Article

Behavioral Public HR: Experimental Evidence on Cognitive Biases and Debiasing Interventions

Paola Cantarelli[1], Nicola Belle[1], and Paolo Belardinelli[2]

Abstract

We draw on behavioral science to investigate a set of decisions that may have an impact on public human resource management, thus affecting public service provision. Our survey-in-the-field-experiment with the nursing personnel of a local health authority showed that respondents' decisions in the area of health care operations management were affected by social pressures (bandwagoning), the presence of a *decoy* option, and the *framing* of alternatives. *Anchoring* and *halo* effects severely biased the assessment of subordinates' performance. Decisions in the domain of health policies were influenced by *denominator neglect* and *zero-risk bias*. Debiasing interventions eliminated the bandwagoning and framing effects. Being midway between abstract and un-testable grand theories and data-driven testable hypotheses, our findings advance behavioral human resource (HR) as a fruitful middle range theory in public personnel administration. Normative implications for scholars and practitioners about the power of the architecture of choices are discussed.

Keywords

cognitive biases, debiasing interventions, behavioral public administration, decision making, public managers and employees, survey-in-the-field experiment

Introduction

Traditional research in human resource (HR) management in governments around the world at different administrative levels includes work on how to select, develop, motivate, assess individual performance, and terminate employment. Abundant scholarship exists on

[1]Scuola Superiore Sant'Anna, Pisa, Italy
[2]Bocconi University, Milan, Italy

Corresponding Author:
Paola Cantarelli, Scuola Superiore Sant'Anna, Piazza Martiri della Libertà, 33, 56127 Pisa, Italy.
Email: paola.cantarelli@santannapisa.it

how to design civil service reforms to improve service delivery (e.g., Battaglio, 2014; Ingraham & Rubaii-Barrett, 2007; Perry, 1993, 2010). The application of behavioral science (e.g., Kahneman, 2002; Thaler, 2017) to the management of public employees is still under-developed. However, scholars in our discipline have argued that "one cannot really understand how organizations operate without a strong sense of how individuals process information and make decisions" (Jones, 2003, p. 401). Indeed, "how our minds work is not a niche interest; it is of wide relevance to many, if not all, aspects of workplace behavior and performance" (Chartered Institute of Personnel and Development [CIPD], 2014, p. 3).

Behavioral public HR can help improve service delivery and provide higher-impact HR strategies by clarifying what affects individuals' thinking in making job-related decisions. Behavioral public HR is also concerned with the designing of an optimum environment for employees' effective thinking, satisfaction and well-being. Taking up the call to apply behavioral science to HR much more widely (CIPD, 2014), this work investigates how public employees and managers actually make decisions. For instance, how do they choose between competing management practices, personnel policies, and programs that have uncertain outcomes? What cognitive biases and debiasing interventions predictably influence their decisions?

Our study is set to make three main contributions. From a descriptive point of view, time seems ripe to test the ecological validity of behavioral science evidence about the effect of systematic decision-making errors in a public context. Although we expected to replicate extant findings, our study departs from previous work for the following reasons. Whereas most available research has explored one systematic error and/or one decision domain at a time, we experimentally investigated the effects of a broad range of cognitive biases (i.e., asymmetric dominance, bandwagoning, attribute framing, anchoring, halo, denominator neglect, and zero-risk) across multiple decision settings (i.e., public management, public personnel management, and public policy). Moreover, we tested whether and to what extent the impact of cognitive biases depends on debiasing interventions because "the sincere desire of many people in this field is to discover flaws not for their own sake, but with the intention of improving decision making" (Larrick, 2004, p. 334). Finally, our work represents one of the few studies that not only employ a sample of real public sector workers but also tailor the research design to the settings of participants' real organization.

The second contribution of our study is theoretical. Our findings can help nurture behavioral HR as a middle range theory in the field of public personnel administration. Middle range theories are particularly useful theory-building strategy for public administration scholarship (e.g., Abner, Kim, & Perry, 2017; Perry, 2010). Our trials seem to meet the requirements that Abner and colleagues (2017) identify to make middle range theories functional to the development of grand theories: enough concreteness to generate testable hypotheses, consistency with reality sustained by the derivation from data rather than pure theorizing, and predisposition to generate results that can be synthesized.

A third contribution is of a normative nature. Normative studies in essence "tell you the right way to think about some problem" (Thaler, 2015, p. 25). Our work suggests

how public managers and policy makers may improve decision making within their organizations. On the one hand, they should recognize how the architecture of choices affects their decisions. On the other hand, they may leverage on the same architecture to encourage desired behaviors and avoid predictable errors. Indeed, our debiasing interventions modify the architecture of choices without limiting the options available or altering economic incentives. This tool seems to be particularly well suited to public personnel administration, which is characterized by low-powered incentives—such as limited use of performance-related pay and job security—but significant decisions' impact.

Theoretical Foundations of Behavioral Public HR

Expected-utility theory (Bernoulli, 1954) used to be the dominant framework to understand decision making under uncertainty. At its core, expected-utility theory features the *homo economicus*: a rational decision maker with a clear and comprehensive knowledge of the environment, a well-organized system of preferences, and excellent computational skills to allow for the selection of optimal solutions. Behavioral sciences—such as the bounded rationality paradigm (Simon, 1956), prospect theory (Kahneman & Tversky, 1979) and the nudge theory (Thaler, 2015; Thaler & Sunstein, 2008)—instead, proposed the existence of the *homo sapiens*, endowed with bounded rationality. The Nobel Prize Winner Herbert Simon (1956) was among the first scientist arguing that individuals are unable to take optimal decisions and rather make satisfying choices in predictable ways. He further envisioned the necessity of debiasing techniques by contending that the core function of government organizations is to design procedures that compensate for employees' essential inability to judge and compute in a complex and uncertain environment. This line of reasoning is still so relevant and powerful that only "few social scientists today would disagree with Simon's premise that a sound organizational theory must rest on a defensible theory of human behavior" (Jones, 2003, p. 401). Indeed, abundant evidence shows that people rely on a limited number of heuristics to translate complex tasks into simpler judgmental operations (Tversky & Kahneman, 1974). Whereas neoclassical economics allows for random mistakes in decision making, behavioral science posits that heuristics bring with them errors that are systematic, and therefore predictable, under certain circumstances (e.g., Ariely, 2010; Gardner, 2009; Gilovich, Griffin, & Kahneman, 2002; Kahneman, 2011).

Evidence of the effects of cognitive biases in judgment and choice is so widespread across decision domains and scientific disciplines that research on debiasing strategies (e.g., Larrick, 2004; Lilienfeld, Ammirati, & Landfield, 2009) and re-biasing interventions (e.g., Thaler & Benartzi, 2001), which use one cognitive bias to offset another, has also blossomed. More broadly, the recent applications of the nudge theory to public decision making show that leveraging on the architecture of choices without limiting the options available and without altering economic incentives can encourage desired behaviors (e.g., Thaler, 2015; Thaler & Sunstein, 2008).

Social science disciplines such as behavioral economics (e.g., Ariely, 2010), political psychology (Taber & Lodge, 2006), applied psychology (e.g., Kahneman, 2011), general management (e.g., Cornelissen & Werner, 2014), and medicine (Blumenthal-Barby & Krieger, 2015; Saposnik, Redelmeier, Ruff, & Tobler, 2016) have a long tradition in the study of cognitive biases. To the contrary, this area of research is still nascent in both public administration (Grimmelikhuijsen, Jilke, Olsen, & Tummers, 2017) and public HR (e.g., Belle, Cantarelli, & Belardinelli, 2017). Scholars in our field have only recently investigated cognitive biases in citizens' judgments of public services (e.g., Andersen & Hjortskov, 2015; Baekgaard & Serritzlew, 2016; Barrows, Henderson, Peterson, & West, 2016; Jilke, Van Ryzin, & Van de Valle, 2016; Marvel, 2015; Moynihan & Lavertu, 2012; Olsen, 2015a, 2017b) and in public employees' assessment of subordinates' performance (Belle et al., 2017). Analyzing 159 case studies from 60 public bodies in 23 states and two international institutions, the Organization for Economic Cooperation and Development (2017) recently reported that attempts to use behavioral insights to inform policies are underway across policy areas such as consumer behavior, education, energy, environment, finance, health and safety, labor market, service delivery, taxes, and telecommunications.

