberis 2013). The principle of reference dependence posits that individuals derive utility from changes relative to some reference point rather than from the absolute value of their wealth. Loss aversion, then, assumes that individuals are more sensitive to losses than to equivalent gains. In other words, individuals tend to be risk averse in the realm of gains and to become risk seekers in the realm of losses. Extant scholarship has pointed to prospect theory as a relevant source for expanding theory in such diverse domains as public administration (Battaglio et al. 2019; Weißmüller, Bouwman, and Vogel 2023), management studies (Cornelissen and Durand 2014; Holmes et al. 2011), and justice (Ganegoda and Folger 2015). Within the scholarship that applies prospect theory across decision domains, the study of how the framing of information influences medical decisions is neither novel nor lagging other professions, though its potential impact on management practices has not yet been fully unleashed (Krishnamurthy, Carter, and Blair 2001). Pioneering work conducted at the Harvard Medical School shows that doctors are more likely to surgically intervene on a patient with lung cancer when they are presented with a 90% chance of survival rather than with the equivalent 10% probability of death (McNeil et al. 1982). Recent systematic reviews report a prevalence of the framing effect among the most common cognitive biases associated with medical decisions (Blumenthal-Barby and Krieger 2015; Gustavo et al. 2016). Well-renowned public management research focused on the impact of equivalence framing across typology of individuals and decision domains has followed that pioneer work (Bellé, Cantarelli, and Belardinelli 2018; Cantarelli, Belle, and Belardinelli 2020; Olsen 2015; Walter and Barton Cunningham 2023).

The last choice architecture intervention that we test is asymmetric dominance, a phenomenon whereby one of two options becomes more popular when a decoy – i.e. a third alternative dominated by that option – is added to the choice set (Joel, Payne, and Puto 1982; Tversky and Simonson 1993). In the simplest case, the target and the competing alternative are compared along two attributes, typically calling for a trade-off. The decoy is equal to the target in one attribute and slightly inferior to the target in the other, or slightly inferior in both. Meanwhile, the decoy is inferior to the competing alternative in one attribute but superior to it in the other. Although the decoy is virtually never selected due to its characteristics compared to the target, it alters individuals' preferences between the other options in the choice set. In a recent test of the asymmetric dominance effect in public healthcare management, the presence of a decoy remarkably reversed choices made by a large sample of nurses in the simulated purchase of one diagnostic instrument (Cantarelli, Belle, and Belardinelli 2020).

Method

Our study consists of six online randomized experiments with 2,647 physicians working for a regional government. Randomized experiments are the gold standard for avoiding the threat of self-selection of subjects into experimental conditions, thus ensuring high internal validity of inference. Based on the taxonomy provided by Harrison and List (2004), our study qualifies as a framed field experiment because the subject pool consists of experts belonging to the actual population of interest – as opposed to students or MTurkers – and the task and information set include field context that was validated as realistic by managers of the partner public institution. Hence, our work follows up on the methodological call to avoid general samples when decisions require professional expertise (Xiaoli, Wang, and Hao 2022).

The content (i.e. typologies of prescriptions) and manipulations (i.e. behaviourally inspired changes to the architecture of choices) of our experiments were designed jointly with the partner institution (i.e. a Regional Government in Italy responsible for organizing and managing health for more than 3.5 million inhabitants) to maximize the relevance of our inference to management interventions, timeliness in tackling real-world pressing challenges, and ecological validity in participants' eyes. Given the content, manipulations, and settings of our study, estimating the impact that factors such as political and institutional forces have on our outcomes is problematic. For instance, system-level elements do not vary because the study is set in one regional government only where the same regulations apply to all institutions and organizational-level components could not be measured or manipulated.

Physicians in our study were asked to imagine that they had just seen a patient and were uncertain about a prescription decision. The content of the prescription could be a test, a drug, or antibiotics. We did not provide any details about the type of test, drug, or antibiotics in order to enhance generalizability of the findings. The baseline scenarios that were presented to participants in the control arms of the six experiments, separately for each typology of prescription unless otherwise specified, are as follows. As to the prescription of a test: '*Imagine having to determine which diagnostic tests to prescribe for a patient you are seeing. You have undoubtedly decided to prescribe two tests that you consider indispensable.* You are undecided whether or not to prescribe a third test. [Experimental treatment, if any]. How likely would you be to prescribe the third test?' As to drug: 'Imagine that you have just finished seeing a patient and have to decide whether to prescribe a drug. This drug is certainly appropriate, but you wonder if perhaps there might be another slightly more effective drug. Identifying this alternative drug would take you some time and would not allow you to prescribe immediately. Postponing the prescription would not pose any risk to the patient; however, the patient should return to you in the next few days. [Experimental treatment, if any]. How likely would you be to prescribe the drug immediately?' As to antibiotic: 'Imagine that you have just finished seeing a patient

Table 1. Experimental procedure.

Experiment	Control Treatment(s)	Outcome
In-group and Out-group	C: Baseline scenario ($n = 135$) T: *A pharmacist from your organization suggests the drug* ($n = 168$) T: *A pharma representative suggests the drug* ($n = 166$) T: *A colleague suggests the drug* ($n = 153$)	Probability of prescribing drug
Past behaviour	C: Baseline scenario ($n = 411$) T: *In the past, in similar circumstances, you usually did not prescribe the test* ($n = 400$) T: *In the past, in similar circumstances, you have usually prescribed the test* ($n = 414$)	Probability of prescribing test
Confirmation bias	C: Baseline scenario ($n = 411$) T: *To prescribe the test it is necessary to type a reason in the electronic medical record* ($n = 421$)	Probability of prescribing test
Loss aversion	C: Baseline scenario ($n = 503$) T: *Data from the European Center for Disease Prevention and Control show that antibiotic resistance caused 671,689 infections and 33,110 deaths in Europe in 2015* ($n = 508$) T: *Each year, the inappropriate prescription of antibiotics causes a loss of many millions of euros, which are subtracted from patient care and clinical research* ($n = 501$) T: *Each year, the appropriate prescription of antibiotics causes a gain of many millions of euros, which are allocated to patient care and clinical research* ($n = 510$)	Probability of prescribing antibiotics
Equivalence framing	C: Baseline scenario ($n = 135$) T: *Based on the scientific evidence, there is a 90% chance that the drug will be effective in this case* ($n = 114$) T: *Based on the scientific evidence, there is a 10% chance that the drug will not be effective in this case* ($n = 128$)	Probability of prescribing drug
Asymmetric dominance	C: *Imagine having to decide which of the following drugs to prescribe. The drugs differ from each other only in the expected effectiveness and in the number of undesirable effects (common or very common).* *Drug X – 95% efficacy; 5 side effects.* *Drug Y – 97% efficacy; 10 side effects.* ($n = 80$. Drug Y serves as the target dug) T: Same as control, with the addition of one of the following decoy options • *Drug W – 96% efficacy; 11 side effects.* ($n = 94$) • *Drug W – 96% efficacy; 10 side effects.* ($n = 81$) • *Drug W – 97% efficacy; 11 side effects.* ($n = 89$)	Choice of drug to prescribe

and are very undecided whether or not to prescribe an antibiotic. [Experimental treatment, if any]. How likely would you be to prescribe an antibiotic immediately?'

Table 1 presents the procedures for each randomized experiment. All trials feature the manipulation of one factor. Subjects in the control arms of experiments 2 and 3 are the same pool. Similarly, experiments 1 and 5 share the same control group. Physicians were prompted to indicate on a scale from 0 to 100 the probability with which they would prescribe the test, drug, or antibiotics. The dependent variable in Randomized Experiment 6, in which clinicians are asked to select a drug to prescribe from a set of drugs, is the only exception. Therefore, the outcome variable is a behavioural intention in Experiments 1 through 5 and a simulated choice in Experiment 6.

As previously explained, through a descriptive rather than normative approach, our study also focuses on level noise. Due to the recency of this construct, no established measure exists that we can use as a general standard. Kahneman et al. (2021) state that 'in statistics, the most common measure of variability is standard deviation' (42). Other viable options might be the kappa statistic, which measures interrater reliability, with level noise being lower when the value of the kappa statistic is higher. In our comprehensive analyses, the representation of level noise will be encapsulated by the nuanced depiction of both the range and interquartile distribution of responses. We firmly believe that our approach, extending beyond the conventional reliance on standard deviation, offers a more holistic comprehension of level noise. By considering not only the dispersion of values but also incorporating the broader context of the range, our measurements afford a more nuanced and insightful examination of the intricacies inherent in level noise. In the forthcoming results section, we will use box plots as a visually compelling and statistically rigorous graphical representation of the identified level noise. These box plots, by illustrating the quartiles, median, and the variability within the dataset, will provide a clear and accessible means for our readers to grasp the intricacies of level noise across the spectrum of responses. This graphical representation is poised to enhance the interpretability of our findings, allowing for a more intuitive understanding of the distributional characteristics and patterns associated with the observed level noise in our study. In short, then, we are convinced that, compared to standard deviation alone, our measurements provide a more comprehensive understanding of level noise because they consider not only the dispersion of values, but also the range.

