



AARHUS UNIVERSITY



Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Bækgaard, M., & Serritzlew, S. (2016). Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension. *Public Administration Review*, 76(1), 73-82. DOI: 10.1111/puar.12406

Publication metadata

Title:	Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension
Author(s):	Bækgaard, M., & Serritzlew, S.
Journal:	<i>Public Administration Review</i> , 76(1), 73-82
DOI/Link:	http://dx.doi.org/10.1111/puar.12406
Document version:	Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension?

Abstract

One main rationale of performance information is to empower citizens to make informed decisions by presenting them with unambiguous information about the performance of institutions. However, even objective, clear and unambiguous performance information is subject to biased interpretation depending on whether the information is consistent with the prior beliefs held by those who receive the information. By integrating the theory of motivated reasoning with the literature on performance information we hypothesize that performance information that is inconsistent with prior beliefs is less likely to be interpreted correctly than belief-consistent information. We show, based on randomized survey experiments in which respondents were presented with quantitative performance data, that subjects systematically interpret performance information in ways that ensure conformity with their prior beliefs. The findings question the assumption that providing performance information automatically increases knowledge about government performance, let alone improves political decisions.

Practitioner points:

- Citizens' interpretations of performance information are systematically biased and depend on their prior beliefs.
- Ideological beliefs are important to how citizens perceive performance of public and private organizations and their performance perceptions are therefore not only a product of actual differences in performance.
- Policy makers should bear in mind that performance information is likely to be systematically misinterpreted by citizens, limiting the payoff from providing citizens performance information.

Public service performance has become such a central topic in the public debate that it is sometimes characterized as a performance movement where one primary concern has been how to measure and compare performance in public organizations (Radin 2006; Van Dooren 2011; Nielsen 2013a). Performance measures may be used for many different reasons, one main reason being to enable citizens to choose between alternative public providers or to pressure decision-makers to improve performance (Van de Walle and Roberts, 2011: 222). Performance information is also assumed to improve political decisions and accountability (Moynihan 2008: 27; Charbonneau & Van Ryzin 2015). Consequently, there has been an expansion in the use of quantified performance information in the public sector (Van de Walle and Roberts 2011: 215).

The value of performance information as a means to enlighten decision makers and citizens has, however, been contested in recent research on comparisons, benchmarks, and quantitative performance information in which it is shown that the way information is presented matters to people's interpretations even in cases where the content of the information is exactly the same. Among the factors relevant to how the information is interpreted is, for instance, the format of the information (Jacobsen et al. 2014), the framing of the information (Olsen 2015), whether it is accompanied by other relevant information (Baekgaard 2015), and whether benchmarks are presented in which performance is compared across different organizations or over time for the same organization (Charbonneau and Van Ryzin 2015).

Similarly, Moynihan (2006) points to the fact that performance information is often ambiguous and open to more than one interpretation. The ambiguity of performance information allows individuals to focus on aspects of the information that are in accordance with their own viewpoints while discordant aspects are ignored or discarded. In Moynihan's words (2006: 155), there can be "incentives for particular actors to advance arguments that reflect their institutional role and context, enhancing the potential for disagreement." Hence, prior beliefs, which can reflect a citizen's

political values, ideology, beliefs about states of the world etc., are likely to have an impact on how ambiguous performance information is interpreted and presented.

In this article we develop a theory of when and how the interpretation of performance information is contingent upon prior beliefs. According to motivated reasoning, a concept from a psychological theory on how information is processed and understood (Kunda 1990; Taber and Lodge 2006; Druckman 2012; Kahan et al. 2013), the interpretation of performance information is contingent upon prior beliefs. We test this idea in two randomized survey experiments of a representative sample of Danish citizens and show that performance information is not merely a technical instrument which automatically increases knowledge of performance. Even unambiguous performance information is subject to interpretation and is distorted by prior beliefs.

The survey experiments differ from other survey experiments of how politicians and citizens respond to performance information (see, for instance, James 2011; Baekgaard 2015; Nielsen and Baekgaard 2015) in that we focus on subjects' ability to correctly assess information rather than on their attitudinal responses. Specifically, our dependent variable measures whether subjects, based on performance information, are able to point out which of two organizations is performing best in a case where the subjects are presented with clear and unambiguous information. To our knowledge this is the first study to systematically examine the relationship between prior beliefs and the ability to correctly interpret performance information. We believe that the results suggest that the literature on performance information should be reoriented to take the impact of prior beliefs into account.

The article proceeds with a theoretical discussion of the effects of prior beliefs on the interpretation of performance information. This is followed by a discussion of why the research question is appropriately addressed by means of a survey experimental design. Next, we discuss the specific design of two survey experiments developed specifically for this purpose. We pay particular attention to how placebo groups can be used to estimate the impact of prior beliefs in cases like this

where it is impossible to randomly assign people to different prior beliefs. Next, we present and discuss the findings of our empirical analysis. The final section concludes and discusses the implications of our study for practice and future research.

Performance information and prior beliefs

The introduction of performance information in public administrations was based on the idea that information on organizational performance may improve decision making and ultimately lead to greater public value for taxpayer money (Kettl 1997: 457; Moynihan 2006: 152). In order to attain this goal, performance information usually takes the form of quantitative indicators of performance although a more qualitative assessment of performance is also often carried out (Heinrich 2012). In practice, performance information may describe outputs, outcomes, and/or responsiveness of public services (Nielsen 2013a: 13).

