

Should we conduct correspondence study field experiments with political elites?

International Political Science Review

I–12

© The Author(s) 2021



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/01925121211026489

journals.sagepub.com/home/ips



Thomas Zittel 

Goethe-University Frankfurt, Germany

Tom Louwerse 

Leiden University, Netherlands

Helene Helboe Pedersen

Aarhus University, Denmark

Wouter Schakel

University of Amsterdam, Netherlands

Abstract

Correspondence study field experiments with political elites are a recent addition to legislative studies research, in which unsolicited emails are sent to elites to gauge their responsiveness. In this article, we discuss their ethical implications. We advance from the viewpoint that correspondence study field experiments involve trade-offs between costs and benefits that need to be carefully weighted. We elaborate this argument with two contributions in mind. First, we synthesize ethical considerations in published work to explore what the specific trade-offs are and how they can be mitigated by experimental design. We conclude that correspondence study field experiments with political elites are worth pursuing given their potential to further good governance. But they also involve distinct trade-offs that are particularly challenging. Second, we draw from our own considerations while designing a comparative correspondence study field experiment and stress challenges resulting from cross-national designs. In sum, we aim to facilitate further reasoned discussion on an important methodological issue.

Keywords

Field experiments, legislators, legislative studies, research ethics, political representation

Corresponding author:

Thomas Zittel, Department of Social Sciences, Goethe-University Frankfurt, Campus Westend PEG – Postbox 40, Frankfurt a.M. 60629, Germany.
Email: zittel@soz.uni-frankfurt.de

Introduction: the costs of correspondence study field experiments with political elites

Correspondence study field experiments (CSFEs) with political elites have proliferated in recent years (for an overview of older work, see Costa, 2017; recent work includes Bol et al., 2021; Breunig et al., 2020; Dinesen et al., 2021; Giger et al., 2020; Habel and Birch, 2019; Landgrave, 2020; Magni and Ponce de Leon, 2020; Rhinehart, 2020; Thomsen and Sanders, 2020; Wiener, 2020). In most of these experiments, unsolicited emails that conceal their true purpose are sent to elected officials. By varying the content of the emails and analyzing the subsequent replies (or the lack thereof), researchers draw inferences about the quality and quantity of individual-level political responsiveness.

CSFEs with political elites raise ethical concerns that require special attention in conducting them (Desposato, 2016; Grose, 2016) and taking their specific professional context into account. The professional context particularly affects how we weight the important principle of informed consent in research with humans. Experimental approaches in biomedical research involve potential violations of universal values such as the physical and psychological integrity of experimental subjects. These high stakes render the principle of informed consent a moral imperative and thus a key baseline against which the ethics of this research need to be evaluated (Israel, 2015; World Medical Association, 2008). In contrast to this, CSFEs with political elites inflict no major risks to the physical and psychological well-being of private citizens. Therefore, the principle of informed consent in these cases must be viewed from a cost-benefit perspective rather than as a moral imperative. However, CSFEs with political elites involve direct interventions in the political process, where obtaining informed consent could result in costs that may stand in the way of realizing their potential benefits. This asks for a careful examination of relevant costs we need to worry about in political science research (Desposato, 2016: 267-289; Whitfield, 2019) and of how to mitigate these costs by experimental design.

The costs that we consider particularly crucial involve collective costs to third parties; that is, negative externalities as compared to individual-level costs (Johnson, 2018; Naurin and Öhberg, 2019; Whitfield, 2019; Zimmermann, 2016). Collective costs can affect two types of communities that we need to pay attention to. First, CSFEs in politics might allocate costs to political communities. Demands that researchers make on legislators' time – that is, excessive individual costs – leave them with less time to perform their actual job of representing real citizens and thus constitute a harm to substantive representation. Furthermore, depending on the nature of the treatment, CSFEs may influence legislators' views and actions and thus also harm substantive representation. Consider, for example, a design in which letters are sent out to a large number of Members of Parliament (MPs) expressing support for a specific policy change. Such a design could affect legislators' perception of public opinion, which they may subsequently consider when deciding on their course of action on the issue in question. This would be particularly problematic when the treatment is not representative of constituents' actual policy views or involves false or misleading information.

Second, CSFEs in politics might allocate costs to other scientists. CSFEs may provoke backlashes among legislators who might feel deceived, tricked or overburdened with emails. This could be aimed at the researchers involved in the CSFE, but could also extend to the wider community of political scientists. This would harm the ties between legislators and the profession and, in that sense, jeopardize research projects that depend on access to MPs.

