

## **Designing Survey Questionnaires**

Should Surveys Ask about Public Servants' Perceptions of Their Organization or Their Individual Experience?

Mikkelsen, Kim Sass; Parker, Camille Mercedes

*Published in:*  
The Government Analytics Handbook

*Publication date:*  
2023

*Document Version*  
Publisher's PDF, also known as Version of record

*Citation for published version (APA):*  
Mikkelsen, K. S., & Parker, C. M. (2023). Designing Survey Questionnaires: Should Surveys Ask about Public Servants' Perceptions of Their Organization or Their Individual Experience? In D. Rogger, & C. Schuster (Eds.), *The Government Analytics Handbook: Leveraging Data to Strengthen Public Administration* (1 ed., pp. 497-523). World Bank Publications. <https://thedocs.worldbank.org/en/doc/3719620d9e14b651c0919fff547660f2-0050042023/original/GAH-CHAPTER-23.pdf>

### **General rights**

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

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

### **Take down policy**

If you believe that this document breaches copyright please contact [rucforsk@kb.dk](mailto:rucforsk@kb.dk) providing details, and we will remove access to the work immediately and investigate your claim.

## CHAPTER 23

# Designing Survey Questionnaires

## Should Surveys Ask about Public Servants' Perceptions of Their Organization or Their Individual Experience?

*Kim Sass Mikkelsen and Camille Mercedes Parker*

### SUMMARY

Civil service surveys are often interested in organizational aggregates and comparisons across organizations. Therefore, the choice of question referent is important in questionnaire design. Should survey questions refer to individual employees or to employees' assessments of their organizations? This chapter provides tools for thinking through this choice. Moreover, experimental evidence from representative public service surveys in Romania and Guatemala shows that the choice of referent matters to how employees respond. Finally, the chapter provides evidence that organizational referents can help reduce socially desirable responding, particularly for highly sensitive questions, and that referent effects may be larger for attitudes and behaviors that are uncommon, but that the size of referent effects beyond this is difficult to predict.

### ANALYTICS IN PRACTICE

- Many civil service surveys are centrally interested in organizational aggregates. Therefore, the choice of question referent is important in questionnaire design. Should questions refer to individual employees or to employees' assessments of their organizations?
- Inside organizations, perceptions of management practices are often only weakly correlated across respondents, suggesting that they are not organizational *constructs*. Organizational referents can—but often do not—better enable survey questions to reflect organizational constructs.

---

Kim Sass Mikkelsen is an associate professor of politics and public administration at Roskilde University. Camille Mercedes Parker is an economist at the United States Agency for International Development.

- Experimental evidence from representative public service surveys in Romania and Guatemala shows that the choice of referent matters to how employees respond.
- We provide evidence that organizational referents may help reduce socially desirable responding, particularly for highly sensitive questions.
- We examine, but uncover little systematic evidence for, a set of other factors that could conceivably influence question-referent effects. We conclude that organizational referents may be less useful in situations where attitudes and behaviors are uncommon because respondents may not have the needed information to answer them. Beyond this, however, the size of referent effects is difficult to predict.

## INTRODUCTION

Many civil service surveys are centrally interested in organizational aggregates. Which surveyed organization has the highest level of job satisfaction among its employees? Which organizations need additional ethics training to keep up with the ethical awareness of employees in other organizations? Questions such as these are core both to internal government benchmarking and, since aggregates attached to recognizable labels (like organization names) are simple to interpret, to government communication of data from civil service surveys.

The focus on organizational aggregates has an intuitive implication for how questions should be asked in civil service surveys: ask civil servants to evaluate their organizations. Indeed, practitioners and academics alike routinely ask civil servants for such evaluations. For example, the United States Office of Personnel Management (OPM) Federal Employee Viewpoint Survey (FEVS) asks its respondents to evaluate the extent to which “employees are protected from health and safety hazards on the job” as a measure of workplace safety (OPM 2018). This practice is sensible. If the target of evaluations is the organization, it seems reasonable to align the *level of measurement*—the level in reference to which respondents are asked to provide answers—with the level at which claims are made (Klein, Dansereau, and Hall 1994). *Referents*, the entities to which a survey question refers, can be sensibly chosen to reflect the entities researchers wish to learn about. The question cited above is an example of the use of *organizational referents* in civil service surveys.

It is not always clear, however, that the organization is the most appropriate or most useful level of measurement. While recognizable labels make organizational comparisons simple and appealing, using organizational referents implies one of two claims: that the question measures the respondent’s *perceptions* of his or her organizational surroundings, or that the subject of the question is an *organization-level phenomenon* (Klein, Dansereau, and Hall 1994; Klein and Kozlowski 2000). In the first instance, *top-down* claims can be made about how respondents react to their perceptions of organizational characteristics, management practices, leadership, culture, and so on. In the second instance, *bottom-up* claims can be made about organization-level phenomena principally detached from any individual public servant’s experiences and beliefs. Both of these claims may be true, but they are infrequently stated explicitly.

Levels of measurement have been subject to contention in leadership research (for example, Schriesheim, Wu, and Scandura 2009), organizational research (for example, Baltes, Zhdanova, and Parker 2009; Chan 1998; Klein and Kozlowski 2000), and survey methods research (for example, Blair, Menon, and Bickart 2004). However, the issue is rarely discussed in the inherently multilevel field focused on civil service surveys. Is the common practice of asking civil service survey questions at the organizational level sensible? Or is the use of *individual referents*, asking respondents to provide information about themselves, a more appropriate strategy? And does the choice matter for survey results?

What is at issue is not whether the level of analysis should match the level at which claims and comparisons are made. There is already good evidence that these levels should match and that the consequence of

mismatches is potentially biased results (for example, Gingerich 2013). Instead, we examine the advantages and disadvantages of using individual versus organizational referents in civil surveys. Should the level of *measurement* match the level at which claims and comparisons are made as well? In which situations should the level of measurement be the individual respondent, and in which should it be individuals' assessments of their employing organizations?

Intuitively, the answer parallels the match between claims and levels of analysis. If one is interested in individual public servants, individual referents should be used. If, by contrast, one is interested in organizations, organizational referents should be used. However, this answer is too simple. It underestimates the complexity of the consequences of choosing question referents. Our chapter describes some of this complexity and provides guidelines for understanding what is at stake in choosing question referents and when to choose which referent.

We are—as far as we know—the first to assess the issue of referent choice in civil service survey design. Yet the public organizational setting likely matters. Organizational referents require information from respondents that civil servants may not possess to the same degree private sector employees do, for instance. Public organizations are frequently very large in terms of both personnel and budget and are often hierarchically organized into relatively segmented and informationally insular parts (for example, Eggers 2007). This can make organizational-referent questions difficult for a public official in one part of an organization to answer due to a simple lack of knowledge about other parts of that same organization (cf. Homburg et al. 2012).

For organizations like ministries, this problem may even grow with managerial reforms that further segment and fragment the ministerial hierarchy into deliberately insular agencies (Dunleavy et al. 2006). In a sense, organizational referents in civil service surveys may have to grapple with issues similar to those that whole-of-government approaches to public sector organizations were intended to solve: information and knowledge can have a hard time traversing the organizations that respondents are asked to evaluate (for example, Christensen and Lægreid 2007).

Our advice to civil service survey designers is not to abandon one question referent in favor of another. Instead, we provide a set of important considerations that designers can use when choosing referents. In particular, designers should consider:

- Whether what they are measuring is, conceptually, an organizational phenomenon. Does it make sense to think of all respondents within an organization rating the same entity when responding? If designers are not measuring an organizational phenomenon, organizational referents are less attractive.
- How sensitive their measures are. Respondents tend to respond as they believe is socially desirable when questions are sensitive, and this effect is more pronounced for questions about them as individuals. If questions are very sensitive, organizational referents may be more attractive.
- How easily accessible to respondents the information required for the measure is. Respondents often have better access to their own experiences, beliefs, and attitudes than those of their colleagues. If questions require information that is not readily available to respondents, organizational referents are less attractive.

These conclusions are based partly on a conceptual discussion and partly on empirical evidence. Empirically, we use experiments embedded in two civil service surveys. We embedded experiments in a survey of more than 6,000 civil servants in Romania's central government, randomly assigning respondents to answer questions about human resource management practices—specifically, recruitment, promotion, dismissal, and turnover intent—using individual or organizational referents. We embedded a similar experiment in a survey of more than 3,000 civil servants in Guatemala's central government.

The basic thrust of the experiment is that, if referent choice matters, otherwise similar questions using different referents will result in different average responses. If referents do not matter, the average employee evaluation of the organization will correspond to the average of the employees' evaluation of themselves. Thus, the experiment can provide evidence that referents matter, the core interest of this chapter. The drawback is that the sources of divergences are harder to determine. We do conduct a series of tests attempting

to determine sources, but the question we can answer most clearly is whether referents matter. Despite its simplicity, a strong answer to this question is useful to survey designers, many of whom do not seem to know whether referents matter or how they matter to the responses they get.

Beyond the questions it can answer, this experimental approach is valuable for the strength of our conclusions. And it sets our study apart from previous examinations of the use of referents in organizational surveys (for example, Baltes, Zhdanova, and Parker 2009; Klein et al. 2001). Prior examinations of referent issues have asked the same respondents to provide information both about themselves (using individual referents) *and* about their organizations (using organizational referents). This is needed, of course, to show that each of the two referents contributes separate information (Klein et al. 2001). However, it creates the risk that respondents anchor their responses to one set of questions to their answers to the other set of questions in order to appear consistent, or that they respond to both sets of questions relative to one another, either to maintain that they are “above average” on relevant metrics (Guenther and Alicke 2010) or because they form their answers relative to comparisons with significant colleagues (Baltes, Zhdanova, and Parker 2009). Thus, responses to questions with individual referents can affect subsequent responses to question sets with organizational referents and vice versa. Our experimental design avoids this issue, permitting a causal assessment of the relative differences between responses stemming from the two referents.<sup>1</sup>

We proceed in four steps. In section 2, we discuss what difference organizational as opposed to individual referents might make theoretically. We focus particularly on concept levels, socially desirable responding, and information availability. In section 3, we describe our survey experiments. Section 4 presents our results. Section 5 contains our discussion of these results for the design of civil service surveys.

## WHAT IS AT STAKE?

In this section, we provide a more detailed account of the already-noted reasons why the choice of question referent might matter. This takes us into the psychology of survey response and questions about levels of theory and measurement from organizational studies. But the point is not the theory. Rather, we aim to provide readers with a rough and simple understanding of the stakes in choosing between individual and organizational referents. Table 23.1 provides an overview of the arguments we discuss. These fall along three main lines: the match between the measure and the target entity of interest, socially desirable responding, and the informational requirements placed on respondents.

### Do Analyses Concern Individuals or Organizations?

At base, the choice of referent should reflect the interest of subsequent analyses. If the interest is in measuring, comparing, or benchmarking organizations, organizational referents appear to be the obvious choice because they create a clear match between the entities in subsequent analyses (the target) and the measure. However, this is not as obvious as it would at first appear. Table 23.2 provides an overview of the discussion.

Table 23.2 distinguishes between referents, the entities referred to in survey questions, and target entities, the entities the survey aims to learn something about. The intuition is that referents should be chosen to match the downward diagonal of the table. Inquiries with an individual focus should ask individual-referent questions, while organizational inquiries should use organizational referents.

The first half of this intuition holds. Inquiries with an individual focus should likely ask about individuals. But the second half of the intuition is more complicated. There are three ways of thinking about settings where questions either use organizational referents or aim to learn about organizations: the *top-down* perspective, which asks about organizations to learn about individual employees, the *bottom-up* perspective, which asks about organizations to learn about organizations (when possible) (Klein and Kozlowski 2000),

**TABLE 23.1 Advantages and Disadvantages of Organizational and Individual Referents When Used for Calculating and Analyzing Organizational Aggregates**

Type of cost or benefit	Organizational referents	Individual referents
Conceptual	+ Match between target and measure – Disagreement	+ No agreement requirement – Possible mismatch between target and measure
Measurement	+ Decreased social-desirability bias – Informational requirements	+ Fewer informational requirements – Social-desirability bias

Source: Original table for this publication.

Note: This table shows a summary of the discussion in the three following subsections. Columns represent question referents (organizational vs. individual). Rows are divided into conceptual concerns (discussed in the first subsection) and measurement concerns (discussed in the second and third subsections). Plus signs indicate competitive advantages relative to the referent in the other column; minus signs indicate competitive disadvantages. Advantages and disadvantages are relative to data used for calculating organizational aggregates. Some points are not relevant in other contexts (for example, “match between target and measure” is not a competitive advantage for organizational-referent questions if an individual’s beliefs are the target, as in the top-down perspective).

