

Experimental evidence on the determinants of citizens' expectations toward public services

Nicola Bellé¹ | Paolo Belardinelli² | Maria Cucciniello³ | Greta Nasi³

¹Management and Health Laboratory, Institute of Management, Scuola Superiore Sant'Anna, Pisa, Italy

²O'Neill School of Public and Environmental Affairs, Indiana University, Bloomington, Indiana, USA

³Covid Crisis Lab and Department of Social and Political Science, Bocconi University, Milan, Italy

Correspondence

Paolo Belardinelli, O'Neill School of Public and Environmental Affairs, Indiana University, 1315 E 10th St, Bloomington, IN 47405, USA.
Email: pbelard@iu.edu

Abstract

We conducted three randomized experiments to investigate whether and to what extent citizens' expectations toward waiting times for public service delivery are influenced by reference points, either in the form of social or numerical references. Consistent with our theoretical expectations, our results provide convergent evidence of reference dependence. Specifically, informing citizens that waiting times are longer (shorter) relative to a social reference causes an increase (decrease) in expected waiting times. Additionally, due to an anchoring bias, priming citizens with a higher numerical value for waiting times extends their expected waiting times. Furthermore, in line with the expectancy-disconfirmation model, citizens' satisfaction with the service is causally impacted by the extent to which actual performance exceeds their expectations.

Evidence for practice

- Public managers should take citizens' expectations toward public services into account in the delivery process, as these expectations ultimately impact citizens' satisfaction.
- Citizens' expectations toward public services can be influenced by different reference points.
- Meaningful reference points, such as social norms, can impact citizens' expectations toward public service delivery, as informing citizens that service performance is higher (lower) than a social reference significantly increases (decreases) their expectations.
- Meaningless reference points, such as numerical anchors, can also affect citizens' expectations toward public service delivery, with higher meaningless numbers increasing citizens' expected waiting times and lower numbers doing the opposite.

INTRODUCTION

Citizens' satisfaction with public services plays a prominent role in citizens-government interactions, particularly in democratic countries where government officials are held accountable by the public and their legitimacy is also determined by the support they receive (Boyne et al., 2009; James, 2011). In this regard, research has consistently demonstrated that citizens' expectations significantly influence how they evaluate government actions and their overall satisfaction (Hjortskov, 2019, 2020;

James, 2009; Morgeson, 2012; Roch & Poister, 2006; Van Ryzin, 2004, 2006, 2013). As a consequence, citizens' expectations have gained substantial interest in the public administration literature, to the extent that one of the major journals in our field has recently dedicated a special issue to this topic (PMR, 2023).

Starting from Van Ryzin (2004), the expectancy-disconfirmation model has been among the most widely adopted to study the relationship between citizens' expectations, public service performance, and citizens' satisfaction with government services (Chen et al., 2022; Favero &

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Public Administration Review* published by Wiley Periodicals LLC on behalf of American Society for Public Administration.

Kim, 2021; Grimmelikhuisen & Porumbescu, 2017; James, 2009; Morgeson, 2012; Petrovsky et al., 2017; Van Ryzin, 2004, 2006, 2013). This model was first introduced in marketing research to study customer satisfaction with private sector goods and services (e.g., Cardozo, 1965; Churchill & Surprenant 1982; Oliver, 1980, 2014). At its core, the model posits that consumers form judgments about private sector goods and services using their prior expectations, their actual experience, and the gap between the two. Likewise, citizens will form their judgment about satisfaction with public services based on their expectations, the experienced public service performance, and the gap between prior expectations and actual performance, that is, disconfirmation.

Whereas initial work in public administration primarily relied on observational studies (James, 2009; Morgeson, 2012; Van Ryzin, 2004, 2006), experimental designs have been adopted to test and replicate the same hypotheses (Chen et al., 2022; Favero & Kim, 2021; Grimmelikhuisen & Porumbescu, 2017; Petrovsky et al., 2017; Van Ryzin, 2013). A recent meta-analysis (Zhang et al. 2022) of 17 studies from the public administration literature largely supports the relationships hypothesized by the expectancy-disconfirmation model. Still, they observe that empirical results might depend on scholars' specific methodological choices. Based on this observation, they identify a few recommendations for future research directions, including the need to delve deeper into how citizens form expectations (Zhang et al. 2022).

We follow up on the call "to unpack the antecedents of expectations" (Zhang et al. 2022, p. 156). In public administration, there is limited research addressing the determinants of citizens' expectations. Existing studies have focused on personality traits (Hjortskov, 2021), past expectations and satisfaction (Hjortskov, 2019), past performance (Hjortskov, 2019; James, 2011), and historical and social reference points (Favero & Kim, 2021). However, further exploration is needed.

Our study contributes to filling this gap by investigating how different reference points can influence citizens' expectations. More specifically, we draw from the focus theory of normative conduct (Cialdini et al., 1991) to examine the roles of descriptive and injunctive social norms. Descriptive norms relate to perceptions of what is commonly done, while injunctive norms pertain to what is commonly considered right or wrong. While considering all necessary theoretical caveats, this distinction seems to resonate with the differentiation between predictive and normative expectations explored recently in public administration literature (e.g., Favero & Kim, 2021; Hjortskov, 2021). Both sets of distinctions involve separating a positive component based on observations or expectations from a normative component referring to what is socially or morally recommended. Although it becomes evident that a parallel exists between these two sets of distinctions, fundamental differences remain. In particular, social norms pertain to judgments that refer to the majority, while expectations refer to individual

beliefs about the future. Social norms have been shown to affect individuals' behavior in a variety of domains that are highly relevant from a public perspective, including enhancing citizens' tax compliance (e.g., Coleman, 1996; Hallsworth et al., 2017; Larkin, Sanders, Andresen, and Algate 2018), sustaining charitable giving in wills (e.g., Behavioral Insight Team, 2013), and increasing vaccination coverage rates among healthcare professionals (Belle & Cantarelli, 2021).

Additionally, we draw from the literature on behavioral public administration and cognitive biases (Battaglio et al. 2019) to investigate whether meaningless numerical anchors can play a role in shaping citizens' expectations. We aim to explore whether this process of expectation determination deviates from rationality. The rationale behind our utilization of a meaningless anchor in our experiment was to investigate whether and how judgments and beliefs can be influenced by supposedly irrelevant factors (Thaler & Sunstein, 2021). Therefore, in the context of our study, a "meaningless" numerical anchor is a "supposedly irrelevant" numerical anchor. Our research question is inspired by Tversky and Kahneman's (1973) work on anchoring heuristics showing that judgments of unfamiliar quantities can be tilted toward irrelevant anchors, as well as recent work by Bordalo et al. (2020) demonstrating that beliefs respond to irrelevant cues.

The contribution of the manuscript is threefold. First, our study provides novel experimental findings that may help understand the role that different reference points, namely social reference points and numerical anchors, play in affecting predictive expectations about public services performance. The second contribution of our study lies in bridging the literature on citizens' expectations (Favero & Kim, 2021; Hjortskov, 2019, 2020; Petrovsky et al., 2017; Van Ryzin, 2004, 2006, 2013; Zhang et al., 2022) with scholarships on focus theory of normative conduct (Cialdini et al., 1991) and on anchoring bias (Tversky & Kahneman, 1974). Thirdly, our study further contributes to the scholarship on citizens' expectations and satisfaction by corroborating earlier work and experimentally testing the expectation-disconfirmation paradigm, while controlling for baseline normative expectations.

THEORETICAL BACKGROUND

Individuals' expectations have been a subject of theoretical exploration and empirical research across social sciences (Hjortskov, 2020). For instance, when focusing on the consumer (dis)satisfaction and service quality literature, a plurality of definitions have been proposed and discussed in terms of their discriminant validity (Teas, 1993). Santos and Boote (2003) identified 56 definitions, which they grouped into nine categories. Among these categories are "should" expectations, reflecting what consumers feel ought to happen, and "predicted" expectations, representing what consumers anticipate will happen. The latter category, which has been later commonly labeled as predictive (but also as

positive or descriptive) expectations, has received significant attention in economics. A debate has emerged within this area concerning whether these predictive expectations can be considered rational or not. On the one hand, for example, Muth (1961) uses a rationalistic approach to advance the hypothesis that expectations are analogous with predictions of the relevant economic theory. On the other hand, Lovell (1986) has reviewed alternative ways to model expectations and concluded that “expectations are a rich and varied phenomenon that is not adequately captured by the concept of rational expectations” (p. 120).