Two characteristics of performance information are particularly important for how it is interpreted by stakeholders. First, for performance information to be interpretable, the disclosed performance level has to be easily comparable to the performance level of similar organizations (“social comparisons”) or the organizational performance in previous years (“historical comparisons”) (Nielsen 2013b; see also Charbonneau and Van Ryzin 2015 for similar arguments but slightly different concepts). Without such benchmarks it is difficult to gauge whether a given performance indicator of the absolute performance level is showing acceptable performance or whether changes are required to increase future performance. Second, performance information may be simple and consist of only one indicator, or it may be complex and cover multiple dimensions or indicators of performance at once. In the former case, it is at the outset easy to interpret performance, but stakeholders may decide to discard the information in case it does not fit with their values based on the reasoning that it does not provide a full picture of performance. In the latter case, the performance information will show in many cases that the organization is performing well on some aspects and less well on other. The very fact that performance in most cases is a multi-dimensional phenomenon means that

stakeholders may decide to emphasize those aspects of performance that fit best with their prior values and beliefs while others are discarded. Thus, regardless of the complexity, an inherent characteristic of performance information is that it is open to political interpretation (Moynihan 2006; 2008).

As argued by Moynihan (2006: 157-58) actors always have existing political values and beliefs. These beliefs may not only affect the tendency to focus on some aspects of performance information or the tendency to point to some dimensions covered by performance information as being the most important while others are discarded. On a more fundamental level, they may also affect individuals' ability to correctly interpret even unambiguous performance information. The rational model of man suggests that actors will use new performance information to update their view of performance. According to this notion, we should not expect prior beliefs to affect how unambiguous performance information is interpreted. A rational actor should be careful to avoid any bias in interpretation. A new piece of information loses value if it is systematically misinterpreted. However, even when actors intend to be rational, and when they believe that they are, the assumption of perfect rationality is not descriptively realistic (Simon, 1957: 241 ff.; Jones, 2001: 54ff.). In particular, prior beliefs tend to systematically bias how new information or arguments are evaluated. According to Taber and Lodge (2006), this is an automatic process. People are simply "largely unaware of the power of their priors" (Taber and Lodge, 2006: 757).

To see how prior beliefs can shape and bias the interpretation of performance information, it is useful to integrate the psychological theory of motivated reasoning with the literature on performance information. Motivated reasoning has proven instrumental to understanding many phenomena in politics. For example, new information on a candidate for public office is likely to be interpreted in light of existing opinions (Druckman, 2012: 201). If the view on the candidate is generally favorable, new information is likely to be interpreted as positive. If not, it is more likely to be interpreted as negative. A similar logic applies to citizens' evaluations of new policy proposals. If a

new policy proposal is sponsored by a party that is preferred by a voter, that voter is more likely to be favorable to the proposal. If the same idea is sponsored by a disliked party, the voter will more likely be skeptical. Common to these examples is that motivated reasoning is key to understanding what is going on: New information is, according to this perspective, not just taken in. It is perceived from the perspective of an individual with prior beliefs, and this affects how the information is perceived and interpreted. In an experimental study of opinion formation, Taber and Lodge (2006) show that prior opinions predict what kind of information actors look for (they choose information consistent with prior beliefs (2006: 764)), and how they deal with it (they spend time and effort refuting arguments that are inconsistent with prior beliefs while they uncritically accept those that are consistent (2006: 762)).

We expect that motivated reasoning is equally important to understand how actors, when they have relevant prior beliefs, interpret performance information, even when the information is completely unambiguous. According to the theory of motivated reasoning, prior beliefs systematically affect how information is understood. Performance information can be seen as new information, or updated information, on the performance of an entity, typically a government institution. Of course, such information cannot be transformed immediately into action. Nor does it have a direct impact on individuals' attitudes. It must be processed and understood, and this happens via reasoning.

However, human reasoning is not neutral. Reasoning always has a motive. One motive could be accuracy. Ask a student to complete a test, and you can expect that the student is motivated to be accurate. Ask an avid smoker for his opinion on the latest study of long-term health effects of smoking and you will often get something different: A smoker may come up with a biased interpretation of the data, perhaps indicating that smoking is not too dangerous. According to the theory of motivated reasoning, humans may process new information by reasoning driven either by accuracy goals or by directional goals (Kunda, 1990: 481-483). These two modes of reasoning involve different mechanisms. When reasoning is driven by accuracy goals, people tend to spend more time

reflecting, to employ more complex strategies in their thinking and to consider more alternative explanations (Kunda, 1990: 482).

However, when reasoning is driven by directional goals, the same people are motivated to arrive at a particular conclusion while still making sense of the information presented to them. This does not require conscious effort; actors can engage in motivated reasoning without being aware of it.

According to Westen et al. (2006: 1947), motivated reasoning can even “be viewed as a form of implicit affect regulation in which the brain converges on solutions that minimize negative and maximize positive affect states”. Kunda (1990: 482) proposes that “people motivated to arrive at a particular conclusion attempt to be rational and to construct a justification of their desired conclusion that would persuade a dispassionate observer”. Motivated (directional) reasoning occurs in different ways (Kunda, 1990: 483-492). First, people may, when induced to perform acts contrary to their attitudes, alter their prior attitudes to obtain consistency. For instance, Linder et al. (1967) show experimentally that subjects after endorsing a law limiting free speech were less opposed to the law than a control group. Second, people can, when presented with information that challenges their beliefs, selectively use statistical heuristics to obtain consistency. For example, Kahan et al. (2013: 18) shows that liberal democrats are more likely than conservative republicans to correctly interpret statistical information indicating that gun control decreases violence. However, if the statistical information indicates that gun control increases violence, conservative republicans are more likely than liberal democrats to correctly interpret the information. Hence, motivated reasoning theory suggests that it is necessary to take prior beliefs and values into account if one wishes to understand how actors interpret new information.

This also applies to performance information on public organizations. The public sector is far from neutral. Most issues in the public sector are or at least can be political and controversial in the sense that people tend to have opinions about them and often disagree on them. Take as an example the issue of whether production should be public or private. Actors have different views on the relative

efficiency of public and private organizations (Andersen and Hvidman 2013), public and private employment represents an important cleavage in voting (Tepe, 2012), and views of the balance between public and private service provision is a classic contested political issue. Hence, public performance information provides information about a politically sensitive issue: Actors are likely to hold strong, and different, prior beliefs. The predictions of the theory of motivated directional reasoning are twofold. First, when confronted with new information that is inconsistent with prior beliefs, actors will tend to selectively use heuristics to interpret the information in a way that is consistent with prior beliefs. Second, when the information is presented by a source with which the actor tends to agree, the actor will be inclined to accept the information and thus interpret it with less bias. In case of general disagreement with the sender, the actor will be more skeptical and hence tend to give the information a biased interpretation. We use the public/private example in the two survey experiments below to test these expectations.