To limit cognitive fatigue, physicians were randomly assigned to no more than five scenarios. Vignettes were presented in a random sequence to guard against any order effect. Of the full sample, 51% were exposed to one scenario only, 22% to two scenarios, 19% to three, 7% to four, and 1% to five. Due to this peculiar characteristic, 'n' represents the number of responses, which does not necessarily correspond to the number of respondents. To account for the panel nature of the dataset in some instances, our empirical strategy for analysing data is twofold. Specifically, within each of our six experiments, we fit an analysis of variance (ANOVA) whenever medical doctors were exposed to only one condition. Otherwise, when physicians could have been exposed to more than one condition, we accommodate panel data by fitting a random-effects generalized least square (GLS) regression.

Results

Of the entire sample of 2,647 physicians, about 45% are female, 43% male, and 12% preferred not to say. As to age, about 14% of doctors are between 30 and 39 years old, 24% between 40 and 49, 28% between 50 and 59, 22% 60 or older, and the remaining 12% did not indicate any age group. About 44% of participants work in hospitals, 17% in ambulatory care settings, 26% in teaching hospitals, and 13% did not provide information about the type of organization in which they work. The remainder of this section presents the results of a series of analyses of variance on the experimental data from our six randomized controlled trials.

Randomized experiment 1 – in-group and out-group

In Randomized Experiment 1, the organization's pharmacist and a colleague are meant to qualify as in-group members and the pharma representative as an out-group member. Based on analysis of variance with Scheffe corrections, in the pharma representative arm, the average probability of prescribing is lower by 9.59% points than in the control ($p = .057$), lower by 10.07% points than in the organizational pharmacist arm ($p = .026$), and lower by 9.79% points than in the organizational colleague arm ($p = .039$). Figure 1 provides a graphical representation of the average treatment effects. The boxplots in Figure 1 provide a graphical representation of the distribution of responses, separately for each experimental arm. The considerable variability between-physicians points to the presence of what Kahneman, Sibony, and Sunstein (2021) define as level noise. This phenomenon is demonstrated by responses covering the entire probability spectrum from 0 to 100 for each of the experimental arms. A visual inspection suggests a great variability in the median. Moreover, it is worth mentioning that the third quarter is relatively narrower when the organization's pharmacist suggests prescribing. This indicates that when the organization's pharmacist suggests the prescription, the level noise among doctors in the quartile just above the median decreases. As a further note, the third boxplot from the left shows that a prescription suggestion from a pharmaceutical representative pushes the second quartile near the lower end of the probability spectrum.

Randomized experiment 2 – past behavior

Figure 2, which is based on a random-effects generalized least square (GLS) regression, reveals that, relative to the baseline condition in which information about past behaviour is not mentioned, the average propensity of prescribing a test is 8.70% points lower when clinicians did not prescribe the test before in similar cases ($p < .0005$) and 8.76% points higher when they previously prescribed the test in comparable situations ($p < .0005$). As to level noise, interquartile ranges appear similar for boxplots in Figure 2, thus indicating that the midspread tends to be consistent across experimental conditions. As in the previous experiment, responses are spread out across the entire probability spectrum for all arms.

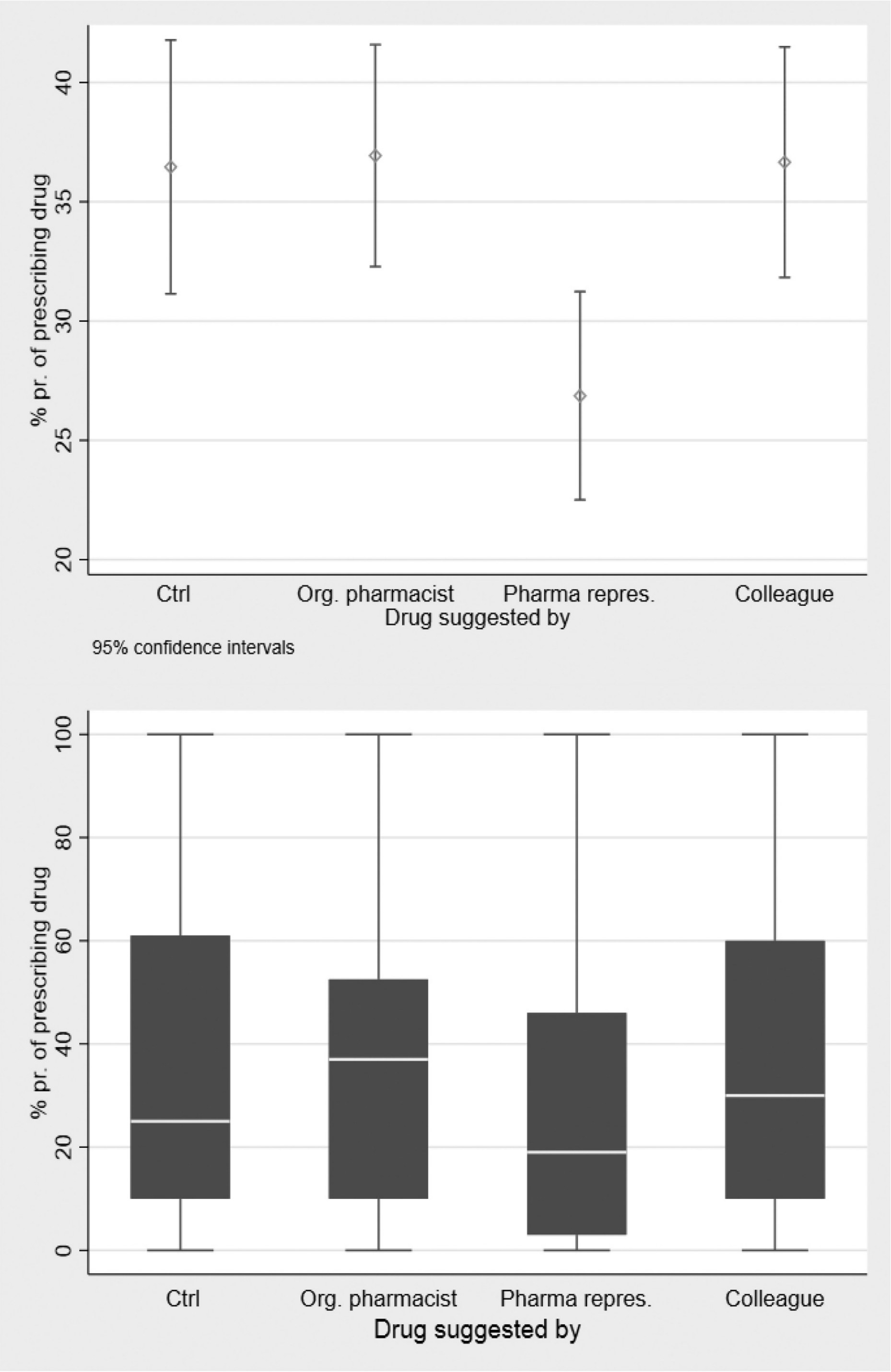

Figure 1. Average percentage probability of prescribing a drug and distribution of responses, by professional suggesting to prescribe (experiment 1 – In-group and out-group).