As citizens’ expectations have been shown to affect how citizens judge government action (Hjortskov, 2019, 2020; James, 2009; Morgeson, 2012; Roch & Poister, 2006; Van Ryzin, 2004, 2006, 2013), the study of expectations within the public administration literature is closely intertwined with the use of the expectancy-disconfirmation model (Chen et al., 2022; Favero & Kim, 2021; Grimmelikhuijsen & Porumbescu, 2017; James, 2009; Morgeson, 2012; Petrovsky et al., 2017; Van Ryzin, 2004, 2006, 2013). Originally introduced in marketing research (e.g., Cardozo, 1965; Churchill and Surprenant 1982; Oliver, 1980, 2014), this model has found extensive application in public administration, particularly to investigate the relationship between citizens’ expectations, public service performance, and citizens’ satisfaction with government services. In this context, public administration studies have primarily focused on predictive expectations (Hjortskov, 2020). Nonetheless, there have been a few attempts to investigate the role of normative expectations (Favero & Kim, 2021; Hjortskov, 2020; Hjortskov, 2021; James, 2009; Petrovsky et al., 2017; Poister & Thomas, 2011) and findings from most recent studies suggest that normative expectations may have a stronger association with satisfaction compared to predictive expectations (Favero & Kim, 2021; Hjortskov, 2020).

The expectancy-disconfirmation model posits that citizens form their judgment about satisfaction with public services based on their expectations, the experienced public service performance, and the gap between prior expectations and actual performance, that is, disconfirmation (Chen et al., 2022; Favero & Kim, 2021; Grimmelikhuijsen & Porumbescu, 2017; James, 2009; Morgeson, 2012; Petrovsky et al., 2017; Van Ryzin, 2004, 2006, 2013). More specifically, expectations are correlated with perceived performance and they both contribute to form disconfirmation, which can be either positive or negative. Positive disconfirmation happens when perceived performance meets or exceeds expectations, while negative disconfirmation is observed when performance falls short of expectations. Whereas the former contributes to increased satisfaction with public services, the latter fuels dissatisfaction with public services. Nonetheless, satisfaction is also directly affected by perceived performance and expectations (Van Ryzin, 2004, 2006, 2013, Zhang et al. 2022). Zhang and colleagues (2022) report results from a meta-analysis of 17 public administration studies that largely support the relationships hypothesized by the expectancy-

disconfirmation model, but they also observe that the scholarship needs “to unpack the antecedents of expectations” (p. 156).

Given that expectations are formalized as exogenous in the expectancy-disconfirmation model, it is pivotal to better understand their origins. Whereas the consumer satisfaction literature has a long history of debating the antecedents of consumers’ expectations, focusing on factors like personal needs, past performance, image, and explicit and implicit service promises in shaping both predictive and normative expectations (Clow et al., 1997; Devlin et al., 2002; Zeithaml et al., 1993), research on determinants of citizens’ expectations in public administration is relatively recent. Hjortskov (2019) tests three different hypotheses on antecedents of expectations, namely prior expectations, prior performance, and prior satisfaction, finding that prior expectations and prior satisfaction are predictors of future expectations, while prior performance is not. This latter result partially contrasts with James (2011), who finds that past performance can function as an antecedent of predictive expectations, but not so much of normative expectations. Most recently, an experimental study by Favero and Kim (2021) delves into the role of historical and social reference points, showing that both of these reference points can determine both predictive and normative expectations. Additionally, an observational study by Hjortskov (2021) explores whether different personality traits can explain predictive and normative expectations and finds that, consistent with theoretical expectations, the former are not affected by personality traits whereas the latter correlate positively with agreeableness, conscientiousness, and openness, and negatively with extraversion. In what follows, drawing from the scholarship on the focus theory of normative conduct (Cialdini et al., 1991) and the scholarship on anchoring bias (Tversky & Kahneman, 1974), we develop hypotheses on the role played by social norms and numerical anchors, respectively, in forming citizens’ expectations.

Social norms

The term ‘norm’ can refer to multiple meanings (Shaffer, 1983), including what is commonly done and what is widely approved. The focus theory of normative conduct (Cialdini et al., 1991) distinguishes between descriptive and injunctive social norms to shed light on how norms can affect human expectations and behavior. Descriptive norms mark perceptions of what most others do and can provide citizens with reference points to confront their own behavior. Injunctive norms mark perceptions of what the majority of others believe is right or wrong to do and entail information about what is normatively approved or disapproved in a given group.

Both types of norms have been shown to predictably influence individuals’ behavior in a variety of domains, including enhancing citizens’ tax compliance

(e.g., Coleman, 1996; Hallsworth et al., 2017; Larkin, Sanders, Andresen, and Algate 2018), promoting residential energy conservation (e.g., Allcott, 2011; Cialdini & Schultz, 2004; Schultz et al., 2007), preserving petrified woods (e.g., Cialdini et al., 2006), increasing curbside recycling (e.g., Schultz, 1999), sustaining charitable giving in wills (e.g., Behavioral Insight Team, 2013), reducing youth initiation to smoking (e.g., Linkenbach & Perkins, 2003), reducing unnecessary prescriptions of antibiotics in general practice (e.g., Hallsworth et al., 2016), increasing mask wearing for COVID-19 reduction (Bokemper et al., 2021), and increasing vaccination coverage rates among health-care professionals (Belle & Cantarelli, 2021).

In the focus theory of normative conduct, Cialdini et al. (1991) posit that norms activate conformant behavior only when subjects' attention is focused on that norm. Belle and Cantarelli (2021) observe that this resonates naturally with two well-established research strands in behavioral science. On the one hand, starting from the classical studies by Asch (1956), research on conformity has shown that individuals generally prefer to conform to the majority rather than being outcasts (e.g., Bond & Smith, 1996; Schultz et al., 2007; Thaler & Sunstein, 2009). On the other hand, foundational studies on heuristics and cognitive biases point to the mechanism through which the impact of social norms plays out: the availability heuristic (e.g., Kahneman, 2011; Tversky & Kahneman, 1973). When relying on heuristics to cope with environmental complexity, thoughts that come more quickly to mind tend to disproportionately influence behavior (Kahneman, 2011). Following the same strand of research, the nudge theory (Thaler & Sunstein, 2009) argues that social norms can trigger such an availability heuristic and nudge individuals to conform to what the majority of other people do or believe is right to do. In their words, 'social norm [is] itself a nudge' (Thaler & Sunstein, 2009, 259).

Within the public administration literature focusing on citizens' expectations and satisfaction, scholars have studied how social comparisons and relative performance information can influence how public managers (Webeck & Nicholson-Crotty, 2020), public employees (Keiser & Miller, 2019), and citizens (Barrows et al., 2016; Olsen, 2017) assess the performance of public services. Favero and Kim (2021) apply the same reasoning to the study of citizens' expectation formation.

We bridge these approaches by studying the effect of distinct social reference points on citizens' expectations, specifically descriptive and injunctive social norms (Cialdini et al., 1991). More specifically, we are interested in studying how predictive expectations change when citizens are provided with comparative performance information with respect to a social reference. This reference can be either purely descriptive (e.g., what happens in the majority of other municipalities) or injunctive in the sense that it entails some degree of normative judgment (e.g., what the majority of citizens think is right). Whereas

we have a clear hypothesis regarding the effect of manipulating comparative performance information, we adopt an explorative approach with respect to the effect caused by a descriptive social reference as opposed to an injunctive social reference on predictive expectations. Being among the first to compare the effects of these two types of norms, we lack both theoretical reasons and evidence on the expected strength of descriptive versus injunctive norms. Therefore, we formulate our social reference dependence hypothesis in expectation formation.

*Social reference dependence hypothesis—*Informing citizens that waiting times are longer [shorter] than a social reference increases [decreases] expected public service waiting times.

Anchoring

Anchoring is the cognitive tendency to estimate unknown values by making adjustments from an initial reference number. As a heuristic and cognitive shortcut used by decision-makers when assessing unknown quantities, anchoring has been shown to bias decisions in various domains because "different starting points yield different estimates, which are biased toward the initial values" (Tversky & Kahneman, 1974, 1128).

Experimental studies across social sciences have consistently detected the anchoring effect in domains as different as general knowledge (e.g., Simmons et al., 2010), human resource management (e.g., Thorsteinson et al., 2008); legal judgments (e.g., Englich et al. 2006, Bennett, 2014), negotiations (e.g., Orr & Guthrie 2005), and economic valuations (e.g., Alevy et al., 2015). In the public administration literature, a systematic review of studies on cognitive biases reports that the anchoring bias is among 'the observable cognitive biases that public administration scholars have investigated the most' (Battaglio et al. 2019, p. 314), especially as applied to public personnel management (Bellé et al., 2017; Nagtegaal et al., 2020; Pandey & Marlowe, 2015).