Table 1 about here

Table 1 shows how, according to motivated reasoning theory, actors will interpret unambiguous performance information, conditional on their prior beliefs about public and private service provision. We expect that actors will systematically misinterpret performance information whenever the performance information is inconsistent with their prior beliefs.

General design considerations

To test whether performance information is subject to systematic misinterpretation according to prior beliefs, we need to distinguish between subjects with different prior beliefs who have been presented with unambiguous performance information. One empirical strategy to obtain such data would be to identify and compare cases in which subjects are presented with information about the performance of real public institutions and then measure how they interpret this information by

surveying their interpretation of the performance. However, three fundamental challenges would be very difficult to overcome in this observational design.

First, we cannot rule out the possibility that selection bias creates an artificial link between subjects' values and their ability to correctly interpret performance information. Thus, any relationship found between values and interpretations of performance information may be due to observed or unobserved characteristics that affect both.¹ If such characteristics are not accounted for, we would risk making false conclusions based on spurious relationships. Second, subjects' ability to correctly interpret performance information may affect their beliefs and political values in case their interpretations consistently show that, for instance, private organizations are doing better than their public counterparts. In such cases we would risk drawing wrong conclusions due to simultaneity bias. Third, at a more basic level, it is likely to be hard to identify a case in which subjects are informed to the same extent, where the information has one and only one correct interpretation, and where we can know that subjects make an active effort to interpret the information.

We therefore turn to randomized survey experiments. Survey experiments (see Sniderman 2011) allow us to survey subjects' prior beliefs and to present all of them with the same unambiguous performance information. Usually, researchers control the independent variable in survey experiments by randomly assigning subjects to treatment and control groups. Any difference in the outcome variable can then be ascribed to the treatment alone and the internal validity of survey experiments is thus usually considered high compared to observational studies. It is, however, impossible to induce prior beliefs by randomly assigning people to treatment and control groups and we therefore rely on an alternative strategy. This strategy involves using placebo groups. The basic idea is that prior beliefs should only be expected to have some impact on subjects' ability to correctly interpret performance information in cases where the issue that the performance information covers to some extent is politicized. By comparing the impact of prior beliefs on performance evaluations in placebo and treatment groups that are identical except for a difference in the extent to which the

issue is politicized, we are able to estimate the causal impact of prior beliefs on subjects' ability to interpret performance information correctly.

This requires politicized and neutral versions of the survey experiments. In the politicized version, we present subjects in the treatment groups with performance information on public and private organizations. As argued above, there is a cleavage between people's attitudes about whether public or private organizations perform best in terms of producing public services. This cleavage is typically strongly correlated with people's more general ideological attitudes (see, for instance, Christensen and Lægreid 2003: 19) and the issue can thus be said to be strongly politicized. In the politically neutral versions, we rely on a comparison between two public organizations. If nothing else is known about the two public organizations than their performance, we would not expect prior beliefs to have any impact on how performance is interpreted. Comparing the extent to which subjects are able to correctly interpret performance information in a case where they are presented with a comparison of the performance of a public and private organization with another scenario in which they are presented with information on two public organizations allows us to test for the causal impact of prior beliefs on interpretations. As subjects in both experiments we use a representative sample of Danish citizens, recruited from an internet panel. We return to this in the following sections.

Design: Experiment 1²

The pool of respondents for the first survey experiment consists of 1,784 Danish citizens selected from an internet panel.³ Stratified random sampling was used to ensure that the citizens were representative of the Danish population at large in terms of gender, age, and geographical location. The survey experiment was conducted as an online survey in May 2014.⁴ The respondents were randomly assigned to one of two treatment groups or one of two placebo groups. A regression

analysis which is reported in the appendix and in which the composition of the four groups is compared one by one with the others shows that the experimental groups balance (gender, age, level of education, party affiliation, and attitudes towards the public sector).

As mentioned in the previous section, the respondents in the treatment groups were presented with performance information for either a public or a private organization. To ensure comparability, we chose hip operations in a public and a private hospital. Hip operation procedures are very alike at public and private hospitals in Denmark (Andersen and Jakobsen, 2011: 963) and performance, measured as re-hospitalizations and post-surgery complications, is, on average, very similar in public and private hospitals when patient composition is accounted for (Andersen and Jakobsen, 2011: 970). Respondents were then asked to evaluate the performance of the organization. In both treatments the respondents were given the following instruction:

“We will now ask you to consider the following constructed example. Below you see a table with information on the quality of hip operations at a public and a private hospital. The patient groups at the two hospitals are very similar. The table shows how often hip operations were performed without and with complications at the two hospitals.

Which of the hospitals has performed best?”

Table 2 about here

“Don’t know” was not a response option in any of the cases. The respondents were thus forced to give their best estimate of which hospital was performing best.⁵

The treatment and placebo groups are shown in Table 2. The only difference between the treatments is that the number of operations with and without complications has been switched. Hence, in treatment 1, complications are much more likely to occur in the public hospital than in the private

hospital (a rate of 25 pct. compared to 16 pct.). Thus, the correct evaluation here would be that the private hospital is doing better.

Based on the responses to the constructed example, we computed the dependent variable which takes the value '1' if the respondent was able to correctly identify the best performing hospital and the value '0' if the wrong hospital was picked. Since the dependent variable is binary, we used logistic regression for the analysis of the experiment.