**TABLE 23.2 Question Referents and Target Entities**

		Target entity	
		Organization	Individual employee
Question referent	Organizational	Bottom-up	Top-down
	Individual	Summary bottom-up	Individual focus

Source: Original table for this publication.

and finally, what we call the *summary bottom-up* perspective, which asks about individuals to learn about organizations through data summaries.

The *top-down* perspective interprets organizational-referent questions as asking about respondents’ perceptions of their working environment. Even questions that appear to be intrinsically at the organizational level may be best thought of at the individual level in terms of definitions, causal efficacy, or both. For instance, Parker et al. (2003, 390) define the *psychological climate* in organizations—a term that, intuitively, has a clear organizational focus, though it does not have this connotation in the relevant literature—as “an individual’s psychologically meaningful representations of proximal organizational structures, processes, and events.” Such representations—including perceptions of management practices and attributions related to those perceptions—are often proposed as causally efficacious for important employee outcomes (for example, Nishii, Lepak, and Schneider 2008). They are related to organizational practices, but they are not themselves organization-level phenomena. Rather, it is employee perceptions or experiences that matter for outcomes. From this perspective, organizational-referent questions are not asking respondents to rate the same entity—indeed, they are, in a sense, not organizational at all. Instead, they are asking about individuals’ representations, beliefs, or experiences. From this perspective, answers to the FEVS question about whether “employees are protected from health and safety hazards on the job” can be interpreted as reflecting individual respondents’ beliefs about health and safety in their workplaces—which can be relevant to understanding their commitment to their workplaces, their job satisfaction, or their turnover intent—but not, strictly, as offering descriptions of their workplaces as they are.

The *bottom-up* perspective is more complicated. It involves interpreting respondents’ evaluations as reflecting genuine organizational constructs—that is, features of the organization—over and above the perspective of the individual respondent. It is not perceptions but features of the organization that are the target of organizational-referent questions, from this perspective. Respondents within an organization are all seen as rating the same entity with the same characteristics.<sup>2</sup> The bottom-up perspective on organizational referents assumes that the characteristic of concern in a question is a characteristic of the organization,



not of the respondent. From this perspective, answers to the question about whether “employees are protected from health and safety hazards on the job” are ratings of the organization; they ask the responding employee to evaluate the organization (principally) as a whole. Consequently, since respondents within an organization are rating the same entity, the bottom-up perspective assumes a substantial level of agreement among respondents in the same organization.

Based on the assumptions behind the bottom-up perspective, it seems reasonable to believe that using organizational referents furthers agreement on responses within organizations because individual respondents are essentially instructed to disregard their personal experiences and report using a *referent shift*. From this perspective, there may be reason to prefer organizational referents because they may further the agreement necessary for the desired bottom-up interpretation of organizational aggregates as reliable descriptions of the organization as one entity evaluated by multiple raters.

But what if respondents within organizations do not agree? The answer can be stated, likely too succinctly: then the measures do not appropriately measure an organization-level characteristic but a construct at a lower level (such as an employee perception). This brings us to the *summary bottom-up* perspective, which construes descriptions of organizations using survey data as summaries of individual perspectives. Employee responses to organizational referents can be thought of as such summaries, but they do not have the advantage of capturing the organization above individual perceptions and experiences. This is because the perspective does not consider employees as rating the organization but as providing their own views.

Uneven implementation within organizations is often proposed as a vehicle for intraorganizational differentiation in civil service management practices when these are measured using organizational referents (Bezes and Jeannot 2018; Meyer-Sahling and Mikkelsen 2016, 2020). From this argument, questions with organizational referents do not necessarily result in organization-level assessments by respondents but rather elicit the experience of respondents in their immediate working environments. The disadvantage is uncertainty about the width of the assessments provided by individual respondents if these are not at the organizational level to which survey items refer. If organizational referents do not prompt consensus on ratings of the same entity, it is not clear what level the questions measure. Instead of capturing their organizational target, organizational aggregates are reduced to *summaries* of features of lower levels, be these sections, teams, or individuals.

Indeed, when organizational aggregates of responses are seen as summaries of individual perspectives, organizations are arguably better described using individual-referent questions: the width of the assessment is determined by the question, and the result is still a useful organizational summary. The cost of this view is that organizational characteristics are redefined to nothing more than aggregates of individual answers. Organizational workplace safety, for example, becomes the proportion of employees who think their work is safe.

In sum, if civil service survey designers are primarily interested in organizational aggregates, should they ask questions with organizational referents to ensure correspondence between levels of measurement and levels of theory? It depends. If respondents’ within-organization responses are strongly correlated, individual and organizational referents are both useful measures of organizational characteristics. While they entail different perspectives on interpreting answers, and the bottom-up perspective has a more intuitive appeal, both kinds of referent can be used.

However, if responses do not strongly correlate within organizations, this indicates that the use of individual referents is preferable on a conceptual basis. The bottom-up perspective, in this situation, does not lend as much analytical leverage as the summary bottom-up perspective because responses do not reflect an organization-level construct; instead, organizational aggregates are more readily understood as summaries of employee information.

In sum, if employees’ beliefs and perceptions are of central interest—as in the right column of table 23.2—the choice of referent is conceptual, not statistical. In that case, organizational referents should be used if respondents’ beliefs about the organization are of central interest, and individual questions should be used if respondents’ own experiences and behaviors are of interest. However, if the organization is the target, the preferable choice of referent is, in part, statistical because organizational-referent

questions impose the requirement of *interrater agreement* among employees of the same organization, while individual-referent questions do not. Even such statistically based choices have conceptual consequences, however, since the bottom-up and summary bottom-up perspectives use different ideas about the composition of individual responses and hence capture somewhat different ideas about what organizational aggregates are (Chan 1998).

### Are Questions Sensitive?

It is often less embarrassing and feels less threatening to respond to a question in a socially undesirable way if the question is not about oneself. “Do you ever steal stationery from work?” is a much more sensitive question on its face than “Do colleagues in your organization ever steal stationery from work?” Consequently, many researchers utilize organizational referents not on conceptual grounds but to limit socially desirable responses. Organizational referents are used to make sensitive questions less sensitive to respondents, on the assumption that they will provide more truthful answers and avoid social-desirability bias (SDB) due to question sensitivity (for example, Graaf, Huberts, and Strüwer 2018; Meyer-Sahling and Mikkelsen 2016).

This assumption is plausible and has been indirectly tested in other fields under labels such as “proxy questioning” (Blair, Menon, and Bickart 2004) and “structured projective questioning” (Fisher 1993). For instance, in marketing, Fisher (1993) studies whether questions that ask for the opinion of others rather than the respondent’s own opinion can reduce SDB. Fisher’s finding accords with the assumptions made in analyses of civil service survey data: indirect questions reduce SDB on questions subject to social influence. Thematically closer to our purpose, Bardasi et al. (2011) find that reported male labor market participation rates dropped substantially when others provided proxy answers, rather than the men themselves. Like these approaches, the use of organizational referents is sometimes interpreted as an indirect question technique because respondents provide information about others, not about themselves.

Questions engender SDB through several channels. Questions can be intrusive, threatening, or socially undesirable (Tourangeau and Yan 2007). Intrusive questions can be seen as offensive, nosy, or taboo. Threatening questions make respondents worry about the disclosure of their responses and the negative consequences that may ensue. Finally, socially undesirable questions are questions for which certain answers violate social norms.

Disclosure threats and socially undesirable answers are particularly relevant to our discussion. In organizational settings, the disclosure of attitudes and behaviors to which colleagues, management, political superiors, the media, or the public will react negatively is a real concern. This is true of questions for which admitting to behaviors can have negative career consequences—such as admitting to kickbacks (Meyer-Sahling and Mikkelsen 2016). And it is true of questions for which agreeing or disagreeing can be seen as negative by colleagues or management and have negative consequences in terms of careers or ostracization at work. *Sensitive* questions—for example, questions about corruption or absenteeism—thus engender one form of SDB, but not the only form. Socially desirable responding can also occur for questions in which anything but a strong endorsement of the question’s content can be seen as undesirable—such as questions about helping colleagues or working hard.

If SDB were all about threats of disclosure, however, anonymity safeguards for individual responses should help the problem. Unfortunately, SDB persists—albeit to varying degrees—even when anonymity is guaranteed (Kreuter, Presser, and Tourangeau 2008).<sup>3</sup> This is why many contemporary studies of very sensitive topics, such as corruption, employ indirect questioning techniques, such as the randomized response method (for example, Gingerich 2013) or conjoint experiments (for example, Schuster, Meyer-Sahling, and Mikkelsen 2020) to protect respondents’ answers. When such techniques are too cumbersome or are not available, the use of organizational referents may be an attractive way to combat residual socially desirable responding by asking respondents about sensitive topics less directly. The cost of doing so, as we discuss below, is that organizational-referent questions on sensitive topics often place strong demands on respondents for information that may not be accessible to them.



In situations where SDB is severe and information is at least somewhat readily available to organizational outsiders, organizational aggregates may even be obtained from raters external to the organization. Such individuals will likely be less affected by SDB, although they may have other interests at stake in responding. However, using their answers comes at the cost of losing access to information from inside organizational boundaries, which may make their assessments noisy or inaccurate (for example, Razafindrakoto and Roubaud 2010). And, of course, this problem is likely to be particularly pernicious for sensitive questions, in which information is likely to be deliberately concealed from external assessment.

In sum, question sensitivity is a common reason for the use of organizational referents. There are good reasons to think this is an effective strategy, but, as far as we know, it has not been empirically examined in the context of civil service surveys. We do so below.

### Is Organizational Information Available to Respondents?

The third topic we cover concerns information. Specifically, in some circumstances, it may be difficult for respondents to have the information that organizational referents ask them to provide. When asked a question with an individual referent, respondents work to retrieve or recall information about the question (Tourangeau, Rips, and Rasinski 2000). For past behaviors, recall involves respondents' remembering what they have previously done. For beliefs or attitudes, following Zaller (1992), we can think of recall as respondents' process of deciding what beliefs or attitudes they hold, which can be either remembered or formed on the spot based on available information.

Recall and introspection are not perfectly reliable, and respondents tend to "fill in" information they are unsure about or do not recall accurately (Tourangeau, Rips, and Rasinski 2000). Yet the difficulties can multiply when questions are posed using organizational referents. Organizational referents impose an additional challenge for respondents. If organizational referents work as intended, respondents rely on different sources of information when answering questions about themselves or about others (cf. Blair, Menon, and Bickart 2004). It is reasonable to believe that information about aggregates, such as organizations, will often be harder for respondents to access, and perhaps harder to recall, than information obtained by introspection (that is, information about themselves). Consequently, respondents' beliefs about their organizational surroundings may be mistaken or biased, which may influence their responses.

When Meyer-Sahling and Mikkelsen (2016), for instance, ask respondents whether "political parties place their supporters in the ministerial structure" as a measure of personnel politicization, they are asking respondents for an evaluation they may not have sufficient information to provide accurately. Did new recruits get their positions due to political influence? Politicization may be hidden, particularly where—as in Central and Eastern Europe, where the authors collect their data—political influence over recruitment often extends to positions formally codified as career posts (for example, Meyer-Sahling 2011).

Due to these difficulties, respondents who are asked questions using organizational referents may get their answers wrong, with consequences for measurement. The literature on *establishment surveys*—instruments in which one respondent replies on behalf of an organization—assumes that respondents use records from the establishment to counteract these difficulties (Edwards and Cantor 2004). However, it is certainly optimistic to expect respondents in civil service surveys to do the same. Even if they could and were willing, many of the topics of central interest to civil service surveys—like politicization—are often not formally recorded. As such, errors rooted in mistaken beliefs are likely to persist.

Moreover, when respondents lack necessary information, they may default to public sector stereotypes or other heuristic shortcuts to construct an answer. If public servants hold views similar to the general public, for instance, they may default to considering their colleagues as stereotypically caring or dedicated (Willems 2020), irrespective of their own concrete knowledge about the caring or dedication of organizational members beyond their immediate coworkers. Or respondents may extrapolate from anecdotes or stories to a systemic evaluation, particularly if they are asked to evaluate questions on topics they view as threatening or emotionally engaging (Freling et al. 2020).