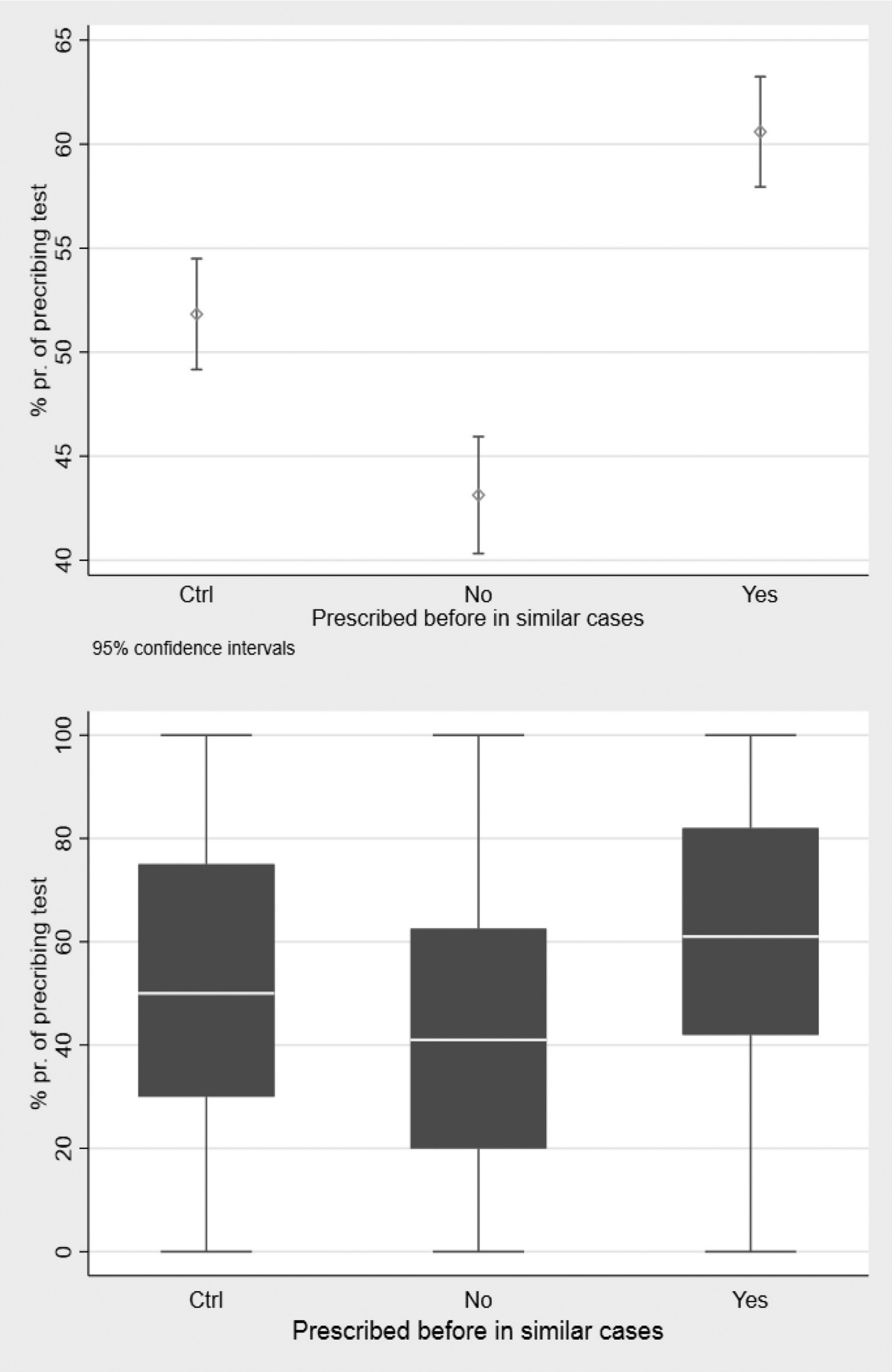

Figure 2. Average percentage probability of prescribing a test drug and distribution of responses, by past behaviour (experiment 2 – past behaviour).

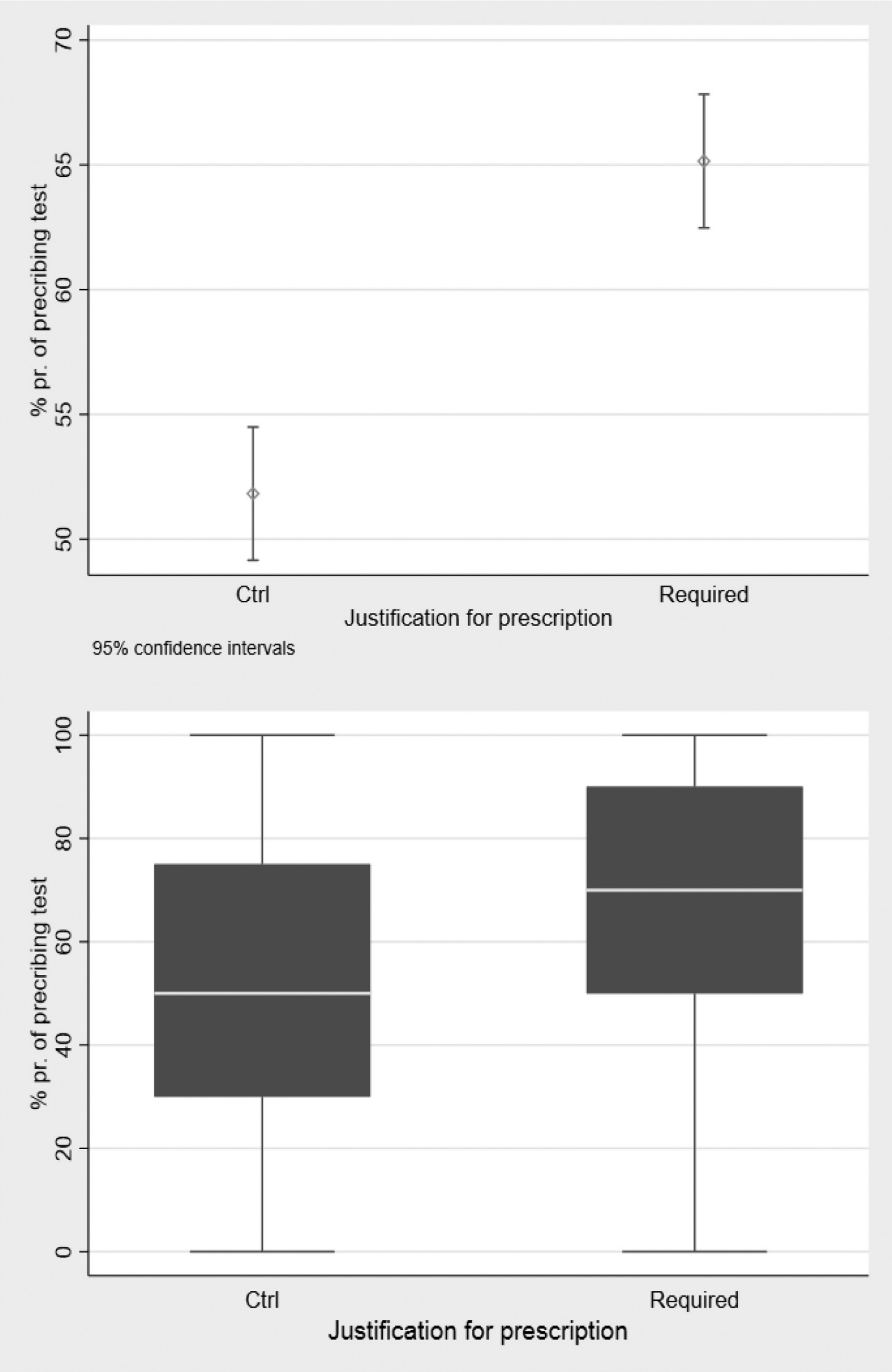

Figure 3. Average percentage probability of prescribing a test and distribution of responses, by request to justify the prescription (experiment 3 – confirmation bias).

Randomized experiment 3 – confirmation bias

As portrayed in Figure 3, estimates from a random-effects regression model show that the request to justify the prescription of a test increases by 13.32% points the average probability that clinicians would prescribe, compared to the control condition in which such a request is not made ($p < .0005$). The boxplots in Figure 3 display a sizable shift of the median in the two experimental arms, with a jump upward when the electronic medical record requires a justification for the decision to prescribe. The lack of a remarkable difference in the interquartile range is accompanied by a decrease of the fourth quarter and an increase of the first quarter in the treated arm relative to the baseline. Level noise, thus, appears comparable in the two quarters around the median but not in those at the extremes.

Randomized experiment 4 – loss aversion

Estimates from a random-effects generalized least square (GLS) regression reveal that, relative to the control condition, the average probability that clinicians would prescribe an antibiotic decreases by 5.54% points when they are reminded about the yearly number of deaths due to antibiotic resistance ($p = .001$) and by 6.41% points when the information is made available to them about the monetary loss due to the inappropriate prescription of antibiotics ($p < .0005$). To the contrary, we fail to find a difference in the average probability of prescribing antibiotics between the control and the monetary gain treatment ($p = .258$) (Figure 4). A series of tests of equality between pairs of coefficients show statistically significant differences in the outcome variable for the number of deaths due to antibiotic resistance condition and the monetary gain condition as well as between the monetary loss and monetary gain conditions. In particular, informing clinicians about the number of deaths generated by resistance to antibiotics decreases the probability they would prescribe them by 3.72% points compared to prompting them to think about the fact that the appropriate prescription of antibiotics causes a gain of many millions of euros that can be dedicated to patient care and clinical research ($p = .021$). Furthermore, nudging clinicians to reflect about how inappropriateness in antibiotic prescription produces monetary losses of funds that are taken away from patient care and clinical research reduces their probability of prescribing antibiotics by 4.59% points with regard to making the monetary gains generated by the appropriateness in antibiotic prescription readily available in their minds ($p = .004$). It is worth noting that the patterns of results partially change when we limit our analysis to the answers subjects provided for the first scenario they were exposed to within Experiment 4. More specifically, an analysis of variance with Scheffe correction indicates that the difference in effect between monetary loss and the monetary gain arms loses statistical significance at the conventional levels (p = .123).

As in previous cases, the range of responses covers the 0 to 100 probability scale for each of the experimental arms (Figure 4). A visual inspection suggests a certain degree of variability in the median. It seems interesting to note that, relative to all other arms, the monetary loss condition features the narrowest second quartile and the largest third quartile. This suggests that highlighting the economic losses caused by the inappropriate prescription of antibiotics decreases the level noise among doctors in the quartile just below the median and increases it in the quartile immediately above the median.