In treatment 2, complications are much more likely to occur at the private hospital (84 pct. compared to 75 pct.). Thus, the difference between success rates is the same in both treatments (9 percentage points). According to the expectation, individuals who are more sympathetic to public service provision should be more likely to misinterpret performance information as favorable to the first hospital in the first treatment (the information shows that the second hospital outperforms the first, but the first hospital is public). Conversely, in the second treatment, where the performance information indicates that the public hospital outperforms the private hospital, supporters of public provision are expected to be more likely to interpret the information correctly.

Experiment 1 was designed in this manner to ensure that the numbers mentioned were kept constant across treatments as are also the order in which the public and private hospitals are mentioned. One shortcoming, however, is that findings might be susceptible to the reversed order of the numbers. Experiment 2 is designed to address exactly this problem, and we will thus discuss this issue in more detail below.

To ensure that our findings from the analysis of treatment groups do in fact reflect a causal impact of prior beliefs on interpretations and are not, for instance, an artifact of the particular performance numbers reported in the treatments, we designed a placebo test. The subjects were presented with the exact same performance information as in the treatments, except that we replaced the words "Public hospital" and "Private hospital" with "Hospital A" and "Hospital B". Apart from this, the

placebo groups are identical to the treatment groups. Hence, while we in the treatment groups expect to see systematic differences in performance evaluations between subjects with different prior beliefs, we expect to see no such differences in the placebo groups, as no reference is made to public or private organizations. Consequently, we would expect prior beliefs (measured as attitudes to public sector service provision) to have a statistically significantly different impact on performance evaluations in comparisons between treatment 1 and placebo 1 and between treatment 2 and placebo 2, respectively. Our independent variable, attitudes to public service provision, is measured by the following questions:

Do you disagree or agree with the following statements?

1. Many public activities could be produced both better and more cheaply by private providers.
2. We should to a larger degree contract out public services (such as child care, elderly care, and hospital treatments).
3. The public sector is best at providing public services.

The response options were: Completely agree, partly agree, neither agree nor disagree, partly disagree, completely disagree. A factor analysis shows that the items load highly on one dimension after the third item has been reversed (factor scores are all above 0.7) and we thus construct an additive index. With a Cronbach's alpha of 0.88 we consider the index highly reliable. The index was rescaled to run between 0-100. Higher values on the index correspond to a stronger preference for public service provision. The index correlates highly (Pearson's $R = 0.53$) with the respondents' self-reported party affiliation (a dummy variable indicating whether they voted for a left-wing party at the latest local government elections in November 2013), and we therefore conclude that the index to a high extent captures important ideological differences between respondents.⁶ Responses to the index cover the whole range from 0-100. With a mean of 61 and a standard deviation of 29 there is a lot of variation in subject attitudes.

Findings: Experiment 1

The analysis of the experiment is presented in Table 3. Since we expect prior beliefs to have a different impact in different experimental arms, we use regression analysis to analyze our data. Model 1 and 2 examine the impact of prior beliefs (attitudes to public service provision) in the two treatment groups, while Model 3 and 4 cover the placebo groups. It appears that subjects' prior beliefs have a negative impact on their ability to interpret performance information correctly if the private hospital is actually performing best (treatment 1, presented in model 1). In terms of substantial significance, simple descriptive statistics show that 86 percent respond correctly among those who are most pro private service provision, while only 46 percent respond correctly among those who are most pro public service provision. Furthermore, a positive impact of a prior positive attitude to public service provision is found if the public hospital performs best (treatment 2, presented in model 2). Again, the findings are substantially significant although the effect is somewhat weaker. Among those who are most negative towards public service provision, 59 percent respond correctly, compared to 75 percent among those who are most positive towards public service provision. Finally, there is no impact of prior beliefs in the placebo groups (models 3 and 4). In sum, the findings thus support the proposition that people are better able to interpret performance information correctly if the information is in accordance with their prior beliefs and less able if the information discords with prior beliefs.

Table 3 about here

Model 5 and 6 compare the impact of attitudes towards public service provision in each of the treatment groups with the impact of attitudes in the corresponding placebo group. It is found in Model 5 that attitudes play a significantly different role in the responses to treatment 1 than in the placebo group. A similar statistically significant difference between treatment and placebo is, however, not identified between the responses to treatment 2 and placebo group 2. The finding that attitudes to service provision do not have a different impact on response to treatment 2 and placebo

group 2 may have to do with the design of these particular groups. For instance, the fact that the number of people who experience post-operative complications is much larger than the number who do not, may introduce a lot of noise in the numbers if people tend to misread the information. In order to examine whether this is the case we conducted a second experiment with stable numbers across groups. Furthermore, information is seldom provided in real world situations without some additional information about who produced the information. As described in the theoretical section, we expect that the attitude towards the sender of the information is important to the degree of misinterpretation of information. The second experiment was therefore designed to also examine this effect.

Design: Experiment 2

This experiment uses one placebo group and four treatment groups. The 1,416 respondents who participated in this experiment in early July 2014 were randomly assigned to the five groups. As the balance test in the appendix shows, the five groups are slightly imbalanced with regard to respondents' age, educational level, and prior beliefs (measured as attitude towards public service provision). We included individual level controls in our analyses to account for these differences. Since the controls do not alter the main findings about the impact of political attitudes on interpretation of performance, we present the analyses without controls.

The subjects in the experiment were given the following instructions where text in normal writing was given to all subjects while italicized text was given only to selected groups as indicated in parentheses (P1= placebo group 1, T1 = treatment group 1 etc.).

“We now ask you to consider the following constructed example. Below you see a table [*produced by the right-wing think tank CEPOS (T4; T5)*]. The table shows the number of students in [*two public schools (P1)/a public and a private school (T2; T3; T4; T5)*] who passed the final exams in Danish and math in ninth grade in 2013. The two schools are very similar in terms of student body composition.

	Passed Danish and math	Failed Danish and math
<i>Public school A (P1)</i> <i>The public school (T2 and T4)</i> <i>The private school (T3 and T5)</i>	548	133
<i>Public school B (P1)</i> <i>The private school (T2 and T4)</i> <i>The public school (T3 and T5)</i>	112	16

What school performed best?”