In the following, we first weight the costs of CSFEs with political elites against their benefits and ask how trade-offs can be mitigated by experimental design. We synthesize ethical considerations in published work to particularly explore contentious as opposed to non-contentious design

choices.¹ Second, we ask how national context matters for the effectiveness of distinct strategies in mitigating trade-offs. To elaborate this issue, we draw from own experiences in designing a collaborative comparative CSFEs involving national legislators in Denmark, Germany, the Netherlands and the United Kingdom. This aims to facilitate further reasoned discussion on an important methodological issue and particularly to guide future efforts to conduct CSFEs outside the United States (US), where most CSFEs with political elites so far have been conducted.

How to maximize the benefits of CSFEs with political elites and keep their costs at bay

Lawmakers in the US have taken note of the potential benefits of experimental research and have issued general guidelines on how to reconcile trade-offs. In this vein, the US 45 Code of Federal Regulations 46.116(f) allows to waive informed consent requirements for cases when the study involves no more than minimal risk to the subject, results in public benefits and could not otherwise practically be carried out. We largely follow this view. Consequently, justifying CSFEs with political elites depends upon our ability to specify potential benefits and then conducting them in ways that maximize benefits while minimizing costs.

CSFEs with political elites result in one important social benefit. They contribute to our knowledge about the quantity (level) and quality of responsiveness among political elites in their interactions with citizens; that is, whether MPs value direct interactions with what kind of citizens and to what end. CSFEs are able to do so in the most unbiased way possible since we need not rely on legislators' own recollections that could be biased by social desirability. Exploring the quantity and quality of legislators' responsiveness is not only a value in itself but has been found to be key to political trust (Esaiasson et al., 2017) and is likely to affect policy outcomes. Conducting research on this topic not only addresses possible shortcomings, but also helps to unveil unsubstantiated fears about the effectiveness of democratic governance and to identify best-practice institutional arrangements. In effect, CSFEs with politicians are audits of the political system just as field experiments are used as audits of labor and housing markets (Auspurg et al., 2018; Riach and Rich, 2004). We do not object to auditing these segments of society to acquire objective insights in the state of social responsiveness and equality; so, why would we not want to care about the same issues in politics?

How do we maximize the stated benefit of CSFEs while keeping collective costs down? In Table 1, we summarize research strategies to achieve positive trade-offs in this regard. We distinguish between strategic goals and distinct measures that we may adopt to further these goals and that we discuss in the following sub-sections in greater detail. In this discussion, we draw from ethical considerations in published CSFEs with legislators to identify tentatively and manifestly contentious measures as compared to non-contentious measures. While non-contentious measures are unambiguously positive, contentious measures might promote one goal but contradict others. Such tensions can be hypothetical in view of the literature, where we might find a clear consensus on what to prefer. But such tensions also can be manifest, where we find different answers in the literature.

Maximizing the societal benefits of CSFEs with political elites

Starting with unambiguous strategies, the literature suggests multiple ways to ensure that CSFEs function effectively as audits of democratic representation. The most important is to use CSFEs to focus on important research questions. To achieve this goal, we recommend to advance from key research questions in the literature that address normatively important issues. The studies on differential responsiveness are exemplary in this regard. Previous CSFEs unearthed substantial variation in the extent to

Table I. Strategies to maximize benefits and minimize costs of correspondence study field experiments (CSFEs) related to audits of the responsiveness of legislators.

Goal	Strategic goals	Specific measures	Contentious
Maximize societal benefits	Ask important research question	Advance from literature and address normatively important issues	No
		Embed CSFE in broader research efforts involving non-experimental data	No
	Avoid social desirability bias	Send realistic emails to avoid exposure	No
		Avoid pre-briefing subjects	Tentative
	Secure strong causal inference	Avoid de-briefing subjects with opt-out option	Tentative
		Embed CSFE in broader research efforts involving non-experimental data	No
	Pre-register experiment	Pre-register experiment	No
		Maximize statistical power	No
		Avoid pre-briefing subjects	Tentative
		Maximize number of observations	Yes
Minimize collective costs	Avoid distorting the representative process	Avoid recruiting confederates to secure control	Yes
		Avoid stating manipulated policy positions in treatments (information deception)	No
		Apply short treatment to pose no burden to Members of Parliament (MPs)	No
		Recruit confederates to inform the experiment	Yes
		Minimize the number of observations	Yes
	Avoid backlash against profession	Subject CSFE to third party ethics review	No
		Do not extract confidential information	No
		Do not report any individual-level findings	No
		Anonymize replication data	No
		Pre-brief subjects	Tentative
		De-brief subjects including opt-out option	Tentative
		Recruit confederates to avoid deception	Yes