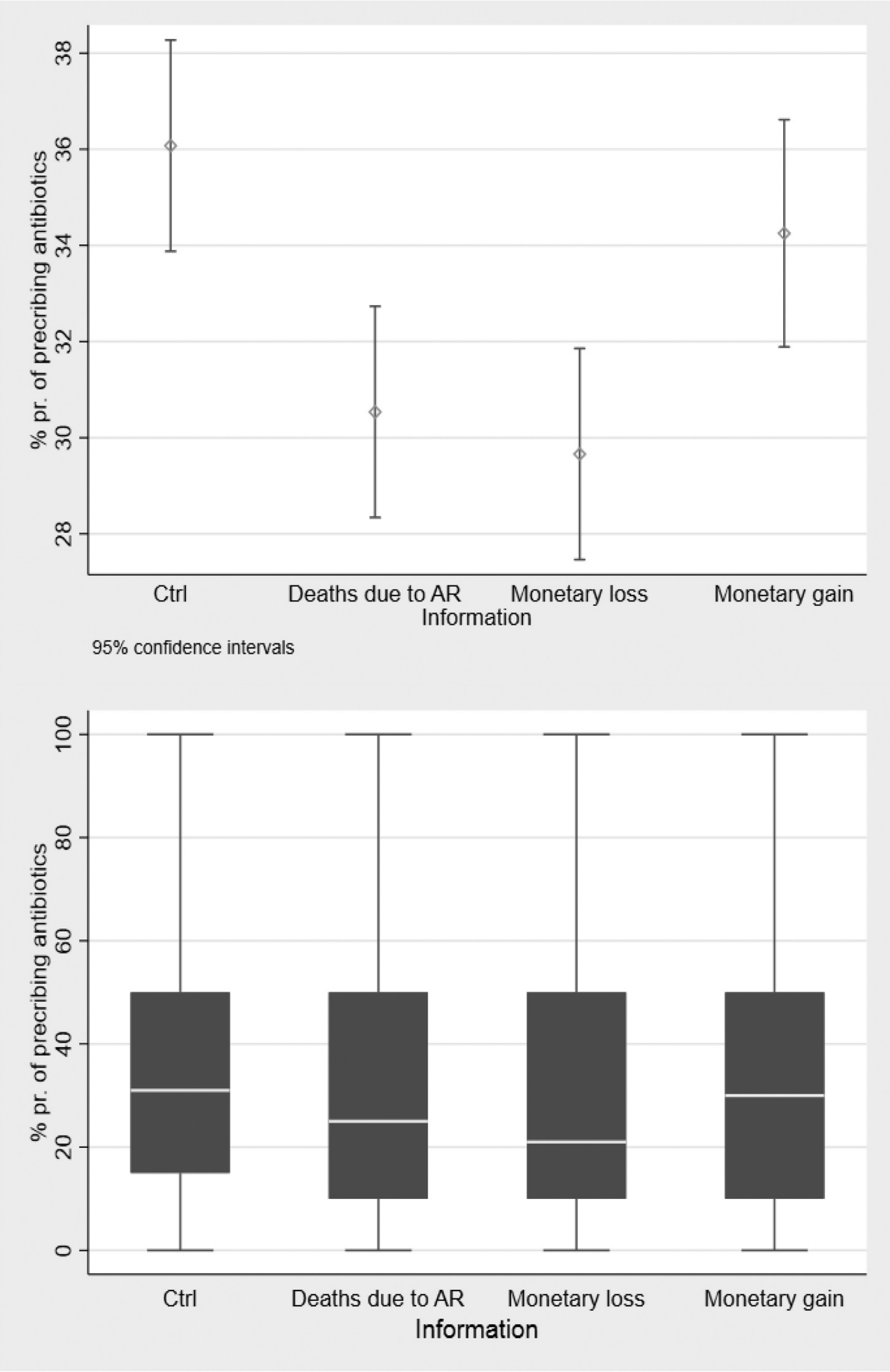

Figure 4. Average percent probability of prescribing antibiotics and distribution of responses, by information made readily available (experiment 4 – loss aversion).

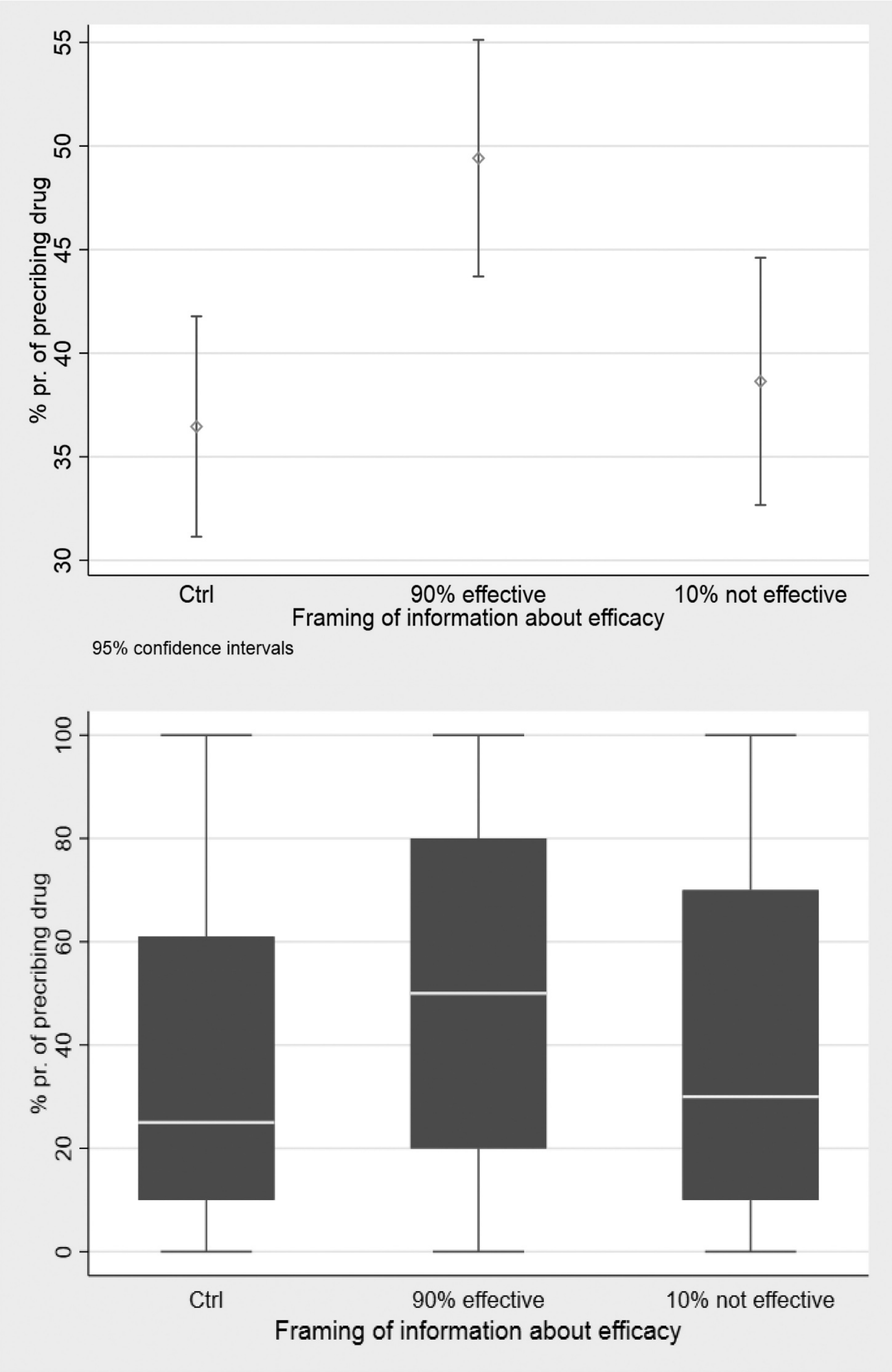

Figure 5. Average percent probability of prescribing a drug and distribution of responses, by framing of information about efficacy (experiment 5 – equivalence framing).

Randomized experiment 5 – equivalence framing

The framing of a drug's efficacy predicts the likelihood that clinicians will prescribe it. Figure 5, which is based on an analysis of variance with Scheffe correction, displays that the likelihood of prescribing a drug is statistically indistinguishable between the control condition (i.e. when no information about efficacy is provided) and the negative framing (i.e. when physicians are informed that the drug is 10% non-effective) (p = *.866*). On the contrary, the propensity to prescribe a drug in the positive framing treatment (i.e. when clinicians are informed that the drug is 90% effective) is 12.95% points higher compared to the control (p=*.005*) and 10.77% points higher compared to the negative framing (p=*.034*).

Boxplots in Figure 5 unveil how the quartile distribution of level noise changes due to the equivalence framing effect. Specifically, compared to the control group, a positive frame enlarges the second quartile and shrinks the third quartile, whereas a negative information frame increases both quartiles around the median value. As in some of the previous experiments, the median value is rather variable in the three arms. Similarly, the minimum and maximum observations fall within the 0 to 100 range.