From the perspective of the response process, the built-in assumption behind the use of organizational referents can easily come to seem somewhat heroic in large and complex organizations. Findings from previous studies do not help. Baltes, Zhdanova, and Parker (2009) propose that respondents may rely on “better-off” or “worse-off” colleagues when responding to questions with organizational referents. This may bias estimates of organizational aggregates because the implicit referents that are actually used are no longer representative of the organization.<sup>4</sup> Similarly, Shah (1998) finds that job-related information is often obtained from people in similar positions, whereas organization-related information is obtained from friends within the organization. This means organizational-referent questions are answered using networked information rather than representative information or simple ratings of features of the organization.

However, organizational information may not be equally difficult to obtain in all organizations or by all public servants. When answering questions about others, respondents may start with themselves and subsequently take in the stories and observed behavior of others (Hoch 1987). This information may be sourced from networks, but there are predictable situations in which it is more likely to accurately represent the organization. In those situations, question referents are likely to matter less for responses, and hence concerns with the information requirements of organizational referents may not matter in practice.

First, drawing information from unrepresentative colleagues, stories, and observed behaviors should matter less when questions concern attitudes or behaviors that are either very rare or very common. In these situations, most colleagues, stories, and observed behaviors will provide the same information: that the attitude or behavior is very rare or very common. This means that while respondents may not, in fact, know the answer to a question using an organizational referent, their assessment is likely to be less affected by how they arrive at it. A similar point holds when most members of an organization hold roughly similar views because networked information in this situation is also more likely to be representative of the common view in the organization.

Learning is another factor that may limit how much questions using organizational referents elicit biased assessments. For instance, years of employment in an organization may improve the accuracy of reports about it (cf. Blair, Menon, and Bickart 2004). That is, respondents may learn to answer questions using organizational referents more accurately after years in an organization because they acquire more information over time.<sup>5</sup>

In sum, questions using organizational referents ask a lot of respondents informationally. Employees are asked to assess the characteristics of large and diverse organizations based on information they may not have. This is concerning because responses may come to rely on unrepresentative information, stories, observed behaviors, networks of colleagues, and social comparisons within public organizations rather than the real features of these organizations. This makes such questions less attractive where information is hard to obtain. The more we know about which respondents in which organizations are most likely to have the necessary information, however, the more we can counteract this disadvantage of organizational referents. In our analysis below, we seek to provide such knowledge, but we find that patterns are difficult to uncover.

To summarize, we arrive at the advantages and disadvantages outlined in table 23.1. Organizational referents have the advantage of matching target to measure when an inquiry is interested in describing organizations. This is the promise of the bottom-up perspective on organizational measurement. The disadvantage is that the perspective underpinning them requires substantial agreement in answers between employees within the same organization. This may not obtain. When agreement does not obtain, as our discussion of the top-down and summary bottom-up perspectives reveals, the conceptual advantage of organizational-referent questions for inquiries interested in organizations diminishes.

Moreover, asking about organizations may decrease SDB but may do so at the cost of placing large informational requirements on responding employees. Conversely, questions about respondents themselves require less external information and no within-organization agreement in responses. But this comes at the cost of greater SDB and of presenting organizational summaries rather than describing organizational features beyond individual respondents’ aggregated perspectives.

## DATA

We rely on two survey experiments to examine the questions we have raised in the previous section. We first describe the surveys in which these experiments were embedded, then the experiments themselves and how they help us gain strong leverage on question referents.

### Surveys

Our experiments were embedded in two surveys of central government public servants. We implemented the first survey in Romania between June 2019 and January 2020. Respondents were randomly assigned to face-to-face or online survey formats and partook in our experiment as part of a longer survey on civil service management practices. In all, we interviewed 3,316 respondents face-to-face (for a response rate of 92 percent) and 2,721 respondents online through Qualtrics (for a 24 percent response rate). The representativeness of our samples and the extent to which it differs according to the survey mode is covered in detail elsewhere in *The Government Analytics Handbook* (chapter 19).

We fielded the second survey in 18 Guatemalan government institutions between October and December 2019. Our experiments were embedded in a longer civil service survey. Respondents were sampled through the sample frame used for the Human Resources National Census, comprising staff lists of 14 central and four decentralized government institutions, and were asked to participate in face-to-face interviews. In all, we interviewed 3,465 respondents (for a 96 percent response rate).<sup>6</sup>

Though both surveys included responses concerning a range of civil service management practices of potential interest for questions surrounding the use of referents, we focus our attention on the analysis of the question-referent experiments. This is, as we explain next, where we get the strongest leverage on question-referent issues.

### Experiments

Our experiments all share the same essential strategy. Each survey respondent was randomly assigned to one of two survey flows. In one flow, the respondent was asked a set of questions (see below) that use organizational referents. In the other flow, the respondent was asked a set of questions differing from the first questions only in their use of individual rather than organizational referents.

Assignment to each survey flow was random for reasons of causal identification: random assignment ensures that the respondents who answered questions using individual referents and those who answered otherwise equivalent questions using organizational referents are identical, on average, on all observed and unobserved characteristics. As a result, any difference between average responses in the two flows must be due to the difference between them: whether question referents are individual or organizational. This ensures that we can causally identify the difference referents make to respondents' answers. It is the experimental setup that enables us to say with confidence that referents matter, how much, and for which organizations or groups of people.

The gist of our argument is this: if we ask some respondents a question on, say, salary satisfaction with reference to themselves and other respondents a salary-satisfaction question with reference to their organization, the average response from all respondents in an organization to each question should be the same if the question referent does not matter. The respondents who answered questions with individual referents are a random sample of all respondents and, thus, representative of them. The respondents who answered questions with organizational referents are also a random sample and, thus, representative in their views of their organization.<sup>7</sup> Therefore, any average difference between respondents assigned to different question referents must be due to the question referents.

In both surveys we fielded, we manipulated different sets of questions in this manner. In the survey in Romania, we assigned respondents to individual- or organizational-referent versions of questions surrounding recruitment (two questions), promotions (two questions), turnover (five questions), and dismissals (two questions). All questions were in five-point Likert scales. Additional follow-up questions on the use of various sources of information in recruitment and the questions asked at recruitment interviews were similarly randomized. We include these only in some of our analyses as they are scaled differently than the questions listed above. Questions were assigned to respondents in groups such that respondents either got all questions using individual referents or all questions using organizational referents. For instance, the group of respondents who received organizational-referent questions was asked the question “Please indicate the extent to which you agree or disagree with the following statements: The promotion process in my institution is fair.” By contrast, the other group of respondents, who received individual-referent questions, was asked the question “Please indicate the extent to which you agree or disagree with the following statements: The promotion process I have to go through in my institution is fair.” Appendix K.1 shows the full lists of questions in both versions.

Similarly, in the survey in Guatemala, we assigned respondents to individual- or organizational-referent versions of questions surrounding promotion confidence (one question), promotion fairness (one question), turnover (three questions), dismissals (two questions), and leadership (nine questions). Appendix K.1 shows the full lists of questions in both versions. Even where themes overlap, questions were formulated somewhat differently in Romania and Guatemala. Consequently, a comparison of results between the two countries should be made with caution.

From a design perspective, the experiments illuminate question-referent effects, but they do share a common drawback: we lack an objective benchmark for the phenomena, behaviors, or attitudes they measure. This means that while we are willing to interpret average higher scores on sensitive questions as diminishing SDB, we are often not strictly able to say whether individual or organizational referents caused the stronger method effect grounded solely in the way the question was posed. This is a weakness shared by most nonlaboratory experiments of this type, but we are still able to examine differences between individual- and organizational-referent questions, which are often informative. With this caveat noted, we proceed to our results.

## RESULTS

There is much we can examine within our framework using our data. Within the confines of this chapter, we cannot address every possible question. Instead, we opt to answer four questions directly related to the issues of substantive interest, information availability, and social desirability that we have outlined. Each subsection poses a question, which is immediately answered before detailed results are provided.

### Do Organizational-Referent Questions Reliably Reflect Organizational Characteristics?

Organizational-referent questions do not generally reflect organizational characteristics, though they often reflect them better than individual-referent questions do. For this reason, individual-referent questions may be preferred on conceptual grounds in many instances, given that organizational referents—while often resulting in increased agreement—are by no means guaranteed to ensure that questions result in clear ratings of organizational characteristics rather than summaries of individual perspectives.

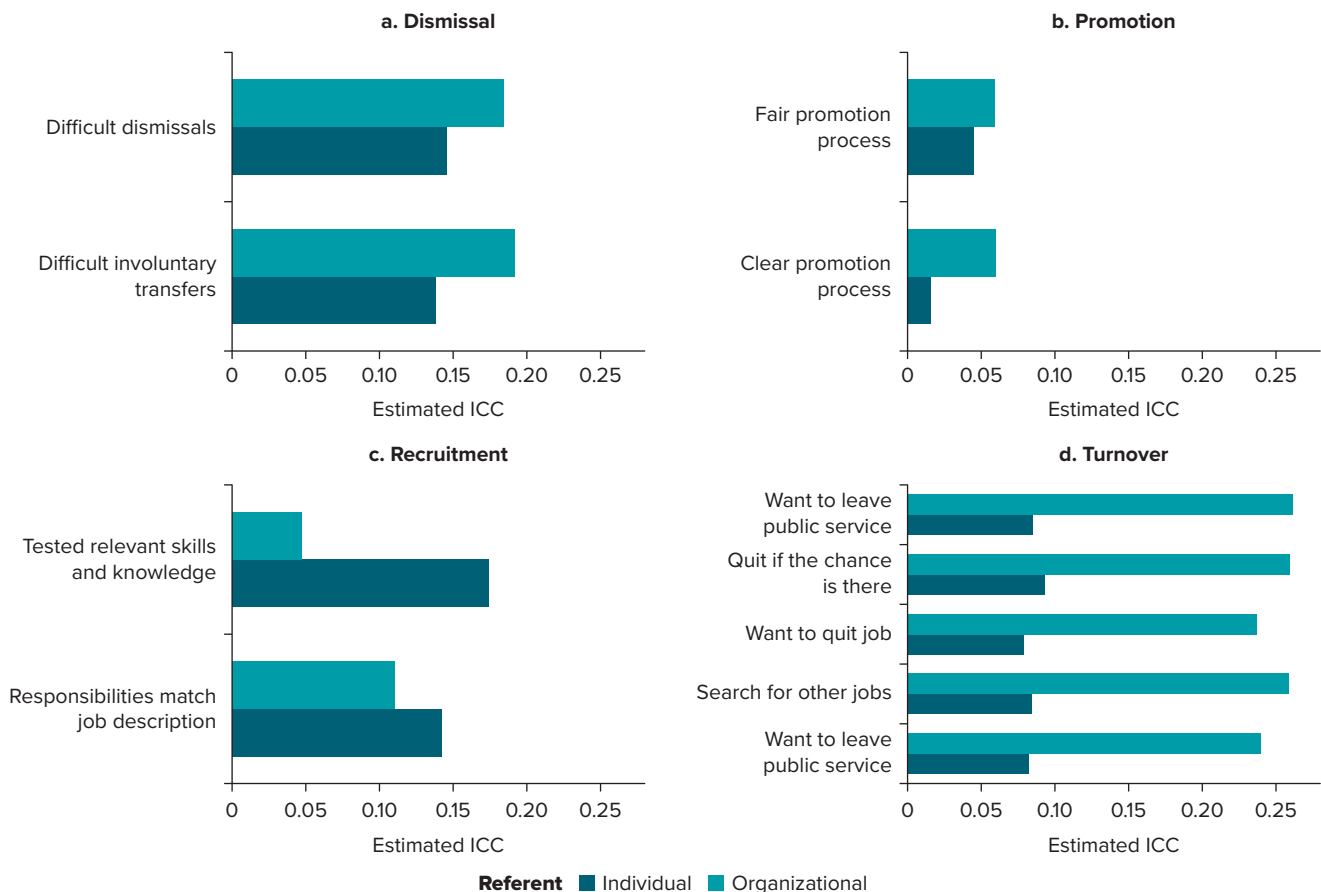
As noted, one central question for the bottom-up perspective on the utility of organizational and individual referents in civil service surveys concerns agreement within organizations. If respondents within an organization tend to agree in their responses to questions about their organization, we can more plausibly claim that their responses evaluate the same organizational phenomenon. If respondents rate the same entity in the world, they should agree in their ratings.

There are many measures of within-group agreement on survey measures. Here we opt for a common and simple measure, intraclass correlation (ICC). ICC is a measure of how much responses to questions rely on respondents' organizational setting. It can be interpreted as the percentage of variation in responses accounted for by organizational level. The higher the ICC, the more responses correlate within organizations—that is, the more respondents within organizations agree on their answers—and the more we can think of measures as reflecting objective organizational characteristics, which are simply observed and reported by respondents.

