

Compared to What?

Reference Points in Performance Evaluation

Asmus Leth Olsen*
Department of Political Science
University of Copenhagen
ajlo@ifs.ku.dk
www.asmusolsen.com

Dansk Selskab for Statskundskab Aarsmøde
Politisk adfaerd og kommunikation 2:
Politiske holdninger og holdningsdannelse, Session I

Abstract

Questions about “good” or “bad” performance are difficult to handle without comparing with something. This paper tests how reference points alter citizens’ evaluation of performance information. Specifically, I test how citizens use historical or social reference points when making relative comparisons. How important is performance relative to past outcomes compared with performance relative to the outcomes obtained by others? The question of comparison is also linked to the often noted negativity bias. Two experiments within a large nationally representative sample are presented (n=3443). The experiments assign citizens to different forms of historical and social reference points with performance data on education and unemployment. Interestingly, in both cases the findings show that social reference points are almost twice as important as historical ones. Mixed evidence on the negativity bias is also found. These strong social comparison effects have implications for studying public opinion and elite framing of performance.

KEYWORDS: performance information · reference points · social comparison

*The experiments conducted in this articles have received funding from the Center for Voting and Parties (CVAP), Department of Political Science, University of Copenhagen, Denmark. I am grateful to Kristian Anker-Moeller, Yosef Bhatti, Peter Thisted Dinesen, Kevin Esterling, Fabrizio Gilardi, Kasper M. Hansen, Oliver James, Cindy Kam, Rasmus Tue Pedersen for comments on earlier drafts on this paper. An earlier version was presented at the Midwest Political Association Meeting. Panel: Political Bias, Misinformation, and Conspiracy Theories. Chicago, 11th of April 2013. A more recent version was presented the 11th Public Management Research Conference in Madison, Wisconsin, June 20-23 2013.

In actual life, men do not usually approach tasks in a vacuum of ignorance about the achievements of others.

– Chapman and Volkmann (1939, 225)

Imagine that the current unemployment rate is about 8 pct. Is that good or bad? In isolation, it becomes difficult to answer without thinking more deeply about the fundamentals of the economy. However, most of us are not economists and citizens do not have the incentives, time, or resources for thinking deeply about this question. Accordingly, when faced with the question most people will immediately search for a point of reference for comparison in order to ease the mental task. For instance, at the same time last year the unemployment rate may have been 7 pct. In other words, the unemployment rate has actually increased. So, clearly, the current unemployment rate is bad, right? On the other hand, what if countries we usually compare ourselves with have an unemployment rate of 9 pct? Then the unemployment rate is actually lower than among our peers.

The above example shows how the question of “good” or “bad” performance is inherently difficult to separate from a point of reference. Today numerical performance information is increasingly informing citizens about government performance. School rankings are offered to guide voice and exit in educational choice. Hospital report cards are offered when critical health care decisions are made. Quarterly GDP data and monthly unemployment numbers are featured prominently in the media to inform the public about everything from local government performance to the “wealth of nations”. These numbers aim for the layman but are, as in the initial example, abstract in their meaning. Is a top-100 school a *good* school? Is a hospital with a 90 percent satisfaction rate doing *great*? Can we be *satisfied* with an unemployment rate above 8 pct? There are no obvious or easy answers to these questions or as noted by Nordhaus (1975, 173), “there is no objective standard by which to measure performance”. Few political scientists would question the importance of these performance numbers for political decision making or voter opinion formation (Downs 1957; Key 1966;

Bendor et al. 2011; James 2011; Ansolabehere, Meredith, and Snowberg 2013). However, at the citizen-level we lack a basic understanding of how citizens actually *make up their mind* about performance based on numerical information. How do citizens map these numbers onto a scale of good and bad performance from which they can make informed decisions?

In this article, I examine this question based on a theory of reference dependence (Tversky and Kahneman 1986; Quattrone and Tversky 1988; Tversky and Kahneman 1991). Reference dependence implies that citizens' evaluation of performance is a matter of losses and gains relative to some reference point (Tversky and Kahneman 1991, 1039). I explore how comparison with a reference point becomes a judgmental shortcut for citizens when they are faced with performance data (Sniderman, Brody, and Tetlock 1991). I aim to answer two important questions derived from a reference-dependent view of citizens' judgment about performance information: First, what type of reference point is most important for citizens' evaluation? When asked to evaluate some element of the social world, many tend to respond: *compared to what?* Reference points are often either social or historical (McDermott, Fowler, and Smirnov 2008, 336). Social reference points stress that comparison happens by looking at the performance of others. It can be the neighboring school, the hospitals in another city, or the employment rate in other countries. Historical reference points denote a temporal comparison between current and previous performance. How did we do last month, in previous years, or in the past decade compared with today? As argued by Kayser and Peress (2012, 680), political science has mostly emphasized historical temporal comparisons while neglecting social comparisons. The focus on historical reference points is reflected in US President Reagan's seminal question: "Are you better off (today) than you were four years ago?" This sentiment is mirrored in retrospective voting (Downs 1957; Key 1966), where citizens are expected to rely on "one comparatively hard bit of data: they know what life has been during the incumbent's administration" (Fiorina 1981, 5). Social reference points seem much more absent from political science. In social psychology and sociology, we find a stronger tradition social reference points in social comparison theory (Festinger 1954; Muss-

weiler 2003), social determinants of aspirational levels (Chapman and Volkmann 1939), and in the work on reference groups (Shibutani 1955). Second, I ask how losses and gains differ when using social and historical reference points? Political science has long been interested in the negativity bias of politics (Lau 1982, 1985). In short it denotes that negative information has more profound effects on attitudes and behavior than positive ones of a similar magnitude (Baumeister et al. 2001; Rozin and Royzman 2001). The negativity bias has also been found relevant for understanding how citizens' respond to performance information in politics (Soroka 2006; James 2011). The question is how the negativity bias works in relation to historical and social reference points.

In order to seek an answer to these two questions, I conduct two survey experiments based on a large representative sample of the adult Danish population ($n = 3443$). The survey experiments ask individuals to evaluate two types of performance information, namely: (a) school performance from grade average information and (b) municipal unemployment performance from local unemployment rates. The results show, across both experiments, that social reference points have a stronger impact on citizens' evaluations than historical reference points. Findings on the negativity bias are more mixed and seem to depend more on type of performance measures than the actual reference point. Findings of strong social comparison effects are in line with studies showing interdependence in the form of yardstick competition or diffusion of policies among political jurisdictions (Besley and Case 1995; Shipan and Volden 2008; Gilardi 2010). By introducing social comparison effects into the study of citizens' performance information, the study can help provide a micro-level foundation for interdependence in political decision-making across organizations and political jurisdictions (Elkins and Simmons 2005, 46). Very few studies have conducted micro-level analysis of social interdependence in political science (Linos 2011, 678). These findings have also implications for how we can expect elites to frame performance (Druckman 2001; Olsen 2013*c,b*) and for the study of subjective vs. objective measures of citizen satisfaction (Van Ryzin 2004). These and further implications are discussed in the final section of the paper.

Theory

Reference-Dependent Evaluation of Performance

The basic argument put forward here will be that citizens' judgment about public services is highly reference dependent. Reference dependence implies that citizens make judgment about a public organization's performance in terms of losses or gains compared to some reference point (Kahneman and Tversky 1979). In other words, reference dependent judgment of performance will have a strong relative component as citizens compare performance relative to something else (i.e. the reference point). The most well-known account of reference dependent judgment is provided in prospect theory (Kahneman and Tversky 1979; Kahneman, Knetsch, and Thaler 1991). Here, individuals can be expected to code performance feedback as either losses or gains relative to some reference point. An implication is that the same piece of absolute information can be evaluated differently depending on the placement of the reference point. In social psychology, many have argued that individuals are deeply inclined to evaluate their own abilities and opinions in relative terms and not in absolute ones. Our perception of everything from light, sound, to colors will have a strong relative component (Kahneman and Tversky 1979). The same holds for the social world where changes matter more to citizens' understanding of value than absolute levels (Markowitz 1952). Changes or differences contain a relative component: a comparison of the value at hand with some other available standard. In fact, across a number of fields we find the notion that performance is coded as either loss or gain, success or failure, negative or positive, or good or bad depending on how the evaluated falls relative to some reference point that informs expectations, aspirations, or norms (Festinger 1954; Simon 1955; March and Simon 1958; Tversky and Kahneman 1974).

Some attribute this binary division of information around a reference point to bounded rationality and the limited cognitive abilities of human beings (Simon 1954). Evaluating performance in relation to some reference point is much less demanding task than evaluating absolute performance. Judgment about an absolute number requires some scale-specific infor-

mation. What are the possible interval of the scale? What is the mean of the scale and what is its distribution? With the help of comparison, the task shifts to comparing the abstract target information with the reference point. In that sense, we can see the reference point as a judgmental shortcut that allows citizens to form evaluations about abstract numerical performance information which they possess no deeper knowledge about (Popkin 1991; Sniderman, Brody, and Tetlock 1991; Ansolabehere, Meredith, and Snowberg 2013; Olsen 2013b). The comparison with reference points can serve as a useful heuristic which allow citizens to simplify the process of making up their mind about a more abstract absolute scale (Mussweiler and Epstude 2009, 2). However, comparison is not just a shortcut. Mussweiler and Epstude (2009) point out that relative comparisons are efficient because they lead to faster decision making without resulting in worse decisions. Comparison has also been found to reduce judgmental uncertainty (Mussweiler and Posten 2011). With regards to performance information, we can also view the reference point as altering or informing the expectations or aspirations which individuals hold to public organizations. In organizational science, reference points can serves as goals or aspirational levels of performance for organizations or individuals (March and Simon 1958; Cyert and March 1963; Heath 1999).¹ Comparing outcomes to some reference point for performance provides a measure of attainment discrepancy and places performance in a domain of either *falling short* or *making the cut* (March 1988). In studies of consumer and citizen satisfaction, reference points are seen as informing the expectations which citizens have to public services (Oliver 1980). These expectations are important as they directly feed into citizens' subjective satisfaction with the public sector. In the widely applied expectations-disconfirmation model, satisfaction is formed by the discrepancy between expectations and the experienced performance (Van Ryzin 2004). Thus, we have multiple accounts of the importance of reference points, but which type of reference points drive these relative comparisons?