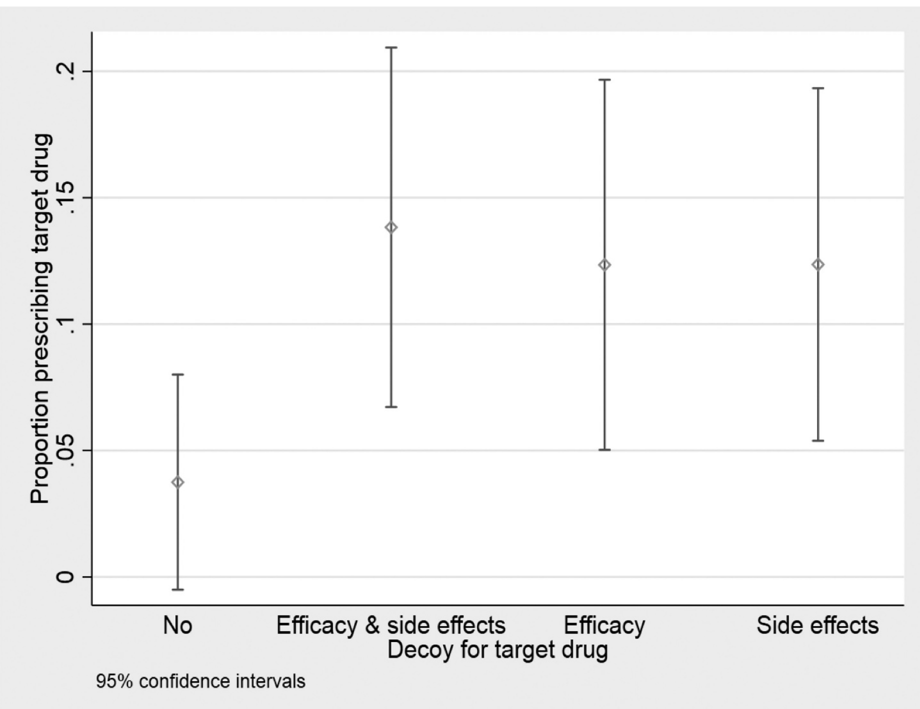

Figure 6. Average proportion of clinicians prescribing a drug, by decoy (experiment 6 – asymmetric dominance).

Randomized experiment 6 – asymmetric dominance

Figure 6 reports the proportion of medical doctors opting for the target drug in the four experimental groups. A logistic regression shows that, compared to participants who were not presented with any decoy, the odds of choosing the target drug were higher by 4.12 times among clinicians presented with a decoy drug that was inferior to the target simultaneously on efficacy and side effects (p=.032), by 3.62 times among subjects presented with a decoy drug that was inferior to the target on efficacy only (p=.058), and 3.62 times among clinicians presented with a decoy drug that was inferior to the target on side effects only (p=.055).

Discussion

Public management scholars have fervently delved into the intricacies of overseeing public healthcare provision and services, thereby enhancing the dynamic research landscape in this domain (Gofen, Meza, and Moreno-Jaimes 2023; Kirkpatrick, Zardini, and Veronesi 2023).

Our study makes a valuable contribution to this stream of work by conducting experimental research on bias and level noise in prescription decisions. This investigation serves as an empirical illustration of decision-making processes in various public service sectors. These include public managers evaluating the performance of their subordinates, recruiters selecting job candidates, judges deliberating sentences and asylum cases, teachers assessing students' tests, examiners adjudicating patent applications, and case managers in child protection agencies determining foster placement for children (Kahneman, Sibony, and Sunstein 2021). More specifically, our work sheds light on the potential for a middle-range theory focused on elucidating specific patterns of belief instability and heterogeneity in decision-making among public professionals. The development of contextualized practice-based theory has proven useful in public administration areas such as public value assessment (Huijbregts, George, and Bekkers 2022; Virtanen and Jalonen 2023), public service logic (Trischler et al. 2023), digital government (Castelnovo and Sorrentino 2018), and innovation (Cinar et al. 2022).

In the historical context of behavioural research within public management and the broader social sciences, predominant attention has been given to individual cognitive biases, often overlooking level-noise across various applications. In their recent call to broaden the perspective on decision-making, Kahneman, Rosenfield, and Blaser (2016, 2021)emphasized the significance of considering both noise and bias. They highlighted their collective impact on overall errors in decision-making processes and their subsequent effects on health, the economy, and the preservation of public values. The examination of noise in public decision-making is particularly crucial due to its association with the impartiality of bureaucracies (Cepiku and Mastrodascio 2021; Frederickson 1971, 2005; Rivera and Connolly Knox 2023). Elevated levels of noise, manifested as between-subjects variability, signify increased volatility in decisions on a single case, stemming from the incidental involvement of a public professional. This results in inequitable access to public services. Notably, to the best of our knowledge, prior works in public administration and management have not thoroughly addressed the measurement of unwanted variability in decision-making. Our study serves as a descriptive demonstration, shedding light on the presence of level-noise in prescription decisions and illustrating how it can be effectively managed to achieve improved outcomes through planned modifications to the architecture of choices. We observe patterns wherein the concentration of decisions in the quartiles shifts based on changes in the decision-making environment.

Some relevant contributions are potentially immediate. This is the case, in particular, with the measurement of noise levels in the outcome variables of our experiments, which are systematically measured and available within most public healthcare organizations. To the extent that prescriptions made by individual doctors are tracked, health care organizations and their managers can measure level noise (Kim, Tong, and Peden 2020). Likewise, whenever healthcare organizations monitor the prescription decisions of a single professional across patients and time, they can readily observe within-subject variability, thus investigating pattern noise. Other interventions aimed at modifying the architecture of choice undoubtedly require a greater investment, but one that is by no means prohibitive and could be widely cost-effective (Benartzi et al. 2017; Beshears and Kosowsky 2020). In this area, for example, Hjortskov et al. (2022) recently demonstrated how a small-scale government information campaigns delivering health messages through information and communication technologies successfully changed citizens' behaviours. Furthermore, because our interventions primarily consist of modifications to the choice architecture that do not involve mandates, they seem well suited to a context in which the professional autonomy of each individual doctor reduces the possibility of controlling prescribing decisions through hierarchical power. Scalability issues should be assessed too.

A second series of contributions of our work is to behavioural science and decision making in public management. The novel experimental evidence we provide can help advance the nascent stream of research studying bias and noise as the two components of judgement error. Hence, our findings deepen understanding of how experts – such as medical doctors – make judgements of likelihood and how decision architectures influence those judgements for better or for worse. As this is a relatively unexplored area of study, our results point to numerous directions for future research. For instance, it remains to be seen to what extent bias and noise are independent of each other, share common sources or even influence each other. From a methodological standpoint, scholars in this area may capitalize on our measure of noise, which goes beyond the mere standard deviation to take into account the range and distribution of outcomes as represented by boxplots.

Our research design is not immune to the same limitations that affect most research of the same kind (DellaVigna and Linos 2022). Consequently, we urge scholars and practitioners alike to interpret our findings in light of the inherent shortcomings. Most notably, the degree to which our temporary and experimentally induced treatment can be generalized to naturally occurring situations is yet to be tested. Although decision makers 'address a plausible, hypothetical problem in much the same way that [they] tackle a real one' (Kahneman, Sibony, and Sunstein 2021, 48), the use of field data would score higher on external validity compared to our study (DellaVigna and Linos 2022). Testing how fieldable our manipulations would be particularly useful in evaluating the possibility of their large-scale implementation. Another potential limitation of our study is the generalizability of results to different categories of prescribers. This threat to external validity is partially mitigated by the heterogeneity of our pool of subjects, which includes clinicians across medical specialities and work settings in the public sector. Moreover, despite being the most efficient tool that behaviouralists can use to have an unbiased estimate of the average effect caused by an intervention of some kind, a randomized experiment is unable to show the mechanisms through which the effect plays out. Although prescription decisions involve a 'complex, multi-stage sequence of mental events that cannot be directly observed' (Kahneman, Sibony, and Sunstein 2021, 168), other research designs – for instance parallel designs, mixed-methods studies, and qualitative inquiries – are superior in addressing the 'how" question. Also, the selection of our six cognitive mechanisms as far as prescription decisions were concerned was a collaborative effort with our partner institution. In particular, we prioritized field feasibility and relevance over exhaustiveness. It is crucial to emphasize that the list of cognitive mechanisms is not exhaustive. Lastly, time seems ripe for future work that complement the study of bias and noise from a micro-level perspective with estimations of how contextual factors impinge on the accuracy of judgements. Contextual factors that might be especially relevant for the field of public administration and management include politics, institutions, and system incentives (A. M. Bertelli and Riccucci 2022; A. Bertelli et al. 2022; Paola, Belle, and Hall 2023). The adoption of study replication in public management, particularly in studies employing experimental research designs, using the protocol developed by Walker et al. (2017), holds significant promise for testing the external validity of inference and illuminating how causal relationships of interest may be moderated by context (Bert et al. 2017). In light of leading-edge research suggesting that selective attention serves as a catalyst for belief instability towards ostensibly irrelevant features of decision contexts (Bordalo et al. 2023), it is imperative for experimental endeavours in this field to prioritize context as a key source of heterogeneity in belief instability.