The differences between the five groups are shown in Table 4.

Table 4 about here

Like experiment 1, this experiment studies the impact of attitudes to public service provision on the ability to interpret the information about public and private organizations correctly and the dependent variable thus also here is binary and takes the value 1 for those respondents who were able to correctly identify the best performer. The main aim of the two experiments is thus identical, but experiment 2 differs from experiment 1 in important respects. First, in contrast to experiment 1 in which we reversed the order of the numbers in comparisons of treatment groups, we in this experiment reversed the order in which the public and private and public organizations were mentioned. Thus, the only difference between treatment 2 and 3 (and between treatment 4 and 5) is that public schools are mentioned in the upper row in treatment 2, while private schools are mentioned in the upper row in treatment 3. This allows us to test whether the pattern found in experiment 1 is a product of reversing the numbers. Second, experiment 2 focuses on schools, whereas experiment 1 focused on hospitals and we furthermore changed the reported performance numbers somewhat to match actual school performance. We decided on these changes to test whether the findings in experiment 1 were unique to this particular design. As for the hospital service in experiment 1, tasks and performance data are highly comparable across public and private schools. In particular, all students are required, with very few exceptions, to finalize ninth grade by passing the same standardized tests in written Danish and math. We thus use actual numbers for

passing scores to create as realistic a setting as possible.⁷ Third, the difference in performance rates between the organizations reported was a little lower in this experiment than in experiment 1 (7 percentage points) with the school with fewest students as the best performer. We should therefore expect that our subjects find it even harder to interpret the information correctly in this experiment. Fourth, treatment 4 and 5 are identical to treatment 2 and 3 with one exception: a sender of the information was included in order to examine the impact of adding a sender of the information to the treatment.⁸ Specifically, we added the pro-market think tank CEPOS as sender of the information. CEPOS' ideological position is reflected in the information it provides, which is typically very much in support of private provision of public services. Based on previous studies of motivated reasoning in the political arena our expectation was that information would be perceived as being less valid among supporters of public service provision if the information shows that the private school is performing better and that their ability to interpret performance information correctly would therefore be even smaller in this case. Furthermore, we expected supporters of private service provision to be even better at interpreting the information in this case because mentioning CEPOS serves as a heuristic which allows them to tease out the right response.

As in experiment 1 we used the index of attitudes towards public service provision consisting of the three items as our independent variable. As in the first experiment, a factor analysis showed that the three items used to construct the index loaded highly on one factor (factor scores above 0.7) and that the index should be considered highly reliable with a Cronbach's alpha of 0.87.

Findings: Experiment 2

Experiment 2 is analyzed with logistic regression. The interaction terms between prior beliefs and the various treatments test whether prior beliefs play a statistically significant larger role in the treatments than in the placebo.

Table 5 about here

Table 5 shows the findings. First, adding information about the sender of the information does not matter for the impact of prior beliefs on interpretations (this conclusion is supported in additional tests in which we compared treatment 2 with treatment 4 only and treatment 3 with treatment 5 only). A plausible interpretation of this null finding would be that the respondents were only presented with rather weak sender treatments. The findings do, however, show some tendency in the expected direction and we are therefore reluctant to completely disregard the idea that information about the sender may moderate the impact of prior beliefs on interpretations although the idea is not supported here.

Second and more importantly, it appears that prior beliefs in two of the four treatments have a significantly different impact from the placebo group on the ability to interpret performance information correctly. Specifically, prior beliefs have a significant, negative impact on the ability to interpret performance correctly in the two treatments where the private school according to the performance information performs best. Conversely, prior beliefs do not have a statistically significant effect on the interpretations in case the public organization performs best. The question is how we can make sense of these findings.

One likely interpretation concerns people's beliefs about the performance of public and private organizations. A descriptive analysis of the extent to which people are able to interpret the performance information provided correctly shows that these rates are markedly higher in the treatments where the private school performs best (T2 = 76%; T4 = 71%), than in treatments where the public school performs best (T3 = 64%; T5 = 66%). This indicates that some people wrongly respond that they think the private organization performs better in treatment 3 and 5 and at the same time report that they prefer publicly provided services. It appears paradoxical that people on the one hand prefer publicly provided services and on the other hand wrongly report that they think that the private school performs better. The explanation may be a fundamental discrepancy between ideological values and perceptions of performance. Our measure of attitudes to public service

provision probably captures broader ideological opinions about how subjects think the public sector should be organized. Furthermore, performance is but one of many aspects of importance to such ideological opinions and other aspects may appear more relevant to especially people with a left-wing point of view. If this is the case, left wingers may be less susceptible to motivated reasoning in the case of performance information because they simply do not have strong opinions about public sector performance. Thus, this reasoning may not help them report the correct answer if they receive treatment 3 or 5. In conclusion, the findings in experiment 2 generally support the idea that prior beliefs affect interpretations of performance information. We infer that the findings in experiment 1 were not just created by the particular design (the varying numbers) since we find a similar effect in experiment 2. However, the results also show that the relationship is complex and that more knowledge is needed to understand it in detail.

Discussion

The study is designed to combine high internal validity from the experimental setting with a large N, and it furthermore aims at improving the external validity of the findings by providing fairly realistic information in the experiments. We believe that the evidence is quite strong that interpretations of performance information are systematically biased by prior beliefs. However, survey experiments are conducted in artificial settings (a survey), with stylized treatments that are less complicated than what people encounter in everyday life. These limitations also apply to some extent in our case and we should expect individuals to be more careful with their interpretations of performance information in real life than the subjects in our experiments. On the other hand, prior beliefs probably also play a more important role in real world situations where actual political outcomes are at stake. Moreover, real performance information may in many cases be much more complicated entailing that people in real life to an even higher extent rely on heuristics in their interpretations of performance information than the subjects in our experiment.