¹Hoppe (1931) introduced the concept of aspirational levels (or *anspruchs niveau*) among individuals as a way of understanding how outcomes are categorized as failure or success.

Historical Reference Points vs. Social Reference Points

There are many sources of reference points (Tversky and Kahneman 1981, 456). Here, we focus on two types of reference points in the form of social reference points (Festinger 1954) and historical reference points (Albert 1977). Across a number of fields, there has been a dual focus on both types of reference points. In psychology, social and historical reference points have been contrasted for individuals' self-evaluation (Smith 1990; Robinson-Welton and Kiecolt-Glaser 1997). Contrasting social and historical reference points for internal organizational performance evaluation has a long tradition in organizational science (Greve 1998). Cyert and March (1963) viewed organizational aspirations as the combined product of the performance of others (social comparison) and an organization's own past performance (i.e. historical comparison). Decision makers are also expected to care about performance relative to that of other organizations as well as the organization's own historical performance (Cyert and March 1963; Levinthal and March 1981).

In historical comparisons, individuals compare current performance with some past historical reference point. Here, I focus explicitly on historical comparison within a given unit in order to draw a sharp distinction with social comparison. At the level of "the compared" historical comparison can be viewed as internal comparison of current and past performance. Findings from a number of fields support the idea that citizens rely on historical reference points for when making comparisons with current performance. The status-quo-bias can be seen as one example of historical comparison (Samuelson and Zeckhauser 1988). Past performance is the status quo which current performance is evaluated against. If current performance is below past performance it will be seen as worse compared to the status quo. Historical reference points emphasize that citizens care about what direction the performance of an organization has been heading most recently. We find this very pronounced in the tradition of retrospective voting studies which emphasizes the calculation of differences between current and past performance as a means for prospective evaluation of the incumbent government (Downs 1957; Fiorina 1981; Key 1966). Hibbs, Rivers, and Vasilatos (1982, 314)

argued strongly for historical comparisons in that citizens' evaluation of performance was driven by comparing the cumulative performance of the current administration with the past performance of the opposition and all previous administrations. From the theory on political business cycles, we learn that voters pay most attention to very recent changes in performance (Nordhaus 1975). Historical reference points are clearly the most broadly used reference point for performance comparison in political science (Kayser and Peress 2012, 680).

Social reference points offer an alternative source of comparison. In social comparisons, individuals compare current performance with the performance simultaneously obtained by others. Whether individuals seek it out or not, the environment will inevitably provide information about the achievement of others (Chapman and Volkmann 1939). We simply cannot ignore the outcomes of similar entities. This broad idea has been framed differently in a number of different fields and research traditions. The idea is reflected in the theoretical traditions of bandwagon effects or "keeping up with the Joneses" in early studies of consumer behavior and public opinion (Pierce 1940; Leibenstein 1950). Reference group theory also denote comparison with *external others* when making judgment about oneself (Shibutani 1955). The research tradition on relative deprivation reflects a similar sentiment (Runciman 1961; Stark and Taylor 1991). Perhaps the most successful formulation of the idea is found in social comparison theory (Festinger 1954). Social comparison theory states that we have an urge to evaluate our own opinions and abilities by comparing them with the opinion and abilities of similar others. In the case of performance measures, the comparative preference and belief evaluations are of particular importance (Suls, Martin, and Wheeler 2000). In studies of health, happiness, and wealth, there is strong evidence of relative comparison with others (Brickman and Campbell 1971; Lyubomirsky and Ross 1997; Diener 2000). The health, happiness, and wealth we observe around us affect our personal self-assessment on these concepts. Initially, social comparison was seen as affected by the availability of absolute information. In other words, social comparison was a second best option in the absence of absolute information. Modern accounts of reference dependence find that relative social com-

parison can be relevant even if absolute measures are available (Moore 2007). In the case of numerical performance information, these forces can be even stronger as numerical measures encourage comparison (Espeland and Sauder 2007, 35).

Goodman and Haisley (2007) pointed out that most social comparison literature focus on how individuals apply information to others. That is, social comparison refers to self-other comparisons. Few studies apply the insights of social comparisons to other levels of analysis such as the organizational or jurisdictional level (Greve 1998, 60). However, social comparison theory point to the importance of general *comparison processes* and not exclusively on individual comparisons (Goodman and Haisley 2007, 115). In economics and political science, we have seen some indirect application of social comparison processes which can be useful for our purpose. Salmon (1987) introduced the idea of social reference points with his yardstick theory for holding government accountable. He argues that historical comparisons will constitute a very noisy reference point for comparison with current performance. Over time, there will be exogenous disturbances. There can be business cycles and long-term trends which cannot be factored in when comparing the present situation with past times. Instead, citizens' are better off by applying neighboring jurisdictions or other similar social reference points as a means for comparison with current performance. These ideas constitute a large empirical field which primarily focus on how policy makers respond to social comparison incentives by mimicking policies in other jurisdictions (Besley and Case 1995).² Thus, while social reference points are less refereed to in political science the above indicates that there are plenty of reasons to believe that they, along with historical reference points, affect citizens evaluation of performance.

²Macro-evidence from US states indicate that unemployment relative to the national average is correlated outcomes in Gubernatorial elections (Cohen and King 2004).

Negativity Bias

The final argument concerns whether relatively good or relatively bad information affects citizens attitudes to the same extent. Drawing on social psychology and prospect theory, we derive a further important implication of reference dependence, namely an asymmetrical effects of performance information around the reference point with larger effects in the domain of losses (Kahneman and Tversky 1979; Kahneman, Knetsch, and Thaler 1991). Across many fields and areas of research, it is found that "bad is stronger than good" (Baumeister et al. 2001). Meta-reviews of this negativity bias finds evidence in human memory, decision making, and behavior (Rozin and Royzman 2001). In political science, this negativity bias has also been noted (Lau 1982, 1985; Boyne et al. 2009; Olsen 2013a). In economic voting studies, there are multiple accounts of a negativity bias in response to historical-temporal reference points. That is, a worsening economy damages the incumbent to a greater degree than an improving economy helps. For instance, Kinder and Kiewiet (1979) found evidence of decreasing voter support for the incumbent only if the unemployment rate went up.

Data and Design: Experiments in a Large Diverse Sample

Given the expectations outlined above, we need a research design which can both (a) compare the relative effect of social reference points with that of historical reference points, and (b) compare relatively negative performance with relatively positive performance of a similar magnitude. In order to do so, we employ a set of experiments which were conducted in a large nationally representative sample of the Danish voting population. The experiments were conducted in a survey with YouGov's Danish online panel, which consists of 40,000 Danes. The panel is recruited via both online, radio and newspaper ads as well as through telephone surveys. The sampling frame for the study was restricted to citizens between the age of 18 to 74. The sample was pre-stratified on gender, region, age, and party choice at the most recent national election (2011). Table 1 below shows descriptive statistics of the sample and

highlights its diversity in terms of age and education. The survey was conducted between the 15th of October and the 22th of October (2012). Respondents were recruited to participate via e-mail with an embedded link to the survey. As follow-up two e-mail reminders were sent out. In total, 3443 respondents participated in the experiments which yielded a response rate of 42%.³

Table 1: Descriptive statistics

Variable	Pct.
Gender (male)	49.8%
Age (mean years)	50.5 (SD=15)
Education	
High school or less	18.4%
Vocational training	24.7%
Short-cycle tertiary	12.9%
Medium-cycle tertiary	28.6%
Long-cycle tertiary	15.5%
Geographical region	
Capital area	24.7%
Sealand	15.3%
Southern Denmark	23.9%
Middle Jutland	23.8%
Northern Jutland	12.3%

Note: n=3443.

The two experiments dealt with school grades and local government unemployment rates. School grades provide an example of a performance measure which represents a specific public service. Education performance information is widely published today in most developed countries (Bowman and Bastedo 2009). Unemployment rates are, on the other hand, an example of a more general economic performance indicator which dominates socio-tropic theories of voter responses to national or local performance. Unemployment measures are often used in analysis of citizens evaluation of the incumbent government (Kramer 1971; Kinder and Kiewiet 1979). Unemployment is often believed to be a very intuitive and accessible measure which is easily communicated in the media (Conover, Feldman, and Knight 1986; Soroka 2006).

³In total 8.204 were invited to participate in the survey.

The actual content and setup of the two experiments differed. However, for both studies the experimental design entails randomized treatments at two levels. The first level of treatments randomly assign subjects to either: (1) absolute performance, (2) absolute performance and an absolute social reference point, and (3) absolute performance and an absolute historical reference point. The first level of treatment allows for both comparing the effects of absolute performance with the effect relative to some reference points, and for determining variations in relative effects for both social and historical reference points. The second level of treatment randomizes the actual numerical content of the performance information which the respondent was provided with. The second level provides a unique setting for understanding how citizens process relatively bad vs. relatively good performance information. It is also an optimal robustness check for understanding sensitivities in evaluations for different degrees of relative performance.⁴

Experiment I: Citizens' Judgment About School Grade Averages

The first experiment deals with the research question in a setting of education performance data. In Denmark, the Ministry of Education has released unadjusted school grade averages for all elementary schools over the course of the last ten years. Respondents were randomly assigned one of three different treatments providing respondents with different absolute performance and/or different reference points. The treatments cover (1) absolute performance, (2) social performance comparison, and (3) historical performance comparison. Overview of the different treatment frames are provided in the table 2 below. Under the first treatment, respondents are only provided with the absolute grade average of an unnamed school (n=1156). This is a form of control state without any reference point information. In the social comparison treatment, the respondents are presented with the absolute grade average of an unnamed school along with a municipal grade average for all schools in the same municipality (n=1148). Finally, for the historical comparison treatment respondents were presented

⁴As the survey experiment comprises multiple experiments the order of experimental treatments was randomized in order to avoid sequence effect across the different treatments.

with the average grade of a school along with the previous year’s average grade level of the same school (n=1139).