Our data permit the examination of two questions regarding ICC. First, are responses within organizations correlated to a high enough degree that we can think of the concepts they measure as genuine organization-level constructs? Second, is the correlation affected by the use of organizational or individual referents? If it is, this could indicate that organizational-referent questions can help survey designers elicit answers that characterize organizations from the appealing bottom-up perspective. Other things being equal, responses to questions about organizations *should* correlate more within organizations than responses to individual-referent questions.

Figures 23.1 and 23.2 examine these questions using the surveys from Romania (figure 23.1) and Guatemala (figure 23.2). Analysis of the Romanian data reveals that there is non-negligible agreement on responses within organizations for several questions but not for others. For some questions, the ICC is low enough that we might ask whether questions using either of the two referents elicit responses that refer to the same underlying phenomenon (rather than reporting two different perspectives).<sup>8</sup>

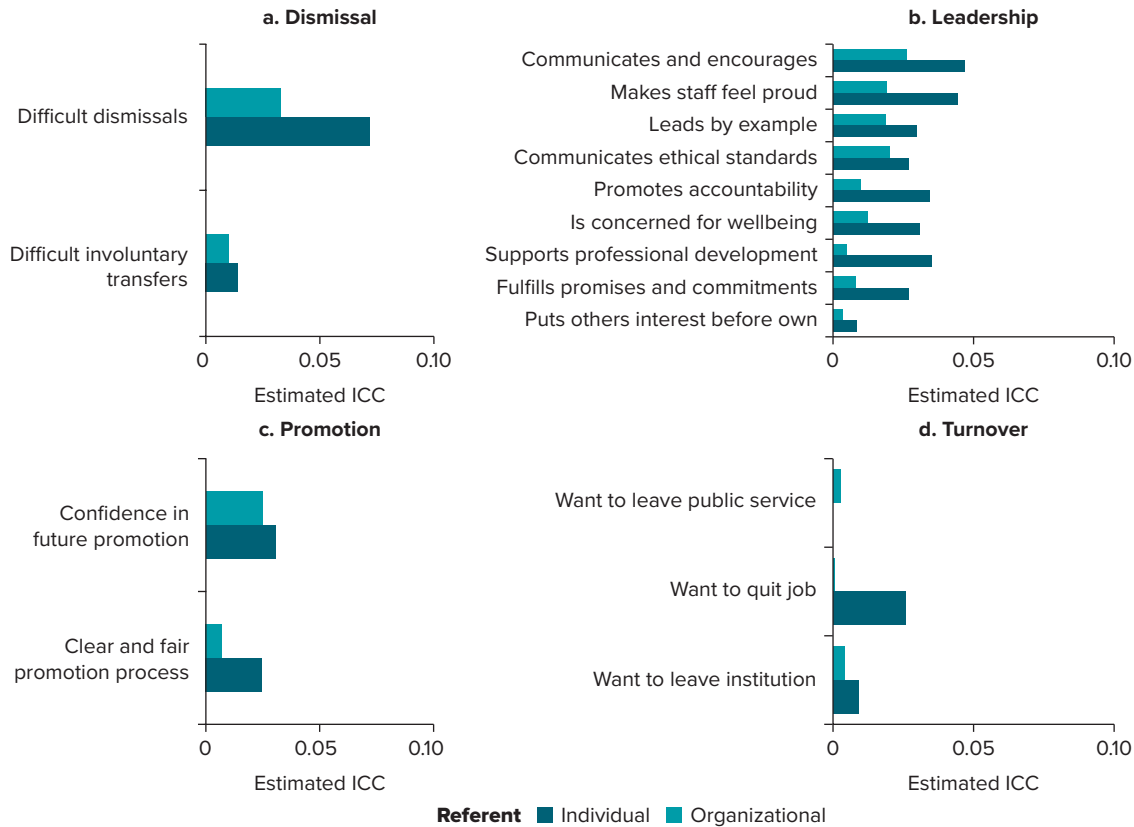
**FIGURE 23.1 Intraclass Correlations for the Romanian Data**



Source: Original figure for this publication.

Note: Bars show the calculated organizational ICC for each variable in the survey experiment in Romania, divided by HR area and treatment status. Positive differences between organizational (light blue) and individual (dark blue) referent questions indicate stronger agreement for the former than for the latter. See appendix K.1 for full items and question labels. ICC = intraclass correlation.

**FIGURE 23.2** Intraclass Correlations for the Guatemalan Data



Source: Original figure for this publication.

Note: Bars show the calculated organizational ICC for each variable in the survey experiment in Guatemala, divided by HR area and treatment status. Positive differences between organizational (light blue) and individual (dark blue) referent questions indicate stronger agreement for the former than for the latter. The horizontal axis is kept on the same scale as in figure 23.1 for ease of comparison. See appendix K.1 for full items and question labels. ICC = intraclass correlation.

Equally important for our purposes, these data show that organizational-referent questions *do* generally correlate more strongly within organizations than questions with individual referents. The expected agreement effect from organizational referents does emerge for some questions. The ICC for questions using organizational referents is higher for all but two recruitment questions in the Romanian data, though some differences are slight.

Question-referent effects are particularly pronounced for the turnover and recruitment questions. For turnover questions, within-organization agreement climbs by a factor of four. One possible explanation for this is social desirability. If respondents differ in their propensity to provide socially desirable answers more than they differ in their views on turnover intention among their colleagues, we could arrive at the pattern we observe. For now, however, this has to be considered speculative.

Somewhat puzzlingly, referent effects on recruitment items are reversed relative to what we would expect. Respondents to individual-referent questions agree more within organizations than respondents to organizational-referent questions. One possible explanation for this is that the questions using individual referent ask about recruitment processes that may have occurred years ago. This could lead to larger differences within organizations that have changed practices over time. However, our data do not reveal substantial differences in estimated ICCs if we split them along years of service.

Another possible explanation is that respondents' beliefs about public sector recruitment generally lead to the underestimation of differences between organizations, which drives down the ICC for questions using organizational referents, while individual-referent questions capture the diversity in recruitment practices.<sup>2</sup>



This is consistent with the fact that the between-organization variance of organizational-referent questions for recruitment is among the lowest in our data (alongside variables related to career advancement).

Analysis of the Guatemalan data reveals a similar pattern, although with a lower ICC across the board (figure 23.2). This offers two important lessons. First, many of the questions we examine do not appear to be statistically sound measures of bottom-up, organization-level constructs in Guatemala. The lower ICCs are due in part to the larger size of Guatemalan institutions, which leads to more variation within them. But this is precisely the point: respondents in these large organizations may be rating effectively different entities. It seems responses in our Guatemalan data are often better seen as employee perspectives, from the summary bottom-up perspective. Second, organizational referents do sometimes, as expected, help consolidate responses around agreeing ratings of organizational constructs, particularly for leadership and turnover.

What does this mean? From the bottom-up perspective of using survey responses to describe organizational characteristics, these analyses are not generally good news. Instead, they indicate that many organizational aggregates are perhaps better thought of, from the summary bottom-up perspective, as data summaries, particularly in Guatemala. That is, the summary bottom-up perspective appears to have more traction here than the pure bottom-up perspective. As noted, there may be good structural reasons for this. Public organizations are large, segmented, and complex entities in which management practices can vary by team, division, or section—particularly where management and human resources tasks are decentralized to line managers. Expecting consistent organizational characteristics to emerge under these conditions is, perhaps, expecting too much. The use of organizational referents does seem to consolidate a unified description by respondents, but to a limited degree, leaving plenty of disagreement behind.

Conceptually, then, while civil service surveys may benefit from the use of organizational referents, the big prize—the reliable description of organizational phenomena as rated by organizational members above and beyond their individual perspectives—appears elusive in our data. Given this conclusion, the question arises whether the use of organizational or individual referents matters to the data summaries both questions can provide.

### Does the Choice of Referent Matter for Responses?

Yes, in most instances, the choice of referent matters for responses, although it matters more for average responses than for relationships between response variables or for the tendency to respond at all. Average responses are sometimes higher and sometimes lower for organizational-referent questions, depending on the question. Similarly, nonresponse is sometimes more common and sometimes less common for organizational-referent questions, depending on the question. There is little systematic evidence that question referents matter to associations between different measures and less evidence still that associations are systematically stronger or weaker.

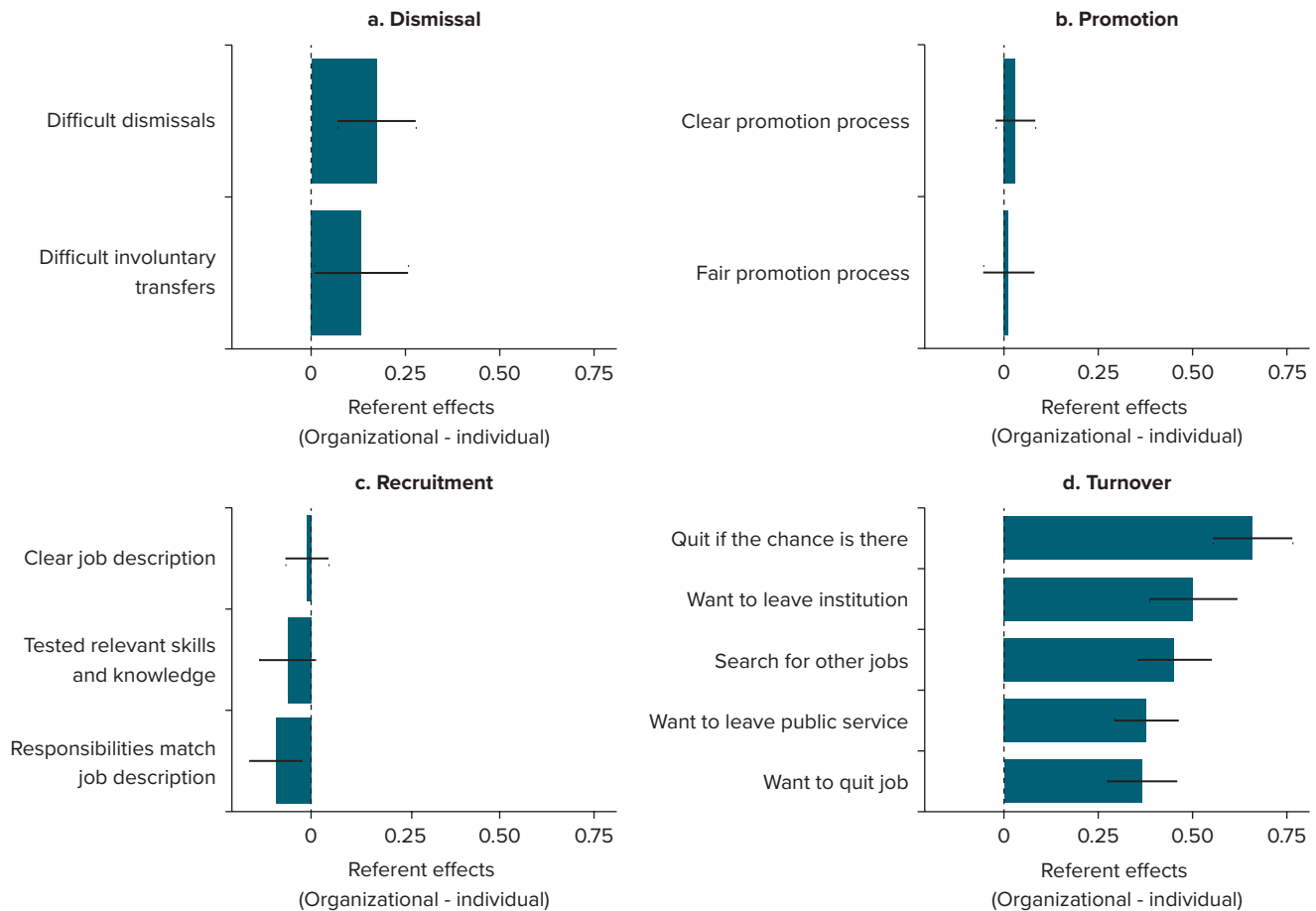
In figure 23.3, we show, using the Romanian data, the differences in average responses to questions on recruitment, turnover, dismissals, and promotion, varying only the use of organizational versus individual referents. As the figure shows, respondents who were asked about themselves rather than their colleagues are, on average,

- More convinced that they are difficult to dismiss or transfer,
- Less convinced that their responsibilities match their job descriptions, and, most markedly,
- Less willing to quit their jobs, organizations, or the public service.

Notably, two recruitment questions and both promotion questions do not show clear evidence that referents matter to responses.