Study Setting, Design, and Participants

Participants in our survey-in-the-field experiment were 602 nurses working at a local health authority (LHA) in Northern Italy. Italian LHAs are responsible for planning and delivering health care and social services to citizens in a specific geographic area. The nursing staff of Italian LHAs falls into three main groups: nursing assistants, nurses, and nursing managers. Nursing assistants are responsible for attending to patients' and clients' needs under the supervision of nurses and doctors; they have successfully completed training programs delivered by Italian Regions. Nurses hold a university degree and perform specialized health-related tasks. A subgroup of the nurses are responsible for coordinating small teams of nurses. Finally, nurse managers hold advanced university degrees, are responsible for managing services and/or personnel, and are rarely in direct contact with patients.

To maximize the ecological validity and contextual relevance of our study, we ran several focus groups with the Head of the nursing staff and a few nurse managers to pinpoint decision-making tasks that were relevant, performed routinely, common enough across medical specialties, and straightforward enough for nurses in our local health authority to understand. Through guided discussions and analyses of any form currently used to make decisions, we jointly identified the following cognitive biases and decision domains as most ecologically valid for the study purposes: asymmetric dominance, bandwagoning, and attribute framing in public management; anchoring and halo in performance appraisal; and denominator neglect and zero-risk in policy making. All the decisions involve the use of tax money and have an impact on the public, whether on the working environment of the public personnel employed by the LHA or on the health care services provided to citizens. Overall, our experiments feature one of the following elements: a novel design (bandwagoning), major innovations

and/or extensions of the original design (denominator neglect and zero risk), or minor variations compared to the original trial (asymmetric dominance, framing, anchoring, and halo).

To further exploit the descriptive approach of this work and to pursue the goal of testing debiasing strategies in addition to empirically demonstrating biases, we also jointly selected a subgroup of the systematic errors for which we designed debiasing interventions. We did so by applying the same criteria of ecological validity and contextual realism that we used for the selection of cognitive biases. Our debiasing strategy featured a definition of the cognitive bias mechanism coupled with its concrete application to the decision that respondents would be exposed to afterwards because extant scholarship on debiasing shows that "the most effective approaches combined an abstract principle with concrete examples" (Larrick, 2004, p. 325). For example, we hypothesized that explaining to subjects that a subordinate's performance score in year X-1 should not affect the performance score in year X on average mitigates the anchoring effect. We employed factorial designs to test the moderating effect of debiasing messages for the following scenarios: bandwagoning, framing, anchoring, and halo.

The final version of our survey-in-the-field experiment consisted in six randomized control trials (RCTs) and a test of the zero-risk bias (Appendix 1). The Head of the nursing staff validated the final survey before agreeing to open the data gathering step.

We administered the experimental survey through Qualtrics. The final sample was composed of 602 respondents (34% response rate). Average age was 46 years (*SD* = 10), although 22% of respondents did not report their age. Females were 63% and males, 15%, with 22% missing values. About 6% of participants were health care assistants, 56% nurses, 11% nurses with coordination responsibilities, and 4% nurse managers, with 23% missing values. About 59% did not manage any subordinates, 5% between 1 and 5, 5% between 6 and 15, 6% between 16 and 30, 2% between 31 and 100, and 1% more than 100, with 22% missing values. We employed pairwise deletion to account for missing responses. Subjects were exposed to the entire set of scenarios described in the remainder of the article in a random order. Within each of the six RCTs, participants were randomly assigned to one experimental arm. We tested the interactions of the cognitive bias manipulation with the debiasing message, when present. As expected, due to the random assignment of respondents to experimental conditions, groups within each of the six RCTs did not statistically differ in terms of our observed demographic variables.

Asymmetric Dominance, Bandwagoning and Framing in Public Management

Asymmetric Dominance—RCT 1

The asymmetric dominance effect, also known as decoy or attraction, causes individuals' preferences between a target option and a competing option to shift toward the target when a decoy, similar to the target but in no way better, is added to the choice set (e.g., Ariely, 2010; Huber, Payne, & Puto, 1982; Tversky & Simonson, 1993). In

the simplest case, options are described by two attributes. The decoy is equal to the target in one attribute and slightly inferior to the target in the other attribute, or slightly inferior in both attributes. Meanwhile, the decoy is inferior to the competing alternative in one attribute but superior to it in the other attribute. These comparative features of the three alternatives in the choice set make the target asymmetrically dominate the decoy. Although the decoy is virtually never selected, it serves the purposes of altering individuals' preferences between the other options in the choice set.

Method

RCT 1 was a variation of Ariely's (2010) *The Economist* study. Public employees of the LHA in our study selected a diagnostic instrument to purchase from a set of instruments that varied along two dimensions: format of the medical record produced and price. Participants in the *no decoy group* made a choice between two diagnostic instruments: one that provided electronic records and cost €7,500 and one that provided both electronic and paper records and cost €9,000. The latter served as the target in our experimental design. The choice set for the respondents in the *decoy group* featured a third (decoy for the target) option. The decoy option produced paper records and costed €9,000.

Results

Figure 1 reports the proportion of respondents opting for the target diagnostic instrument in the two experimental groups, .61 in the *no decoy condition* ($N = 246$) and .84 in the *decoy condition* ($N = 251$). A logistic regression showed that the odds of choosing the target option (i.e., the diagnostic instrument providing both paper and electronic records and costing €9,000) were 3.42 times greater among participants presented with the decoy option (i.e., the diagnostic instrument providing paper records and costing €9,000) as compared to participants who were not presented with the decoy option ($p < .001$).

Bandwagoning and Isomorphism—RCT 2

Although bandwagoning and isomorphism are native to various social sciences disciplines, the two phenomena share key elements as far as systematic errors in decision making are concerned. Bandwagoning is grounded in the research of group think and conformity in social psychology (e.g., Colman, 2003): the rate of acceptance of beliefs and behaviors increases the more that they have already been adopted by others, regardless of individuals' own opinions. Cohen and Rothschild (1979) described the bandwagons of medicine as "the overwhelming acceptance of unproved but popular ideas" (Cohen & Rothschild, 1979, p. 531). The authors further argued that "some of these ideas eventually prove valid, and their uncritical acceptance is belatedly justified. More often, however, they are disproved and abandoned, or replaced by another bandwagon" (Cohen & Rothschild, 1979, p. 531).
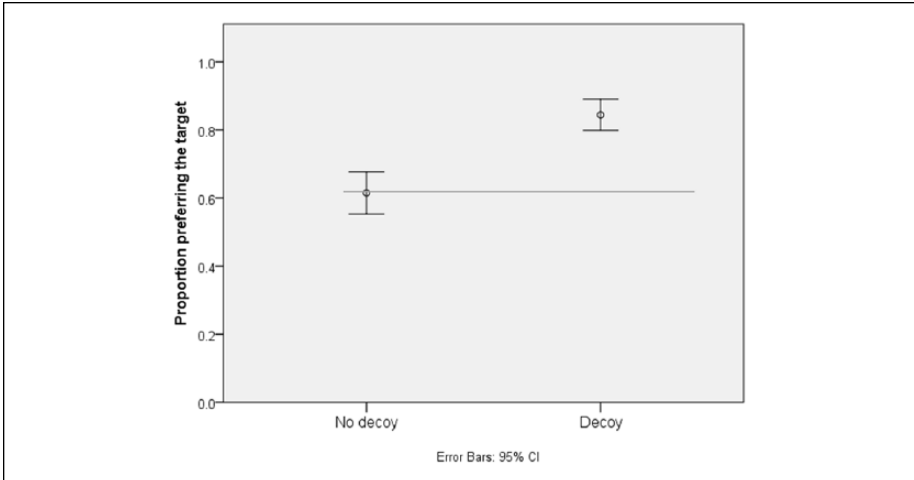
Figure 1. Proportion of participants preferring the target diagnostic instrument, with and without a decoy diagnostic instrument available (experiment 1).
Note. The proportions are different ($p < .001$), as suggested by the fact that neither CI includes the mean value of the other group (see the horizontal line). CI = Confidence Interval.

Rooted in institutional theories, isomorphism is the tendency of institutions to become alike one another because the decision makers in one organization adopt the same policies, procedures, or arrangements that decision makers in another organization have already adopted. Under coercive pressures, institutions become more alike because of the formal and informal authority exerted by parent organizations, government, or society. Mimetic isomorphism, then, results from coping with uncertainty by imitating organizations that are deemed successful or have a good reputation. Finally, normative isomorphic pressures tend to make organizations more homogeneous because of the similar professionalization and socialization mechanisms in the professional associations to which employees belong (DiMaggio & Powell, 1983).