Operationalizing the historical treatment is more straightforward than the social comparison as it is only a matter of determining the temporal lag on the reference point. Here, a one year lag is used as it resembles the actual rate of publication of grade averages in Denmark. The social comparison case is trickier as comparison with others can happen in numerous ways. In social comparison theory, it is argued that individuals seek out comparison with others that are believed to have similar characteristics which affect performance (Garcia, Tor, and Gonzalez 2006; Goethals and Darley 1977). In studies of policy diffusion and interdependence, there has been some emphasis on spatial proximity (Gilardi 2010). Here, I apply schools from the same municipality which both are likely to share some common characteristics and which are also located close to each other.⁵

Table 2: Experimental design: performance evaluation from school grade averages

Baseline question	Frame treatment	Question wording	Numerical treatment
Each year, the Ministry of Education releases a grade average for all schools in the country. How well do you think this school is doing?	<i>Absolute level only</i>	The school has a grade average of x .	$x \in N(\mu = 6.5, \sigma = 1.0)$
	<i>Social comparison</i>	The school has a grade average of x_1 . The grade average for all schools in the same municipality is x_2 .	$x_1 \in N(\mu = 6.5, \sigma = 1.0)$ $x_2 \in N(\mu = 6.5, \sigma = 1.0)$ $x_1 \perp x_2$
	<i>Historical comparison</i>	The school has a grade average of x_1 . The schools grade average was last year x_2 .	$x_1 \in N(\mu = 6.5, \sigma = 1.0)$ $x_2 \in N(\mu = 6.5, \sigma = 1.0)$ $x_1 \perp x_2$

Note: 6.5 corresponds to the national grade average for all schools in 2011.

At the second stage of treatment, the actual school grade values are drawn from a normal distribution with an average of 6.5 and a standard deviation of 1.0. Higher grade averages

⁵In Denmark, public schools are governed by municipalities and school choice is most likely to happen within the boundaries of the same municipality.

correspond to better results.⁶ The mean grade average corresponds to the actual grade average obtained for all schools in 2011. The distribution reflects, to some degree, the school grade averages people would experience in their everyday life. The averages are presented with one decimal which corresponds to how the government and the media would normally report them. School grade averages will therefore constitute a performance measure which most individuals have some familiarity with.

For each respondent, a grade value is drawn independently from a distribution with the above parameters. That is, for the absolute treatment respondents are presented with only one value drawn randomly from this distribution. For the two reference point treatments, respondents are provided with two values randomly drawn from the same distribution: one for the absolute level of the unnamed target school, and one for the social or historical reference point. Often survey experiments only present respondents with a few sets of various values or written descriptions. Here, respondents are presented with a large number of unique numerical values. For the unnamed target school the absolute grade average takes on around 60 uniquely different values across the three treatments. For all questions, the respondents should score their response on a slider scale from 0 to 100 where 0 was labeled *very bad* and 100 was labeled *very good*. Respondents were not informed about the underlying metric but the slider clearly indicated that respondents could place themselves at any point in between. A graphical presentation of the scale is provided in figure 1.



Figure 1: Screen caption of the exact response scale used in the experiment. The scale varies from "very bad" ("meget dårligt", 0) to "very good" ("meget godt", 100).

⁶The Danish grade scale is a 7-point scale with values from worst to best: -3, 0, 2, 4, 7, 10, and 12. It is directly comparable to an American scale where 12 corresponds to an A, and 0 and -3 both represent an F.

Experiment II: Citizens' Judgment of Municipal Unemployment Rates

The second experiment is similar to the first one in terms of treatments. However, the substantive setting is different as citizens are asked to evaluate municipal unemployment performance based on information about local unemployment rates. The wording of the questions and frames are outlined in table 3. In the baseline treatment, respondents are only provided with the absolute unemployment rate in an unnamed municipality (n=1155). In the social comparison treatment, the respondents are presented with the absolute unemployment rate of the unnamed municipality along with a hypothetical national unemployment rate (n=1142). Finally, in the historical reference point treatment, respondents were provided the unemployment rate of the unnamed municipality along with previous years' absolute unemployment rate in the same municipality (n=1146).

Table 3: Experimental design: Performance evaluation from local unemployment rates

Baseline question	Frame treatment	Question wording	Numerical treatment
How well do you think that this municipality is doing in terms of unemployment?	<i>Absolute level only</i>	The municipality has an unemployment rate of $x\%$.	$x \in N(\mu = 6.3, \sigma = 1.0)$
	<i>Social comparison</i>	The municipality has an unemployment rate of $x_1\%$. The unemployment rate in the rest of the country is $x_2\%$.	$x_1 \in N(\mu = 6.3, \sigma = 1.0)$ $x_2 \in N(\mu = 6.3, \sigma = 1.0)$ $x_1 \perp x_2$
	<i>Historical comparison</i>	The municipality has an unemployment rate of $x_1\%$. The municipality had last year an unemployment rate of $x_2\%$.	$x_1 \in N(\mu = 6.3, \sigma = 1.0)$ $x_2 \in N(\mu = 6.3, \sigma = 1.0)$ $x_1 \perp x_2$

Note: 6.3% was the average national unemployment rate during the survey period.

At the second stage of treatment, respondents are assigned a randomly drawn unemployment rate. For the unemployment rates, the values are drawn from a normal distribution with a mean of 6.3% and a standard deviation of 1.0%. The average corresponds to the national

unemployment rate in Denmark available at the time of the study. For each treatment status, unemployment rates are drawn independently from a distribution with the above parameters. For the relative frames, two independent values were drawn from this distribution: one for the unnamed target municipality, and one for the national average (social reference point) or last year's unemployment rate (historical comparison).

Empirical Findings

Historical vs. social comparison

Results from both experiments are shown in table 4 below. Column A shows the effect of treatment A where respondents were only provided the absolute level of the organization being evaluated. In column B I-II the models for the social reference point treatment are provided. Finally, columns C I-II provide the results from the historical reference point treatment. The school grade experiment shows the following findings: In the absolute performance treatment in column I (treatment A) the pure absolute grade average is highly positively correlated with citizens' evaluation of school performance. This simply indicates that the respondents evaluated the unnamed school around 8.5 points higher for each average grade point. It shows that the respondents react to grade averages in the expected way.

For the social comparison frame the absolute effect is even stronger. However, this is partly due to that the absolute measure now also reflects some unmeasured relative effects from the reference point. In model II (treatment B), the social reference point information is added which markedly improves the explanatory power of the model as measured by the adjusted R^2 . Introducing the social reference point of the absolute performance of the other schools in the municipality reveals a strong negative effect. A one grade point higher social reference points decreases citizens' evaluation of the unnamed target school by around 7.5 points. For the historical comparison treatment we also find a strong effect of relative performance. However, the coefficient for the historical reference point is significantly and

substantially lower than in the social comparison case with an effect of only 4.3 points (cf. column C II). This is also reflected in the explanatory power of the models which is much lower. It clearly indicates that citizens' provided with a social reference point reacted much stronger to relative performance than those assigned with a historical reference point for school performance.

Next we turn to the unemployment experiment⁷: As expected the absolute treatment in column A I shows a significant negative effect of the unemployment rate on citizens' evaluation of municipal unemployment performance. However, the effect of about 4.6 is much lower than we saw in the school case.⁸ Exposure to the absolute measures of unemployment rate had a much more noise effect on subsequent evaluation of the unnamed municipality. In the social comparison case, the reference point of the national unemployment rate shows a positive effect of the same magnitude as the negative effect of the municipality's own unemployment rate. That is, citizens' evaluate the target municipality's unemployment performance better as the the national unemployment rate increases. A similar pattern is found for the historical comparison treatment where the respondents were exposed to last years unemployment rate for the target municipality. Citizens' respond negatively to higher unemployment in the target municipality and positively to higher levels of past unemployment. Interestingly, we see a similar pattern for the school case: the coefficient for the relative social comparison reference point is almost twice as large as the relative historical comparison. Accordingly, the explanatory power of the social comparison frame is also a lot better than in the historical treatment.

The increases in explanatory power is not due to differences in how much respondents consider each treatment. Trimmed mean response times for the social grade average treatment was 17.9 seconds, and 18.5 seconds for the historical case. In the unemployment experiment the response time was 14.7 seconds for the social treatment and 15.1 seconds for the historical

⁷Note that the interpretation of the coefficients is reversed as they now measure a negative measure (unemployment rates) as opposed to the positive measures of school grade averages.

⁸The coefficients are comparable as the independent variables have approximately the same means and standard deviations.

treatment. In other words: the social reference points provided slightly faster response times and much less noise in evaluations. This is in line with psychological research showing faster and more efficient judgment via social comparison (Mussweiler and Epstude 2009; Mussweiler and Posten 2011).