Managerial implications and conclusions

The results of our six randomized experiments demonstrate that prescribing decisions made by physicians may be subject to considerable errors of judgement, which come in the two forms of systematic deviations from rationality (i.e. bias) or unwanted variability (i.e. noise). In other words, we observe systematic violations of predictions based on standard economics models and subjects who tend to make predictably irrational decisions. In fact, we find that the probability of prescribing is lower if an out-group member suggests it or if loss aversion is triggered. The likelihood of prescribing instead is higher when the doctor has made a prescription in similar cases before, justification is required, information is positively framed, or a decoy is added to the choice-set. As far as noise is concerned, in all our experiments, we find evidence of significant between-subject variability (i.e. level noise), as demonstrated by the range of responses covering the entire probability spectrum for each of our experimental arms. Another noteworthy finding is that the distribution of level noise, in terms of the interquartile distribution of responses, differs significantly between the experimental arms.

This evidence merits further considerations. Firstly, our findings of in-group favouritism raise a bold call to investigate groupthink (Kaba et al. 2016) in prescription decisions. Psychological research has long demonstrated that overconfidence tends to be higher for tasks that are more difficult when the predictability of forecasts is lower, and in the absence of timely and clear feedback (Fischhoff, Slovic, and Lichtenstein 1977; Griffin and Tversky 1992; Lichtenstein, Baruch, and Lawrence 1977). Prescribing seems to have several of these characteristics because it is a difficult task, with low predictability, and one in which feedback tends to be delayed and noisy.

Overconfidence and reputational concerns have been shown to be present among doctors (Baumann, Deber, and Thompson 1991; Christensen-Szalanski and Bushyhead 1981; Grant 2016), and the tendency to seek consensus instead of fostering dissent is unlikely to be absent in medical judgements. Organizational procedures aimed at reducing groupthink, mitigating overconfidence, and encouraging appropriate dissent before making a decision might include adding a second independent opinion on a single clinical case or asking multidisciplinary or multiprofessional groups of healthcare experts to analyse options and make choices collectively (Freeman, Robinson, and Scholtes 2021; Mitchell et al. 2014). Furthermore, extant scholarship seems to provide divergent findings as to whether requiring a justification for a prescription is likely to increase or decrease the probability of prescribing (Meeker et al. 2016). Given the relative ease with which such an intervention could be implemented in the hospital setting, it is desirable that our results are replicated in the field.

Another choice architecture that healthcare organizations should manage more consciously is the status quo. Because our results demonstrate that prescribers, like other experts (Baekgaard and Serritzlew 2016; Bellé, Cantarelli, and Belardinelli 2018), tend to be creatures of habit and to prescribe in the present as they have in the past, it seems essential to detect unwanted reference points that could fuel vicious prescription cycles. Considering the abundance of evidence about reference dependence (Baekgaard et al. 2019; Bellé, Cantarelli, and Belardinelli 2018; Blumenthal-Barby and Krieger 2015; Cantarelli, Belle, and Belardinelli 2020; Gustavo et al. 2016; Javier, Van Ryzin, and Leth Olsen 2021; Olsen 2015), it seems surprising that healthcare organizations do not yet fully leverage the potential impact of loss aversion and framing effects on enhancing the appropriateness of prescriptions.

Lastly, to the best of our knowledge, no previous work has rigorously tested the asymmetric dominance effect in medical prescriptions. In addition to extending the number of domains in which previous literature has already documented the impact of adding a decoy option into a choice set (Cantarelli, Belle, and Belardinelli 2020; Joel, Payne, and Puto 1982; Sebastian, Van Ryzin, and Van de Walle 2016; Tversky and Simonson 1993), our experiment urges healthcare managers to identify situations in which the inclusion of a decoy may reverse preferences and make clinicians and/or patients worse off.

In conclusion, our work might add value to public management scholarship that has fervently delved into the intricacies of overseeing public healthcare provision and services (Gofen, Meza, and Moreno-Jaimes 2023; Kirkpatrick, Zardini, and Veronesi 2023) to pursue the impartiality and equity in accessibility that are core tenet of public bureaucracies (Frederickson 1971, 2005).

Acknowledgments

This research is an integral component of a collaborative project conducted in partnership with the Italian regional government Regione Toscana, for whose invaluable support and collaboration we extend our sincere gratitude.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Nicola Belle http://orcid.org/0000-0003-0873-7253
Paola Cantarelli http://orcid.org/0000-0002-9914-2548

References

Arad, Ayala. 2013. "Past Decisions Do Affect Future Choices: An Experimental Demonstration." *Organizational Behavior and Human Decision Processes* 121 (2): 267–277. https://doi.org/10.1016/j.obhdp.2013.01.006.

Baekgaard, Martin, Nicola Belle, Søren Serritzlew, Mariafrancesca Sicilia, and Ileana Steccolini. 2019, "Performance Information in Politics: How Framing, Format, and Rhetoric Matter to Politicians' Preferences." *Journal of Behavioral Public Administration* 2 (2). https://doi.org/10.30636/jbpa.22.67.

Baekgaard, Martin, and Søren Serritzlew. 2016. "Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension." *Public Administration Review* 76 (1): 73–82. https://doi.org/10.1111/puar.12406.

Barberis, Nicholas C. 2013. "Thirty Years of Prospect Theory in Economics: A Review and Assessment." *Journal of Economic Perspectives* 27 (1): 173–196. https://doi.org/10.1257/jep.27.1.173.

Barnes, Andrew J., Michael Karpman, Sharon K. Long, Yaniv Hanoch, and Thomas Rice. 2021. "More Intelligent Designs: Comparing the Effectiveness of Choice Architectures in US Health Insurance Marketplaces." *Organizational Behavior and Human Decision Processes* 163 (March): 142–164. https://doi.org/10.1016/j.obhdp.2019.02.002.

Battaglio, R. Paul, Paolo Belardinelli, Nicola Bellé, and Paola Cantarelli. 2019. "Behavioral Public Administration Ad Fontes: A Synthesis of Research on Bounded Rationality, Cognitive Biases, and Nudging in Public Organizations." *Public Administration Review* 79 (3): 304–320. https://doi.org/10.1111/puar.12994.

Baumann, Andrea O., Raisa B. Deber, and Gail G. Thompson. 1991. "Overconfidence Among Physicians and Nurses: The 'Micro-Certainty, Macro-Uncertainty' Phenomenon." *Social Science & Medicine* 32 (2): 167–174. https://doi.org/10.1016/0277-9536(91)90057-J.

Belle, Nicola, and Paola Cantarelli. 2021. "Nudging Public Employees Through Descriptive Social Norms in Healthcare Organizations." *Public Administration Review* 81 (4): 589–598. https://doi.org/10.1111/puar.13353.

Bellé, Nicola, Paola Cantarelli, and Paolo Belardinelli. 2018. "Prospect Theory Goes Public: Experimental Evidence on Cognitive Biases in Public Policy and Management Decisions." *Public Administration Review* 78 (6): 828–840. https://doi.org/10.1111/puar.12960.

Belle, Nicola, Giorgio Giacomelli, Sabina Nuti, and Milena Vainieri. 2022. "Factoring in the Human Factor: Experimental Evidence on How Public Managers Make Sense of Performance Information." *International Public Management Journal* 25 (2): 178–191. https://doi.org/10.1080/10967494.2021.1975856.

Benartzi, Shlomo, John Beshears, Katherine L Milkman, Cass R Sunstein, Richard H Thaler, Maya Shankar, Will Tucker-Ray, William J Congdon, and Steven Galing. 2017. "Should Governments Invest More in Nudging?" *Psychological Science* 28 (8): 1041–1055. https://doi.org/10.1177/0956797617702501.

Bertelli, Anthony M, and Norma M. Riccucci. 2022. "What is Behavioral Public Administration Good For?" *Public Administration Review* 82 (1): 179–183. https://doi.org/10.1111/puar.13283.

Bertelli, Anthony, Norma Riccucci, Paola Cantarelli, Maria Cucciniello, Christian Grose, Peter John, Elizabeth Linos, Anjali Thomas, and Martin Williams. 2022. "The (Missing?) Role of Institutions in Behavioral Public Administration: A Roundtable Discourse." *Journal of Behavioral Public Administration* 5 (1). https://doi.org/10.30636/jbpa.51.304.

Beshears, John, and Harry Kosowsky. 2020. "Nudging: Progress to Date and Future Directions." *Organizational Behavior and Human Decision Processes* 161 (November): 3–19. https://doi.org/10.1016/j.obhdp.2020.09.001.

Blumenthal-Barby, J. S., and Heather Krieger. 2015. "Cognitive Biases and Heuristics in Medical Decision Making: A Critical Review Using a Systematic Search Strategy." *Medical Decision Making* 35 (4): 539–557. https://doi.org/10.1177/0272989X14547740.

Bordalo, Pedro, John J. Conlon, Nicola Gennaioli, Spencer Yongwook Kwon, and Andrei Shleifer. 2023. "How People Use Statistics." Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w31631.