Method

RCT 2 employed a 4 (isomorphic pressures) × 2 (debiasing) between-subjects design. Respondents were asked to indicate which one of two diagnostic instruments they would recommend their organization to purchase. They read that the two diagnostic instruments had the same price and that one was slightly superior to the other in terms of accuracy, user-friendliness, and durability. Whereas participants in the *control group* read only this information, respondents in the other three conditions were exposed to isomorphic pressures encouraging the adoption of the inferior diagnostic instrument. More precisely, subjects in the *coercive pressure group* read that the regional government guidelines suggested the adoption of the inferior instrument. Participants in the *mimetic pressure group*, instead, were informed that all the best performing local health

authorities in the nation would adopt the inferior option. Finally, respondents in the *normative pressure group*, were told that the professional association to which they belonged suggested the adoption of the inferior option. We designed the isomorphic pressures by drawing on DiMaggio and Powell's (1983) conceptualization of institutional isomorphism and applying it to the relationships of formal and informal power that are customary for our sample. Indeed, not only is the relationship between local health authorities and their regional government asymmetric in terms of power/legitimacy and resources but also regional governments typically use guidelines to provide recommendations to local health authorities. Furthermore, intra-organizational comparisons are increasingly performed among local health authorities across the nation to raise the standard of health care services, set benchmarks, and share best practices. Finally, professional associations are very common in the health care industry.

The debiasing manipulation consisted of exposing half the respondents to an explanation of the cognitive bias. Before being randomly assigned to one of the conditions described above, subjects in the *debiasing group* read that we may prefer services and goods that are inferior to others only because we are influenced by the guidelines of superordinate organizations, or by the recommendations of a professional association of which we are a member, or because we emulate the decisions of organizations that have a positive reputation.

Results

Figure 2 displays the two-way interaction of isomorphic pressures and debiasing intervention on the preference for the inferior diagnostic instrument. When not exposed to the explanation of the effects of isomorphism, the percentages of respondents opting for the worse option were as follows: 15% in the control group (i.e., in the absence of any isomorphic pressures; $N = 66$); 30% in the coercive condition (i.e., when the inferior diagnostic instrument was suggested by guidelines issued by the Regional Government; $N = 57$); 23% in the mimetic condition (i.e., when the inferior diagnostic instrument was adopted by the local health authorities with the highest reputation nationwide; $N = 60$); and 11% in the normative condition (i.e., when the inferior diagnostic instrument was recommended by a professional association to which respondents belonged; $N = 62$). The corresponding percentages for subjects exposed to the debiasing strategy were as follows: 11% in the control condition ($N = 63$), 11% in the coercive condition ($N = 62$), 8% in the mimetic condition ($N = 60$), and 10% in the normative condition ($N = 61$).

Results of a logistic regression with interaction terms showed that the odds that employees would choose the inferior diagnostic instrument were 2.38 times higher ($p = .054$) in the coercive condition as compared to the control condition. The effect of guidelines issued by the Regional Government on the selection of the worse option disappeared when subjects were provided with the explanation of the cognitive bias ($p = .241$). The odds that respondents would choose the worse option over the better option were not statistically different in the mimetic and normative groups as compared to the control group.
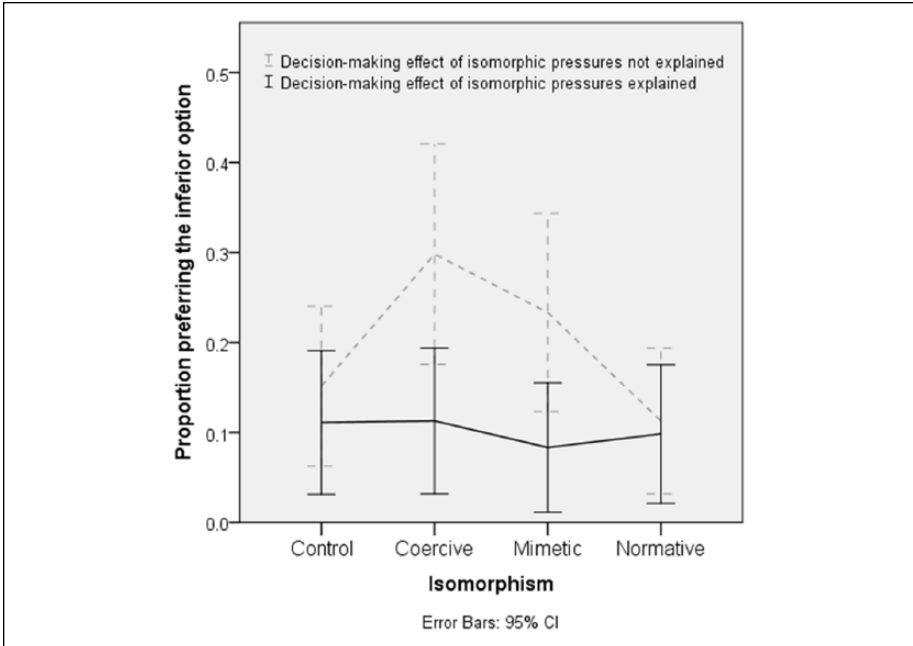
Figure 2. Two-way interaction of isomorphic pressure and debiasing strategy on the proportion of participants preferring the inferior diagnostic instrument (experiment 2).
Note. CI = Confidence Interval.

Attribute Framing—RCT 3

Attribute framing in one of the main typologies of the framing effect (e.g., Levin, Schneider, & Gaeth, 1998), typically studied under the prospect theory mantra that "losses loom larger than gains" (Kahneman & Tversky, 1979, p. 279). When individuals fall prey to attribute (or equivalence) framing, they react differently to objectively equivalent information, depending on whether the information is presented positively or negatively (e.g., Olsen, 2015a, 2015b, 2017b). More precisely, individuals' behavioral outcomes tend to differ systematically when they are exposed to a 90% survival rate rather than the equivalent 10% mortality rate, or when they are informed that a food is 95% fat-free rather than 5% fat (e.g., Kahneman, 2011; McNeil, Pauker, Sox, & Tversky, 1982; Tversky & Kahneman, 1981).

Method

The decision scenario in RCT 3 was the adoption of an email software for inter-organizational communication. Respondents indicated their propensity to adopt the software on a 0-100 scale. The framing manipulations refer to whether respondents were informed that 20% of the users are dissatisfied with the software (*negative framing group*), 80% of the users are satisfied with the software (*positive framing group*);
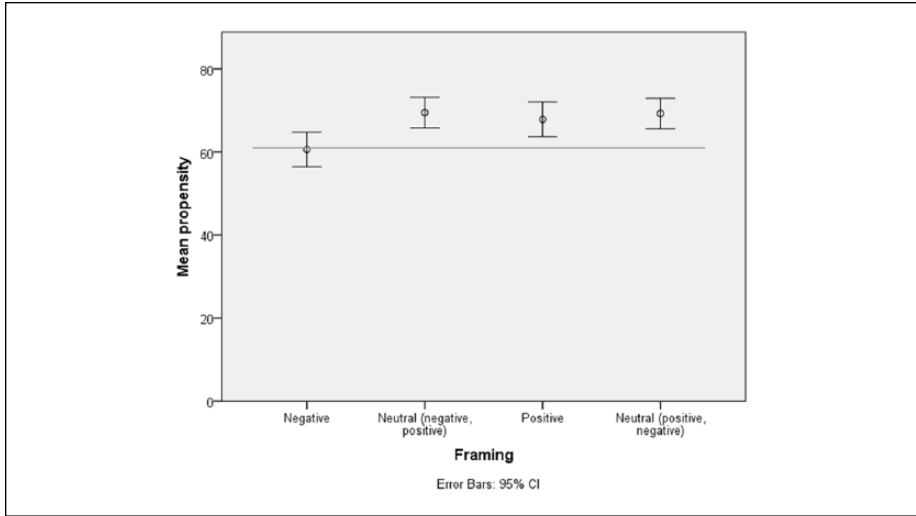
Figure 3. Propensity to purchase the email software Xmail, by framing (experiment 3).
Note. The mean propensity in the negative condition differs from the mean propensities in all other conditions ($p < .01$) as suggested by the fact the CIs of the neutral and positive groups do not include the mean score of the negative group (see the horizontal line). CI = Confidence Interval.

or if they were presented with both percentages in random orders (i.e., 20% dissatisfied and 80% satisfied or 80% satisfied and 20% dissatisfied; *neutral framing group*). To avoid potential ambiguity about the number of satisfaction categories, all participants read that software users could express their opinion by selecting one of two options: satisfied or dissatisfied. In RCT 3, the neutral framing manipulation serves as debiasing intervention.