Table 4: Experiments: Social vs. Historical Performance Information

	<i>Treatment A:</i>		<i>Treatment B:</i>		<i>Treatment C:</i>	
	<i>No comparison</i>		<i>Social comparison</i>		<i>Historical comparison</i>	
	I		I	II	I	II
<i>Exp. 1: Grades</i>						
School average	8.46**		10.14**	10.45**	9.21**	9.09**
	(0.48)		(0.55)	(0.51)	(0.53)	(0.51)
Reference point average				-7.51**		-4.29**
				(0.52)		(0.50)
Intercept	-3.57		-10.66**	35.86**	-7.66**	20.87**
	(3.18)		(3.61)	(4.65)	(3.45)	(4.70)
<i>N</i>	1156		1148	1148	1139	1139
adj. R^2	0.21		0.23	0.34	0.21	0.26
Resid. sd	16.45		19.16	17.65	17.72	17.18
<i>Exp. 2: Unemployment</i>						
Municipal rate	-4.62**		-7.60**	-7.76**	-5.69**	-5.79**
	(0.61)		(0.59)	(0.54)	(0.58)	(0.56)
Reference point rate				7.16**		4.34**
				(0.52)		(0.55)
Intercept	79.34**		97.76**	53.64**	88.24**	61.57**
	(3.90)		(3.78)	(4.75)	(3.66)	(4.91)
<i>N</i>	1155		1142	1142	1146	1146
adj. R^2	0.05		0.13	0.25	0.08	0.12
Resid. sd	21.24		19.08	17.67	19.66	19.15

Coefficients obtained by OLS estimation. Standard errors in parentheses.

Significance levels correspond to: * $p < .05$ and ** $p < .01$ (all two-tailed tests).

Reference Dependence and Negativity Bias

Next we turn to the question of a negativity bias in response to positive and negative information. In figures 2 and 3 the findings are plotted. The x-axis shows the relative grade difference between the target and the reference point. Negative values denote that the target is lower than the reference point. The straight lines represent linear regressions of the relative performance on performance evaluations. The other lines are smooth lowess fits. Separate lines are plotted for relatively good and bad performance to allow for differences in slope for “relatively good” and “relatively bad” performance. The plots also contain three means with 95%-confidence intervals. The mean at zero represent the average response for those who got a treatment where the performance information is exactly the same in the target as in the reference points. This forms a neutral treatment which cannot be definitively described as relatively bad or relatively good performance. The two other means represent the average response for those who got a target either below or above the reference point. The means are placed at the average treatment distance between target and reference point (approximately $-/+ 1.1$).

First, we turn to the grade experiment in figure 2. In the social reference point treatment to the left there is a stronger reaction in the negative domain indicated by a more steep trend line below the reference point as compared to above the reference point. For relatively good grades (i.e. school grades above the reference point) the effect of an extra grade point is 5.0 points. However, for relatively bad grades (i.e. school grades below the reference point) an extra grade point affects evaluations with 8.8 points. This difference in slopes is highly significant ($p < 0.01$) and is in line with the idea of loss aversion (Tversky and Kahneman 1991). As a more general indication of the negativity bias, we see that average evaluations below the reference point are 15.7 points lower than the neutral treatment while average evaluations above the reference point are only 7.5 points higher. This clearly reflects an asymmetrical response to relative performance in line with the negativity bias. For the case of the historical reference point the findings are more mixed. Average evaluations on

either side are very similar in their distance to the neutral treatment. And there is only weak evidence of loss aversion with a difference in slopes of 2.2 points ($p = 0.07$). However, the historical comparison experiment shows clear signs of decreasing sensitivity to relative performance as the difference in performance increases, which is in line with the prediction of decreasing marginal effects in reference dependent evaluations (Tversky and Kahneman 1991).

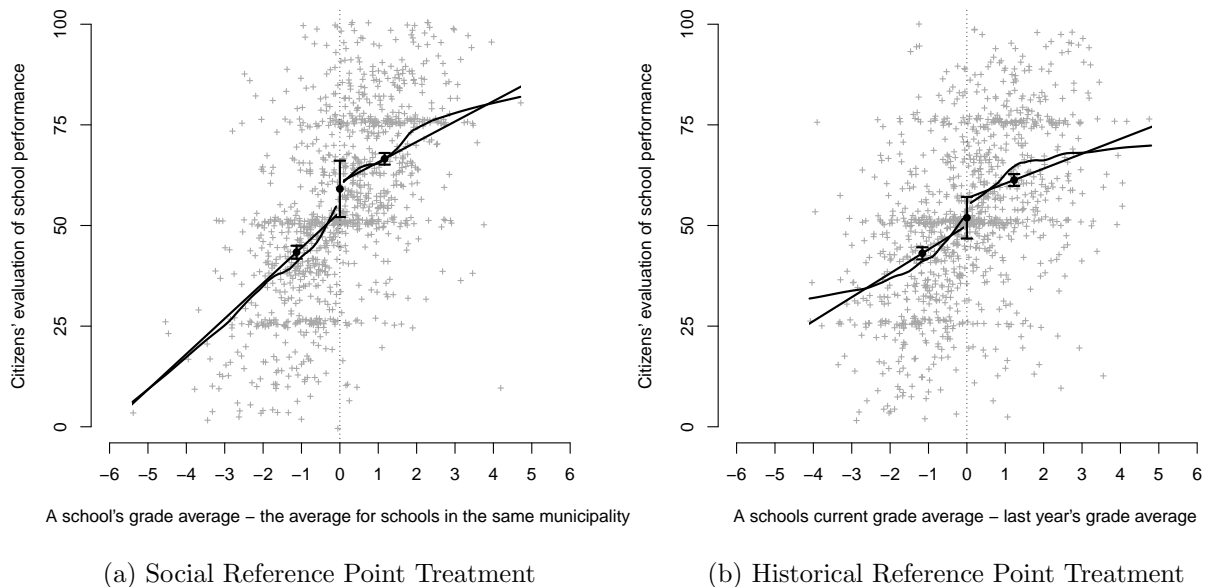


Figure 2: School Grade Average Experiment

Next, we turn to the unemployment experiment reported in figure 3. Here negative values indicate that the municipality's unemployment rate is lower than the reference point. In the social comparison case the difference in slopes is less than 1 point and insignificant ($p = 0.44$). There is also no indication of a difference in average evaluations for relatively bad or relatively good performance if compared with the neutral case. In the historical case there is weak evidence of loss aversion with a difference in slopes of -2.3 points ($p = 0.09$). In fact if a municipality's unemployment rate is lower than the reference point, then there is no significant correlation between distance to the reference point and evaluations ($\beta =$

$-1.1; p = 0.26; n = 548$). In addition, there is great sensitivity around the reference point: moving from slight below to slight above the reference point changes evaluations by about 10 points. However, this cannot be attributed to a negativity bias but rather a simultaneous positive jump for unemployment rates below the reference point.

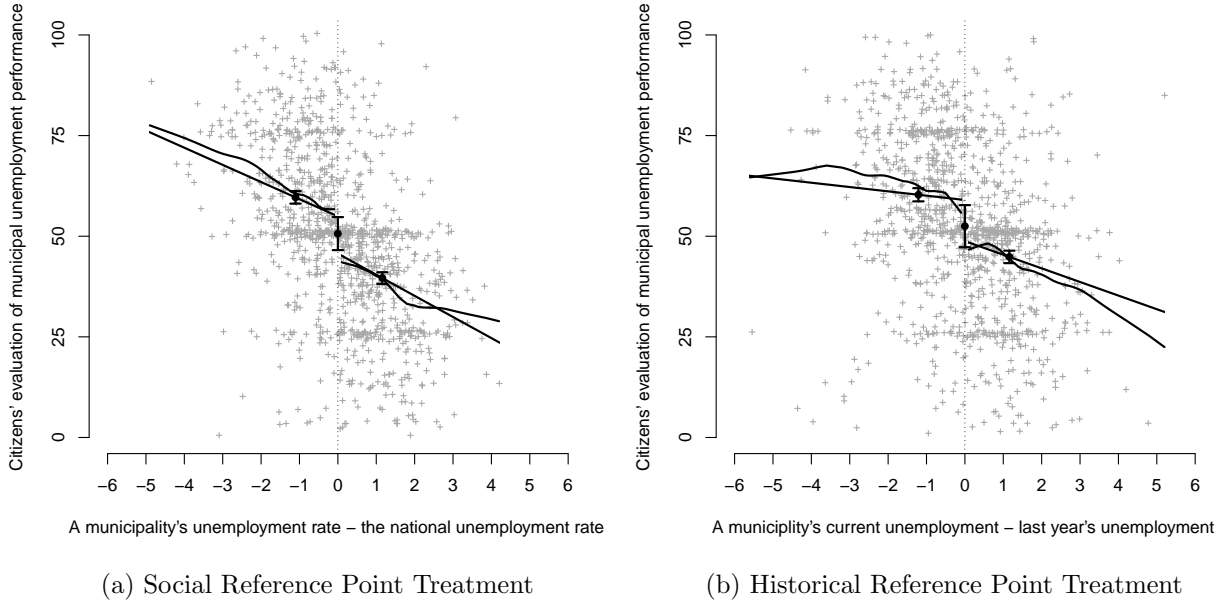


Figure 3: Municipal Unemployment Rate Experiment

Discussion and Conclusion

This study aimed to understand how citizens make up their mind about the performance of public organizations when faced with numerical performance information. The increasing amount of numerical performance information in the public sector was the substantial backdrop for the study. The experiments conducted here shed some important light on the role of reference points in this process. The analysis showed that not all reference points are of equal importance. Specifically, social reference points were found to affect citizens' evaluation to a much greater extent than historical-temporal reference points. This was both the case when the social reference point was local (other schools in the same municipality) as well as when

the social reference point was national (i.e. the national unemployment rate). In both cases, the relative social reference point effect was about twice as big as the historical reference point which compared performance for the organization one year earlier. “Are you better off today than you were four years ago?” is indeed a relevant question. It captures a fundamental relative historical performance evaluation which clearly affects citizens’ evaluation of performance. However, the analysis pointed out that the question needs to be supplemented with the more important question of: ”Are you better off than the people around you?”

The analysis found mixed evidence of a negativity bias in citizens’ evaluations. In one of the four experiments, there was a much larger marginal response in evaluations if performance fell narrowly below the reference point. This supports the idea of loss aversion and the negativity bias in citizens evaluations. There was also some indication of decreasing sensitivity to relative performance further away from the reference point. In addition, respondents’ evaluation of performance react very sharply in the close vicinity of the reference point - in particular in the unemployment case. This speaks to the importance of reference points.