The anchoring effect holds even when the numerical reference points are random and, therefore, meaningless. For example, in a classical study, students were asked to indicate the percentage of African Nations in the United Nations. The median estimate was 45% when students were exposed to the spin of a wheel of fortune that landed on the number 65, significantly higher than the 25% median estimate of those exposed to the number 10. In another study by Englich and colleagues (2006), participating judges anchored their sentencing decisions to the roll of a pair of dice that were loaded so that they always showed either the numbers 1 and 2 (low anchor) or the numbers 3 and 6 (high anchor). Judges exposed to the high anchor gave higher final sentences, that is, about 8 months, than those confronted with a low anchor, that

is, about 5 months. Neither the wheel of fortune in the former study nor the pair of dice in the latter study could possibly provide any useful information about anything (Tversky & Kahneman, 1974). In other words, ‘anchors that are obviously random can be just as effective as potentially informative anchors’ (Kahneman, 2011, p. 125).

Whereas citizens’ expectations are influenced by numerical anchors in the form of past performance (Hjortskov, 2019; James, 2011), historical and social reference points (Favero & Kim, 2021), in this study, we investigate whether and how meaningless numerical anchors affect citizens’ expectations. The rationale behind our focus on meaningless anchors is to investigate whether and how judgments and beliefs, in the form of expectations, can be influenced by supposedly irrelevant factors (Thaler & Sunstein, 2021). Our hypothesis is inspired by Tversky and Kahneman’s (1973) work on anchoring heuristics showing that judgments of unfamiliar quantities can be tilted toward irrelevant anchors as well as recent work by Bordalo et al. (2020) demonstrating that beliefs respond to irrelevant cues. We, therefore, formulate our *anchoring hypothesis* in expectation formation.

Anchoring hypothesis—Exposing citizens to meaningless numerical anchors affect their predictive expectations so that a higher anchor increases expected waiting times in processing citizens’ applications.

METHODS

Experiment 1

We first ran a two-step factorial vignette experiment with 2438 subjects that reasonably resembled the Italian working-age population. Among our participants, 52% were female, about 76% were less than 50 years old, and 39% had at least one degree from post-secondary education. Table A1 in the Appendix reports demographic characteristics of our sample, compared to the Italian population. Participants were recruited in November 2021 through a professional service company. The research design was approved by the Ethics Committee of the university of one of the authors.

Step 1. Manipulation of expectations

Participants in our factorial experiment were first asked to imagine that they had to use the registry service of the municipality where they reside to request an electronic ID card. We provided participants with relative performance information in the form of waiting times for the electronic ID card. This piece of performance information was the object of our manipulations. Respondents ($n = 2438$) were randomly assigned to one of four experimental

conditions, resulting from the combination of two factors, each manipulated at two levels as follows:

- Factor 1. Relative performance—Expected waiting times for the appointment for the electronic ID card
 - Level 1. Much shorter than [social reference]
 - Level 2. Much longer than [social reference]
- Factor 2. Social reference
 - Level 1. In the majority of other municipalities (descriptive social norms)
 - Level 2. What the majority of citizens think is right (injunctive social norms)

The first experimental factor aimed at manipulating expected performance in the form of waiting times. The second factor was a manipulation of the social reference point. While what *the majority of other municipalities do* reflects an operationalization of a descriptive social norm, *what the majority of citizens think is right* speaks to an injunctive social norm. Based on the information provided, participants were asked to indicate how many days they expected to wait for the electronic ID card.

Step 2. Satisfaction

Once respondents indicated their waiting time expectations, they were asked to imagine that the waiting time for the first appointment for the electronic ID card was known. Here we manipulated the actual number of waiting days, and respondents were randomly exposed to one of two experimental conditions: either 27 or 55 days. We chose these numbers based on actual data from the municipality of Milan during the years 2020 and 2019, respectively. We create our disconfirmation variable as the difference between predictive expectations (i.e., the outcome of the first step of the experimental procedure) and the randomly assigned actual performance. Finally, respondents were asked to state their satisfaction with the service on a scale from 0 to 10. The experimental vignette is reported in the online Appendix.

Experiment 2

In a second survey with 538 subjects recruited in February 2022 through a professional service company and reasonably representative of the Italian working-age population, we sequentially ran two additional experiments: *Experiment 2* and *Experiment 3*. Among our participants, 47% were female, about half of them were less than 50 years old, and 34% had at least one degree from post-secondary education. Table A2 in the Appendix reports demographic characteristics of our sample, compared to the Italian population. The research was approved by the same Ethics Committee that approved the first survey. In *Experiment 2*, subjects were first asked to imagine that

Avatar is a municipality where citizens, to use public services, forward their requests to public offices and await their processing. We then randomly assigned participants to three experimental arms: namely control, low anchor, and high anchor. Participants were forced to spend 10 s on either a blank screen (control), a screen showing the number 50 (low anchor), or a screen showing the number 500 (high anchor), depending on the randomly assigned experimental condition. The blank screen of the control condition was to ensure that the only varying element of the treatment groups was the numerical anchor. Then, citizens in our sample were asked to indicate the expected waiting times for processing citizens' applications for municipal services in Avatar.

Experiment 3

Experiment 3 is an exact replication of *Experiment 1*. However, along with the same demographic items accompanying *Experiment 1*, at the beginning of this survey, we included an additional item on how many days the participants deemed it acceptable to wait for the registry service of their municipality to issue an electronic ID card; this was our measure of normative expectations (Favero & Kim, 2021; Petrovsky et al., 2017) and allowed us to experimentally study the role of predictive expectations in shaping citizens' satisfaction while controlling for normative expectations.

RESULTS

Experiment 1

Table 1 reports descriptive statistics of our two main outcomes of *Experiment 1*, namely (i) subjects' predictive expectations, that is, how many days citizens expect to wait for the electronic ID card and (ii) satisfaction with the registry service.

Table 2 reports estimates from three ordinary least squares models predicting the number of expected waiting days. The coefficient of the relative length of wait time in Model 1 indicates that informing citizens that wait times are relatively longer rather than shorter than a social reference increases their expected waiting days by 3.97 days on average ($p < .01$). The coefficient of the type of social norm in Model 2 indicates that, when averaging across the two levels of the first factor, expected waiting days are 1.7 fewer when the social norm is descriptive rather than injunctive. However, this coefficient becomes not significant in the full Model 3. Table A3 in the Appendix reports the results of analyses of variance (ANOVA) on the same data. Whereas our manipulation of estimated waiting times does affect citizens' expectations, these are not influenced by the type of social reference, that is, descriptive or injunctive norms. In other words, it is the

relative performance but not the type of social reference that matters in forming citizens' expectations. Figure A1 (panel A) in the Appendix provides a visual representation of these findings.

When we inform subjects about actual performance at step 2, we find that—other things being equal—having to wait 27 rather than 55 days significantly increases satisfaction with the service by 1.17 points, which is almost half a standard deviation (see Model 1 in Table 3). The coefficient of predictive expectations in Model 2 indicates that—other things being equal—satisfaction increases by .06 scale points per each extra day that subjects are expecting to wait. In other words, satisfaction tends to be higher when the expected performance is lower (i.e., longer waiting times). Because the disconfirmation variable was created as the difference between predictive expectations (i.e., the outcome of the first step of the experimental procedure) and the randomly assigned actual performance, a positive value of disconfirmation entails actual performance is better than expected and a negative value means that actual performance is worse than expected. In other words, disconfirmation indicates the extent to which actual performance exceeds expectations. Therefore, the larger the disconfirmation, the better. Model 3 indicates a positive impact of disconfirmation on satisfaction. More precisely, other things being equal, satisfaction increases by 0.05 scale points as the difference between the number of expected waiting days and the number of actual waiting days increases by one unit. Full Model 4, which is also reported in Figure A1 (panel B) in the Appendix and provides a complete test of our expectancy disconfirmation model, indicates that a one

TABLE 1 Outcomes' descriptive statistics.

<i>N</i> = 2438	Mean	SD	Min	Max
Predictive expectations (waiting days)	29.2	24.79	0	100
Satisfaction	3.9	2.91	0	10

TABLE 2 Treatment effects on expected waiting days-OLS.

	Model 1	Model 2	Model 3
	<i>N</i> = 2438	<i>N</i> = 2438	<i>N</i> = 2438
Estimated waiting times:	3.97***		3.93***
Longer versus shorter	(1.001)		(1.422)
Social norm: Descriptive versus Injunctive		−1.70* (1.004)	−1.48 (1.405)
Longer estimated waiting time × Descriptive social norm			−0.06 (2.005)
Cons	27.23*** (0.702)	30.04*** (0.713)	28.02*** (1.022)

Note: Standard errors in parentheses.

****p*-value <.01; ***p*-value <.05;

**p*-value <.1.

TABLE 3 Treatment effects on satisfaction/Test of expectancy-disconfirmation model.

	Model 1	Model 2	Model 3	Model 4	Model 4 (std coefficients)
	<i>N</i> = 2438	<i>N</i> = 2438	<i>N</i> = 2438	<i>N</i> = 2438	<i>N</i> = 2438
Better performance (27 days)	1.17*** (0.116)				
Predictive expectations (manipulated)		0.06*** (0.002)		0.02*** (0.004)	0.16
Disconfirmation ^a			0.05*** (0.002)	0.04*** (0.004)	0.39
Cons	3.32*** (0.082)	2.20*** (0.079)	4.53*** (0.054)	3.81*** (0.166)	

Note: Standard errors in parentheses.