Cantarelli, Paola, Nicola Belle, and Paolo Belardinelli. 2020. "Behavioral Public HR: Experimental Evidence on Cognitive Biases and Debiasing Interventions." *Review of Public Personnel Administration* 40 (1): 56–81. https://doi.org/10.1177/0734371X18778090.

Cantarelli, Paola Nicola Belle, and Jeremy L. Hall. 2023. "Information Use in Public Administration and Policy Decision-Making: A Research Synthesis." *Public Administration Review* 83 (6): 1667–1686. https://doi.org/10.1111/puar.13735.

Castelnovo, Walter, and Maddalena Sorrentino. 2018. "The Digital Government Imperative: A Context-Aware Perspective." *Public Management Review* 20 (5): 709–725. https://doi.org/10.1080/14719037.2017.1305693.

Cepiku, Denita, and Marco Mastrodascio. 2021. "Equity in Public Services: A Systematic Literature Review." *Public Administration Review* 81 (6): 1019–1032. https://doi.org/10.1111/puar.13402.

Christensen-Szalanski, Jay J. J., and James B. Bushyhead. 1981. "Physicians' Use of Probabilistic Information in a Real Clinical Setting." *Journal of Experimental Psychology: Human Perception and Performance* 7 (4): 928–935. https://doi.org/10.1037//0096-1523.7.4.928.

Cinar, Emre, Christopher Simms, Paul Trott, and Mehmet Akif Demircioglu. 2022. "Public Sector Innovation in Context: A Comparative Study of Innovation Types." *Public Management Review* 26 (1): 265–292. https://doi.org/10.1080/14719037.2022.2080860.

Cornelissen, Joep P., and Rodolphe Durand. 2014. "Moving Forward: Developing Theoretical Contributions in Management Studies: Theoretical Contributions in Management Studies." *Journal of Management Studies* 51 (6): 995–1022. https://doi.org/10.1111/joms.12078.

DellaVigna, Stefano, and Elizabeth Linos. 2022. "RCTs to Scale: Comprehensive Evidence from Two Nudge Units." *Econometrica* 90 (1): 81–116. https://doi.org/10.3982/ECTA18709.

Dinhof, Katharina, Sheeling Neo, Isa Bertram, Robin Bouwman, Noortje de Boer, Gabriela Szydlowski, Jurgen Willems, and Lars Tummers. 2023. "The Threat of Appearing Lazy, Inefficient, and Slow? Stereotype Threat in the Public Sector." *Public Management Review* 0 (0): 1–22. https://doi.org/10.1080/14719037.2023.2229326.

European Centre for Disease Control. 2022. "35 000 Annual Deaths from Antimicrobial Resistance in the EU/EEA. Press Release." Accessed December, 2023. https://www.ecdc.europa.eu/en/news-events/eaad-2022-launch.

Fischhoff, Baruch, Paul Slovic, and Sarah Lichtenstein. 1977. "Knowing with Certainty: The Appropriateness of Extreme Confidence." *Journal of Experimental Psychology* 3 (November): 552–564. https://doi.org/10.1037//0096-1523.3.4.552.

Frederickson, H. George. 1971. "Organization Theory and New Public Administration." In *Toward a New Public Administration, the Minnowbrook Perspective*, edited by Frank Marini, 309–331. 3rd ed. Scranton, Philadelphia: Chandler Press.

Frederickson, H. George. 2005. "Public Administration and Social Equity." In *Public Administration and Law*, edited by Julia Beckett, Heidi O. Koenig, 209–221. 1st ed. New York: Routledge.

Freeman, Michael, Susan Robinson, and Stefan Scholtes. 2021. "Gatekeeping, Fast and Slow: An Empirical Study of Referral Errors in the Emergency Department." *Management Science* 67 (7): 4209–4232. https://doi.org/10.1287/mnsc.2020.3711.

Fuenzalida, Javier Gregg G. Van Ryzin, and Asmus Leth Olsen. 2021. "Are Managers Susceptible to Framing Effects? An Experimental Study of Professional Judgment of Performance Metrics." *International Public Management Journal* 24 (3): 314–329. https://doi.org/10.1080/10967494.2020.1752338.

Ganegoda, Deshani B., and Robert Folger. 2015. "Framing Effects in Justice Perceptions: Prospect Theory and Counterfactuals." *Organizational Behavior and Human Decision Processes* 126 (January): 27–36. https://doi.org/10.1016/j.obhdp.2014.10.002.

Ganju, Kartik K., Hilal Atasoy, Jeffery McCullough, and Brad Greenwood. 2020. "The Role of Decision Support Systems in Attenuating Racial Biases in Healthcare Delivery." *Management Science* 66 (11): 5171–5181. https://doi.org/10.1287/mnsc.2020.3698.

George,Bert Sebastian Desmidt, Poul A. Nielsen, and Martin Baekgaard. 2017. "Rational Planning and Politicians' Preferences for Spending and Reform: Replication and Extension of a Survey Experiment." *Public Management Review* 19 (9): 1251–1271. https://doi.org/10.1080/14719037.2016.1210905.

Gofen, Anat, Oliver Meza, and Carlos Moreno-Jaimes. 2023. "Frontline Organizations as Experimental Settings for Policy Change: Why Public Management Matters Even More." *Public Management Review* 0 (0): 1–22. https://doi.org/10.1080/14719037.2023.2171095.

Grant, Adam M. 2016. *Originals: How Non-Conformists Move the World*. New York: Viking Penguin.

Griffin, Dale, and Amos Tversky. 1992. "The Weighting of Evidence and the Determinants of Confidence." *Cognitive Psychology* 24 (3): 411–435. https://doi.org/10.1016/0010-0285(92)90013-R.

Grima, François, Irène Georgescu, and Lionel Prud'homme. 2020. "How Physicians Cope with Extreme Overwork: An Exploratory Study of French Public-Sector Healthcare Professionals." *Public Management Review* 22 (1): 27–47. https://doi.org/10.1080/14719037.2019.1638440.

Grimmelikhuijsen, Stephan, Sebastian Jilke, Asmus Leth Olsen, and Lars Tummers. 2017. "Behavioral Public Administration: Combining Insights from Public Administration and Psychology." *Public Administration Review* 77 (1): 45–56. https://doi.org/10.1111/puar.12609.

Gurmankin, Levy Andrea, and John C. Hershey. 2006. "Distorting the Probability of Treatment Success to Justify Treatment Decisions." *Organizational Behavior and Human Decision Processes* 101 (1): 52–58. https://doi.org/10.1016/j.obhdp.2006.04.003.

Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42 (4): 1009–1055.

Hjortskov, Morten, Simon Zacher Kjeldsen, and Emil Sydendal Hansen. 2022. "Receiving and Engaging: Can a Simple ICT Delivered Government Message Change Citizen Health Behavior? A Field Experiment." *Public Management Review* 0 (0): 1–21. https://doi.org/10.1080/14719037.2022.2108126.

Holmes, R. Michael, Philip Bromiley, Cynthia E. Devers, Tim R. Holcomb, and Jean B. McGuire. 2011. "Management Theory Applications of Prospect Theory: Accomplishments, Challenges, and Opportunities." *Journal of Management* 37 (4): 1069–1107. https://doi.org/10.1177/0149206310394863.

Hong, Sounman, Sun Hyoung Kim, and Jieun Son. 2020. "Bounded Rationality, Blame Avoidance, and Political Accountability: How Performance Information Influences Management Quality." *Public Management Review* 22 (8): 1240–1263. https://doi.org/10.1080/14719037.2019.1630138.

Huber, Joel John W. Payne, and Christopher Puto. 1982. "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis." *Journal of Consumer Research* 9 (1): 90–98. https://doi.org/10.1086/208899.

Huijbregts, Rowie, Bert George, and Victor Bekkers. 2022. "Public Values Assessment as a Practice: Integration of Evidence and Research Agenda." *Public Management Review* 24 (6): 840–859. https://doi.org/10.1080/14719037.2020.1867227.

Jilke, Sebastian Gregg G. Van Ryzin, and Steven Van de Walle. 2016. "Responses to Decline in Marketized Public Services: An Experimental Evaluation of Choice Overload." *Journal of Public Administration Research & Theory* 26 (3): 421–432. https://doi.org/10.1093/jopart/muv021.

Kaba, Alyshah, Ian Wishart, Kristin Fraser, Sylvain Coderre, and Kevin McLaughlin. 2016. "Are We at Risk of Groupthink in Our Approach to Teamwork Interventions in Health Care?" *Medical Education* 50 (4): 400–408. https://doi.org/10.1111/medu.12943.

Kahneman, Daniel, Andrew M Rosenfield, and Tom Blaser. 2016. "Noise: How to Overcome the High, Hidden Cost of Inconsistent Decision Making." *Harvard Business Review* 94 (10): 38–46.

Kahneman, Daniel, Olivier Sibony, and Cass Sunstein. 2021. *Noise: A Flaw in Human Judgment*. New York: Little, Brown Spark; Hachette Book Group.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47 (2): 263. https://doi.org/10.2307/1914185.