The findings of a strong social comparison adds to recent studies which highlight yardstick competition or diffusion of policies among political jurisdictions (Besley and Case 1995; Shipan and Volden 2008; Gilardi 2010). In a world of social comparison, reelection-minded politicians are well advised to keep in line with the performance that citizens observe in similar or adjacent political units. The findings here provide micro-level support for interdependence in political decision-making between political jurisdictions. While this may seem as an obvious and uncontroversial observation, most empirical work done in political science is actually ignoring social interdependence (Lazer 2005). The findings on social reference points are also in line with recent observational studies, which find that citizens factor in the global economy in elections (Kayser and Peress 2012). It fits the idea that not all governments are voted out of office during the recent global economic crisis. From a historical reference point all governments in power around 2008-2009 are doing worse today because of the world-wide recession. However, voters factor in the performance they observe elsewhere which in a global

economic crisis will tend to be equally bad (Kayser and Peress 2012, 661). Relative social comparison is thus a likely explanation of why incumbent governments can survive in a less fortunate global economy.

The importance of social reference points also raises the question about the role of political elites in the framing performance information (Druckman 2001; Olsen 2013*c,b*). To what degree do political elites affect the reference point by which their own performance is compared against? One line of argument could be that social reference points offer a larger pallet of selection than historical ones. Political elites will have a harder time forming their own history of performance than they will have pointing to politically helpful points of social reference. Combined with the findings of the negativity bias in social comparisons, we may expect that political elites aim for downward comparisons, i.e. to point to the performance of political jurisdictions doing worse than their own. The strong social comparison effect might also help explain why objective measures of performance can show improvements without being reflected in subjective perceptions of performance at the citizen level (Van Ryzin 2004). If performance evaluations have a strong relative component, then the provision of performance information can induce a form of 'hedonic treadmill' as has been found for perceptions of personal wealth and happiness (Brickman and Campbell 1971).

In real world settings, citizens will often be exposed to multiple reference points for the same target object. The experimental method used here does raise some concerns about external and ecological validity. Often there will be multiple opportunities for various simultaneous historical and social comparisons. What if an organization has improved its performance compared to last year but is still below in a salient social comparison? Maybe citizens' then judgment is likely to be ambivalent and characterized by mixed feelings (Kahneman and Miller 1986; Kahneman 1992). Future studies should seek to disentangle the effects of multiple reference points and move the question of reference points into more realistic and diverse treatment settings.

References

- Albert, S. 1977. "Temporal comparison theory." *Psychological Review* 84(6): 485–503.
- Ansolabehere, Stephen, Marc Meredith, and Erik Snowberg. 2013. "Asking About Numbers: Why and How." *Political Analysis* 21(1): 48–69.
- Baumeister, R. F., E. Bratslavsky, C. Finkenauer, and K. D. Vohs. 2001. "Bad is stronger than good." *Review of general psychology* 5(4): 323–370.
- Bendor, Jonathan, Daniel Diermeier, David A. Siegel, and Michael M. Ting. 2011. *A Behavioral Theory of Elections*. Princeton and Oxford: Princeton University.
- Besley, Timothy, and Anne Case. 1995. "Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition." *The American Economic Review* 85(1): 25–45.
- Bowman, N. A., and M. N. Bastedo. 2009. "Getting on the front page: Organizational reputation, status signals, and the impact of US News and World Report on student decisions." *Research in Higher Education* 50(5): 415–436.
- Boyne, George A., Oliver James, Peter John, and Nicolai Petrovsky. 2009. "Democracy and Government Performance: Holding Incumbents Accountable in English Local Governments." *The Journal of Politics* 71(04): 1273–1284.
- Brickman, P., and D. T. Campbell. 1971. "Hedonic relativism and planning the good society." In *Adaptation-level theory: A symposium*, ed. M. H. Apley. Michigan: Academic Press, 287–305.
- Chapman, D. W., and J. Volkmann. 1939. "A social determinant of the level of aspiration." *The Journal of Abnormal and Social Psychology* 34(2): 225–238.
- Cohen, J. E., and J. D. King. 2004. "Relative unemployment and gubernatorial popularity." *Journal of Politics* 66(4): 1267–1282.

- Conover, P. J., S. Feldman, and K. Knight. 1986. "Judging inflation and unemployment: The origins of retrospective evaluations." *Journal of Politics* 48(3): 565–588.
- Cyert, R. M., and J. G. March. 1963. *A behavioral theory of the firm*. Englewood Cliffs, New Jersey: Prentice-Hall International Series in Management and Behavioral Sciences in Business Series.
- Diener, E. 2000. "Subjective well-being: The science of happiness and a proposal for a national index." *American psychologist* 55(1): 34–43.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. 1st ed. Harper and Row.
- Druckman, J. N. 2001. "The implications of framing effects for citizen competence." *Political Behavior* 23(3): 225–256.
- Elkins, Zachary, and Beth Simmons. 2005. "On Waves, Clusters, and Diffusion: A Conceptual Framework." *The ANNALS of the American Academy of Political and Social Science* 598(1): 33–51.
- Espeland, Wendy N., and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 133(1): 1–40.
- Festinger, L. 1954. "A theory of social comparison processes." *Human Relations* 7(2): 117–140.
- Fiorina, M. P. 1981. *Retrospective voting in American national elections*. New Haven: Yale University Press.
- Garcia, Stephen M., Avishalom Tor, and Richard Gonzalez. 2006. "Ranks and Rivals: A Theory of Competition." *Personality and Social Psychology Bulletin* 32(7): 970–982.
- Gilardi, Fabrizio. 2010. "Who Learns from What in Policy Diffusion Processes?" *American Journal of Political Science* 54(3): 650–666.

- Goethals, G. R., and J. M. Darley. 1977. "Social comparison theory: An attributional approach." In *Social comparison processes: Theoretical and empirical perspectives*, eds. J. M. Suls, and R. L. Miller. Washington, DC:: Hemisphere. , 259–278.
- Goodman, P., and E. Haisley. 2007. "Social comparison processes in an organizational context: New directions." *Organizational Behavior and Human Decision Processes* 102(1): 109–125.
- Greve, Henrich R. 1998. "Performance, Aspirations, and Risky Organizational Change." *Administrative Science Quarterly* 43(1): 58–86.
- Heath, C. 1999. "Goals as Reference Points." *Cognitive Psychology* 38(1): 79–109.
- Hibbs, D. A., R. D. Rivers, and N. Vasilatos. 1982. "On the demand for economic outcomes: macroeconomic performance and mass political support in the United States, Great Britain, and Germany." *The Journal of Politics* 44(2): 426–462.
- Hoppe, Ferdinand. 1931. "Untersuchungen zur Handlungs- und Affektpsychologie." *Psychological Research* 14(1): 1–62.
- 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.
- Kahneman, D., and A. Tversky. 1979. "Prospect theory: An analysis of decision under risk." *Econometrica: Journal of the Econometric Society* 47(2): 263–291.
- Kahneman, D., and D. T. Miller. 1986. "Norm theory: Comparing reality to its alternatives." *Psychological review* 93(2): 136–153.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler. 1991. "Anomalies: The endowment effect, loss aversion, and status quo bias." *The Journal of Economic Perspectives* 5(1): 193–206.

- Kahneman, Daniel. 1992. "Reference points, anchors, norms, and mixed feelings." *Organizational Behavior and Human Decision Processes* 51(2): 296–312.
- Kayser, and M. Peress. 2012. "Benchmarking across Borders: Electoral Accountability and the Necessity of Comparison." *American Political Science Review* 106(03): 661–684.
- Key, V. O. 1966. *The Responsible Electorate: Rationality in Presidential Voting, 1936-1960*. Cambridge, MA: Belknap Press.
- Kinder, D. R., and D. R. Kiewiet. 1979. "Economic discontent and political behavior: The role of personal grievances and collective economic judgments in congressional voting." *American Journal of Political Science* 23(3): 495–527.
- Kramer, G. H. 1971. "Short-term fluctuations in US voting behavior, 1896-1964." *American Political Science Review* 65(1): 131–143.
- Lau, R. R. 1985. "Two explanations for negativity effects in political behavior." *American Journal of Political Science* 29(1): 119–138.
- Lau, Richard R. 1982. "Negativity in Political Perception." *Political Behavior* 4(4): 353–377.
- Lazer, David. 2005. "Regulatory Capitalism as a Networked Order: The International System as an Informational Network." *The ANNALS of the American Academy of Political and Social Science* 598(1): 52–66.
- Leibenstein, H. 1950. "Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand." *The Quarterly Journal of Economics* 64(2).
- Levinthal, Daniel, and James G. March. 1981. "A model of adaptive organizational search." *Journal of Economic Behavior & Organization* 2(4): 307–333.
- Linos, Katerina. 2011. "Diffusion through Democracy." *American Journal of Political Science* 55(3): 678–695.

- Lyubomirsky, S., and L. Ross. 1997. "Hedonic consequences of social comparison: a contrast of happy and unhappy people." *Journal of Personality and Social Psychology* 73(6): 1141–1157.
- March, J. G., and H. A. Simon. 1958. *Organizations*. New York: Wiley.
- March, James G. 1988. "Introduction: A Cronicle of speculations about organizational decision-making." Oxford, UK: .
- Markowitz, H. 1952. "The utility of wealth." *The Journal of Political Economy* 60(2): 151–158.
- McDermott, Rose, James H. Fowler, and Oleg Smirnov. 2008. "On the Evolutionary Origin of Prospect Theory Preferences." *The Journal of Politics* 70(02): 335–350.
- Moore, D. A. 2007. "Not so above average after all: When people believe they are worse than average and its implications for theories of bias in social comparison." *Organizational Behavior and Human Decision Processes* 102(1): 277–291.
- Mussweiler, T. 2003. "Comparison processes in social judgment: mechanisms and consequences." *Psychological review* 110(3): 472–489.
- Mussweiler, T., and A. C. Posten. 2011. "Relatively certain! Comparative thinking reduces uncertainty." *Cognition* 122: 236–240.
- Mussweiler, T., and K. Epstude. 2009. "Relatively fast! Efficiency advantages of comparative thinking." *Journal of Experimental Psychology: General* 138(1): 1–21.
- Nordhaus, William D. 1975. "The Political Business Cycle." *The Review of Economic Studies* 42(2): 169–190.
- Oliver, R. L. 1980. "A cognitive model of the antecedents and consequences of satisfaction decisions." *Journal of marketing research* 42: 460–469.