****p*-value <.01; ** *p*-value <.05;

* *p*-value <.1.

^aWe measure disconfirmation as the difference between predictive expectations (i.e., the outcome of the first step of the experimental procedure) and the randomly assigned actual performance (Disconfirmation = Number of expected waiting days – Number of actual waiting days).

TABLE 4 Outcome's descriptive statistics.

<i>N</i> = 538	Mean	SD	Min	Max
Expectations (waiting days)	23.4	73.75	0	500

standard deviation increase in the disconfirmation variable causes an increase of 0.39 standard deviations in subjects' satisfaction. In addition, one additional standard deviation in the expectations variable positively affects satisfaction by 0.16 standard deviations. Therefore, according to our data, expectations and disconfirmation simultaneously affect satisfaction with the registry service.

Experiment 2

Table 4 reports descriptive statistics of our outcome for *Experiment 2*: (i) how many days citizens expect to wait for the Avatar administration to process citizens' applications for municipal services.

We find that exposing citizens in our sample to a meaningless number has a sizable impact on their expectations. Table 5 reports the mean number of expected waiting days for the three experimental arms, namely the control group, in which participants were asked how many days they expected to wait for the Avatar administration to process citizens' applications for municipal services without being shown any number beforehand, the low anchor condition in which participants were shown the number 50, and the high anchor condition in which subjects were shown the number 500. Table A4 displays the results of an analysis of variance showing that subjects who were exposed to a high anchor reported expecting to wait about 34 days longer ($p < .001$) than their counterparts in the low anchor condition and some

TABLE 5 Mean expected waiting days, by experimental arm.

Control (<i>N</i> = 217)	11.10 (16.541)
Low anchor (<i>N</i> = 173)	16.02 (17.173)
High anchor (<i>N</i> = 148)	50.02 (134.620)
Total (<i>N</i> = 538)	23.39 (73.752)

Note: Standard deviations in parentheses.

39 days more ($p < .001$) than citizens in the control arm. These effects are driven by a difference across experimental conditions in the proportion of respondents answering 500. It is worth noting that the number of days citizens expect to wait is not statistically different between the low anchor and the control group. On this latter point, post-hoc power calculations reveal that, given our sample size and the observed means and standard deviations of predictive expectations, power is equal to 30%. Therefore, we cannot rule out the possibility of type II error in testing for a difference between the low anchor and the control condition. Figure A2 in the Appendix visualizes estimates from an OLS regression.

Experiment 3

Table 6 reports descriptive statistics of our two main outcomes of *Experiment 3*, namely (i) subjects' predictive expectations, that is, how many days citizens expect to wait for the electronic ID card and (ii) satisfaction with the registry service. The same table reports our measure of

normative expectations, that is, how many days respondents think it is acceptable to wait for the electronic ID card, which was included in this replication of *Experiment 1*. It is worth noticing again that, differently from descriptive expectations, normative expectations were not

manipulated in the experiment, but were measured through an item in the beginning of the survey.

Table 7 mirrors Table 2 and reports estimates from three ordinary least squares models predicting the number of expected waiting times. As was the case in *Experiment 1*, the coefficient of the relative length of wait time in Model 1 indicates that informing citizens that wait times are relatively longer rather than shorter than a social reference increases their expected waiting days by 4.89 days on average ($p < .01$). However, different from *Experiment 1*, the coefficient of the type of social norm in Model 2 indicates that the type of social reference does not significantly impact citizens' expectations. These results hold in the full Model 3, which is also plotted in Figure A3 (panel A). Table A5 reports the results of analyses of variance (ANOVA) on the same data. Whereas our manipulation of estimated waiting times does affect citizens' expectations, these are not influenced by the type of social reference, that is, descriptive or injunctive norms. In other words, these findings confirm *Experiment 1* results in that relative performance but not the type of social reference matters in forming citizens' expectations.

When we inform subjects about actual performance at step 2, we find that—other things being equal—having to wait 27 rather than 55 days significantly increases satisfaction with the service by 0.41 points, which is almost half a standard deviation (see Model 1 in Table 8). Modules 2 and 3, respectively, indicate that—other things being equal—satisfaction increases by 0.07 scale points per each extra day that subjects are expecting to wait (predictive expectations-manipulated), as well as per each additional day subjects think it is acceptable to wait

TABLE 6 Outcomes' descriptive statistics.

<i>N</i> = 538	Mean	SD	Min	Max
Predictive expectations (waiting days)	16.7	19.33	0	100
Normative expectations (waiting days)	14.2	18.90	0	100
Satisfaction	2.3	2.47	0	10

TABLE 7 Treatment effects on expected waiting days-OLS.

	Model 1 <i>N</i> = 538	Model 2 <i>N</i> = 538	Model 3 <i>N</i> = 538
Estimated waiting times: Longer versus shorter	4.89*** (1.655)		6.49*** (2.329)
Social norm: Descriptive versus Injunctive		−0.95 (1.668)	0.75 (2.321)
Longer estimated waiting time × Descriptive social norm			−3.28 (3.314)
Cons	14.30*** (1.160)	17.17*** (1.173)	13.92*** (1.647)

Note: Standard errors in parentheses.

****p*-value <.01; ** *p*-value <.05;

* *p*-value <.1.

TABLE 8 Treatment effects on satisfaction/Test of expectancy-disconfirmation model.

	Model 1 <i>N</i> = 538	Model 2 <i>N</i> = 538	Model 3 <i>N</i> = 538	Model 4 <i>N</i> = 538	Model 5 <i>N</i> = 538	Model 5 (std coefficients)	Model 6 <i>N</i> = 538	Model 6 (std coefficients)
Better performance (27 days)	0.41* (0.212)							
Predictive expectations (manipulated)		0.07*** (0.005)			0.05*** (0.008)	0.35	0.01 (0.009)	0.1
Normative expectations (observed)			0.07*** (0.005)				0.04*** (0.007)	0.34
Disconfirmation ^a				0.05*** (0.004)	0.02*** (0.006)	0.2	0.02*** (0.006)	0.22
Cons	2.07*** (0.151)	1.17*** (0.12)	1.28*** (0.113)	3.55*** (0.135)	2.05*** (0.284)		2.01*** (0.273)	
<i>R</i> -squared	0.01	0.27	0.29	0.24	0.29		0.34	

Note: Standard errors in parentheses.

****p*-value <.01; ** *p*-value <.05;

* *p*-value <.1.

^aWe measure disconfirmation as the difference between predictive expectations (i.e., the outcome of the first step of the experimental procedure) and the randomly assigned actual performance (Disconfirmation = Number of expected waiting days – Number of actual waiting days).

(normative expectations-observed). In other words, satisfaction tends to be higher when the predicted performance is lower (i.e., longer waiting times) and when the appropriate performance, according to citizens' judgment, is lower (i.e., longer waiting times). Moving to Model 4, our experimental data indicate a positive impact of disconfirmation (i.e., the difference between predictive expectations and actual performance) on satisfaction. More precisely, other things being equal, satisfaction increases by 0.05 scale points when the difference between the number of expected waiting days and the number of actual waiting days increases by one unit. Model 5 in Table 8 provides a complete test of the expectancy disconfirmation model that does not control for normative expectations. Results hold, indicating that a one standard deviation increase in the disconfirmation variable causes a 0.20 standard deviation increase in subjects' satisfaction. In addition, one additional standard deviation in our predictive expectation variable positively affects satisfaction by 0.35 standard deviations. Therefore, as in Experiment 1, predictive expectations and disconfirmation seem to affect satisfaction with the registry service simultaneously. However, the coefficients included in the full Model 6 indicate that predictive expectations become insignificant when controlling for normative expectations. At the same time, a one standard deviation increase in the normative expectations variable positively affects subjects' satisfaction by 0.34 standard deviations. In addition, one additional standard deviation in the disconfirmation variable causes a 0.22 standard deviation increase in subjects' satisfaction. This can be clearly seen in Figure A3 (panel B) in the Appendix. In a nutshell, when keeping fixed the number of days citizens think is right to wait, the number of days they predict to wait does not affect their satisfaction anymore, but disconfirmation does. Table A6 included in the Appendix reports additional models testing the robustness of these findings.