Note: a measure denoted as 'Yes' with regard to whether it must be considered contentious or not involves debates in the relevant literature: A measure that we denote as 'Tentative' in this category does raise conflicting concerns, but these are not subject to debate.

which MPs are responsive to distinct social groups (Costa, 2017: 247) and helped to identify those groups most at risk. Both are normatively important insights. In this regard, studies have found evidence for ethnic discrimination in the US (Butler and Broockman, 2011: 248; Costa, 2017; Landgrave, 2020). Subsequent research has addressed the mechanisms behind ethnic discrimination (Broockman, 2013; Mendez and Grose, 2018), class-based discrimination (Butler, 2014), discrimination on the basis of gender (Rhinehart, 2020; Thomsen and Sanders, 2020) and differential responsiveness outside the US (Dinesen et al., 2021; Habel and Birch, 2019; Magni and Ponce de Leon, 2020).

To further augment the relevance of the question asked, we also suggest imbedding the CSFE in a broader research effort including careful pre-experimental investigation of citizens' perceptions of and preferences for political representation. With this latter recommendation, we do not argue that relevance is wholly determined by majority opinion but merely that it becomes more pertinent to explore a certain question if citizens express strong feelings towards it (see e.g. Wiener, 2020). Clearly, while having an important research question is beneficial for all studies, the costs imposed by CSFEs make it a particularly important goal, arguably even an essential one.

The social benefits of CSFEs with political elites – that is, their effectiveness as an audit instrument – also depends on good science; that is, how robust and sound their findings are. We are able to exploit their full potential as long as we eliminate social desirability biases that are likely to affect other methods when used to explore the responsiveness of legislators. Social desirability biases in representation research result from the fact that high and unbiased responsiveness are widely viewed as desirable qualities. It follows from this that using, for example, survey questions to gauge the responsiveness of political actors will likely paint an overly rosy picture since respondents avoid reporting low and biased responsiveness. Independent of strategic considerations, respondents might not be aware of their biases and thus might misreport their behavior and attitudes (Lensvelt-Mulders, 2008).

CSFEs only help to avoid social desirability biases as long as legislators are not aware they are taking part in an experiment; that is, as long as researchers are not exposed. To achieve this, the treatment emails should mimic real citizen emails as closely as possible. While this sounds simple, the challenge is to draft a message that is specific enough to pass for a genuine request, yet general enough to be relevant for all MPs to answer. Here, too, it is helpful to conduct exploratory research on legislators' constituency communication beforehand. This can involve asking citizens about emails they have written to their representatives or let them draft emails that can be used as sources while drafting the actual mails. This can also involve tapping into available constituency questions (see e.g. Gell-Redman et al., 2018) or conduct exploratory interviews with legislators in this regard, which is what we applied in our own research. Pre- or de-briefing subjects with opt-in or opt-out options should be avoided if we wish to out rule social desirability biases. However, such strategies could tentatively be promoted to minimize collective costs, which will be discussed further in the section on contentious measures.

The societal benefits that we are able to extract from using CSFEs as an audit instrument in legislative studies research also depends upon the robustness of the causal inferences that we can draw from it. Assessments, overall, take an optimistic stance on this issue and stress the promises of CSFEs compared to survey-based or observational research. However, they also highlight problems that we need to be aware of while designing experimental setups (Gerber and Green, 2012: chap. 1; Grose, 2014). The main concerns result from the fact that CSFEs aimed at testing theories of political behavior take place in real-world settings that complicate causal identification (Carlson, 2020), even if we account for the fact that randomization helps keeping omitted variable bias at bay.

To increase the validity of the experimental findings, we suggest embedding the CSFE in a larger project and collecting individual-level but also context-level data. This helps to enrich the data analysis by improving covariate balance, which is of special importance with a limited number of observations. Embedding the experiment in a larger research project also addresses issues of external validity since this helps to better contextualize the results emerging from the experiment. Last but not least, paying attention to issues of statistical power and opting for an adequately powered design is also unambiguously beneficial. Two ways to achieve this are to use block assignment (where randomization takes place within separate blocks of, for instance, male and female politicians) and a within-subject design (where the same politician is contacted several times). Both produce a narrower sampling distribution and thus a more precise estimate of the causal effect (Gerber and Green, 2012: 71–79, 273–276). Table 1 lists three further measures that should be avoided if we wish to secure strong causal inference. But again, such strategies could be promoted to minimize collective costs, which will be discussed further in the section on contentious measures.

Minimizing the collective costs of CSFEs with political elites

Avoiding costs that result in negative externalities is a key challenge while conducting CSFEs with political elites; that is, costs that result in disruptions of the representative process and a backlash

against the profession. To avoid the former, researchers should be careful when stating policy positions on behalf of voters in their treatments. In particular, treatments should not include false information about policy positions of larger groups of citizens, which may induce MPs to form unsubstantiated understandings of constituency sentiments and result in information deception (Landgrave, 2020). In this vein, our own experimental design deliberately stressed the individual nature of the email sent and refrained from articulating distinct policy positions.