- Olsen, A. L. 2013a. “Framing Performance Information: An Experimental Study of the Negativity Bias.” Unpublished paper, Paper to be presented at the 11th Public Management Research Conference Madison, Wisconsin, June 20-23 2013. Panel: Measurement and Methods in Public Management Research.
- Olsen, Asmus L. 2013b. “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 L. 2013c. “The politics of digits: evidence of odd taxation.” *Public Choice* 154(1-2): 59–73.
- Pierce, Walter M. 1940. “Climbing on the Bandwagon.” *The Public Opinion Quarterly* 4(2): 241–243.
- Popkin, S. L. 1991. *The reasoning voter: Communication and persuasion in presidential elections*. Chicago: University of Chicago.
- Quattrone, George A., and Amos Tversky. 1988. “Contrasting Rational and Psychological Analyses of Political Choice.” *The American Political Science Review* 82(3): 719–736.
- Robinson-whelen, S., and J. Kiecolt-glaser. 1997. “The Importance of Social Versus Temporal Comparison Appraisals Among Older Adults1.” *Journal of Applied Social Psychology* 27(11): 959–966.
- Rozin, P., and E. B. Royzman. 2001. “Negativity Bias, Negativity Dominance, and Contagion.” *Personality and Social Psychology Review* 5(4): 296–320.
- Runciman, W. G. 1961. “Problems of Research on Relative Deprivation.” *European Journal of Sociology* 2(2): 315–323.
- Salmon, Pierre. 1987. “Decentralisation as an Incentive Scheme.” *Oxford Review Of Economic Policy* 3(2): 24–42.

- Samuelson, W., and R. Zeckhauser. 1988. "Status quo bias in decision making." *Journal of risk and uncertainty* 1(1): 7–59.
- Shibutani, Tamotsu. 1955. "Reference Groups as Perspectives." *American Journal of Sociology* 60(6): 562–569.
- Shipan, Charles R., and Craig Volden. 2008. "The Mechanisms of Policy Diffusion." *American Journal of Political Science* 52(4): 840–857.
- Simon, H. A. 1954. "Bandwagon and underdog effects and the possibility of election predictions." *Public Opinion Quarterly* 18(3): 245–253.
- Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *The Quarterly Journal of Economics* 69(1): 99–118.
- Smith, P. 1990. "The use of performance indicators in the public sector." *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 153(1): 53–72.
- Sniderman, P. M., R. A. Brody, and P. E. Tetlock. 1991. *Reasoning and choice: Explorations in political psychology*. Cambridge University Press.
- Soroka, Stuart N. 2006. "Good News and Bad News: Asymmetric Responses to Economic Information." *Journal of Politics* 68(2): 372–385.
- Stark, O., and J. E. Taylor. 1991. "Migration incentives, migration types: The role of relative deprivation." *The economic journal* 101(408): 1163–1178.
- Suls, Jerry, René Martin, and Ladd Wheeler. 2000. "Three Kinds of Opinion Comparison: The Triadic Model." *Personality and Social Psychology Review* 4(3): 219–237.
- Tversky, A., and D. Kahneman. 1974. "Judgment under uncertainty: Heuristics and biases." *Science* 185(4157): 1124–1131.

Tversky, A., and D. Kahneman. 1981. “The framing of decisions and the psychology of choice.” *Science* 211(4481): 453–458.

Tversky, A., and D. Kahneman. 1991. “Loss aversion in riskless choice: A reference-dependent model.” *The Quarterly Journal of Economics* 106(4): 1039–1061.

Tversky, Amos, and Daniel Kahneman. 1986. “Rational Choice and the Framing of Decisions.” *The Journal of Business* 59(4): 251–278.

Van Ryzin, G. G. 2004. “Expectations, performance, and citizen satisfaction with urban services.” *Journal of Policy Analysis and Management* 23(3): 433–448.

For online appendix

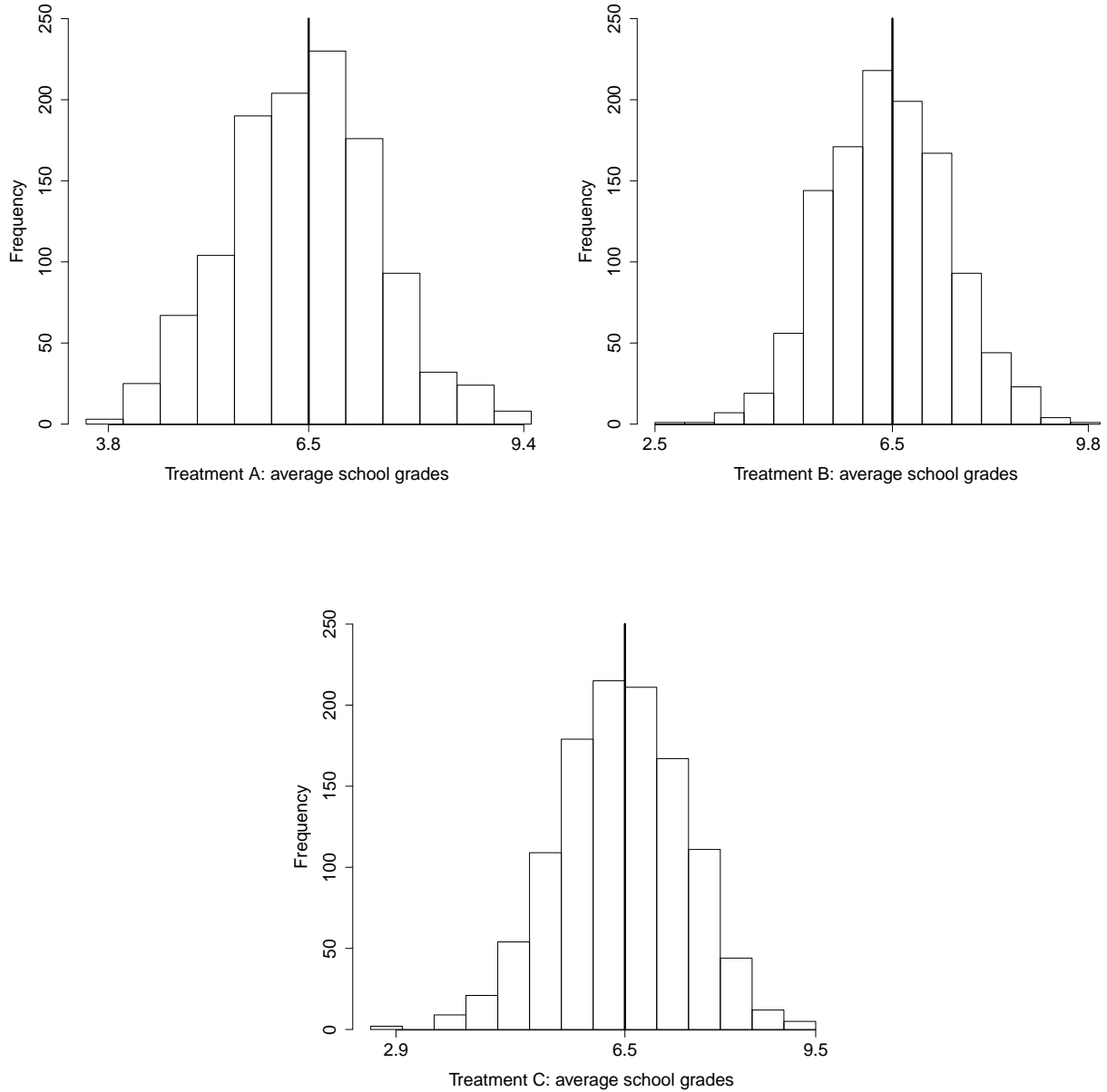


Figure 4: The figures show the distributions for which the absolute grade treatments were drawn under the three treatment conditions A, B, and C. All three distributions share the properties of $\mu = 6.5$ and $\sigma = 1.0$). The sample size of the distributions in the final data set were respectively (a) 1156, (b) 1148, and (c) 1139. The x-axis shows the min, mean, and max values drawn for each distribution.