Hence, arguments point in both directions in terms of whether findings are either stronger or weaker in actual evaluations of performance information than in our experiment. What we can say for certain is that our survey experiments demonstrate that prior beliefs can play a role in how real performance is interpreted and thus are likely to do so in reality. A main contribution of the analysis is the finding that people are more inclined to interpret unambiguous, clear, and valid performance information wrongly if the information disagrees with their prior values. Furthermore, prior beliefs not only affect the interpretation of ambiguous but also interpretations of unambiguous performance information.

The findings relate to the extant literature on how benchmarks and performance information are interpreted (James 2011; Jacobsen et al. 2014; Baekgaard 2015; Charbonneau and Van Ryzin 2015; Olsen 2013; 2015) in two important ways. First, different interpretations of performance information based on different political points of view are not only a matter of conscious political conflict but also take place on a subconscious level. In that sense, performance information is more than a technical device that can – provided that valid, clear and unambiguous information is collected and distributed – automatically increase awareness about real performance. We therefore suggest, as does previous research (Stewart and Walsh 1994; Moynihan 2006; 2008; Johnsen 2011), that performance information, at least partly, is a political instrument, which is prone to interpretation and discussion. This implies that receivers of performance information do not uncritically adopt the information. They sometimes use performance information to support their own prior beliefs and, perhaps, political agenda.

Second, it follows that an underlying assumption of previous research (see, for instance, Jacobsen et al. 2014; Charbonneau and Van Ryzin 2015) can be questioned. Producers of performance information are not able to automatically shape how recipients of the information will interpret it by manipulating the way the information is presented. Rather, the findings suggest that recipients will tend to interpret the information in a way that is consistent with their prior beliefs regardless of how

the information is presented. On this basis, we suggest that future research should take recipient prior beliefs into account and study under what conditions performance information might lead to interpretations in discord with prior beliefs.

The findings are also important to the broader literature on performance information use (for an overview see Kroll 2014). Performance information is often thought of as a tool to inform citizens, guide decision makers and thereby improve decision making (Kettl 1997: 457; Moynihan 2006: 152). The fact that performance information is systematically misinterpreted even in a stylized case like this where nothing political is at stake and where the information is much simpler than much actual information, which might include complicated rankings on several parameters, questions the potential of performance information.

The findings may also help explain why the introduction of performance information systems in many cases has limited effects on the actual performance of government institutions. This is consistent with a classic account of how policy decisions are made: Lindblom famously argued more than 50 years ago that policy decisions are not made by undertaking “systematic comparisons of multitudes of alternatives to determine which attains the greatest amount of values” (1959: 79). Policymaking is rather a result of ‘muddling through’ by relying “heavily on the record of past experience” (1959: 79; see also Simon 1945; Wildavsky 1966). Decisions are not made by objective analysis of all available facts. Prior practices (and in this context: prior beliefs) play an important role in decision making (Bendor, 2015).

Ensuring that performance information is provided does not guarantee that prior beliefs, existing political values, or prejudices are changed. Since the message of the performance information does not always come through, it is perhaps not so surprising that it does not always have effects on actual performance. This in turn raises a number of questions for future research. These are questions like for whom the misinterpretations matter the most, the extent to which misinterpretations matter for actual decisions, and whether misinterpretations can be overcome by,

for instance, presenting performance information in a different way than here. Furthermore, future research may find it worthwhile to study in more detail to what extent and under what conditions the findings travel to real decision making processes. One idea would be to study whether people adapt incrementally to performance information over time in cases where the information discords with their prior values.

Conclusion

This article contributes to the literature on the interpretation of performance information (Moynihan 2006; 2008; James 2011; Olsen 2013; 2015; Charbonneau & Van Ryzin 2015) by showing that people's interpretations of even unambiguous performance information are likely to be systematically biased in accordance with their prior beliefs. This is demonstrated in two large-scale survey experiments in which subjects are asked to interpret a simple piece of performance information which has one – and only one – correct interpretation. To our knowledge, the study is the first to systematically examine in large-N experiments how prior beliefs relate to people's ability to interpret performance information correctly.

We conclude that interpretations of objective and clear performance information will be biased in the direction of prior beliefs held by those who receive the information. This questions the assumption that providing performance information automatically increases knowledge about government performance, let alone improves political decisions. This in turn raises a number of questions for future research. These are questions like for whom the misinterpretations matter the most, the extent to which misinterpretations matter for actual decisions, and whether misinterpretations can be overcome by, for instance, presenting performance information in a different way than here. Another important issue is whether there are ways to provide performance information that reduces the potential for biased interpretation. For instance, prospect theory

implies that presenting information in the domain of losses compared to the domain of gains have strong effects on how information is perceived (Kahneman and Tversky, 1979). Relatedly, Olsen (2015) shows that the valence framing of performance information affects how performance is perceived. A next step could be to consider under which circumstances the amount of bias is reduced. Furthermore, future research may find it worthwhile to study in more detail to what extent and under what conditions the findings travel to real decision making processes. One idea is to study whether people adapt incrementally to performance information over time in cases where the information discords with their prior values.

Appendix

Table A1 shows the composition of the groups in experiment 2 in terms of gender, age, level of education, job status, and political affiliation (measured by the voting behavior at the latest round of municipal elections in November 2013).

Table A1 about here

Acknowledgments

The authors would like to thank Asmus Leth Olsen, Rune Slothuus, the anonymous reviewers and the editor for valuable comments. They also wish to acknowledge the inspiration from Kahan et al. (2013), especially regarding their experiment.

Notes

1. A relevant characteristic in this particular case would be length of education. It is likely that political values differ between people with different education, and it seems plausible that people with longer educations on average are better able to interpret performance information than people with shorter educations.
2. This experiment draws heavily on Kahan et al.s design (2013). The main deviation is the focus on performance information and the use of comparable organizations in both treatment and placebo groups.
3. The panel is run by and the data collected by Userneeds, a Nordic survey company.
4. In both experiments a link to the survey was sent to the respondents who were given two full weeks to complete it.
5. The dependent variable in both experiments is thus essentially based on the ability to solve a math problem and some respondents may not have bothered computing the percentages. However, there is no reason to believe that the inclination to do so should differ between experimental arms and we therefore consider this an issue which should not affect the validity of our findings.
6. We are able to replicate all findings reported in the article if we use this alternative and more conventional measure (party affiliation) instead. We have chosen to use our index of

attitudes to public service provision in this article because we consider it to be a more direct and more fine-grained measure of the attitudes that we want to capture here.