Kim, Song-Hee, Jordan Tong, and Carol Peden. 2020. "Admission Control Biases in Hospital Unit Capacity Management: How Occupancy Information Hurdles and Decision Noise Impact Utilization." *Management Science* 66 (11): 5151–5170. https://doi.org/10.1287/mnsc.2019.3491.

Kirkpatrick, Ian, Alessandro Zardini, and Gianluca Veronesi. 2023. "Management Reforms, Re-Stratification and the Adaptation of Professional Status Hierarchies: The Case of Medicine in Publicly Owned Hospitals." *Public Management Review* 0 (0): 1–24. https://doi.org/10.1080/14719037.2023.2222124.

Krishnamurthy, Parthasarathy, Patrick Carter, and Edward Blair. 2001. "Attribute Framing and Goal Framing Effects in Health Decisions." *Organizational Behavior and Human Decision Processes* 85 (2): 382–399. https://doi.org/10.1006/obhd.2001.2962.

Lee, Seulki, and Minjung Kim. 2023. "Public Perceptions of Cross-Sector Collaboration and Sector Bias: Evidence from a Survey Experiment." *Public Management Review* 0 (0): 1–23. https://doi.org/10.1080/14719037.2023.2273316.

Lepore, Walter and J. Barton Cunningham. 2023. "Making Choices in Addressing Sustainability Problems: A Link to Framing Effects and Protected Values." *Public Management Review* 0 (0): 1–25. https://doi.org/10.1080/14719037.2023.2180656.

Lichtenstein, Sarah, Fischhoff Baruch, and D. Phillips. Lawrence 1977. "Calibration of Probabilities: The State of the Art." In *Decision Making and Change in Human Affairs: Proceedings of the Fifth Research Conference on Subjective Probability, Utility, and Decision Making, Darmstadt, 1–4 September, 1975*, edited by Helmut Jungermann and Gerard De Zeeuw, 275–324. Theory and Decision Library. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-010-1276-8_19.

Liu, Bingsheng, Zengqiang Qin, and Jinfeng Zhang. 2022. "Cognitive Bias in Citizens' Perceptions of Government Performance in Response to COVID-19: Evidence from a Large-Scale Survey Experiment in China." *International Public Management Journal* 25 (7): 1072–1093. https://doi.org/10.1080/10967494.2022.2056276.

Lu, Xiaoli Weijie Wang, and Xu. Hao. 2022. "Who Can Serve as the Proxy for Public Employees in Public Administration Experiments? A Cross-Sample Comparison." *Public Management Review* 24 (6): 819–839. https://doi.org/10.1080/14719037.2020.1864014.

Marvel, John D. 2015. "Public Opinion and Public Sector Performance: Are Individuals' Beliefs About Performance Evidence-Based or the Product of Anti–Public Sector Bias?" *International Public Management Journal* 18 (2): 209–227. https://doi.org/10.1080/10967494.2014.996627.

McNeil, Barbara J., G. Pauker Stephen, Harold C. Jr Sox, and Tversky. Amos. 1982. "On the Elicitation of Preferences for Alternative Therapies." *The New England Journal of Medicine* 306 (2): 1259–1262. https://doi.org/10.1056/NEJM198205273062103.

The Medicines Utilisation Monitoring Centre. 2023. "National Report on Medicines Use in Italy. Year 2022." Rome: Italian Medicines Agency. Accessed December , 2023. https://www.aifa.gov.it/documents/20142/1967301/Rapporto-OsMed-2022.pdf.

Meeker, Daniella, Jeffrey A. Linder, Craig R. Fox, Mark W. Friedberg, Stephen D. Persell, Noah J. Goldstein, Tara K. Knight, Joel W. Hay, and Jason N. Doctor. 2016. "Effect of Behavioral Interventions on Inappropriate Antibiotic Prescribing Among Primary Care Practices: A Randomized Clinical Trial." *JAMA* 315 (6): 562–570. https://doi.org/10.1001/jama.2016.0275.

Mitchell, Rebecca, Vicki Parker, Michelle Giles, and Brendan Boyle. 2014. "The ABC of Health Care Team Dynamics: Understanding Complex Affective, Behavioral, and Cognitive Dynamics in Interprofessional Teams." *Health Care Management Review* 39 (1): 1–9. https://doi.org/10.1097/HCM.0b013e3182766504.

Nagtegaal, Rosanna, Lars Tummers, Mirko Noordegraaf, and Victor Bekkers. 2019, 2. "Nudging Healthcare Professionals Towards Evidence-Based Medicine: A Systematic Scoping Review." *Journal of Behavioral Public Administration* 2 (2). https://doi.org/10.30636/jbpa.22.71.

Nagtegaal, Rosanna, Lars Tummers, Mirko Noordegraaf, and Victor Bekkers. 2020. "Designing to Debias: Measuring and Reducing Public Managers' Anchoring Bias." *Public Administration Review* 80 (4): 565–576. https://doi.org/10.1111/puar.13211.

Neo, Sheeling Isa Bertram, Gabriela Szydlowski, Robin Bouwman, Noortje de Boer, Stephan Grimmelikhuijsen, M. Jae Moon Étienne Charbonneau, and Lars Tummers. 2023. "Working 9 to 5? A Cross-National Analysis of Public Sector Worker Stereotypes." *Public Management Review* 0 (0): 1–30. https://doi.org/10.1080/14719037.2023.2254306.

Nickerson, Raymond S. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises." *Review of General Psychology* 2 (2): 175–220. https://doi.org/10.1037/1089-2680.2.2.175.

Olsen, Asmus Leth. 2015. "Citizen (Dis)satisfaction: An Experimental Equivalence Framing Study." *Public Administration Review* 75 (3): 469–478. https://doi.org/10.1111/puar.12337.

Olsen, Asmus Leth. 2017. "Compared to What? How Social and Historical Reference Points Affect Citizens' Performance Evaluations." *Journal of Public Administration Research and Theory* 27 (4): 562–580. https://doi.org/10.1093/jopart/mux023.

Organization for Economic Cooperation and Development. European Union. 2022. *Health at a Glance: Europe 2022: State of Health in the EU Cycle*. Paris: OECD Publishing. Accessed December , 2023. https://doi.org/10.1787/507433b0-en.

Rivera, Jason D., and Claire Connolly Knox. 2023. "Bureaucratic Discretion, Social Equity, and the Administrative Legitimacy Dilemma: Complications of New Public Service." *Public Administration Review* 83 (1): 65–77. https://doi.org/10.1111/puar.13550.

Roberts, Patrick S., and Kris Wernstedt. 2019. "Decision Biases and Heuristics Among Emergency Managers: Just Like the Public They Manage For?" *The American Review of Public Administration* 49 (3): 292–308. https://doi.org/10.1177/0275074018799490.

Samuelson, William, and Richard Zeckhauser. 1988. "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty* 1 (1): 7–59. https://doi.org/10.1007/BF00055564.

Saposnik, Gustavo Donald Redelmeier, Christian C. Ruff, and Philippe N. Tobler. 2016. "Cognitive Biases Associated with Medical Decisions: A Systematic Review." *BMC Medical Informatics and Decision Making* 16 (1): 138. https://doi.org/10.1186/s12911-016-0377-1.

Simon, Herbert A. 1947. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*. New York: Macmillan.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, US: Yale University Press.

Trischler, Jakob, Maria Røhnebæk, Bo Edvardsson, and Bård Tronvoll. 2023. "Advancing Public Service Logic: Moving Towards an Ecosystemic Framework for Value Creation in the Public Service Context." *Public Management Review* 0 (0): 1–29. https://doi.org/10.1080/14719037.2023.2229836.

Turner, Jhon C., and Katherine J. Reynolds. 2001. "The Social Identity Perspective in Intergroup Relations: Theories, Themes, and Controversies." *Blackwell Handbook of Social Psychology: Intergroup Processes* 4:133–152.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–1131.

Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (4481): 453–458. https://doi.org/10.1126/science.7455683.

Tversky, Amos, and Itamar Simonson. 1993. "Context-Dependent Preferences." *Management Science* 39 (10): 1179–1189. https://doi.org/10.1287/mnsc.39.10.1179.

Virtanen, Petri, and Harri Jalonen. 2023. "Public Value Creation Mechanisms in the Context of Public Service Logic: An Integrated Conceptual Framework." *Public Management Review* 0 (0): 1–24. https://doi.org/10.1080/14719037.2023.2268111.

Walker, Richard M., Oliver James, and Gene A. Brewer. 2017. "Replication, Experiments and Knowledge in Public Management Research." *Public Management Review* 19 (9): 1221–1234. https://doi.org/10.1080/14719037.2017.1282003.

Weißmüller, Kristina S. 2022. "Publicness and Micro-Level Risk Behaviour: Experimental Evidence on Stereotypical Discounting Behaviour." *Public Management Review* 24 (4): 601–630. https://doi.org/10.1080/14719037.2020.1862287.

Weißmüller, Kristina S., Robin Bouwman, and Rick Vogel. 2023. "Satisficing or Maximizing in Public–Private Partnerships? A Laboratory Experiment on Strategic Bargaining." *Public Management Review* 25 (7): 1282–1308. https://doi.org/10.1080/14719037.2021.2013072.