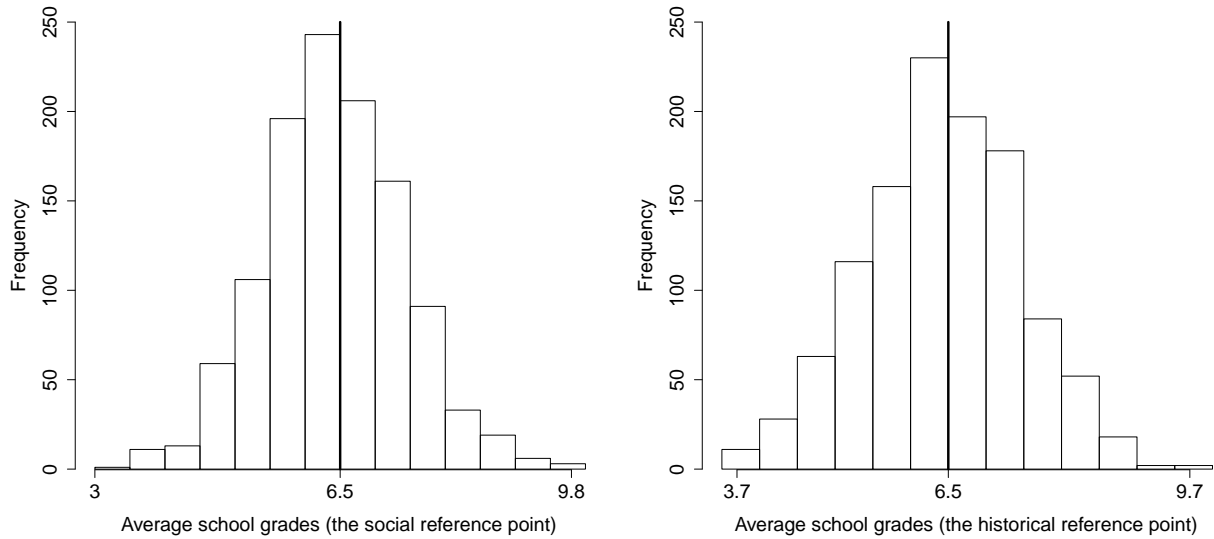


Figure 5: The figures show the distributions for the reference point grade in the treatment conditions B and C. Both distributions share the properties of $\mu = 6.5$ and $\sigma = 1.0$). The x-axis shows the min, mean, and max values drawn for each distribution.

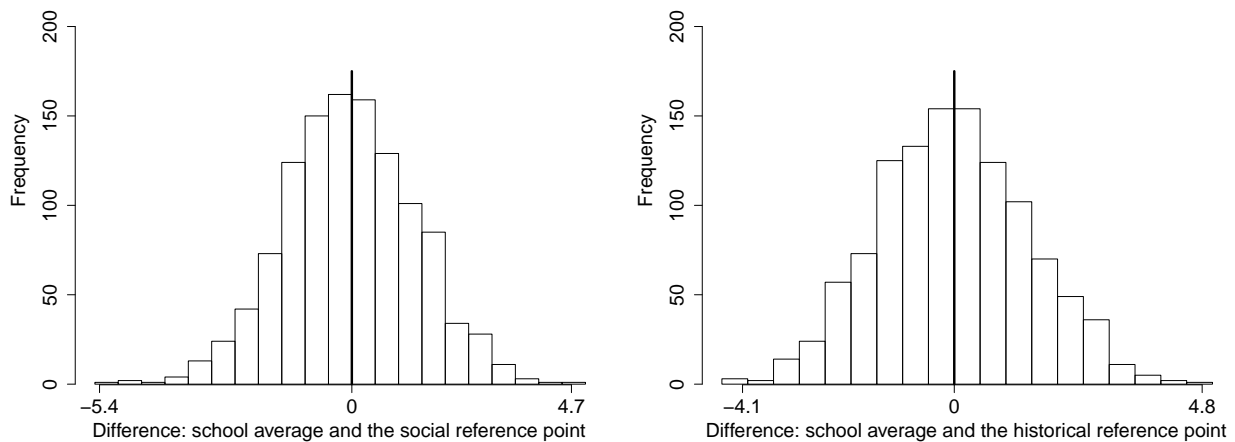


Figure 6: The figures show the relative distributions of grades in treatment conditions B and C. These relative distributions are given by subtracting the reference point grade from the targets schools average grade, i.e., *Absolute School Grade Average* – *Absolute Reference Point Grade Average*. The x-axis shows the min, mean, and max values for each distribution.

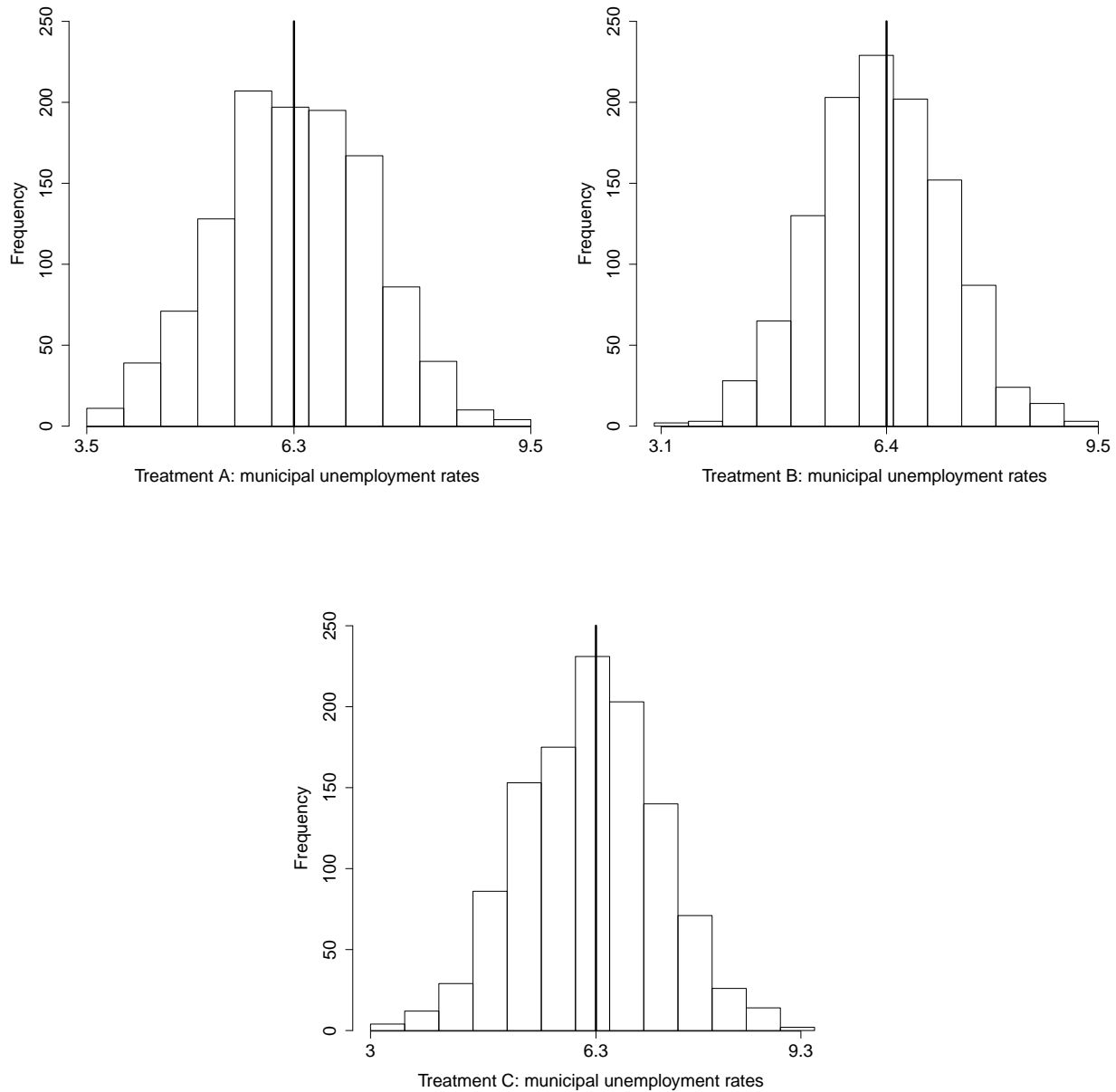


Figure 7: The figures show the distributions for which the absolute unemployment treatments were drawn under the three treatment conditions A, B, and C. All three distributions share the properties of $\mu = 6.3$ and $\sigma = 1.0$. The sample size of the distributions in the final data set were respectively (a) 1155, (b) 1142, and (c) 1146. The x-axis shows the min, mean, and max values drawn for each distribution.

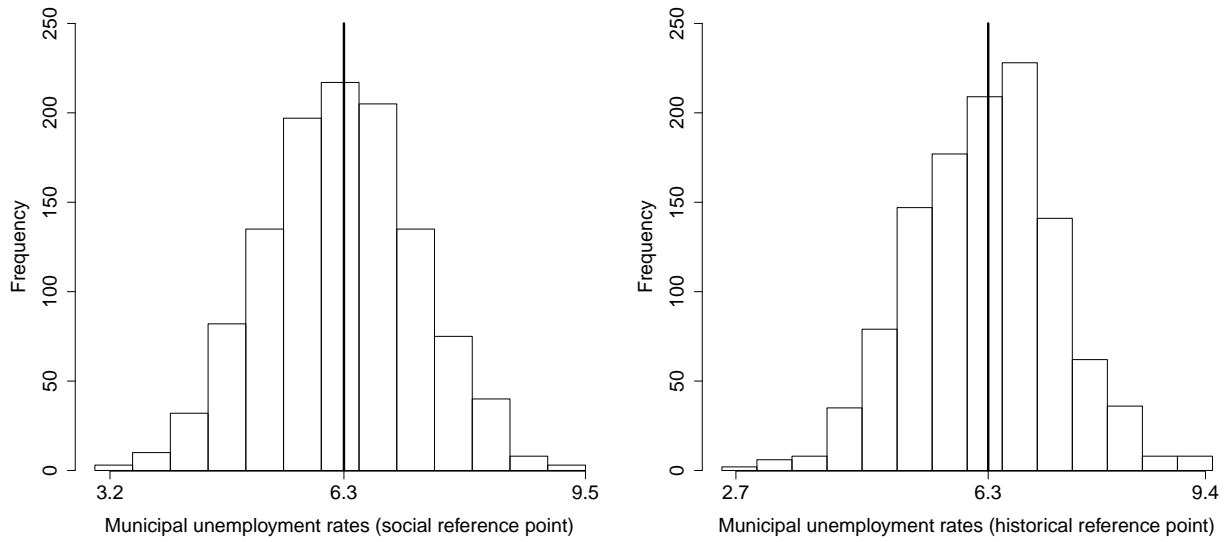


Figure 8: The figures show the distributions for the reference point unemployment in the treatment conditions B and C. Both distributions share the properties of $\mu = 6.3$ and $\sigma = 1.0$). The x-axis shows the min, mean, and max values drawn for each distribution.