Results

Figure 3 shows the average propensity by experimental intervention, on a 0-100 scale, to purchase the email software. Employees who read that 80% of users were satisfied with the software (i.e., positive framing) tended to report a higher propensity to purchase ($M = 67.87$, $SD = 22.95$, $N = 127$) compared to their peers who read that 20% of users were dissatisfied (i.e., negative framing; $M = 60.02$, $SD = 23.00$, $N = 123$), ($p < .01$). The mean propensity to purchase the software in the two neutral framing groups were statistically indistinguishable from each other ($M = 69.59$, $SD = 19.79$, $N = 114$ in the condition in which subjects read the dissatisfaction percentage first, and $M = 69.40$, $SD = 20.70$, $N = 129$ in the condition in which subjects read the satisfaction percentage first), statistically undistinguishable from the positive framing group, and higher than in the negative framing group ($p < .01$).

Anchoring and Halo in Public Personnel Performance Appraisal

Anchoring—RCT 4

Anchoring is the cognitive tendency to estimate unknown quantities by making adjustments from an initial value, even if it is unmistakably arbitrary (e.g., Tversky & Kahneman, 1974). Extant randomized trials consistently have found that individuals required to assess unknown quantities and to select a certain number for their evaluations tend to provide final estimates that are insufficient adjustments of the initial value. Incomplete computations, numbers generated randomly in the presence of the decision maker, and numbers provided with a priming role all function as anchors and lead to the anchoring effect (e.g., Blumenthal-Barby & Krieger, 2015; Furnham & Boo, 2011; Kahneman, 2011; Saposnik et al., 2016). As suggested by an anonymous Reviewer, research on comparisons (e.g., Charbonneau & van Ryzin, 2015; James, 2010; Nielsen, 2014; Olsen, 2017a) may come out from extant evidence of the anchoring bias.

Method

RCT 4 employed a 2 (anchors: low vs. high) × 2 (ratee: subordinate vs. self) between-subjects design in the context of individuals' performance appraisals on the job. Experiment 4 expands on the work of Belle et al. (2017) by testing the external validity of their findings and adding an experimental factor. Subjects read that the ratee had reached satisfactory results and had demonstrated a discrete degree of openness toward colleagues. We selected these performance dimensions from the actual performance evaluation form used in the local health authority. As in Belle et al. (2017) the last year's performance rating served as the anchor: respondents in the *low anchor* group read that last year, the ratee received a performance rating of 51/100 whereas those in the *high anchor* group read that last year, the ratee received a performance rating of 91/100. Unlike Belle et al. (2017), participants had to evaluate a subordinate in the *evaluation of a subordinate group* and themselves in the *self-evaluation group*. All subjects indicated the *performance score* that they would assign to the ratee for this year on a 0-100 scale.

Results

Figure 4 shows the mean performance rating that respondents assigned to the ratee, separately for the four groups. A linear regression showed that, everything else being equal, the performance rating was on average 13.82 scale-points higher for employees in the high anchor group ($N = 232$) as compared to their counterparts in the low anchor group ($N = 262$; $p < .001$). Furthermore, keeping everything else constant, the mean performance rating was 3.67 scale-points higher for subjects who self-evaluated themselves ($N = 168$) rather than a hypothetical subordinate ($N = 326$; $p = .022$).
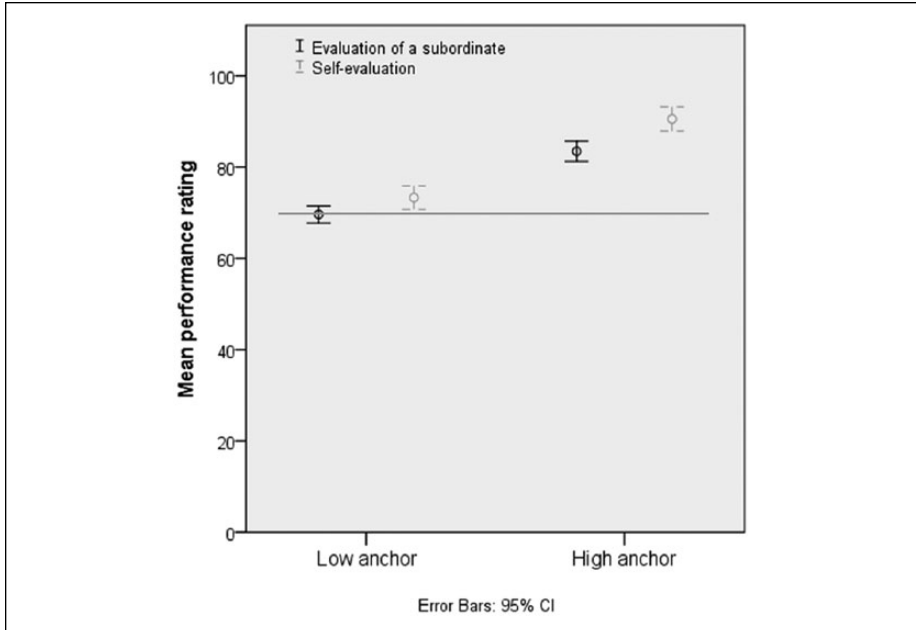
Figure 4. Mean performance rating, by anchor (i.e., last year performance score) and ratee's identity (self vs. subordinate; experiment 4).

Note. The mean performance ratings in the four experimental arms are different, as indicated by the fact that none of the CIs contains the mean performance rating of the other experimental conditions (see the horizontal line). CI = Confidence Interval.

The debiasing message, explaining that last year's performance score should not influence the performance score for this year, was ineffective in mitigating the anchoring effect. The mean rating that respondents assigned to the ratee was influenced by the anchor, regardless of the exposure or lack thereof to the explanation of the anchoring effect in performance appraisal.

Halo—RCT 5

The halo effect refers to "how judgments about some aspects of an object may influence how other aspects of an object are judged" (CIPD, 2015, p. 24). When raters fall prey to the halo effect, they transfer their assessment of a ratee's domain to another by providing consistently high (or low, or average) ratings across performance dimensions, even in the presence of disconfirming information (e.g., Borman, 1975; Nisbett & Wilson, 1977). Early work on halo showed that when superiors in the army had to provide a score for officers' performance on four different dimensions, they assigned ratings whose "correlations are too high and too even" (Thorndike, 1920, p. 27). Halo effects in performance evaluation has been shown in different professions: among public sector managers and non-managers (e.g., Belle et al., 2017), among post

commanders and sergeants in a state police agency (e.g., King, Hunter, & Schmidt, 1980), among workers in a manufacturing company (e.g., Holzbach, 1978), and among students assessing faculty members (e.g., Jacobs & Kozlowski, 1985).

Method

Drawing on Belle et al. (2017), participants in RCT 5 were asked to imagine being a rater who had to evaluate a ratee along two dimensions: technical skills in carrying out job duties and interpersonal skills in communicating with patients. The former dimension was our independent variable, which we manipulated at two levels: poor versus excellent. In other words, a random half of the subjects read that the ratee's technical skills were poor, whereas the other half read that the ratee's technical skills were excellent. All subjects, then, were told that the ratee had good abilities to interact with patients and were finally asked to indicate on a 0-100 point-scale their evaluation of the ratee along the two performance dimensions. The rating on technical skills served as the manipulation check, which we included to make sure that our experimental treatment produced the intended effect. The rating on interpersonal skills was our dependent variable. As in RCT 4, we selected the performance dimensions described in RCT 5 from the actual forms used at the local health authority.

Results

A two-sample mean comparison test showed that the average rating assigned to the ratee's technical skills was lower in the poor technical skills group ($M = 36.86$, $SD = 26.52$, $N = 243$) as compared to the excellent technical skills group ($M = 85.31$, $SD = 16.12$, $N = 255$; $p < .001$). Therefore, our experimental manipulation was effective.

Figure 5 reports the mean rating that participants assigned to the ratee's interpersonal skills in communicating with patients, by ratee's technical skills on the job. On average, raters in the poor technical skills condition scored ratee's interpersonal skills lower ($M = 71.78$, $SD = 17.92$) than raters in the excellent technical skills condition did ($M = 75.11$, $SD = 15.71$; $p = .028$). As expected, our findings suggest that employees can be systematically prone to halo effects in assessing different dimensions of ratees' performance. Indeed, the mean score assigned to interpersonal abilities was contingent upon the manipulation of technical skills.

The tendency to judge interpersonal skills higher for ratees with excellent rather than poor technical skills was robust across the debiasing intervention. In fact, subjects who read that the performance score on technical competences should not influence the rating of interpersonal abilities were as much prone to the halo effect as their peers who did not read the explanation of the bias.