These results provide the minimally expected result that different question referents result in different responses. Moreover, they are our first indication that the use of organizational referents really does make respondents more willing to admit to sensitive attitudes and behaviors, such as turnover intentions, as well as

**FIGURE 23.3 Organizational and Individual Referents in the Romanian Data**



Source: Original figure for this publication.

Note: Bars show estimated differences between organizational- and individual-referent questions in the survey experiment in Romania with 95 percent confidence intervals based on cluster-robust standard errors. Bars left of zero on the horizontal axis indicate higher scores on the individual-referent version of the question, whereas bars right of zero indicate higher scores on the organizational-referent version. All variables are scaled on the same 1–5 Likert scale. See appendix K.1 for full items and question labels.

slightly less prone to exaggerate their views on dismissals and their job descriptions. We return to this issue in more detail below.

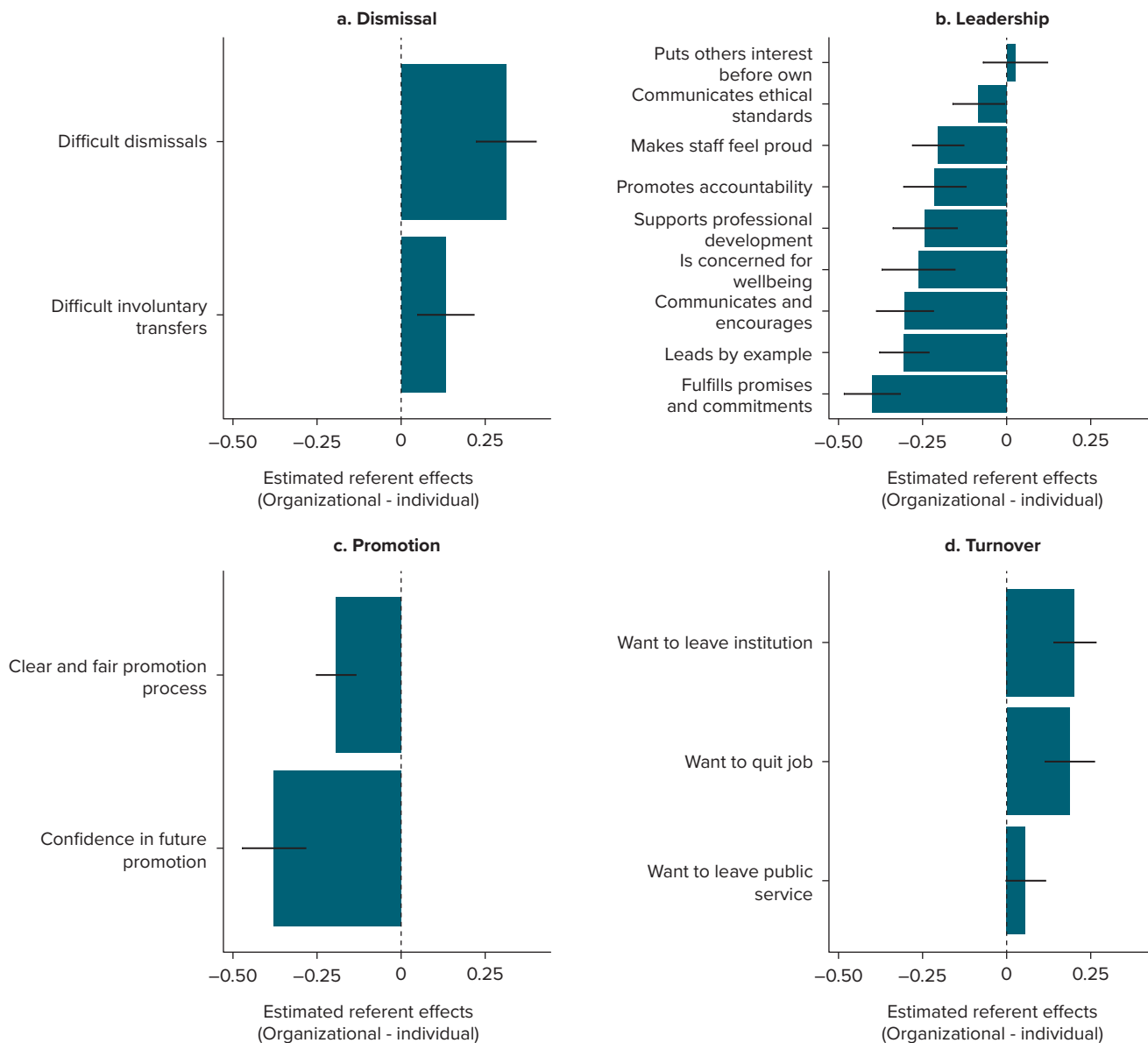
Figure 23.4 shows the results of a similar analysis using the Guatemalan data. In these data, the use of organizational versus individual referents matters more to average responses than in the Romanian data. Individual referents make respondents

- More likely to report that their direct managers are more transformational and ethical in their leadership styles on nearly any measure,
- Less prone to report turnover intentions,
- Less concerned about involuntary dismissals and transfers, and
- More convinced that promotions are within reach and that the process for achieving promotion is fair.

We can conclude at this stage that the choice of referent often matters to average responses—sometimes not a lot, but substantially for some questions. We return to plausible determinants of when referent choice matters below. Qualitatively speaking, however, we can already establish that referents do matter.

The average responses provided to survey questions matter a great deal, not least because they feed the organizational descriptive statistics commonly used in benchmarking organizations (about which, more shortly).

**FIGURE 23.4 Organizational and Individual Referents in the Guatemalan Data**



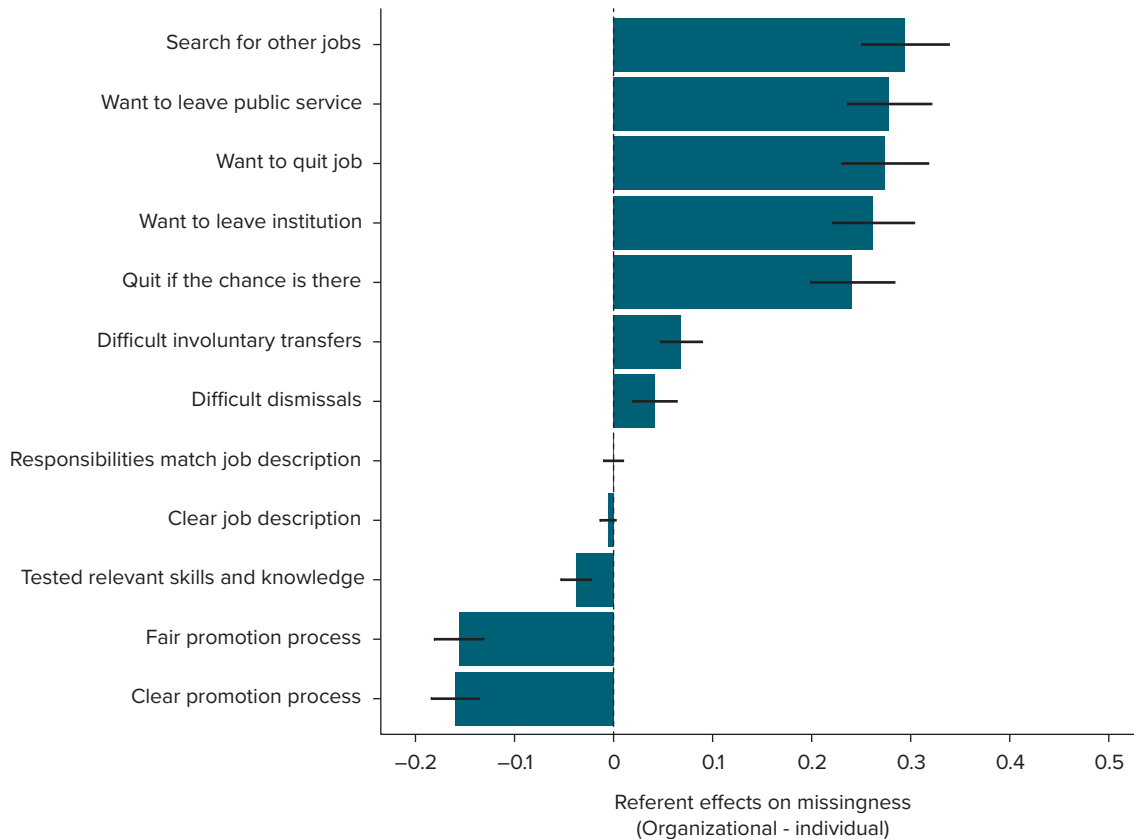
Source: Original figure for this publication.

Note: Bars show estimated differences between organizational- and individual-referent questions in the survey experiment in Guatemala with 95 percent confidence intervals based on cluster-robust standard errors. Bars left of zero on the horizontal axis indicate higher scores on the individual-referent version of the question, whereas bars right of zero indicate higher scores on the organizational-referent version. All variables are scaled on the same 1–5 Likert scale. See appendix K.1 for full items and question labels.

Yet average responses are not the only quantity that question referents may affect. It is possible, for instance, that organizational-referent questions are harder for respondents to understand, prompting item nonresponse—that is, respondents' not responding to individual items (see chapter 22).

Figure 23.5 examines this question using our Romanian data. Using a set of linear probability models with institution fixed effects, we find evidence of substantial nonresponse effects, particularly for questions relating to turnover. For each turnover question, the estimated probability of respondents not responding to individual-referent questions is increased by more than 20 percent relative to otherwise identical organizational-referent questions. This effect is substantial and worth considering. It is also worth noting, however, that less-sensitive questions on dismissal show much smaller effects, and questions on recruitment show no clear evidence of an effect at all. Moreover, individual referents substantially *reduce* nonresponse

**FIGURE 23.5** Estimates of Referent Effects on the Likelihood of Item Nonresponse



Source: Original figure for this publication.

Note: Bars show linear probability estimates of the differences between organizational- and individual-referent questions in the survey experiment in Romania with 95 percent confidence intervals based on cluster-robust standard errors. Bars left of zero on the horizontal axis indicate a higher probability of missingness on the individual-referent version of the question, whereas bars right of zero indicate a higher probability of missingness on the organizational-referent version. All variables are scaled on the same 0–1 scale, where 1 indicates “missing.” See appendix K.1 for full items and question labels.

relative to organizational referents for questions relating to promotion. One explanation for this finding may be that questions surrounding promotion processes are difficult to answer on behalf of the organization as a whole, leading respondents to nonresponse as a way of indicating they do not know the answer (see chapter 22). We return to the consequences of these findings below.

A final question we can examine is whether there are referent effects not on responses to individual survey variables but on relationships between survey variables. It is possible, for instance, that respondents fall back on their general opinions about the organization when asked for specific information about it, forming their attitudes as they go. This could result in increased statistical relationships between variables because they all tap into the same overarching attitude.

Table 23.3 examines this question using the leadership questions from the Guatemalan data. The table shows differences in statistical association between respondents who answered individual-referent questions and respondents who answered organizational-referent questions. Positive values indicate that organizational-referent questions correlate more strongly than similar individual-referent questions.

Table 23.3 does give some indication that variables covary differently when using organizational- rather than individual-referent questions. The effects we find indicate that the relevant relationships are generally—though not always—stronger when organizational referents are used. The differences vary in size, and not all are substantial. However, qualitative conclusions about the relationships between factors do sometimes hinge on the choice of referent. For example, when using our leadership, recruitment, and promotion variables to

**TABLE 23.3** Estimated Differences in Relationships between Leadership Variables for Different Referents, Guatemala (Organizational—Individual)

	Communicates and encourages	Communicates ethical standards	Fulfills promises and commitments	Is concerned for wellbeing	Leads by example	Makes staff feel proud	Promotes accountability	Puts others interest before own	Supports professional development
Communicates and encourages		−0.064* (0.028)	−0.046 (0.033)	−0.027 (0.022)	−0.085*** (0.019)	−0.037 (0.022)	−0.069* (0.031)	0.095* (0.036)	−0.042‡ (0.021)
Communicates ethical standards	−0.092** (0.026)		−0.122** (0.034)	−0.130*** (0.028)	−0.125*** (0.030)	−0.108*** (0.021)	−0.138*** (0.031)	0.059 (0.034)	−0.128*** (0.029)
Fulfills promises and commitments	0.012 (0.040)	−0.027 (0.043)		0.020 (0.041)	−0.055 (0.032)	−0.017 (0.033)	−0.007 (0.044)	0.179*** (0.037)	0.020 (0.026)
Is concerned for wellbeing	−0.033 (0.025)	−0.117** (0.036)	−0.035 (0.026)		−0.086* (0.034)	−0.021 (0.021)	−0.056‡ (0.031)	0.161*** (0.031)	−0.038* (0.017)
Leads by example	−0.037 (0.036)	−0.044 (0.037)	−0.073* (0.034)	−0.032 (0.028)		0.000 (0.025)	−0.046 (0.036)	0.154** (0.042)	−0.028 (0.027)
Makes staff feel proud	−0.038 (0.031)	−0.078* (0.028)	−0.086* (0.033)	−0.022 (0.026)	−0.056* (0.022)		−0.075** (0.022)	0.126*** (0.033)	−0.035 (0.032)
Promotes accountability	−0.013 (0.028)	−0.032 (0.027)	−0.009 (0.038)	−0.003 (0.025)	−0.035 (0.030)	−0.007 (0.023)		0.144*** (0.036)	−0.003 (0.028)
Puts others interest before own	0.137** (0.047)	0.117* (0.048)	0.208** (0.056)	0.191*** (0.044)	0.194** (0.056)	0.162*** (0.038)	0.176** (0.049)		0.156** (0.042)
Supports professional development	−0.019 (0.030)	−0.089* (0.038)	−0.009 (0.021)	−0.004 (0.027)	−0.049 (0.028)	−0.006 (0.025)	−0.032 (0.031)	0.150*** (0.037)	

Source: Original table for this publication.