DISCUSSION

We explored the impact of different reference points, in the form of social norms and meaningless numerical anchors, on predictive expectations across three online randomized surveys. Taken together, our results show that informing citizens about the relative performance of a service, compared to a social reference, significantly impacts their expectations. More precisely, knowing that waiting times in their municipality are relatively lower (higher) increases (decreases) citizens' expectations. Interestingly, descriptive and injunctive social norms do not seem to have different impacts. Furthermore, our experimental data suggest that reference dependence may extend beyond social references. More precisely, citizens' predictive expectations about waiting times for processing applications for public services were significantly influenced by exposure to meaningless numbers, with a

higher anchor increasing the expected number of waiting days, compared to a lower anchor. Besides investigating these antecedents of expectations, we found that the more actual performance exceeds the expected performance, the greater the satisfaction. Our conclusions hold true even when controlling for normative expectations, that is, what citizens think government performance ought to be. Beyond disconfirmation, consistent with previous experimental studies (Filtenborg et al., 2017; Grimmelikhuijsen & Porumbescu, 2017), expectations seem to have a significant direct impact on citizens' satisfaction. This significant relationship may be explained by the assimilation effect, according to which, similar to confirmation bias, citizens might not be able to precisely judge public service performance and might assimilate their satisfaction to their prior expectations about it (Van Ryzin, 2004, 2006). This risk might be particularly high in judgments of public service, which can be deeply rooted in individuals' general political attitudes toward government.

The first contribution of our experiments lies in the exploration of reference dependence in forming citizens' expectations from multiple perspectives and through different mechanisms. Our study provides novel experimental findings that may enhance our understanding of reference points' role in affecting predictive expectations regarding public service performance. On the one hand, social norms that include comparative performance information have been found to impact citizens' expectations, which corroborates previous work on the effect of social references on expectations (Favero & Kim, 2021). This finding bears relevance in light of the widespread adoption or use of benchmarking in the public sector (McGuire et al., 2021). Interestingly, we did not find any significant difference between descriptive and injunctive norms. As noted by one anonymous reviewer, a possible explanation for this null effect may be given by our low-intensity treatment. On the other hand, to the best of our knowledge, this is the first study to experimentally test whether and to what extent meaningless numerical anchors can influence expectations. In a debate often dominated by a focus on rational explanations, which may stem from the past emphasis on predictive expectations over other types of expectations (Hjortskov, 2020), our evidence suggests the importance of considering how expectations can be shaped by cognitive biases. In this respect, our study is inspired by Tversky and Kahneman's (1973) work on anchoring heuristics showing that judgments of unfamiliar quantities can be tilted toward irrelevant anchors as well as recent work by Bordalo et al. (2020) demonstrating that beliefs respond to irrelevant cues. Our findings enlarge a nascent stream of work in public administration that explores how performance evaluation is affected by anchoring bias across different domains. Whereas previous work in this area has mostly focused on civil servants (Bellé et al., 2017, 2018), our results extend exploration to incorporate citizens' evaluation of public

services. Overall, our evidence could prove highly valuable for policymakers and public managers, especially considering the expanding volume and diversity of performance information accessible to the public, which increases the likelihood of comparisons and potential exposure to inaccurate or irrelevant data.

A second contribution of our study involves the application of the expectancy-disconfirmation framework (Favero & Kim, 2021; Petrovsky et al., 2017; Van Ryzin, 2004, 2006, 2013; Zhang et al., 2022) to bridge various streams of research within the public administration literature, namely the scholarships on both the focus theory of normative conduct (Belle & Cantarelli, 2021; John et al., 2019) and anchoring bias (Bellé et al., 2017, 2018; Nagtegaal et al., 2020). By factoring in descriptive and injunctive social norms, our study provides a more granular understanding of how social references influence the formation of citizens' predictive expectations. As we have observed, the distinction between descriptive and injunctive norms seems to naturally parallel the one between predictive and normative expectations. We have provided with a first empirical inquiry that encompasses both sets of distinctions.

Thirdly, and relatedly, our evidence supports recent research suggesting that normative expectations seem to have a more significant impact on satisfaction than predictive expectations (Petrovsky et al., 2017; Favero & Kim, 2021, Zhang et al. 2022). Specifically, when controlling for normative expectations, we did not observe a significant effect of predictive expectations on satisfaction. In other words, when keeping fixed the number of days citizens think is right to wait, the number of days citizens predict to wait no longer affects their satisfaction.

Our results should be interpreted in light of a few limitations. Firstly, our research design does not allow a direct comparison between the impact of predictive and normative expectations. Indeed, predictive expectations resulted from our manipulations, whereas normative expectations were measured at baseline and were thus immune to any experimental manipulation. Secondly, different from the manipulation of anchors, the one of social norms does not entail a control group. Therefore, our manipulation of social reference points allows testing only for the difference between descriptive and injunctive social norms, but not for the impact of social norms with respect to a baseline. Thirdly, the information included in our experimental vignette was not comprehensive and we cannot be sure that our results would hold beyond the specific experimental units, treatments, operations, and settings (Shadish et al., 2002). For example, in Experiment 2, we cannot know how citizens would respond when considering a specific public service. Furthermore, we did not follow recent best practices as we did not pre-registered our experiments; however, ex-post, we maintained full transparency by reporting all the details of the experimental procedure. An additional limitation of our study lies in the risk of a false negative due to underpowering when testing for the difference between the low anchor and the

control group in Experiment 2. As pointed out in the results section, post-hoc power analysis revealed that for that specific comparison, we reached a power of 30%. As a final limitation, given our data collection process, we are unable to test for potential causes of differences in attrition rates across experimental arms, which seem particularly relevant in Experiment 2.

Future research shall address these limitations by varying the research design and testing our findings against alternative experimental units, operations, and settings. For example, scholars in our field might investigate the effect of meaningless numerical anchors on normative expectations, as these might be more stable and difficult to manipulate (Favero & Kim, 2021). Beyond meaningless numerical anchors and supposedly irrelevant factors, the effect of meaningful numerical anchors in shaping expectations might also be explored. In this respect, a promising idea might be testing the effect of comparative performance information in the form of numerical anchors, instead of social norms. As for social norms, it seems equally important to replicate our findings by slightly changing the experimental design in order to test the effect of descriptive and injunctive social norms with respect to a baseline. Furthermore, to gain a richer understanding of the dynamics underlying the effects we measured experimentally, qualitative work could be conducted in the form of focus groups. One first step in this direction would be to investigate the assimilation effect that has been argued to explain the direct relationship between expectations and satisfaction. Additionally, given that our experimental design does not illuminate some of the mechanisms that might be driving our descriptive findings, future work should address this limitation by employing experimental designs aimed at testing competing explanations. For example, in our experiments, social norms may influence outcomes through citizens' aversion to unequal treatment when interacting with public services, but also through the availability heuristic. From a methodological standpoint, future replications should aim for even balance across experimental arms. Finally, whereas our use of actual performance data from an actual municipality increases the ecological validity of our research design, it also restricted our capacity to investigate the effects of more extreme values in performance and disconfirmation. Scholars interested in the expectancy-disconfirmation model might design new experiments with more extreme performance values to give nuances on the potential effects of positive disconfirmation.

FUNDING INFORMATION

US Mission to Italy Alumni Small Grant Program 2020.

CONFLICT OF INTEREST STATEMENT

The authors report there are no competing interests to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available online in the Harvard Dataverse, at <https://doi.org/10.7910/DVN/IV3X2J>.