Denominator Neglect and Zero-Risk Effects in Public Policy

Denominator Neglect—RCT 6

The denominator neglect (or ratio bias) is the tendency to pay attention to the number of times a target event has happened (the numerator of a ratio) neglecting the overall
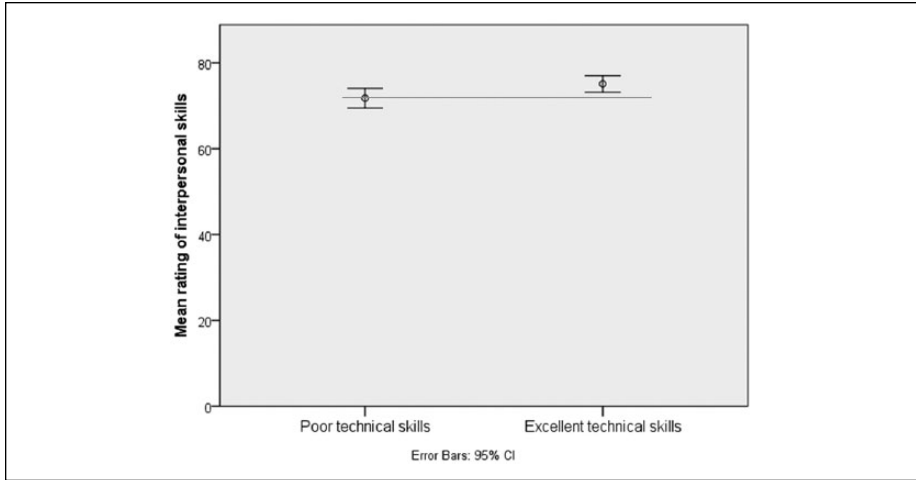
Figure 5. Mean rating that respondents assigned to the hypothetical ratee's interpersonal skills, by ratee's technical skills (experiment 5).

Note. The mean ratings of interpersonal skills are different (*p* < .001), as suggested by the fact than neither one of the CIs contains the mean performance rating of the other experimental arm (see the horizontal line). CI = Confidence Interval.

number of opportunities for the event to happen (the denominator of a ratio; e.g., Alonso & Fernandez-Berrocal, 2003; Epstein, 1994; Pacini & Epstein, 1999; Pedersen, 2017; Reyna & Brainerd, 2008). In a classic example, when asked to choose between a bowl containing one red bean out of 10 and a bowl containing nine red beans out of 100, subject preferred the latter, ignoring the fact that it offered a smaller probability of picking a red bean (9% vs. 10%; Denes-Raj & Epstein, 1994).

The cognitive-experiential self-theory (e.g., Epstein, 1994) suggests that the numerosity heuristic and the small-numbers effect nurture this systematic error. On one side, individuals better understand frequencies rather than ratios. Indeed, whereas absolute frequencies are numbers, relative frequencies entail relationships among numbers. On the other side, individuals better understand smaller rather than larger numbers. When the probability of an event is low, the numerosity heuristic and the small-numbers effect influence preferences in the same direction. For instance, when individuals are asked to choose between a 1:10 ratio and a 10:100 ratio, both facets incline them toward the 10:100 ratio because 10 is a small enough number. To the contrary, when the probability of an event is high, the two facets operate in opposite directions. If individuals are asked to select either one of a 9 in 10 ratio and a 90 in 100 ratio, the numerosity heuristic would incline them toward the ratio 90:100, whereas the small-numbers effect would induce them to choose 9:10 (e.g., Kirkpatrick & Epstein, 1992).

Experimental work has tested the denominator neglect effect on tasks such as risk understanding for medical treatments (e.g., Garcia-Retamero, Galesic, & Gigerenzer,

2010; Okan, Garcia-Retamero, Cokely, & Maldonado, 2012), disease risk rating (e.g., Yamagishi, 1997), and preferences for education policies (e.g., Pedersen, 2017).

Method

RCT 6 tests the main, simple, and interactive effects of denominator neglect and equivalence framing. Subjects were randomly assigned to one of four conditions. All respondents were asked to indicate on a 0-100 scale their propensity to implement a generic health care project. Participants randomly assigned to the *smaller absolute number of instances and positive framing group* read that in the past the project had been successful in 75 out of 100. Their counterparts in the *larger absolute number of instances and positive framing group*, instead, were told that the project had been successful in 1,500 of 2,000 health care organizations. Thus, the relative frequency of instances was held constant but the absolute frequency varied. Subjects in the *negative framing* conditions read the logically equivalent information: the project has failed in 25 out of 100 and 500 out of 2,000 health care organizations, respectively.

Results

Figure 6 shows the effects of our manipulations of the absolute frequency of instances, while keeping the relative frequency constant, and of the framing of information on respondents' propensity to implement the project, N(smaller absolute frequency; negative framing) = 142; N(smaller absolute frequency; positive framing) = 114; N(larger absolute frequency; negative framing) = 111; N(larger absolute frequency; positive framing) = 129. A two-way analysis of variance (ANOVA) showed that the propensity to implement the program was influenced by the variation of the absolute number of instances ($p = .022$) and by the framing of the information ($p < .001$) and was immune to the interactions of the two factors. More precisely, everything else being equal, the propensity to implement the project was on average 6.48 scale-points lower for employees exposed to a total number of 2,000 cases of local health authorities that have implemented the project in the past as compared to employees primed to consider 100 cases. Furthermore, keeping everything else constant, exposure to the number of local health authorities in which the project has failed in the past reduced participants' propensity to implement the project by 20.20 scale-point on average as compared to exposure to the complementary number of local health authorities in which the project has been implemented successfully, regardless of the fact that the relative frequencies were the same.

Test of the Zero-Risk Bias

Zero-risk bias refers to irrational behaviors that are triggered by the opportunity to eliminate rather than mitigate a risk. In the classic example of hazardous waste cleanup that generated cancer cases, individuals preferred the complete elimination of the risk, even when the alternative options would have produced a greater reduction in risk
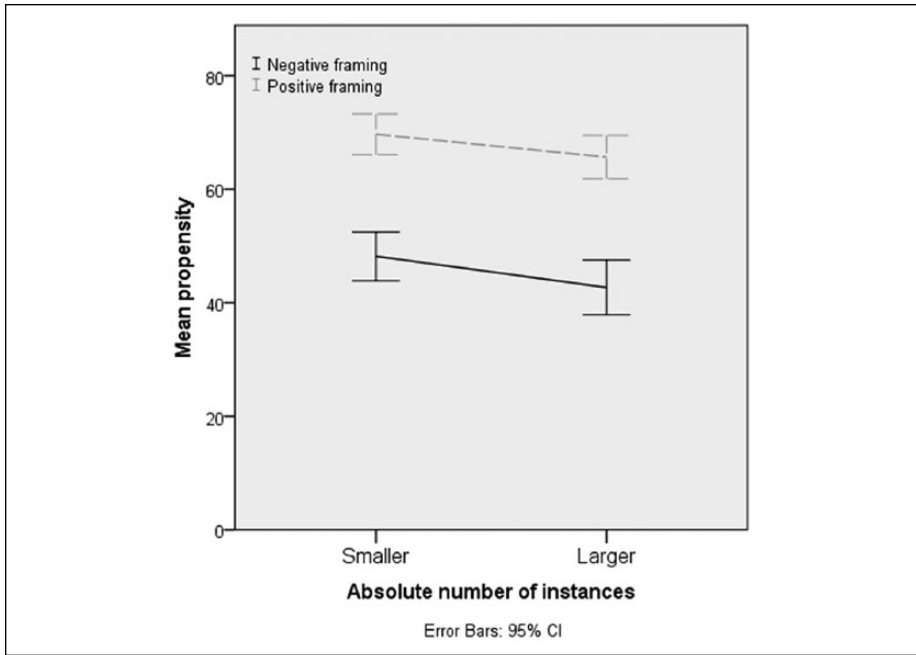
Figure 6. Two-way interaction of absolute number of instances (smaller vs. larger) and framing (negative vs. positive) on the mean propensity to implement project Beta; the relative frequency of instances is held constant across conditions (experiment 6).
Note. CI = Confidence Interval.

overall (Baron, Gowda, & Kunreuther, 1993). In another classic example of health risk, parents were willing to pay on average an additional US$2.38 to reduce the risk of insect spray inhalation poisoning and child poisoning by two-thirds and more than three times as much (i.e., US$8.09) to eliminate the same risks completely (Viscusi, Magat, & Huber, 1987). When the elimination of risks creates benefits to others, individuals' decisions may be moderated by the degree to which they are prosocially motivated. Indeed, abundant evidence has shown that prosocial motives and impact affect choices and performance in tasks that enhance others' well-being (e.g., Bellé, 2013, 2014; Bolino & Grant, 2016; Grant, 2007; Perry, Hondeghem, & Wise, 2010).