Note: Results from ordinary least squares models with institution fixed effects and standard errors clustered by institution. Each cell in the table is the estimated interaction between our experimental treatment and the question in the cell's row in a model predicting the question in the cell's column. All variables are scaled on the same 1–5 Likert scale. See appendix K.1 for full items and question labels. *p*-values: ‡ *p* < 0.100, \* *p* < 0.050, \*\* *p* < 0.010, \*\*\* *p* < 0.001.

predict turnover variables in the Guatemalan data, 13 percent of estimated associations have different signs depending on the referent used.<sup>10</sup>

In sum, the choice of question referent matters. We find often small but sometimes substantial referent effects on the average responses to most questions we examine. Given our experimental setup, these differences must be due to the way we pose our questions. Hence, average differences are, in most cases, plausibly interpreted as being due to the question referent. We also find substantial referent effects on nonresponse patterns, but without a single direction of the effect. Whether referents make people respond more or less often appears to hinge on the question, its sensitivity, and how difficult it is to respond to. Finally, we find referent effects on relationships between some variables, but not in any clear direction.

### Can Organizational Referents Limit Social-Desirability Bias?

Yes, organizational referents limit SDB, but mostly for strongly sensitive items. We find evidence that more-sensitive questions show larger differences between individual- and organizational-referent questions in our experiment. This likely indicates that organizational referents can help limit SDB in civil service surveys. We find indications that this effect may be particularly pronounced for very sensitive questions.

As noted above, combatting SDB is a sensible reason for the use of organizational referents. To examine this question in more depth, we coded our individual questions in the Romania and Guatemala experiments for their sensitivity (see chapter 22 for details on the procedure). For the sake of statistical power in the analyses that follow, we now include the follow-up questions on the use of various sources of information in recruitment and the questions asked at recruitment interviews from the Romania questionnaire we have excluded from our analysis up to this point.

We regress this measure on the absolute difference between average responses to questions using individual and organizational referents (the referent effect), which we standardize to make our different response scales comparable. We run regressions with two sets of observations: one (model 1 in table 23.4) in which each observation is a question—from either survey—with its associated referent effect and sensitivity score, and one (model 2) in which each observation is an organizational aggregate for a question.

If organizational referents guard against SDB, we would expect a positive association between the sensitivity of questions and referent effect sizes because a reduction in SDB for sensitive questions would increase the difference between responses using organizational and individual referents. In our analysis, we find evidence for this assertion. In model 1, the expected positive association is significant only at the 10 percent level due to a low number of observations. In the more well-powered model 2, the expected association is highly significant. As expected, sensitive questions see larger question-referent effects, indicating that organizational referents may diminish socially desirable responding. It is worth noting, however, that this analysis

**TABLE 23.4** Standardized Question-Referent Effects, by Sensitivity

	Model 1 (Questions as observations)	Model 2 (Institution aggregates as observations)
Sensitivity	0.107 <sup>‡</sup> (0.056)	0.079*** (0.016)
(Intercept)	0.201*** (0.043)	0.328*** (0.013)
<i>N</i>	40	2,664
<i>R</i> -squared adjusted	0.065	0.008

Source: Original table for this publication.

Note: Results from ordinary least squares models. Each observation in model 1 is a question; each observation in model 2 is a question aggregate from an institution. The dependent variable is the absolute referent effect—the absolute difference in average responses between individual- and organizational-referent questions—standardized to account for the different scales of the included variables. See appendix K.2 for model results using other measures.

<sup>‡</sup>  $p < 0.100$ ; \*  $p < 0.050$ ; \*\*  $p < 0.010$ ; \*\*\*  $p < 0.001$ .



cannot leverage randomization to the same extent that our previous analyses do and, consequently, that it cannot be conclusively established whether the associations we document are due to sensitivity.

However, this analysis masks an additional finding: some very sensitive questions do appear to display larger differences than less sensitive questions. To see this, consider the violin plot in figure 23.6, showing the distribution of standardized referent effects for nonsensitive and sensitive questions (the thicker the “violin” at a certain height, the more questions have referent effects at the corresponding value on the second axis).

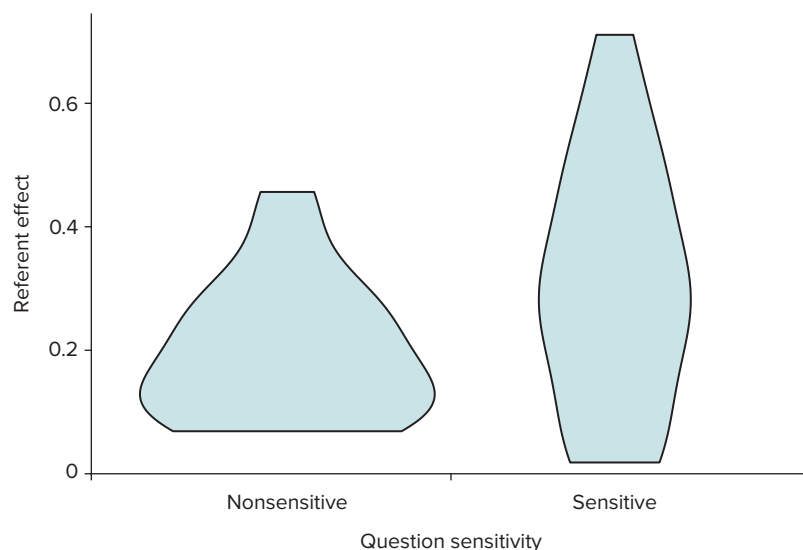
As the figure shows, the top of the referent effects distribution is far above the sensitivity effect observed in the table above, indicating that some sensitive questions have larger-than-predicted referent effects. On a qualitative inspection, these turn out to be very sensitive questions—particularly concerning turnover. This is a valuable conclusion. When examining sensitive issues—particularly highly sensitive issues such as corruption, politicization, or absenteeism—organizational referents appear to be able to combat SDB. For nonsensitive issues, the difference organizational referents make is more limited. The implication is that if the use of individual referents is preferred on other grounds, shifting to organizational referents may be justified on the grounds of SDB if questions are highly sensitive.

### Does Information Availability Matter?

Yes, information availability matters, but not in all the ways one might think. We find evidence that referent effects are smaller for very common attitudes and behaviors. However, we find no statistically clear evidence that respondents who have served longer in their organizations are less prone to referent effects.

As discussed, the availability of information may determine how much question referents matter if they are partly rooted in information availability. In these instances, we would expect smaller referent effects for questions about attitudes or behaviors that are either very common or very uncommon in respondents’ surroundings. Respondents are less (more) likely to report rare (common) behaviors about themselves by definition, but they are also less (more) likely to report rare (common) behaviors about their organizations because they encounter them rarely (commonly). By contrast, attitudes and behaviors that some hold but others do not can give rise to substantial referent effects, particularly if they are unevenly distributed within organizations.

**FIGURE 23.6** Distributions of Referent Effects for Sensitive and Nonsensitive Questions



Source: Original figure for this publication.

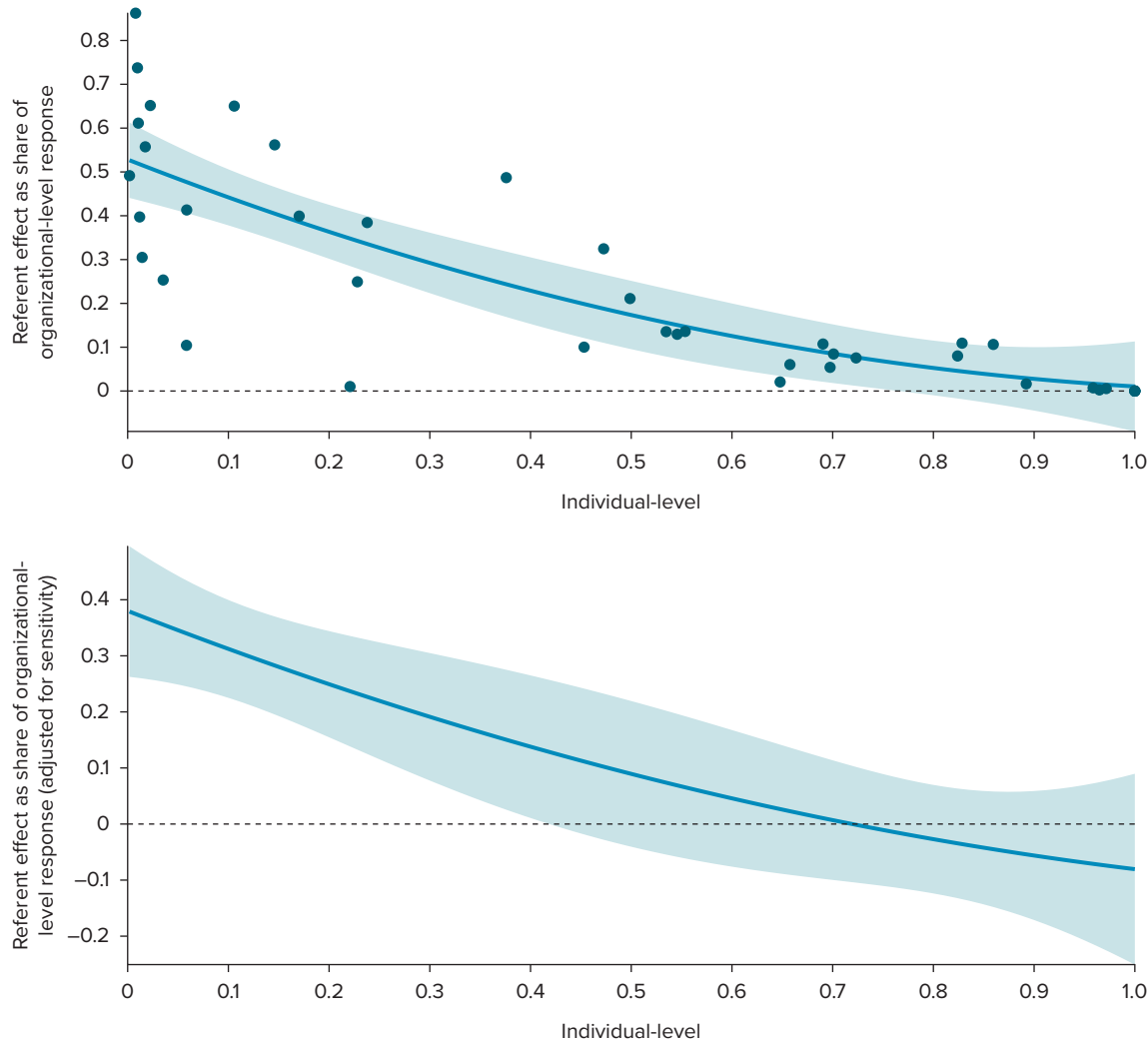
Note: The figure shows the distributions of referent effects split by question sensitivity. The width of the “violins” indicates the number of referent effects at or around the size indicated on the vertical axis. Thus, sensitive questions have a smaller range of referent effects, with the largest and smallest referent effects larger and smaller than for nonsensitive questions in our sample.