REFERENCES

- Alevy, J. E., C. E. Landry, and J. A. List. 2015. "Field Experiments on the Anchoring of Economic Valuations." *Economic Inquiry* 53(3): 1522–38.
- Allcott, H. 2011. "Social Norms and Energy Conservation." *Journal of Public Economics* 95(9–10): 1082–95.
- Asch, S. E. 1956. "Studies of Independence and Conformity: I. A Minority of One against a Unanimous Majority." *Psychological Monographs: General and Applied* 70(9): 1–70.
- Barrows, S., M. Henderson, P. E. Peterson, and M. R. West. 2016. "Relative Performance Information and Perceptions of Public Service Quality: Evidence from American School Districts." *Journal of Public Administration Research and Theory* 26(3): 571–583.
- Battaglio, R. P., Jr., P. Belardinelli, N. Bellé, and P. Cantarelli. 2019. "Behavioural Public Administration Ad Fontes: A Synthesis of Research on Bounded Rationality, Cognitive Biases, and Nudging in Public Organizations." *Public Administration Review* 79(3): 304–320.
- Behavioural Insights Team. 2013. *Applying Behavioural Insights to Charitable Giving*. London: Cabinet Office.
- Belle, N., and P. Cantarelli. 2021. "Nudging Public Employees through Descriptive Social Norms in Healthcare Organizations." *Public Administration Review* 81(4): 589–598.
- Bellé, N., P. Cantarelli, and P. Belardinelli. 2017. "Cognitive Biases in Performance Appraisal: Experimental Evidence on Anchoring and Halo Effects with Public Sector Managers and Employees." *Review of Public Personnel Administration* 37(3): 275–294.
- Bellé, N., P. Cantarelli, and P. Belardinelli. 2018. "Prospect Theory Goes Public: Experimental Evidence on Cognitive Biases in Public Policy and Management Decisions." *Public Administration Review* 78(6): 828–840.
- Bennett, M. W. 2014. "Confronting Cognitive Anchoring Effect and Blind Spot Biases in Federal Sentencing: A Modest Solution for Reforming a Fundamental Flaw." *Journal of Criminal Law and Criminology* 104: 489.
- Bokemper, S. E., M. Cucciniello, T. Rotesi, P. Pin, A. A. Malik, K. Willebrand, Elliott E. Paintsil, Saad B. Omer, Gregory A. Huber, and A. Melegaro. 2021. "Experimental Evidence that Changing Beliefs about Mask Efficacy and Social Norms Increase Mask Wearing for COVID-19 Risk Reduction: Results from the United States and Italy." *PLoS One* 16(10): e0258282.
- Bond, R., and P. B. Smith. 1996. "Culture and Conformity: A Meta-Analysis of Studies Using Asch's (1952b, 1956) Line Judgment Task." *Psychological Bulletin* 119(1): 111–137.
- Bordalo, P., N. Gennaioli, and A. Shleifer. 2020. "Memory, Attention, and Choice." *The Quarterly Journal of Economics* 135(3): 1399–1442.
- Boyne, G. A., O. James, P. John, and N. Petrovsky. 2009. "Democracy and Government Performance: Holding Incumbents Accountable in English Local Governments." *The Journal of Politics* 71(4): 1273–84.
- Cardozo, R. N. 1965. "An Experimental Study of Customer Effort, Expectation, and Satisfaction." *Journal of Marketing Research* 2(3): 244–49.
- Chen, W., B. Dong, C. W. Hsieh, M. J. Lee, N. Liu, R. M. Walker, Yao Wang, et al. 2022. A Replication of 'An Experimental Test of the Expectancy-Disconfirmation Theory of Citizen Satisfaction'. *Public Administration* 100(3): 778–791.
- Churchill, G. A., and C. Surprenant. 1982. "An Investigation into the Determinants of Consumer Satisfaction." *Journal of Marketing Research* 19: 495–504.
- Cialdini, R., and W. Schultz. 2004. *Understanding and Motivating Energy Conservation Via Social Norms*. Report Prepared for the William and Flora Hewlett Foundation.
- Cialdini, R. B., L. J. Demaine, B. J. Sagarin, D. W. Barrett, K. Rhoads, and P. L. Winter. 2006. "Managing Social Norms for Persuasive Impact." *Social Influence* 1(1): 3–15.
- Cialdini, R. B., C. A. Kallgren, and R. R. Reno. 1991. "A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior." In *Advances in Experimental Social Psychology*, Vol 24 201–234. Academic Press.
- Clow, K. E., D. L. Kurtz, J. Ozment, and B. Soo Ong. 1997. "The Antecedents of Consumer Expectations of Services: An Empirical Study across Four Industries." *Journal of Services Marketing* 11(4): 230–248.
- Coleman, S. 1996. *The Minnesota Income Tax Compliance Experiment: State Tax Results*.
- Devlin, J. F., A. L. Gwynne, and C. T. Ennew. 2002. "The Antecedents of Service Expectations." *Service Industries Journal* 22(4): 117–152.
- Englich, B., T. Mussweiler, and F. Strack. 2006. "Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on experts' Judicial Decision Making." *Personality and Social Psychology Bulletin* 32(2): 188–200.
- Favero, N., and M. Kim. 2021. "Everything Is Relative: How Citizens Form and Use Expectations in Evaluating Services." *Journal of Public Administration Research and Theory* 31(3): 561–577.
- Filtenborg, A. F., F. Gaardboe, and J. Sigsgaard-Rasmussen. 2017. "Experimental Replication: An Experimental Test of the Expectancy Disconfirmation Theory of Citizen Satisfaction." *Public Management Review* 19(9): 1235–50.
- Grimmelikhuisen, S., and G. A. Porumbescu. 2017. "Reconsidering the Expectancy Disconfirmation Model. Three Experimental Replications." *Public Management Review* 19(9): 1272–92.
- Hallsworth, M., T. Chadborn, A. Sallis, M. Sanders, D. Berry, F. Greaves, Lara Clements, and S. C. Davies. 2016. "Provision of Social Norm Feedback to High Prescribers of Antibiotics in General Practice: A Pragmatic National Randomised Controlled Trial." *The Lancet* 387(10029): 1743–52.
- Hallsworth, M., J. A. List, R. D. Metcalfe, and I. Vlaev. 2017. "The Behavioralist as Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance." *Journal of Public Economics* 148: 14–31.
- Hjortskov, M. 2019. "Citizen Expectations and Satisfaction over Time: Findings from a Large Sample Panel Survey of Public School Parents in Denmark." *The American Review of Public Administration* 49(3): 353–371.
- Hjortskov, M. 2020. "Interpreting Expectations: Normative and Predictive Expectations from the citizens' Viewpoint." *Journal of Behavioral Public Administration* 3(1).
- Hjortskov, M. 2021. "Personality Traits in Citizen Expectations towards Public Services." *Journal of Behavioral Public Administration* 4(1).
- James, O. 2009. "Evaluating the Expectations Disconfirmation and Expectations Anchoring Approaches to Citizen Satisfaction with Local Public Services." *Journal of Public Administration Research and Theory* 19(1): 107–123.
- James, O. 2011. "Managing citizens' Expectations of Public Service Performance: Evidence from Observation and Experimentation in Local Government." *Public Administration* 89(4): 1419–35.
- John, P., M. Sanders, and J. Wang. 2019. "A Panacea for Improving Citizen Behaviors? Introduction to the Symposium on the Use of Social Norms in Public Administration." *Journal of Behavioral Public Administration* 2(2).
- Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Keiser, L. R., and S. M. Miller. 2019. "The Power of Comparisons: Peer Information as a Management Tool." *Public Policy and Administration* 34(4): 494–516.
- Larkin, C., M. Sanders, I. Andresen, and F. Algate. 2018. *Testing Local Descriptive Norms and Salience of Enforcement Action: A Field Experiment to Increase Tax Collection Available at SSRN 3167575*.
- Linkenbach, J., and Perkins, H. W. 2003. "MOST of us Are Tobacco Free: An Eight-Month Social Norms Campaign Reducing Youth Initiation of Smoking in Montana." In *The Social Norms Approach to*