Method

Our test of the zero-risk bias draws on Baron et al. (1993) to test individuals' preference for an option that would eliminate the risk of deaths, even when the number of avoided death was smaller compared to another option that did not reduce the risk to zero. All respondents read the same scenario and picked one of two intervention plans. The scenario informed that two developing countries had been hit by an infectious disease and that the doses of the perfectly effective vaccination were limited. One

intervention plan consisted in a partial vaccination coverage in both nations. This plan would reduce deaths from 8,000 to 4,000 in one country and from 4,000 to 2,000 in the other country. The other intervention plan entailed a full vaccination coverage in one nation and a partial vaccination coverage in the other nation. This would reduce deaths from 8,000 to 7,000 in one country and from 4,000 to 0 in the other country.

Building on Bolino and Grant (2016), prosocial motivation (PSM) was measured with a 7-point Likert-type type scale (1 = *completely disagree*, 7 = *completely agree*) on several items (i.e., I put effort in my job to help others; I put effort in my job to have a positive impact on others; I am fully aware of the ways in which my job create benefit for others; I think that my job makes a positive difference in the life of others).

Results

Figure 7 reports the distribution of respondents by their preference for the intervention that did not eliminate the risk in any states but would avoid a larger number of deaths over the intervention eliminating the risk in one state and avoiding a smaller number of deaths. About 71% (N = 353) of the subjects chose the former option while 29% (N = 142) of the subjects would implement the latter option (i.e., the irrational intervention). The proportion of respondents that fell into the zero-risk trap was statistically different from the expected proportion of zero, $p < .001$.

A series of two-sample comparisons between respondents in the two groups revealed that they were statistically the same across our observed demographic variables (i.e., average age, proportion of females, distribution by number of subordinates, and job duties). PSM, instead, was on average marginally higher among employees who selected the rational intervention (i.e., the one that saved a larger number of people) than among their peers who selected the irrational intervention, $p = .010$. Furthermore, the proportion of respondents whose self-reported PSM was above the entire sample median value was larger among those who preferred to avoid a larger rather than smaller number of deaths (.58 and .44, respectively, $p = .010$) (Figure 8).

General Discussion and Implications

This study investigated the effects of a broad range of cognitive biases and debiasing interventions on public employees' decisions across realistic managerial tasks and policy areas to advance behavioral HR for the management of public organizations and their employees. Our results demonstrated that decisions were highly dependent on systematic patterns of deviation from rationality and that debiasing interventions were only sometimes successful in eliminating such deviations (Table 1).

RCTs 1 through 3 showed the effects of cognitive biases within the realm of public management decisions. In particular, in RCT 1, subjects changed their preferences toward diagnostic instruments depending on whether an irrelevant alternative, that is, the decoy, was part of the choice set, in which case the target option became more attractive as expected. In RCT 2, public workers in our sample were more likely to suggest the purchasing of an inferior diagnostic instrument when exposed to either a coercive or a
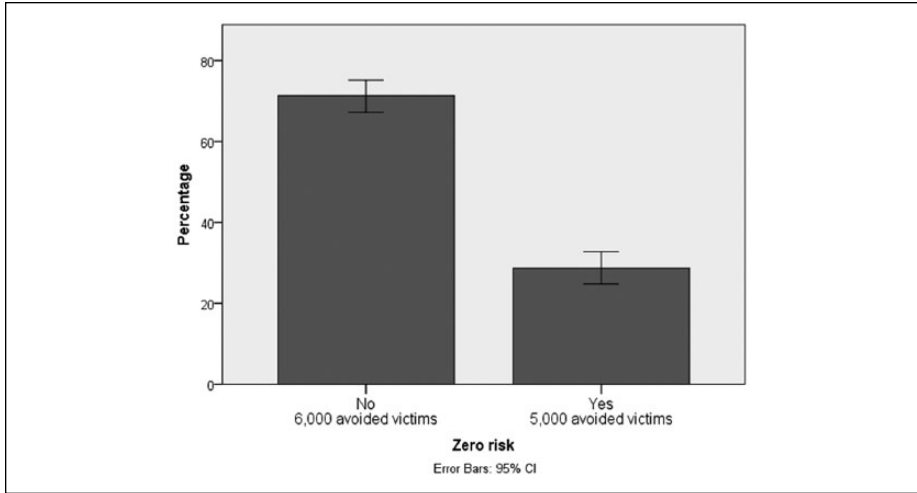
Figure 7. Distribution of respondents by preference for the elimination of risk and number of deaths avoided overall (experiment 7).
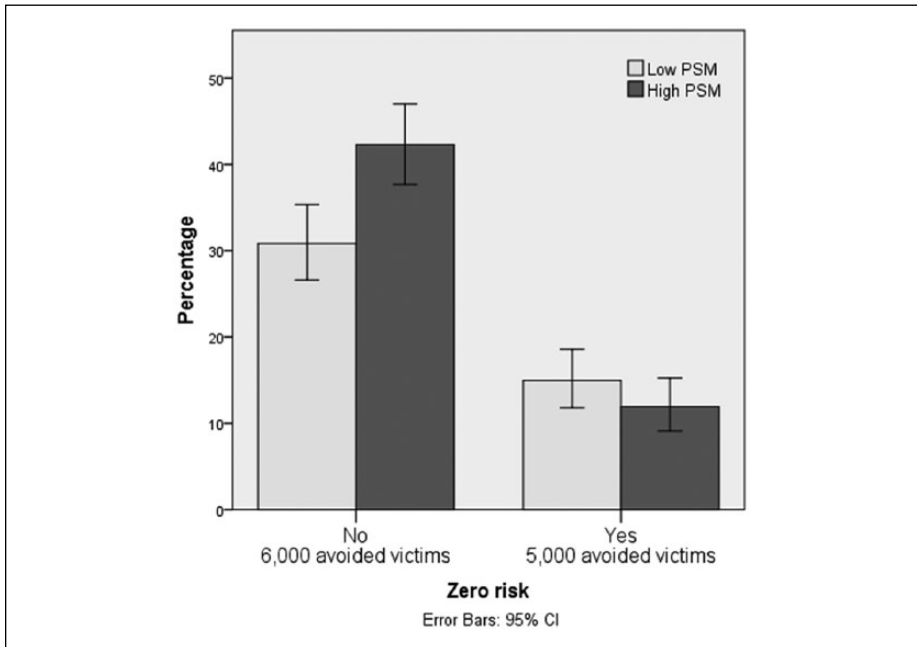*Note.* CI = Confidence Interval.

Figure 8. Distribution of respondents by (i) preference for the elimination of risk and number of deaths avoided overall and (ii) PSM.
Note. PSM = prosocial motivation; CI = Confidence Interval.

Table 1. Overview of the Study Design and Main Findings.

Cognitive bias	Evidence of predictably biased decisions	Debiasing intervention
Asymmetric dominance (Experiment 1): "An asymmetrically dominated alternative is dominated by one item in the set but not by another. Adding such an alternative to a choice set can increase the probability of choosing the item that dominates it." (Huber, Payne, & Puto, 1982, p. 90)	Yes	NA
Bandwagoning and isomorphism (Experiment 2): respectively, "Overwhelming acceptance of unproved but popular ideas" (Cohen & Rothschild, 1979, p. 531) and "An understanding of the manner in which fields become more homogenous would prevent policymakers and analysts from confusing the disappearance of an organizational form with its substantive failure" (DiMaggio & Powell, 1983, p. 158)	Yes	Successful
Framing (Experiment 3): "Surgery appeared to be much more attractive when the outcomes were framed in terms of the probability of survival rather than in terms of the probability of death" (McNeil, Pauker, Sox, & Tversky, 1982, p. 1262)	Yes	Successful
Anchoring (Experiment 4): "Different starting points yield different estimates, which are biased toward the initial values" (Kahneman & Tversky, 1974, p. 1128)	Yes	Not successful
Halo (Experiment 5): "Global evaluations alter evaluations of attributes about which the individual has information fully sufficient to allow for an independent assessment" (Nisbett & Wilson, 1977, p. 255)	Yes	Not successful
Denominator neglect (Experiment 6): "Belief that the occurrence of an event of low probability is more probable when the ratio is presented making use of large numbers rather than of small numbers" (Alonso & Fernandez-Berrocal, 2003, p. 1538)	Yes	NA
Zero risk (Experiment 7): "Preference for cleaning up a waste site completely, even if the total number of lives saved as a result of this strategy is lower than the number that would be saved if the identical resources had been expended toward cleaning up two different waste sites partially" (Baron, Gowda, & Kunreuther, 1993, p. 190)	Yes	NA