We examine this prediction by looking at patterns in referent effects. If very common or very rare attitudes and behaviors give rise to smaller referent effects, we should expect the referent effects, relative to reported commonality on organizational- (individual-) referent measures, to depend on how commonly the attitude or behavior in question is reported by respondents who are asked individual- (organizational-) referent questions. Specifically, we would expect an inverted-U relationship, in which referent effects are smaller for very rare or very common attitudes or behaviors.

Figure 23.7 speaks to this prediction. The figure plots the organizational proportion of affirmative responses to each question in our Romanian experiment (using individual referents) against the absolute difference between individual- and organizational-referent questions as a proportion of responses to the organizational-referent question.<sup>11</sup> Affirmative responses are interpreted as responses scoring on the upper two quintiles of the possible answers for scale questions (for example, “Strongly agree” and “Agree” on a Likert scale) and affirmative answers to follow-up questions, where respondents could indicate “yes” or “no.”

As the top panel in the figure shows, differences are not generally smaller for questions where scores are generally very low or very high, behaviors or practices are very rare or very common, and information is

**FIGURE 23.7** Response Score and Question-Referent Effects in the Romanian Data



Source: Original figure for this publication.

Note: The figure shows the average referent effect—here defined as the absolute difference between organizational- and individual-referent versions of each question in the experiment as a proportion of the score of the organizational-referent version—as a polynomial regression function of responses to the individual-referent version of the questions. The top panel shows the raw association, with individual questions plotted as points. The lower panel shows the association adjusted for question sensitivity.

more readily available. Instead, it shows referent effects declining as a function of commonality. One interpretation of this aligns, albeit asymmetrically, with information availability: it is easier for respondents to provide information about their organizations if they experience the relevant attitudes or behaviors around them, and this renders relative referent effects smaller for questions about common attitudes and behaviors than for questions where attitudes and behaviors are less common.

One obvious objection to this finding is that sensitive questions often result in indications that behaviors are rare, either because the behaviors in question *are* rare or because of SDB. As a result, the association depicted in the top panel of figure 23.7 could reflect sensitivity rather than information availability. To examine this issue, the lower panel in figure 23.7 shows the same association adjusted for question sensitivity. Indeed, the identified referent effects are weaker, but the pattern holds: referent effects appear to be smaller for questions targeting attitudes and behaviors that are common. Of course, one cannot definitively conclude from this simple analysis that greater information availability to respondents either will or will not result in smaller question referent effects. But the analysis does suggest that the use of organization-level referents may require caution when targeting rare behaviors or attitudes.

In our data, at least, we are not able to further pin down plausible determinants of information availability that give rise to the predicted changes in referent effects. To exemplify, further analysis of our two data sets (not shown) shows that organization size does not appear to matter for respondents' reactions to organizational versus individual referents, although one might expect smaller organizations to be easier to rate for respondents who use the information available to them, all else being equal. Moreover, as shown in table 23.5, the effect of using organizational referents in our Romanian sample does not generally vary with years of service. The exception is recruitment, where a negative referent effect grows with years of service (contrary to the idea that organizational experience would facilitate learning and diminish information-based referent effects). This effect could reflect changing recruitment practices over time, which would be consistent with the finding not being recovered when limiting the sample to relatively recently recruited public servants (model 6 in table 23.5).

However, some organizational characteristics do matter for referent effect sizes. If split by organization, the average referent effect size in the Romanian data is 0.15 (standardized across all experimental questions), but effect sizes range widely from one organization to another, from 0.02 to 0.31, the latter being a moderately sized effect, whereas the former is negligible.

The conclusion, then, is that if information availability matters in our data, we are not able to get very far in pinpointing its determinants. We can offer two suggestive conclusions, however. First, referent effects appear smaller for questions targeting attitudes that are very common. Second, arguing when information is available is no simple matter and is not a function of simple structural characteristics, such as organization size, or respondent characteristics, such as years of service.

**TABLE 23.5 Question-Referent Effects, by Years of Service, Romania**

	Model 3 (Dismissals)	Model 4 (Recruitment)	Model 5 (Turnover)	Model 6 (Recruitment, <5 years)
Organizational level	0.148 (0.091)	0.005 (0.041)	0.453*** (0.062)	-0.009 (0.073)
Years of service	-0.000 (0.004)	0.001 (0.002)	-0.003 (0.002)	-0.002 (0.023)
Organizational level × Years of service	-0.001 (0.005)	-0.005* (0.002)	0.002 (0.004)	-0.011 (0.021)
<i>N</i>	3,016	3,298	2,898	656
<i>R</i> -squared adjusted	0.137	0.088	0.216	0.212

Source: Original table for this publication.

Note: Results from ordinary least squares models with cluster-robust standard errors by institution. Each observation is a respondent in the Romania data set. The dependent variable is the indexes for our experimental measures of dismissals (model 3), recruitment (models 4 and 6), and turnover (model 5). All are kept on the same 1–5 scale as their items. Years of service is a single-item measure of how long respondents have served in public administration (measured in years). See appendix K.2 for model results using other measures.

‡  $p < 0.100$ ; \*  $p < 0.050$ ; \*\*  $p < 0.010$ ; \*\*\*  $p < 0.001$ .

## DISCUSSION AND CONCLUSION

Where do our experiments leave us? What do we learn from them? While they do give valuable insights on the effects of individual versus organizational referents in civil service surveys, they also raise new and interesting questions to which we do not yet have the answers.

The primary lesson is that the choice of referent matters. Using organizational referents often leads to more agreement between respondents in the same organization—a finding consistent with Glick (1985) and Klein et al. (2001). Yet in our measures, this agreement is often too low for responses to reliably track bottom-up, organization-level features above and beyond the perspective of respondents. The use of organizational referents can, however, provide summaries of respondents' perspectives and experiences, which are also, per the summary bottom-up perspective, valuable organizational metrics.

Moreover, average responses to survey questions often change when question referents change. For some questions, respondents report stronger agreement when asked about their organizations than when asked about themselves. For other questions, the pattern is reversed. In general, these effects are of modest size, but for some questions, they are substantial—and predicting for which questions referents will matter the most is not straightforward. Similarly, we find substantial question-referent effects on nonresponse but without uncovering one clear direction. For some questions, organizational referents substantially reduce nonresponse; for others, they exacerbate it. We also find some evidence that relationships between variables are affected by referents. But not all associations between variables are clearly impacted by the choice of referent, and we cannot propose a general direction of effects when they are.

We have examined the determinants of referent effect sizes: when does the choice between individual and organizational referents matter the most? From our analyses, we can draw only a few lessons about the question of referent effect size. First, referent effects seem to be larger for (highly) sensitive questions. This is consistent with organizational referents' ability to mitigate SDB for sensitive questions. Second, referent effects seem to be larger for attitudes, behaviors, and practices that are not common among respondents. This is consistent with the view that organization-level questions can pose higher informational demands than respondents can meet. It is also notable that question-referent effects are stronger in some organizations than others, but it is not clear which organizational characteristics drive these differences. And question-referent effects are not negatively associated with experience in the organization, suggesting that learning may have limited consequences for their size.

What does all this mean for civil service survey designers? It means they must be aware of the referents used in the questions they include in their surveys. Using organizational referents, as is common practice today, is not uniformly preferable on conceptual grounds—since responses often track but do not directly reflect organizational characteristics over and above respondents' perspectives. However, using individual referents is not uniformly preferable either. Particularly on measurement grounds, there is evidence that individual referents may suffer from SDB both for sensitive questions and for questions for which respondents wish to positively manage impressions.

Beyond awareness, we can make a few recommendations for more specific situations. First, a survey designer including very sensitive questions in a survey should consider posing these questions using organizational referents to combat SDB. It is important to recognize the limitations of this advice, however. Our analysis shows that more sensitive behavior is reported when using organizational referents. Yet this does not mean organizational referents provide an accurate estimate of how frequently the sensitive behavior or attitude occurs.

Moreover, using organizational referents comes at a heavy conceptual cost if the survey is interested in anything more than organizational aggregates. Predicting individual behaviors and attitudes with individual responses to sensitive organizational-referent questions implies a shift in what is studied (Klein and Kozlowski 2000). There is a difference between saying that a respondent's manager is abusive and that managers in the organization generally are abusive. Predicting sensitive organizational-referent questions with individual attitudes and experiences is often problematic because it tends to operationally conflate beliefs

about the organizational collective with individual attitudes and behaviors. If survey designers want to know why individual public servants behave and think as they do, the conceptual cost of organizational referents may be higher than the measurement gain, even for sensitive topics.

Second, survey designers should consider how the information needed to answer a question will be acquired by respondents. If using an organizational referent, can individual respondents reasonably be expected to know the answer? Individual referents are preferable if introspection provides more or more-reliable information than beliefs and available information about the organization. Our findings indicate few systematic patterns in which questions are most affected by this or in which respondents are most prone to provide the needed information accurately, rather than information infused with impressions, rumors, and beliefs. However, this does not mean that information availability can be glossed over by survey designers. Instead, it highlights the need for more measurement studies specifically targeting information availability and its determinants.

Third, our results may help survey designers think about utilizing other levels of measurement than individual or organizational. Of course, this implication is somewhat speculative, and more data are needed. Consider the conceptual issue with an organizational-referent question that elicits low levels of intraorganizational agreement in response. This means that respondents perceive their organization differently even though they all work within it. As noted, the usual interpretation of this occurrence is that organizational practices differ, that implementation of policies and procedures is uneven, and that management and leadership matter to how organizational practices are felt by public servants. There is nothing intrinsically wrong with this interpretation—but it is uncertain. After all, respondents were asked about their organization, not their section, team, manager, or other lower-level entities. It is not clear from our responses which level respondents draw on the most for information. This is an important weakness of organizational-referent questions in such a situation.

The interpretation gives rise to a question we cannot examine in detail using our data. Would it be a better strategy to use team referents or section referents rather than organizational or individual ones? Is it possible that using team referents would combat socially desirable responding without posing too high of informational demands on respondents? Our results cannot speak directly to this question. They do suggest that the answer is likely contingent on the type of question. Teams are often psychologically closer to people than whole organizations (Riketta and Dick 2005), which might mean that for some questions, team referents will do little to combat SDB. Similarly, some information can be difficult or impossible to access even within teams. Yet is likely to be more easily accessible within teams than for the entire organization. As such, team referent measures may be preferred to organizational-referent questions on measurement grounds if questions are not too sensitive. On the other hand, civil service survey designers may be less interested in reporting team aggregates to decision-makers or other audiences. And aggregating team aggregates to the organizational level is not likely to resolve the issues we discuss in this chapter.

Let us end with a few open questions for which both research and practice would benefit from systematic answers. We know much, both in conceptual and measurement terms, about multilevel theory, measurement, and modeling (Humphrey and LeBreton 2019; Klein and Kozlowski 2000). However, the literature on referent choice is limited, seemingly on the assumption that matching to the level of stated claims is all there is to it. This is sensible enough if one requires organizational measures to reflect organizational characteristics over and above respondents' perspectives in order to be useful. Yet such a perspective is overly limiting, not least for the practice of civil service survey design. For many variables, including management practices, perspectives on leadership, human resources functions, and more, data summaries of employee views and perspectives—interpretable from what we have referred to as the summary bottom-up perspective—can be valuable forms of decision support.

If we accept that organizational—or other higher-level—measurement referents can be useful even if respondents do not strongly align in response to them, our analyses point to a series of underexamined questions. First, which questions are particularly exposed to referent effects? We have found very sensitive questions to be affected, but much more knowledge is needed to reliably provide the type of advice survey designers want. Second, we have scarcely any evidence on whether the choice of referent affects different

survey respondents differently. We have not found any such effects in a few exploratory analyses, but this does not mean they do not exist. Third, it appears in our data that organizations affect the size of referent effects. We note that organization size does not appear to matter systematically, but we can see in our data that *something* about organizations does. Yet again, much more knowledge is needed on this issue.

The fact that our findings are not straightforward should highlight for both interested academics and survey designers that the choice of levels of measurement is a complicated issue, and, as we have shown, it is a choice that matters more than current practice seems to be aware.