- Preventing School and College Age Substance Abuse: A Handbook for Educators, Counselors, and Clinicians*, edited by H. W. Perkins, 224–234. San Francisco, CA: Jossey-Bass.
- Lovell, M. C. 1986. “Tests of the Rational Expectations Hypothesis.” *The American Economic Review* 76(1): 110–124.
- McGuire, L., E. Prior Jonson, S. Perryman, and T. McKeown. 2021. “Benchmarking Government: Report on Government Services (RoGS)—25 Years on.” *Australian Journal of Public Administration* 80(4): 987–1001.
- Morgeson, F. V. 2012. “Expectations, Disconfirmation, and Citizen Satisfaction with the US Federal Government: Testing and Expanding the Model.” *Journal of Public Administration Research and Theory* 23(2): 289–305.
- Muth, J. F. 1961. “Rational Expectations and the Theory of Price Movements.” *Econometrica: Journal of the Econometric Society* 29: 315–335.
- Nagtegaal, R., L. Tummers, M. Noordegraaf, and V. Bekkers. 2020. “Designing to Debias: Measuring and Reducing Public Managers’ Anchoring Bias.” *Public Administration Review* 80(4): 565–576.
- Oliver, R. L. 1980. “A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions.” *Journal of Marketing Research* 17(4): 460–69.
- Oliver, R. L. 2014. *Satisfaction: A Behavioral Perspective on the Consumer*. New York: Routledge.
- Olsen, A. L. 2017. “Compared to What? How Social and Historical Reference Points Affect citizens’ Performance Evaluations.” *Journal of Public Administration Research and Theory* 27(4): 562–580.
- Orr, D., and C. Guthrie. 2005. “Anchoring, Information, Expertise, and Negotiation: New Insights from Meta-Analysis.” *Ohio State Journal on Dispute Resolution* 21: 597.
- Pandey, S. K., and J. Marlowe. 2015. “Assessing Survey-Based Measurement of Personnel Red Tape with Anchoring Vignettes.” *Review of Public Personnel Administration* 35(3): 215–237.
- Petrovsky, N., J. Y. Mok, and F. León-Cázares. 2017. “Citizen Expectations and Satisfaction in a Young Democracy: A Test of the Expectancy-Disconfirmation Model.” *Public Administration Review* 77(3): 395–407.
- PMR. 2023. “Call for Papers for a Special Issue: The Next Generation of Citizen Satisfaction Research: Theoretical and Methodological Advancements.” *Public Management Review*.
- Poister, T. H., and J. C. Thomas. 2011. “The Effect of Expectations and Expectancy Confirmation/Disconfirmation on Motorists’ Satisfaction with State Highways.” *Journal of Public Administration Research and Theory* 21(4): 601–617.
- Roch, C. H., and T. H. Poister. 2006. “Citizens, Accountability, and Service Satisfaction: The Influence of Expectations.” *Urban Affairs Review* 41(3): 292–308.
- Santos, J., and J. Boote. 2003. “A Theoretical Exploration and Model of Consumer Expectations, Post-Purchase Affective States and Affective Behaviour.” *Journal of Consumer Behaviour: An International Research Review* 3(2): 142–156.
- Schultz, P. W. 1999. “Changing Behavior with Normative Feedback Interventions: A Field Experiment on Curbside Recycling.” *Basic and Applied Social Psychology* 21(1): 25–36.
- Schultz, P. W., J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius. 2007. “The Constructive, Destructive, and Reconstructive Power of Social Norms.” *Psychological Science* 18(5): 429–434.
- Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Mifflin and Company: Houghton.
- Shaffer, L. S. 1983. “Toward Pepitone’s Vision of a Normative Social Psychology: What Is a Social Norm?” *The Journal of Mind and Behavior* 4(2): 275–293.
- Simmons, J. P., R. A. LeBoeuf, and L. D. Nelson. 2010. “The Effect of Accuracy Motivation on Anchoring and Adjustment: Do People Adjust from Provided Anchors?” *Journal of Personality and Social Psychology* 99(6): 917–932.
- Teas, R. K. 1993. “Consumer Expectations and the Measurement of Perceived Service Quality.” *Journal of Professional Services Marketing* 8(2): 33–54.
- Thaler, R. H., and C. R. Sunstein. 2009. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New York: Penguin.
- Thaler, R. H., and C. R. Sunstein. 2021. *Nudge: The Final Edition*. New Haven: Yale University Press.
- Thorsteinson, T. J., J. Breier, A. Atwell, C. Hamilton, and M. Privette. 2008. “Anchoring Effects on Performance Judgments.” *Organizational Behavior and Human Decision Processes* 107(1): 29–40.
- Tversky, A., and D. Kahneman. 1973. “Availability: A Heuristic for Judging Frequency and Probability.” *Cognitive Psychology* 5(2): 207–232.
- Tversky, A., and D. Kahneman. 1974. “Judgment under Uncertainty: Heuristics and Biases: Biases in Judgments Reveal some Heuristics of Thinking under Uncertainty.” *Science* 185(4157): 1124–31.
- Van Ryzin, G. G. 2004. “Expectations, Performance, and Citizen Satisfaction with Urban Services.” *Journal of Policy Analysis and Management* 23(3): 433–448.
- Van Ryzin, G. G. 2006. “Testing the Expectancy Disconfirmation Model of Citizen Satisfaction with Local Government.” *Journal of Public Administration Research and Theory* 16(4): 599–611.
- Van Ryzin, G. G. 2013. “An Experimental Test of the Expectancy-Disconfirmation Theory of Citizen Satisfaction.” *Journal of Policy Analysis and Management* 32(3): 597–614.
- Webeck, S., and S. Nicholson-Crotty. 2020. “How Historical and Social Comparisons Influence Interpretations of Performance Information.” *International Public Management Journal* 23(6): 798–821.
- Zeithaml, V. A., L. L. Berry, and A. Parasuraman. 1993. “The Nature and Determinants of Customer Expectations of Service.” *Journal of the Academy of Marketing Science* 21: 1–12.
- Zhang, J., W. Chen, N. Petrovsky, and R. M. Walker. 2022. “The Expectancy-Disconfirmation Model and Citizen Satisfaction with Public Services: A Meta-Analysis and an Agenda for Best Practice.” *Public Administration Review* 82: 147–159.

AUTHOR BIOGRAPHIES

Nicola Bellé is an Associate Professor at Scuola Superiore Sant’Anna (Management and Healthcare Laboratory, Institute of Management). His research focuses on behavioral and experimental public administration. nicola.belle@santannapisa.it

Paolo Belardinelli is an Assistant Professor in the O’Neill School of Public and Environmental Affairs at Indiana University. His research focuses on behavioral public policy and administration. pbelard@iu.edu

Maria Cucciniello is an Associate Professor in the Department of Social and Political Sciences at Bocconi University. Her research focuses on transparency and service design oriented to value creation. maria.cucciniello@unibocconi.it

Greta Nasi is an Associate Professor in the Department of Social and Political Sciences at Bocconi University. Her research focuses on public services and administration. greta.nasi@unibocconi.it

How to cite this article: Bellé, Nicola, Paolo Belardinelli, Maria Cucciniello, and Greta Nasi. 2023. “Experimental Evidence on the Determinants of Citizens’ Expectations toward Public Services.” *Public Administration Review* 1–16. <https://doi.org/10.1111/puar.13775>

APPENDIX A

A.1 | Experimental vignettes-English translation

A.1.1. | Experiments 1 and 3

Step 1: Imagine having to use the registry service of the municipality in which you reside to request an electronic identity card. According to perfectly reliable estimates, waiting times for the appointment for the electronic identity card will be *much shorter (much longer)* than in the majority of other municipalities (what the majority of citizens think is right). Based on this information, how many days do you expect to wait for the first appointment for your electronic ID card? (1–100 slider).

Step 2: Now imagine that the waiting times for the first appointment for your electronic ID card are 27 (55) days. On a scale of 0 to 10, how satisfied are you?

A.1.2. | Experiment 2

Avatar is a typical municipality where citizens, in order to use public services, submit requests to the public administration and wait for these to be processed.

10 s of: a blank page/a page showing the number 50/a page showing the number 500.

How long do you expect that Avatar citizens wait, on average, for applications to the Avatar administration to be processed? Please enter the number of days in digits.

TABLE A1 Sample descriptives, by experimental group-Experiment 1.

	Italian population	Overall sample	Shorter ET	Longer ET	Injunctive SN	Descriptive SN
<i>N</i>	59 mln	2438	1241	1197	1209	1229
Female	0.51	0.52	0.51	0.52	0.53	0.50
Age bracket						
18–24	0.08	0.17	0.16	0.17	0.18	0.16
25–34	0.12	0.21	0.20	0.22	0.21	0.21
35–49	0.23	0.40	0.40	0.39	0.39	0.40
50–64	0.28	0.20	0.21	0.19	0.19	0.21
Over 65	0.28	0.03	0.02	0.03	0.03	0.03
Post-secondary education	0.21	0.39	0.38	0.39	0.38	0.39
Geographic area						
North	0.47	0.52	0.54	0.51	0.52	0.53
Center	0.20	0.17	0.17	0.18	0.17	0.17
South and Islands	0.34	0.30	0.29	0.31	0.31	0.30

TABLE A2 Sample descriptives, by experimental group-Experiments 2 and 3.

	Italian population	Overall sample	Shorter ET	Longer ET	Injunctive SN	Descriptive SN	No anchor	Low anchor	High anchor
<i>N</i>	59 mln	538	274	264	272	266	217	173	148
Female	0.51	0.47	0.49	0.45	0.47	0.47	0.48	0.43	0.51
Age bracket									
18–24	0.08	0.12	0.12	0.13	0.13	0.12	0.12	0.10	0.14
25–34	0.12	0.12	0.12	0.13	0.13	0.11	0.12	0.12	0.12
35–49	0.23	0.24	0.26	0.22	0.25	0.23	0.28	0.24	0.18
50–64	0.28	0.23	0.23	0.24	0.22	0.24	0.21	0.27	0.22
Over 65	0.28	0.28	0.28	0.28	0.27	0.30	0.26	0.27	0.34
Post-secondary education	0.21	0.34	0.35	0.33	0.34	0.34	0.34	0.28	0.41
Geographic area									
North	0.47	0.44	0.45	0.44	0.43	0.45	0.44	0.43	0.45
Center	0.20	0.20	0.22	0.18	0.21	0.19	0.20	0.23	0.16
South and Islands	0.34	0.36	0.33	0.39	0.36	0.36	0.35	0.34	0.39

TABLE A3 Treatment effects on expected waiting days-ANOVA.

	Partial SS	df	MS	F	p
Model	10998.857	3	3666.29	6	.0005
Estimated waiting times: Longer versus shorter	9228.1256	1	9228.13	15.11	.0001
Social norm: Descriptive versus Injunctive	1386.8521	1	1386.85	2.27	.132
Longer estimated waiting time × Descriptive social norm	0.49596363	1	0.49596	0	.9773
Residual	1486600.8	2434	610.765		
Total	1497599.7	2437	614.526		

Note: Root MSE 24.71; $N = 2438$; R -squared = 0.0073; Adj R -squared = 0.0061.

TABLE A4 ANOVA.

	Partial SS	df	MS	F	p
Model	147115.41	2	73557.71	14.19	.0000
Anchor	9228.1256	2	73557.71	14.19	.0000
Residual	2773832.6	535	5184.7339		
Total	2920948	537	5439.3818		

Note: Root MSE 72.0051; $N = 538$; R -squared = 0.0504; Adj R -squared = 0.0468.