mimetic isomorphic pressure fostering that choice. Our debiasing intervention was effective, as the effect disappeared when participants were informed that people, in making decisions, might surrender to isomorphic pressures. As suggested by an anonymous Reviewer, we are unable to test whether subjects' behavior is not irrational but rather is triggered by a rationality that is alternative to purchasing the superior instrument available. For instance, one may argue that it is rational to seek legitimization through conformity to social norms. Future studies contrasting hypotheses of rationality in reacting to isomorphic pressures and/or employing mixed-methods research designs that use qualitative inquiry to better understand the causal mechanisms behind experimental results are needed. In RCT 3, negatively-framed information decreased participants' propensity to adopt a software with respect to both positively and neutrally framed information. RCTs 4 and 5 revealed that deviations from rational predictions held true in the performance appraisal domain. In the former, the average performance scores assigned to both a subordinate and oneself were higher when raters were exposed to a high, rather than low, anchor. In the latter, the average score assigned to interpersonal skills was higher when ratee's technical skills were indicated as excellent, rather than poor. Debiasing strategies were ineffective in both RCTs about performance appraisal decisions. A word of caution is required about the halo effect because one's ability to engage with patients and one's ability to perform technical tasks may neither be highly correlated nor fully independent. Nonetheless, our findings are in line with previous evidence showing that the two skills turned out to be too highly and evenly linked in respondents' estimates (e.g., Belle et al., 2017; CIPD, 2015; Holzbach, 1978; Jacobs & Kozlowski, 1985; King et al., 1980). To conclude, RCT 6 and test of the zero-risk bias demonstrated that, maybe even more worryingly, public servants are prone to cognitive fallacies when deciding about public policies. In RCT 6, indeed, the propensity to implement a new program was lower for employees exposed to a higher rather than lower absolute frequency of cases of local health authorities that have implemented the project in the past, regardless of the fact that relative frequencies were held constant. In the test of the zero-risk bias, when deciding between two vaccination plans for two nations, a significant proportion of subjects chose the irrational intervention that avoided a smaller number of deaths in total but entailed a full vaccination coverage in one nation.

Our work has relevant implications for both public administrators and policy makers who are interested in improving decision making for managing mission-driven organizations and their employees. Behavioral HR as a middle range theory in our discipline "is far from the only perspective of relevance but it is another significant string to our bow" (CIPD, 2014, p. 7). Evidence of the asymmetric dominance effect should encourage public sector professionals to analyze carefully whether a decoy is present among the options at their disposal and consider how this inferior alternative can systematically influence their decisions. Also, professionals and policy makers should recognize situations in which social pressures to conform may affect behaviors and be aware that explaining this mechanism may prevent falling prey to bandwagoning and isomorphism. Framing effects, then, can be expected anytime individuals have to base their decisions on information that can be framed equivalently in positive or negative terms and only one of the two is presented. Indeed, exposure to both types of

information eliminated the impact of attribute framing. Furthermore, an initial piece of information can significantly bias subsequent judgments through anchoring and/or halo mechanisms. Decision makers should also recognize situations that may trigger the human tendency to focus on numerators and neglect denominators, or the natural preference for the complete elimination of a risk, even when alternative options would produce a greater reduction in risk overall.

Additional efforts to design successful debiasing techniques are needed. Motivational strategies rest on the assumption that when individuals are provided with incentives that are large enough, they will pay more attention and slow-thinking will kick in. Cognitive strategies include asking individuals to consider the opposite and training, which can enhance the strategies in the slow-thinking system and help individuals understand when to use them as a fast-thinking process. Technological strategies entail supporting individuals through external tools such as decision models, decision-making software or group decision making (Larrick, 2004). More broadly, additional experimental and mixed-methods studies should investigate how re-biasing interventions and the architecture of choices can avoid or mitigate predictably irrational decisions. In this respect, the nudge theory (Thaler, 2015; Thaler & Sunstein, 2008) particularly seems to suit behavioral public HR and public administration, where high-powered incentives are rarely available. Complementing and nurturing scholarly efforts, public organizations and their managers around the world can follow the steps of the U.K. Behavioral Insight Team and create units dedicated to improving policies and services through the use of behavioral HR and the broader behavioral science.

We fully acknowledge that our work is not immune to limitations and that our results should be interpreted accordingly. For example, although we tried to be as comprehensive as possible with regard to the typology of cognitive biases, our selection might have left out other systematic errors that have the potential to impinge on rational health care decision making. Also, this study is unable to disentangle the micro-mechanisms underlying the effects of cognitive biases on public decision making. From the methodological standpoint, this work is subject to the same general limitations that affect most survey experiments. For instance, although survey experiments in real organizations score high on external validity relative to laboratory experiments, there may be legitimate concerns about whether and to what extent our findings would be replicated in more naturally occurring settings. Natural field experiments would allow better generalizability of results. On a related note, the pattern of results that we observed in our sample may vary across different types of units, treatments, operations, and settings.

Conclusion

Behavioral science suggests that judgments are systematically biased under certain circumstances and debiasing techniques may improve decision making. Our work provides three main contributions. First, from a descriptive perspective, our experimental survey-in-the-field shows that nursing personnel's decisions in the area of health care operations tended to be less rational due to the desire to conform to social pressures

(bandwagoning), when a decoy was included in the choice set, and when the framing of options was negative rather than positive. Anchoring and halo effects severely biased the assessment of subordinates' performance. Decisions in the domain of health policies were irrationally influenced by the tendency to focus on numerators and neglect denominators (denominator neglect) and by the preference for the complete elimination of a risk, even when alternative options would produce a greater reduction in risk overall (zero-risk bias). Debiasing strategies eliminated the bandwagoning and framing effects. Second, from a theoretical standpoint, our study advances behavioral HR as a middle range theory in our discipline. In fact, our endeavor is midway between grand theories that are often abstract and un-testable, at one end, and testable hypotheses that derive from empirical observations and are amenable to being rejected, at the other end. Third, from a normative viewpoint, our RCTs inform meaningful implications for practice. Policy makers and scholars alike can accumulate and systematize knowledge about the effects of cognitive biases and the architecture of choices in a public context to improve governments and civil service.

Declaration of Conflicting Interests

Funding

Supplementary Material

Supplementary material is available for this article online.

References

Abner, G. B., Kim, S. Y., & Perry, J. L. (2017). Building evidence for public human resource management: Using middle range theory to link theory and data. *Review of Public Personnel Administration*, *37*, 139-159.

Alonso, D., & Fernandez-Berrocal, P. (2003). Irrational decisions: Attending to numbers rather than ratios. *Personality and Individual Differences*, *35*, 1537-1547.

Andersen, S. C., & Hjortskov, M. (2015). Cognitive biases in performance evaluations. *Journal of Public Administration Research and Theory*, *26*, 647-662.

Ariely, D. (2010). *Predictably irrational, revised and expanded edition: The hidden forces that shape our decisions*. London, England: HarperCollins.

Baekgaard, M., & Serritzlew, S. (2016). Interpreting performance information: Motivated reasoning or unbiased comprehension. *Public Administration Review*, *76*, 73-82.

Baron, J., Gowda, R., & Kunreuther, H. (1993). Attitudes toward managing hazardous waste: What should be cleaned up and who should pay for it? *Risk Analysis*, *13*, 183-192.

Barrows, S., Henderson, M., Peterson, P. E., & West, M. R. (2016). Relative performance information and perceptions of public service quality: Evidence from American school districts. *Journal of Public Administration Research and Theory*, *26*, 571-583.

Battaglio, R. P. (2014). *Public human resource management: Strategies and practices in the 21st century*. Thousand Oaks, CA: SAGE.

Belle, N. (2013). Experimental evidence on the relationship between public service motivation and job performance. *Public Administration Review*, *73*, 143-153.

Belle, N. (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*, 109-136.

Belle, N., Cantarelli, P., & Belardinelli, P. (2017). Cognitive biases in performance appraisal: Experimental evidence on anchoring and halo effects with public sector managers and employees. *Review of Public Personnel Administration*, *37*, 275-294. doi:10.1177/07343 71X17704891

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica: Journal of the Econometric Society*, *22*, 23-36.

Blumenthal-Barby, J. S., & Krieger, H. (2015). Cognitive biases and heuristics in medical decision making: A critical review using a systematic search strategy. *Medical Decision Making*, *35*, 539-557.