7. On average approximately 16 percent of Danish students fail their final exams in Danish and math.

8. For ethical reasons, we explicitly mentioned in the experimental treatments (as we also did in the first experiment) that the example was constructed. Furthermore, the subjects were debriefed after the experiment in order to avoid that they were left with the impression that CEPOS had actually produced the information that they were presented with.

References

- Andersen, Lotte Bøgh and Mads Leth Jakobsen (2011). "Does ownership matter for the provision of professionalized services? Hip operations at publicly and privately owned clinics in Denmark", *Public Administration*, 89 (3): 956-974.
- Andersen, Simon Calmar and Ulrik Hvidman (2013). "Perceptions of Public and Private Performance. Evidence from a Survey Experiment", paper presented at the Public Management Research Association (PMRA) Conference, June 2013.
- Baekgaard, Martin (2015). "Performance Information and Citizen Service Attitudes: Do Cost Information and Service Use Affect the Relationship?" *International Public Management Journal*, (forthcoming).
- Bendor, Jonathan (2015). "Incrementalism: Dead yet Flourishing", *Public Administration Review*, 75 (2): 194-205.
- Charbonneau, Étienne, & Van Ryzin, Gregg G. (2015). "Benchmarks and Citizen Judgments of Local Government Performance: Findings from a Survey Experiment", *Public Management Review*, 17 (2): 288-304.
- Christensen, Tom and Per Lægreid (2003). *Trust in Government – the Significance of Attitudes Towards Democracy, the Public sector and Public Sector Reforms*, Bergen University: Stein Rokkan Centre for Social Studies.
- Druckman, James N. (2012). "The Politics of Motivation", *Critical Review: A Journal of Politics and Society*, 24 (2):199-216.
- Heinrich, Carolyn J. (2012). "How Credible is the Evidence, and Does it matter? An Analysis of the Program Assessment Rating Tool", *Public Administration Review*, 72 (1): 123-134.
- Jacobsen, Rebecca, Jeffrey W. Snyder, and Andrew Saultz (2014). "Informing or Shaping Public Opinion? The Influence of School Accountability Data Format on Public Perceptions of School Quality", *American Journal of Education*, 121 (1): 1-27.

- James, Oliver (2011). "Performance Measures and Democracy: Information Effects on Citizens in Field and Laboratory Experiments", *Journal of Public Administration Research and Theory*, 21 (3): 399-418.
- Johnsen, Åge (2011). "Performance Information and Educational Policy Making", pp. 161-177 in Wouter van Dooren and Steven van de Walle (eds.): *Performance Information in the Public Sector: How it is Used*, Houndmills, Basingstoke: Palgrave Macmillan.
- Jones, Bryan (2001). *Politics and the Architecture of Choice*, Chicago: Chicago University Press.
- Kahan, Dan M., Ellen Peters, Erica Cantrell Dawson, and Paul Slovic (2013). "Motivated Numeracy and Enlightened Self-Government", unpublished working paper.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2319992. Paper last accessed April 27, 2015.
- Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, 47 (2): 263-292.
- Kettl, Donald (1997). "The Global Revolution in Public Management: Driving Themes, Missing Links", *Journal of Policy Analysis and Management*, 16 (3): 446-62.
- Kroll, Alexander (2014). "Drivers of Performance Information Use: Systematic Literature Review and Directions for Future Research", *Public Performance & Management Review* (forthcoming).
- Kunda, Ziva (1990). "The case for motivating reasoning", *Psychological Bulletin*, 108 (3): 480-498.
- Lindblom, Charles E. (1959). "The Science of 'Muddling Through'", *Public Administration Review*, vol 19 (2): 79-88.
- Linder, Darwyn E.; Joel Cooper; Edward E. Jones (1967). "Decision freedom as a determinant of the role of incentive magnitude in attitude change", *Journal of Personality and Social Psychology*, 6 (3): 245-254
- Moynihan, Donald P. (2006). "What Do We Talk About When We Talk About Performance? Dialogue Theory and Performance Budgeting." *Journal of Public Administration Research and Theory*, 16 (2): 151-168.

- Moynihan, Donald P. (2008). *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, DC: Georgetown University Press.
- Nielsen, Poul Aaes (2013a). *Performance Information in Politics and Public Management: Impacts on Decision Making and Performance*, PhD Dissertation, Aarhus: Politica.
- Nielsen, Poul A. (2013b). "Learning from Performance Feedback: Performance Information, Aspiration Levels and Managerial Priorities." *Public Administration*, 92 (1): 142-160.
- Nielsen, Poul Aaes and Martin Baekgaard (2015). "Performance Information, Blame Avoidance, and Politicians' Attitudes to Spending and Reform: Evidence from an Experiment", *Journal of Public Administration Research and Theory*, 25 (2): 545-569.
- Olsen, Asmus Leth (2013). "Leftmost-digit-bias in an enumerated public sector? An experiment on citizens' judgment of performance information." *Judgment and Decision Making*, 8 (3): 365-371.
- Olsen, Asmus Leth (2015). "Citizen (Dis)satisfaction: An Experimental Equivalence Framing Study", *Public Administration Review*, 75(3): 469-478.
- Radin, Beryl A. (2006). *Challenging the Performance Movement: Accountability, Complexity, and Democratic Values*, Washington DC: Georgetown University Press.
- Simon, Herbert A. (1945). *Administrative Behavior*, New York: The Free Press.
- Simon, Herbert A. (1957). *Models of Man*, Oxford: Wiley.
- Sniderman, Paul M. (2011). "The Logic and Design of the Survey Experiment: An Autobiography of a Methodological Innovation", in James N. Druckman, Donald P. Green, James H. Kuklinski and Arthur Lupia (eds.), *Cambridge Handbook of Experimental Political Science*. Cambridge: Cambridge University Press.
- Stewart, John and Kieron Walsh (1994). "Performance measurement: When performance can never be finally defined", *Public Money & Management*, 14 (2): 45-49.
- Taber, Charles S. and Milton Lodge (2006). "Motivated Skepticism in the Evaluation of Political Beliefs", *American Journal of Political Science*, 50 (3): 755-769.