## NOTES

1. An alternative design could randomly assign respondents a question order, with one group being asked individual-referent questions before organizational-referent questions and another group being asked organizational-referent questions before individual-referent questions. This would permit estimation of the average anchoring effect. However, as there is likely to be substantial heterogeneity in this effect, adjusting for the effect can become challenging. For this reason, we opt to ask each respondent only one set of questions.
2. For attitudinal variables, the equivalent of this perspective is that survey aggregates capture shared attitudes in the organization (Chan 1998).
3. This is true, in part, because impression management—wanting to control how one is viewed normatively—concerns both others (impression management proper) and oneself (self-deception) (for example, Millham and Kellogg 1980; Paulhus 1986).
4. The findings of Baltes, Zhdanova, and Parker (2009) suggest that organizational aggregates may depend on the unknown mixture of respondents using “upward” and “downward” comparisons to arrive at their answers. (Their findings also suggest that downward comparison is more common in their sample, but they are unable to assess the specific mixture.)
5. Respondents who have served longer in organizations have been shown in previous studies to be less prone to using heuristics in their decision-making because they can substitute their experience (cf. Pedersen, Stritch, and Thuesen 2018). Translated into the survey-response setting, more experienced personnel may not need to rely on stories and other heuristic devices when assessing their organizations.
6. Some respondents were interviewed even though they were not included on the original staff lists, meaning this number is somewhat inflated relative to those staff lists.
7. Note the assumption behind this null hypothesis is that respondents aggregate information in a way that approximates averaging when responding with reference to their organization. If this assumption does not hold, it poses an additional problem for organizational-referent questions because the aggregation used by respondents is then both unknown and does not approximate common-sense (though not the only sensible) aggregation procedures. Theoretically, this simply adds complexity to the information-processing discussion already noted.
8. This is because the organizational construct assessment of the ICC treats it as a measure of reliability. One way to think of this is to consider each respondent a rater of his or her organization. From this perspective, if at most 15 percent of variance is accounted for by organizations, for an ICC of 0.15, and at least 85 percent is accounted for by the raters, this does not indicate a reliable assessment of organizational characteristics. Raters affect responses too much.
9. We are grateful to an external reviewer for pointing us to this possibility and regret we have no better options available for examining it.
10. This figure includes only associations where effects in at least one direction are statistically significant at the 5 percent level. In none of the included cases are effects in both directions both statistically different from zero.
11. We thank a reviewer for pointing us in this direction. We originally considered simply presenting the absolute differences between answers to questions using different referents, but this created downward trends on the extremes of figure 23.7, consistent both with the prediction and a methodological artifact related only to question scaling.

## REFERENCES

- Baltes, Boris B., Ludmila S. Zhdanova, and Christopher P. Parker. 2009. “Psychological Climate: A Comparison of Organizational and Individual Level Referents.” *Human Relations* 62 (5): 669–700. <https://doi.org/10.1177/0018726709103454>.



- Bardasi, Elena, Kathleen Beegle, Andrew Dillon, and Pieter Serneels. 2011. "Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania." *The World Bank Economic Review* 25 (3): 418–47. <https://doi.org/10.1093/wber/lhr022>.
- Bezes, Philippe, and Gilles Jeannot. 2018. "Autonomy and Managerial Reforms in Europe: Let or Make Public Managers Manage?" *Public Administration* 96 (1): 3–22. <https://doi.org/10.1111/padm.12361>.
- Blair, Johnny, Geeta Menon, and Barbara Bickart. 2004. "Measurement Effects in Self vs. Proxy Response to Survey Questions: An Information-Processing Perspective." In *Measurement Errors in Surveys*, edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, 145–66. Hoboken, NJ: Wiley. <https://doi.org/10.1002/9781118150382.ch9>.
- Chan, David. 1998. "Functional Relations among Constructs in the Same Content Domain at Different Levels of Analysis: A Typology of Composition Models." *Journal of Applied Psychology* 83 (2): 234–46. <https://doi.org/10.1037/0021-9010.83.2.234>.
- Christensen, Tom, and Per Lægrend. 2007. "The Whole-of-Government Approach to Public Sector Reform." *Public Administration Review* 67 (6): 1059–66. <https://doi.org/10.1111/j.1540-6210.2007.00797.x>.
- Dunleavy, Patrick, Helen Margetts, Simon Bastow, and Jane Tinkler. 2006. "New Public Management Is Dead—Long Live Digital-Era Governance." *Journal of Public Administration Research and Theory* 16 (3): 467–94. <https://doi.org/10.1093/jopart/mui057>.
- Edwards, W. Sherman, and David Cantor. 2004. "Toward a Response Model in Establishment Surveys." In *Measurement Errors in Surveys*, edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, 211–33. Hoboken, NJ: Wiley. <https://doi.org/10.1002/9781118150382.ch12>.
- Eggers, William D. 2007. *Government 2.0: Using Technology to Improve Education, Cut Red Tape, Reduce Gridlock, and Enhance Democracy*. Lanham, MD: Rowman & Littlefield.
- Fisher, Robert J. 1993. "Social Desirability Bias and the Validity of Indirect Questioning." *Journal of Consumer Research* 20 (2): 303–15. <https://doi.org/10.1086/209351>.
- Freling, Traci H., Zhiyong Yang, Ritesh Saini, Omar S. Itani, and Ryan Rashad Abualsamh. 2020. "When Poignant Stories Outweigh Cold Hard Facts: A Meta-Analysis of the Anecdotal Bias." *Organizational Behavior and Human Decision Processes* 160: 51–67. <https://doi.org/10.1016/j.obhdp.2020.01.006>.
- Gingerich, Daniel W. 2013. "Governance Indicators and the Level of Analysis Problem: Empirical Findings from South America." *British Journal of Political Science* 43 (3): 505–40. <https://doi.org/10.1017/S0007123412000403>.
- Glick, William H. 1985. "Conceptualizing and Measuring Organizational and Psychological Climate: Pitfalls in Multilevel Research." *Academy of Management Review* 10 (3): 601–16. <https://doi.org/10.2307/258140>.
- Graaf, Gjal de, Leo Huberts, and Tebbine Strüwer. 2018. "Integrity Violations and Corruption in Western Public Governance: Empirical Evidence and Reflection from the Netherlands." *Public Integrity* 20 (2): 131–49. <https://doi.org/10.1080/10999922.2017.1350796>.
- Guenther, Corey L., and Mark D. Alicke. 2010. "Deconstructing the Better-Than-Average Effect." *Journal of Personality and Social Psychology* 99 (5): 755–70. <https://doi.org/10.1037/a0020959>.
- Hoch, Stephen J. 1987. "Perceived Consensus and Predictive Accuracy: The Pros and Cons of Projection." *Journal of Personality and Social Psychology* 53 (2): 221–34. <https://doi.org/10.1037/0022-3514.53.2.221>.
- Homburg, Christian, Martin Klarmann, Martin Reimann, and Oliver Schilke. 2012. "What Drives Key Informant Accuracy?" *Journal of Marketing Research* 49 (4): 594–608. <https://doi.org/10.1509/jmr.09.0174>.
- Humphrey, Stephen E., and James M. LeBreton, eds. 2019. *The Handbook of Multilevel Theory, Measurement, and Analysis*. Washington, DC: American Psychological Association.
- Klein, Katherine J., Amy Buhl Conn, D. Brent Smith, and Joann Speer Sorra. 2001. "Is Everyone in Agreement? An Exploration of Within-Group Agreement in Employee Perceptions of the Work Environment." *Journal of Applied Psychology* 86 (1): 3–16. <https://doi.org/10.1037/0021-9010.86.1.3>.
- Klein, Katherine J., Fred Dansereau, and Rosalie J. Hall. 1994. "Levels Issues in Theory Development, Data Collection, and Analysis." *Academy of Management Review* 19 (2): 195–229. <https://doi.org/10.5465/amr.1994.9410210745>.
- Klein, Katherine J., and Steve W. J. Kozlowski, eds. 2000. *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions*. San Francisco: Jossey-Bass.
- Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity." *Public Opinion Quarterly* 72 (5): 847–65. <https://doi.org/10.1093/poq/nfn063>.
- Meyer-Sahling, Jan-Hinrik. 2011. "The Durability of EU Civil Service Policy in Central and Eastern Europe after Accession." *Governance: An International Journal of Policy, Administration, and Institutions* 24 (2): 231–60. <https://doi.org/10.1111/j.1468-0491.2011.01523.x>.
- Meyer-Sahling, Jan-Hinrik, and Kim Sass Mikkelsen. 2016. "Civil Service Laws, Merit, Politicization, and Corruption: The Perspective of Public Officials from Five East European Countries." *Public Administration* 94 (4): 1105–23. <https://doi.org/10.1111/padm.12276>.

- Meyer-Sahling, Jan-Hinrik, and Kim Sass Mikkelsen. 2020. "Codes of Ethics, Disciplinary Codes, and the Effectiveness of Anti-Corruption Frameworks: Evidence from a Survey of Civil Servants in Poland." *Review of Public Personnel Administration* 42 (1): 142–64. <https://doi.org/10.1177/0734371X20949420>.
- Millham, Jim, and Richard W. Kellogg. 1980. "Need for Social Approval: Impression Management or Self-Deception?" *Journal of Research in Personality* 14 (4): 445–57. [https://doi.org/10.1016/0092-6566\(80\)90003-3](https://doi.org/10.1016/0092-6566(80)90003-3).
- Nishii, Lisa H., David P. Lepak, and Benjamin Schneider. 2008. "Employee Attributions of the 'Why' of HR Practices: Their Effects on Employee Attitudes and Behaviors, and Customer Satisfaction." *Personnel Psychology* 61 (3): 503–45. <https://doi.org/10.1111/j.1744-6570.2008.00121.x>.
- OPM (Office of Personnel Management). 2018. *Governmentwide Management Report: Results from the 2018 Federal Employee Viewpoint Survey*. Washington, DC: US Office of Personnel Management, US Government. <https://www.opm.gov/fevs/reports/governmentwide-reports/governmentwide-reports/governmentwide-management-report/2018/2018-governmentwide-management-report.pdf>.
- Parker, Christopher P., Boris B. Baltes, Scott A. Young, Joseph W. Huff, Robert A. Altmann, Heather A. Lacost, and Joanne E. Roberts. 2003. "Relationships between Psychological Climate Perceptions and Work Outcomes: A Meta-Analytic Review." *Journal of Organizational Behavior* 24 (4): 389–416. <https://doi.org/10.1002/job.198>.
- Paulhus, Delroy L. 1986. "Self-Deception and Impression Management in Test Responses." In *Personality Assessment via Questionnaires*, edited by Alois Angleitner and Jerry S. Wiggins, 143–65. Berlin: Springer-Verlag. [https://doi.org/10.1007/978-3-642-70751-3\\_8](https://doi.org/10.1007/978-3-642-70751-3_8).
- Pedersen, Mogens Jin, Justin M. Stritch, and Frederik Thuesen. 2018. "Punishment on the Frontlines of Public Service Delivery: Client Ethnicity and Caseworker Sanctioning Decisions in a Scandinavian Welfare State." *Journal of Public Administration Research and Theory* 28 (3): 339–54. <https://doi.org/10.1093/jopart/muy018>.
- Razafindrakoto, Mireille, and François Roubaud. 2010. "Are International Databases on Corruption Reliable? A Comparison of Expert Opinion Surveys and Household Surveys in Sub-Saharan Africa." *World Development* 38 (8): 1057–69.
- Riketta, Michael, and Rolf van Dick. 2005. "Foci of Attachment in Organizations: A Meta-Analytic Comparison of the Strength and Correlates of Workgroup versus Organizational Identification and Commitment." *Journal of Vocational Behavior* 67 (3): 490–510. <https://doi.org/10.1016/j.jvb.2004.06.001>.
- Schriesheim, Chester A., Joshua B. Wu, and Terri A. Scandura. 2009. "A Meso Measure? Examination of the Levels of Analysis of the Multifactor Leadership Questionnaire (MLQ)." *The Leadership Quarterly* 20 (4): 604–16. <https://doi.org/10.1016/j.leaqua.2009.04.005>.
- Schuster, Christian, Jan-Hinrik Meyer-Sahling, and Kim Sass Mikkelsen. 2020. "(Un)principled Principals, (Un)principled Agents: The Differential Effects of Managerial Civil Service Reforms on Corruption in Developing and OECD Countries." *Governance: An International Journal of Policy, Administration, and Institutions* 33 (4): 829–48. <https://doi.org/10.1111/gove.12461>.
- Shah, Priti Pradhan. 1998. "Who Are Employees' Social Referents? Using a Network Perspective to Determine Referent Others." *Academy of Management Journal* 41 (3): 249–68. <https://doi.org/10.2307/256906>.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.
- Tourangeau, Roger, and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133 (5): 859–83. <https://doi.org/10.1037/0033-2909.133.5.859>.
- Willems, Jurgen. 2020. "Public Servant Stereotypes: It Is Not (At) All about Being Lazy, Greedy and Corrupt." *Public Administration* 98 (4): 807–23. <https://doi.org/10.1111/padm.12686>.
- Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. Cambridge, UK: Cambridge University Press.