Bolino, M. C., & Grant, A. M. (2016). The bright side of being prosocial at work, and the dark side, too: A review and agenda for research on other-oriented motives, behavior, and impact in organizations. *Academy of Management Annals*, *10*, 599-670.

Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. *Journal of Applied Psychology*, *60*, 556-560.

Charbonneau, É., & van Ryzin, G. G. (2015). Benchmarks and citizen judgments of local government performance: Findings from a survey experiment. *Public Management Review*, *17*, 288-304.

Chartered Institute of Personnel and Development. (2014). *Our minds at work: Developing the behavioural science of HR* (Championing better work and working lives). Retrieved from https://www.cipd.co.uk/Images/our-minds-at-work_2014-developing-behavioural-science-hr_tcm18-9731.pdf

Chartered Institute of Personnel and Development. (2015). *A head for hiring: The behavioural science of recruitment and selection* (Championing better work and working lives). Retrieved from https://www.cipd.co.uk/Images/a-head-for-hiring_2015-behavioural-science-of-recruitment-and-selection_tcm18-9557.pdf

Cohen, L., & Rothschild, H. (1979). The bandwagons of medicine. *Perspectives in Biology and Medicine*, *22*, 531-538.

Colman, A. (2003). Group think: Bandwagon effect. In Andrew Colman (Ed.), *Oxford dictionary of psychology* (p. 77). Oxford, UK.

Cornelissen, J. P., & Werner, M. D. (2014). Putting framing in perspective: A review of framing and frame analysis across the management and organizational literature. *Academy of Management Annals*, *8*, 181-235.

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology*, *66*, 819-829.

DiMaggio, P., & Powell, W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*, 147-160.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*, 709-724.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*, 35-42.

Garcia-Retamero, R., Galesic, M., & Gigerenzer, G. (2010). Do icon arrays help reduce denominator neglect? *Medical Decision Making*, *30*, 672-684.

Gardner, D. (2009). *Risk: The science and politics of fear*. Random House. London, UK: Virgin Books.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, UK: Cambridge university press.

Grant, A. M. (2007). Relational job design and the motivation to make a prosocial difference. *Academy of Management Review*, *32*, 393-417.

Grimmelikhuijsen, S., Jilke, S., Olsen, A. L., & Tummers, L. (2017). Behavioral public administration: Combining insights from public administration and psychology. *Public Administration Review*, *77*, 45-56.

Holzbach, R. L. (1978). Rater bias in performance ratings: Superior, self-, and peer ratings. *Journal of Applied Psychology*, *63*, 579-588.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*, 90-98.

Ingraham, P. W., & Rubaii-Barrett, N. (2007). *Foundations of Public Administration Series: Human resource management as a core dimension of public administration*. Washington, DC: American Society for Public Administration.

Jacobs, R., & Kozlowski, S. W. (1985). A closer look at halo error in performance ratings. *Academy of Management Journal*, *28*, 201-212.

James, O. (2010). Performance measures and democracy: Information effects on citizens in field and laboratory experiments. *Journal of Public Administration Research and Theory*, *21*, 399-418.

Jilke, S., Van Ryzin, G. G., & Van de Walle, S. (2016). Responses to decline in marketized public services: An experimental evaluation of choice overload. *Journal of Public Administration Research and Theory*, *26*, 421-432.

Jones, B. D. (2003). Bounded rationality and political science: Lessons from public administration and public policy. *Journal of Public Administration Research and Theory*, *13*, 395-412.

Kahneman, D. (2002, December). Nobel prize lecture: Maps of bounded rationality: A perspective on intuitive judgment and choice. *Nobelprize.org*. Retrieved from https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2002/kahnemann-lecture.pdf (accessed 11 May 2018).

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Macmillan.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, *47*, 263-292.

King, L. M., Hunter, J. E., & Schmidt, F. L. (1980). Halo in a multidimensional forced-choice performance evaluation scale. *Journal of Applied Psychology*, *65*, 507-516.

Kirkpatrick, L. A., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: Further evidence for two conceptual systems. *Journal of Personality and Social Psychology*, *63*, 534-544.

Larrick, R. P. (2004). Debiasing. In Rachel A. Powsner & Edward R. Powsner (Eds.), *Blackwell handbook of judgment and decision making* (pp. 316-338).

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, *76*, 149-188.

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science*, *4*, 390-398.

Marvel, J. 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*, 209-227.

Marvel, J. D. (2016). Unconscious bias in citizens' evaluations of public sector performance. *Journal of Public Administration Research and Theory*, *26*, 143-158.

McNeil, B. J., Pauker, S. G., Sox, H. C., Jr., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *The New England Journal of Medicine*, *306*, 1259-1262.

Moynihan, D. P., & Lavertu, S. (2012). Does involvement in performance management routines encourage performance information use? Evaluating GPRA and PART. *Public Administration Review*, *72*, 592-602.

Nielsen, P. A. (2014). Learning from performance feedback: Performance information, aspiration levels, and managerial priorities. *Public Administration*, *92*, 142-160.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*, 250-256.

Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2012). Individual differences in graph literacy: Overcoming denominator neglect in risk comprehension. *Journal of Behavioral Decision Making*, *25*, 390-401.

Olsen, A. L. (2015a). Citizen (dis)satisfaction: An experimental equivalence framing study. *Public Administration Review*, *75*, 469-478.

Olsen, A. L. (2015b, September). *Negative performance information causes asymmetrical evaluations and elicits strong responsibility attributions*. San Francisco, California: Annual meeting of the American Political Science Association.

Olsen, A. L. (2017a). Compared to what? How social and historical reference points affect citizens' performance evaluations. *Journal of Public Administration Research and Theory*, *27*, 562-580.

Olsen, A. L. (2017b). Human interest or hard numbers? Experiments on citizens' selection, exposure, and recall of performance information. *Public Administration Review*, *77*, 408-420.

Organization for Economic Cooperation and Development. (2017). *Behavioural insights and public policy: Lessons from around the world*. Paris, France: Author. doi:10.1787/9789264270480-en

Pacini, R., & Epstein, S. (1999). The interaction of three facets of concrete thinking in a game of chance. *Thinking & Reasoning*, *5*, 303-325.

Pedersen, R. T. (2017). Ratio bias and policy preferences: How equivalency framing of numbers can affect attitudes. *Political Psychology*, *38*, 1103-1120.

Perry, J. L. (1993). Strategic human resource management. *Review of Public Personnel Administration*, *13*(4), 59-71.

Perry, J. L. (2010). A strategic agenda for public human resource management research. *Review of Public Personnel Administration*, *30*, 20-43.

Perry, J. L., Hondeghem, A., & Wise, L. R. (2010). Revisiting the motivational bases of public service: Twenty years of research and an agenda for the future. *Public Administration Review*, *70*, 681-690.

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, *18*, 89-107.

Saposnik, G., Redelmeier, D., Ruff, C. C., & Tobler, P. N. (2016). Cognitive biases associated with medical decisions: A systematic review. *BMC Medical Informatics and Decision Making*, *16*(1), Article 138.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129-138.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*, 755-769.

Thaler, R. (2015). *Misbehaving: The making of behavioral economics*. New York, NY: W.W. Norton.

Thaler, R. (2017, December). Nobel prize lecture: From cashews to nudges: The evolution of behavioral economics (Nobel Media AB 2014). *Nobelprize.org*. Retrieved from http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2017/thaler-lecture.html

Thaler, R., & Benartzi, S. (2001). *Save more tomorrow: Using behavioral economics to increase employee saving*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.153.5567

Thaler, R., & Sunstein, C. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. London, UK: Penguin Books.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*, 25-29.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453-458.

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, *39*, 1179-1189.

Viscusi, W. K., Magat, W. A., & Huber, J. (1987). An investigation of the rationality of consumer valuations of multiple health risks. *The RAND Journal of Economics*, *18*, 465-479.

Yamagishi, K. (1997). When a 12.86% mortality is more dangerous than 24.14%: Implications for risk communication. *Applied Cognitive Psychology*, *11*, 495-506.

Author Biographies

Paola Cantarelli is a postdoctoral scholar at the Scuola Superiore Sant'Anna (Pisa, Italy). Her current research focuses on work motivation and behavioral HR in public organizations.

Nicola Belle is an assistant professor at the Scuola Superiore Sant'Anna (Pisa, Italy). His research focuses on behavioral HR and behavioral management in mission-driven organizations.

Paolo Belardinelli is a PhD candidate at Bocconi University (Milan, Italy). His research focuses on behavioral public administration.