TABLE A5 Treatment effects on expected waiting days-ANOVA.

	Partial SS	df	MS	F	p
Model	3672.6778	3	1224.23	3.32	0.0197
Estimated waiting times: Longer versus shorter	3161.7599	1	3161.76	8.57	0.0036
Social norm: Descriptive versus Injunctive	107.55312	1	107.553	0.29	0.5895
Longer estimated waiting time × Descriptive social norm	362.39892	1	362.399	0.98	0.3221
Residual	197005.7	534	368.925		
Total	200678.4	537	373.703		

Note: Root MSE 19.2074; $N = 538$; R -squared = 0.0183; Adj R -squared = 0.0128.

TABLE A6 Treatment effects on satisfaction/Additional tests of expectancy-isconfirmation model.

	Model 1 (std coefficients)		Model 2 (std coefficients)		Model 3 (std coefficients)		Model 4 (std coefficients)	
	$N = 538$		$N = 538$		$N = 538$		$N = 538$	
Better performance (27 days)					0.61*** (0.181)	0.12	0.66*** (0.174)	0.13
Predictive expectations	0.04*** (0.007)	0.28	0.04*** (0.007)	0.28	0.07*** (0.005)	0.53	0.04*** (0.007)	0.28
Normative expectations	0.04*** (0.007)	0.33	0.02** (0.009)	0.16			0.04*** (0.007)	0.34
(Normative) Disconfirmation ^a			0.02*** (0.006)	0.22				
Cons	1.07*** (0.117)		2.01*** (0.273)		0.84*** (0.154)		0.71*** (0.149)	
R -squared	0.32		0.33		0.28		0.34	

Note: Standard errors in parentheses.

*** p -value < .01; ** p -value < .05; * p -value < .1.

^aWe measure '(Normative) Disconfirmation' as the difference between normative expectations and the randomly assigned actual performance (Disconfirmation = Number of days is right to wait – Number of actual waiting days).

A.2 | Experimental vignettes-Original version in Italian

A.2.1. | Experiments 1 and 3

Step 1: Immagina di dover usufruire del servizio anagrafe del tuo Comune per il rilascio della carta d'identità elettronica. Stime perfettamente attendibili indicano che i tempi di attesa per il primo appuntamento per la carta d'identità elettronica saranno *molto più brevi [molto più lunghi]* che *nella maggioranza dei Comuni [la maggioranza dei cittadini ritiene giusti]*. Sulla base di queste informazioni, quanti giorni prevedi di dover attendere per il primo appuntamento per la tua carta d'identità elettronica? (1–100 slider).

Step 2: Immagina ora che i tempi di attesa necessari per il primo appuntamento per la tua carta d'identità

elettronica siano pari a 27 [55] giorni. Su una scala da 0 a 10, quanto sei soddisfatto?

A.2.2. | Experiment 2

Avatar è un tipico Comune nel quale i cittadini, per usufruire dei servizi comunali, inoltrano delle pratiche e attendono che queste siano evase dall'amministrazione.

10 s of: a blank page/a page showing the number 50/a page showing the number 500.

Quanti giorni ti aspetti che i cittadini di Avatar attendano, in media, perché le pratiche che inoltrano al Comune siano evase? Digita il numero di giorni in cifre.

Additional analyses.

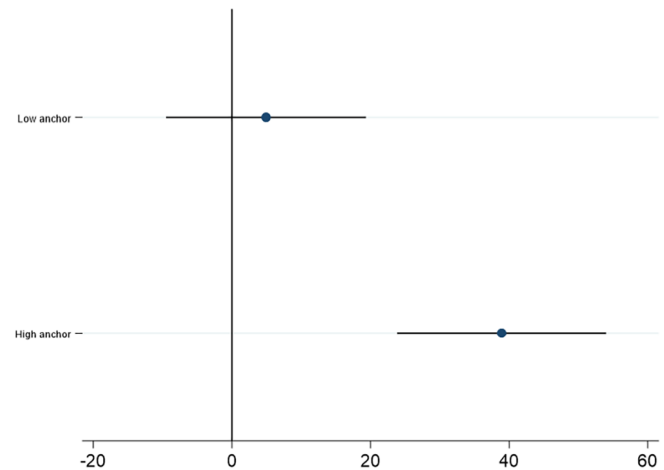
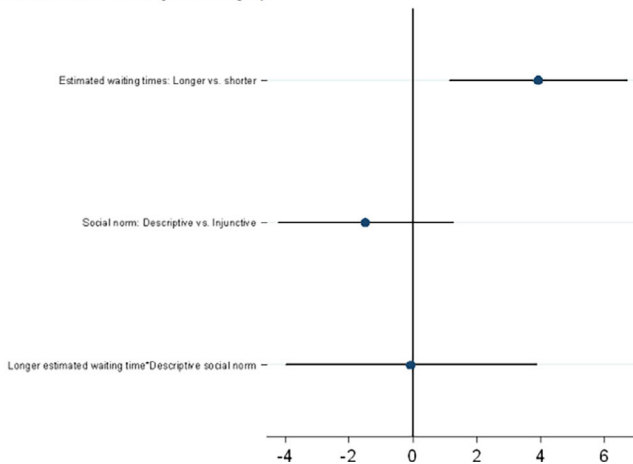


FIGURE A2 Treatment effects on expected waiting days-Experiment 2. 95% confidence intervals.

Panel A. Treatment effects on expected waiting days



Panel B. Treatment effects on satisfaction / Test of expectancy-disconfirmation model

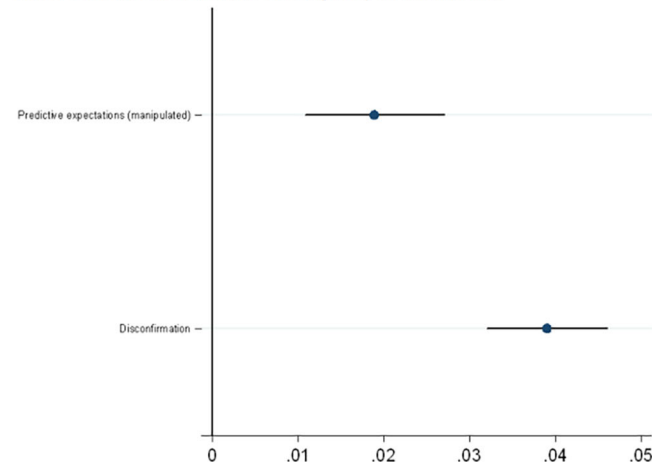
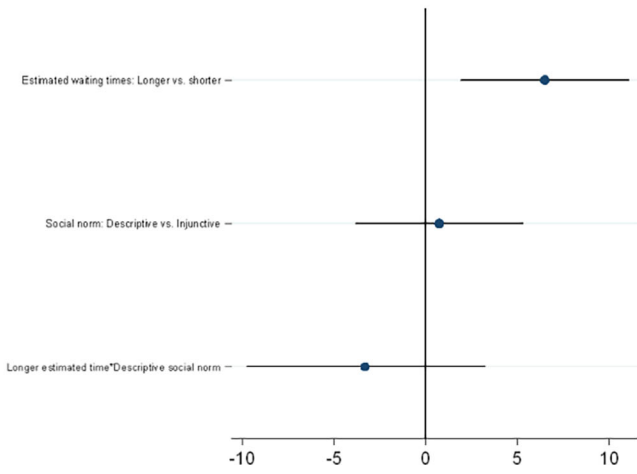


FIGURE A1 Treatment effects from Experiment 1. 95% confidence intervals.

Panel A. Treatment effects on expected waiting days



Panel B. Treatment effects on satisfaction / Test of expectancy-disconfirmation model

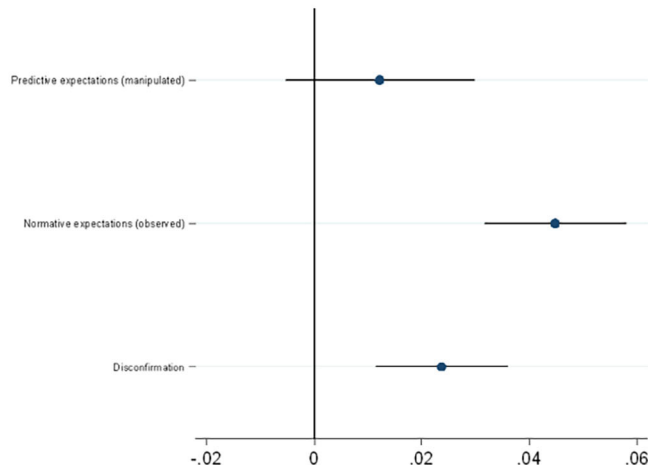


FIGURE A3 Treatment effects from Experiment 3. 95% confidence intervals.