- Tepe, Markus (2012). "The Public/Private Sector Cleavage Revisited: The Impact of Government Employment on Political Attitudes and Behaviour in 11 West European Countries", *Public Administration*, 90 (1): 230-261.
- Van de Walle, Steven and Alasdair Roberts (2011). "Publishing Performance Information: An Illusion of Control?", pp. 215-230 in Wouter van Dooren and Steven van de Walle (eds.): *Performance Information in the Public Sector: How it is Used*, Houndmills, Basingstoke: Palgrave Macmillan.
- Van Dooren, Wouter (2011). "Nothing New Under the Sun? Change and Continuity in the Twentieth-Century Performance Movements", pp. 15-27 in Wouter van Dooren and Steven van de Walle (eds.): *Performance Information in the Public Sector: How it is Used*, Houndmills, Basingstoke: Palgrave Macmillan.
- Westen, Drew, Pavel S. Blagov, Keith Harenski, Clint Kilts and Stephan Hamann (2006). "Neural Bases of Motivated Reasoning: An fMRI Study of Emotional Constraints on Partisan Political Judgment in the 2004 U.S. Presidential Election", *Journal of Cognitive Neuroscience*, 18 (11): 1947-1958.
- Wildavsky, Aaron (1966). "The Political Economy of Efficiency: Cost Benefit Analysis, Systemsanalysis, and Program Budgeting", *Public Administration Review*, vol. 26 (4), pp. 292-310.

Table 1: Interpretation of performance information according to motivated reasoning

	Public organization scores best	Private organization scores best
Public preferred	Actor will tend to interpret performance information correctly (supporting prior beliefs)	Actor will tend to interpret performance information wrongly (supporting prior beliefs)
Private preferred	Actor will tend to interpret performance information wrongly (supporting prior beliefs)	Actor will tend to interpret performance information correctly (supporting prior beliefs)

Table 2: Treatment and placebo groups in experiment 1

Treatment 1/placebo 1:		
	Operations without complications	Operations with complications
Public hospital/ Hospital A	203	68
Private hospital/ Hospital B	47	9

Treatment 2/placebo 2:		
	Operations without complications	Operations with complications
Public hospital/ Hospital A	68	203
Private hospital/ Hospital B	9	47

Note: The information was presented as showing the performance of public and private hospitals in the treatments and as showing the performance of hospital A and B in the placebos.

Table 3: Findings from experiment 1. Logistic regression analyses of correct identification of best performing hospital

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Data	T1: Priv. hosp. best	T2: Pub. hosp. best	P1: Hosp. B best	P2: Hosp. A best	T1;P1	T2;P2
Pro public service provision	-0.019** (0.004)	0.009* (0.003)	0.001 (0.003)	0.005 (0.004)	0.001 (0.003)	0.005 (0.004)
Treatment dummy					0.869* (0.345)	-0.255 (0.329)
Pro public service provision * treatment					-0.020**	0.004 (0.005)
Chi ²	28.43**	6.27*	0.14	1.77	35.64**	8.38*
N	459	434	448	443	907	877

Notes: Intercepts were included in the models but are not presented in the models. Entries are logistic regression coefficients; robust standard errors in parentheses. **: p < 0.01; *: p < 0.05.

Table 4: Differences between groups in the second experiment

	P1	T2	T3	T4	T5
CEPOS mentioned as sender of the information	No	No	No	Yes	Yes
Organizations compared	Public vs. public	Public vs. private	Public vs. private	Public vs. private	Public vs. private
Organization mentioned first	---	Public first	Private first	Public first	Private first

Table 5: Findings from experiment 2. Logistic regression analyses of correct identification of best performing school

	Model 1:
P1	Ref.
T2 (private school performing best)	0.923 (0.407)*
T3 (public school performing best)	-0.559 (0.390)
T4 (private school performing best + Pro market sender of information mentioned)	0.997 (0.415)*
T5 (public school performing best + Pro market sender of information mentioned)	-0.571 (0.407)
Pro public service provision	0.006 (0.005)
T2 * pro public service provision	-0.014 (0.006)*
T3 * pro public service provision	0.001 (0.006)
T4 * pro public service provision	-0.021 (0.006)**
T5 * pro public service provision	0.002 (0.006)
Chi ²	38.05**
N	1,416
Data	All observations

Notes: Intercepts were included in the models but are not presented in the models. Entries are logistic regression coefficients; robust standard errors in parentheses. **: p < 0.01; *: p < 0.05.

Table A1: Balance checks of randomization on observable characteristics of respondents

	Experiment 1:				Experiment 2				
	T1	T2	P1	P2	T2	T3	T4	T5	P1
Women (percent)	49	52	51	46	48	56	50	49	55
Average age (years)	44	43	44	45	45	43**	46	47*	46
Higher education ¹ (percent)	52	52	55	57	55	53	50*	65**	54
Voted for socialist party at latest municipal elections ² (percent)	49	52	50	45	44	45	41	44	41
Pro public service provision	61	61	62	60	56	58	56	61*	55

Notes: Significant differences are identified by comparing either of the groups with the others in two-tailed t-tests. **: p < 0.01; *: p < 0.05. ¹: At least a short university degree or alike. ²: Voted for the Social Democrats, the Socialist People's Party, or the Red-green Alliance.