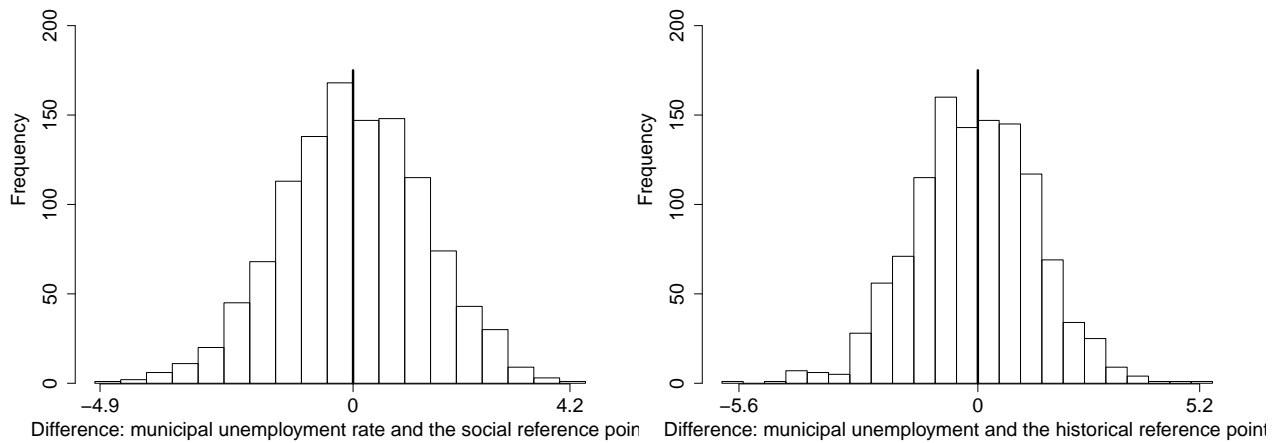
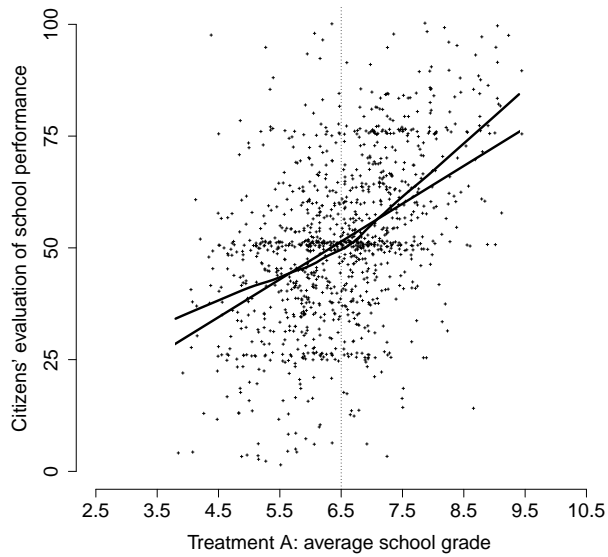
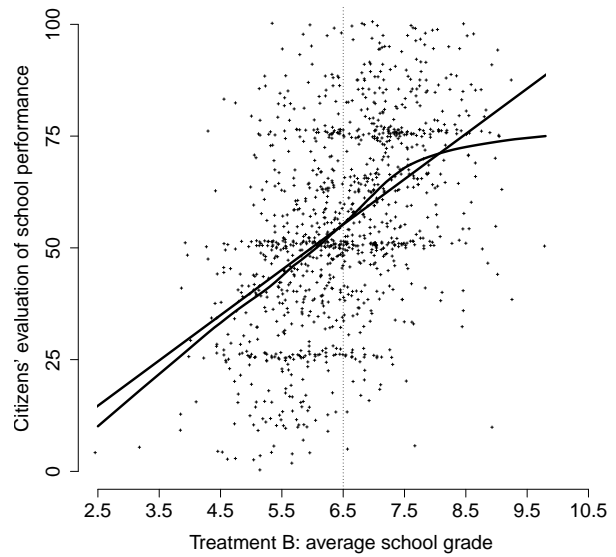


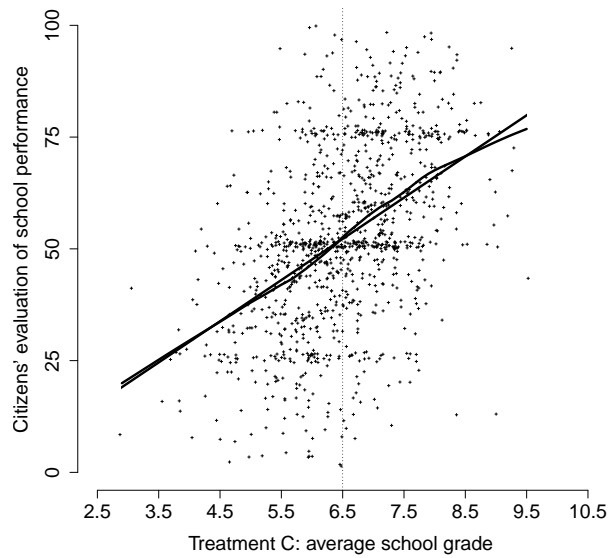
Figure 9: The figures show the relative distributions of unemployment rates in treatment conditions B and C. These relative distributions are given by subtracting the reference point unemployment from the targets municipal's unemployment rate, i.e., *Absolute Municipal Unemployment Rate* – *Absolute Reference Point Unemployment Rate*. The x-axis shows the min, mean, and max values drawn for each distribution.



(a)



(b)



(c)

Figure 10: School grade experiment: the plots show the correlation between the absolute grade for the target school under the three treatment frames and the evaluation provided by the citizens. The straight fit line is from a simple OLS regression while the flexible fit is a lowess estimation.

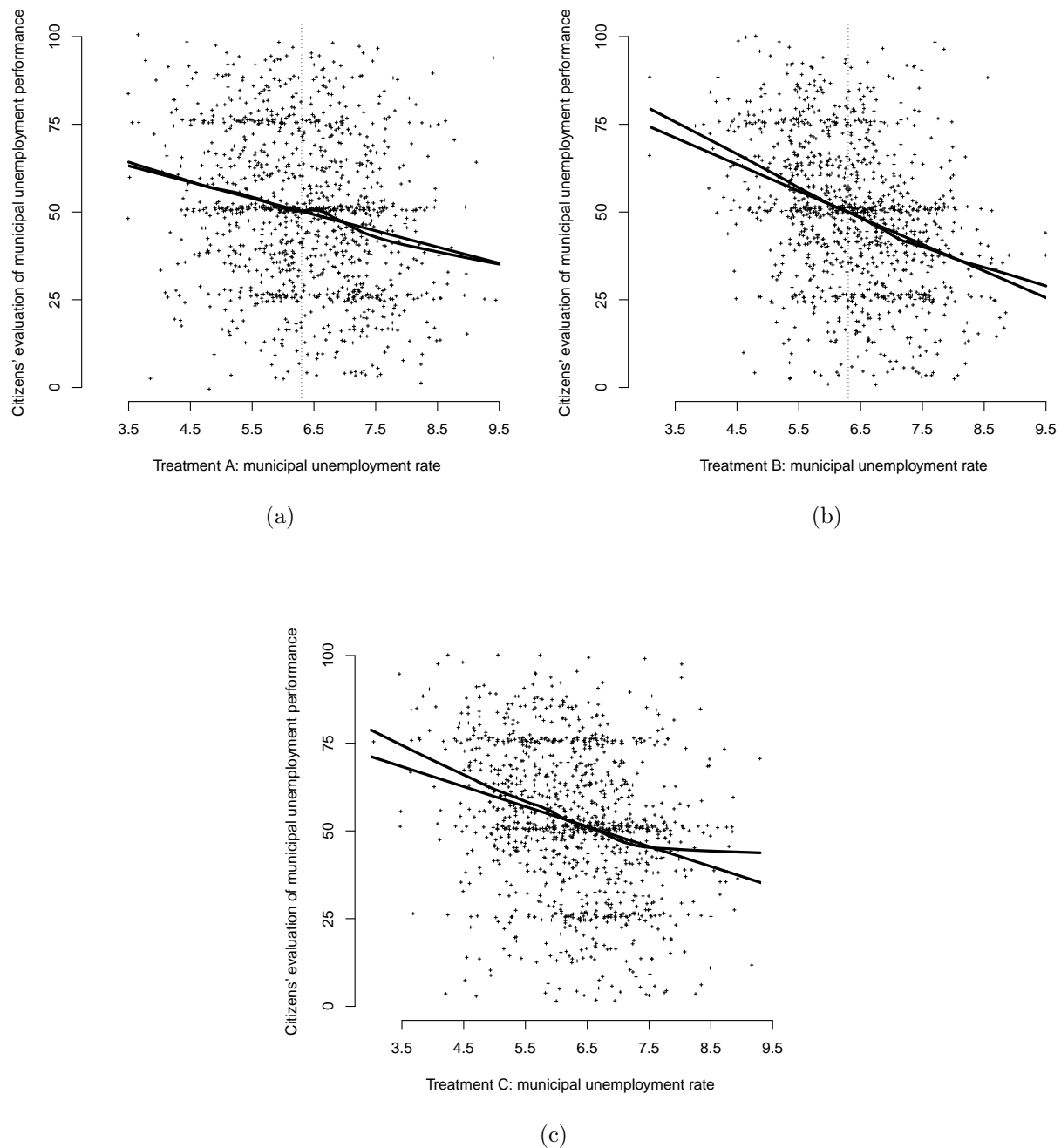


Figure 11: Unemployment rate experiment: the plots show the correlation between the unemployment rate for the target municipality under the three treatment frames and the evaluation provided by the citizens. The straight fit line is from a simple OLS regression while the flexible fit is a lowess estimation.