Distortions of the representative process can also result from excessive burdens that CSFEs place on legislators in terms of time commitments. It is widely agreed upon that treatments need to be short and simple to answer, to not distract legislators and their staff from performing their actual responsibilities in their representative function. Published experiments aimed to achieve this goal by focusing on commonplace easy-to-answer service-related questions such as requests to provide help with voter registration procedures. Since our own design aimed to tap into legislators' policy responsiveness, we decided to focus on a salient issue, where answers should be readily available, and also to provide leeway in how to respond by asking a general question. In this particular question, we inquired about how legislators and their parties aimed to solve problems surrounding the COVID-19 pandemic. Legislators should have readily available positions on such salient issues that do not require extra work while drafting an answer.

With regard to a possible backlash against the profession, subjecting CSFEs to rigid ethics audits is an important measure to bring down potential risk. This renders backlashes against the research community less likely since scholars are able to show that CSFEs are in accordance with ethical guidelines that receive impartial third-party support. This would also most likely decrease the number of CSFEs conducted, thereby decreasing individual costs in terms of political elites' time and efforts and thus help to accommodate political concerns. Ethics audits are common practice in published research, which shows that ethical standards are indeed taken seriously among those conducting CSFEs. However, ethics audits in comparative CSFEs raise special challenges, which we further discuss in the section on comparative approaches.

As we stressed above, CSFEs with political elites inflict minimal personal harm to individual legislators. However, another issue is avoiding harm from reputational damage that could trigger a backlash. This can easily be avoided through a number of strategies. The potential for reputational damage is greatly reduced as long as researchers avoid asking legislators for confidential information and restrict their treatments to information that is publicly available. In addition, findings that allow individuals to be identified should not be reported. Doing so would not only be ethically problematic, it would also be theoretically uninteresting, since, by default, we are only interested in aggregate-level patterns of responsiveness. Party-level results also should only be made available if this is of theoretical interest. A third strategy to avoid reputational damage is to anonymize replication data.²

Contentious strategies

In a next step, we identify those strategies that have both positive and negative effects and that render CSFEs with political elites particularly challenging. To further discuss the most intricate choices in this regard, we distinguish between tentatively contentious and manifestly contentious strategies, where only the latter are subject to controversy in the literature. Trade-offs involved in the former need to be kept in mind but are commonly resolved in one or the other way.

One tentatively contentious strategy is to pre-brief those political elites that are to become subjects to a CSFE. Hereby, the research heeds the principle of informed consent to keep collective costs at bay. The briefing can take several forms. One is to inform subjects of the upcoming experiment and ask them to opt in; another possibility is to inform subjects and allow them to opt out.

This appears to minimize the risk of professional backlash. However, pre-briefing also has substantial downsides. The very goal of this strategy is to inform subjects of the treatment, which almost certainly re-introduces social desirability bias into their responses. Furthermore, pre-briefing impairs the causal inferences we can make with CSFEs, because it potentially leads to a lower number of observations, which may result in bias and inefficiency. In addition, a common awareness of the CSFE among legislators may lead to a situation in which the response from one legislator is influenced by the treatment which another legislator has received. This violates one of the core assumptions of experiments, namely that of non-interference (or stable unit treatment values) (Gerber and Green, 2012: 43–44). In sum, pre-briefing results in a negative cost-benefit score and therefore should be avoided.

Another tentatively contentious strategy is to de-brief legislators by informing them of the nature of the experiment after it has taken place. This might avoid distorting the relationship between voters and politicians; that is, that legislators will act on the information provided in the treatment. But as we stressed above, this should be precluded by design in any event and thus should be of no concern in CSFEs with political elites. The downside of de-briefing is that it confronts subjects with the inherent identity deception of the experiment in a very direct way, amplifying its otherwise minor individual costs, and thus leading to professional backlash (McClendon, 2012). In sum, de-briefing results in a negative cost-benefit score and is not promoted in the published literature. If it is seen as critical in ethics audits, we would at least suggest delaying this until the result of a study can be included in the de-brief. In that way, respondents are adequately informed about the reasons for and benefits of the field experiment, which may counterbalance any negative reactions.

A strategy that is manifestly contentious involves the use of confederates while sending out emails to politicians. Proponents of this strategy argue that we might be able to diminish deception in that way since the emails are sent out by actual citizens. Consequently, this may reduce a possible backlash against the profession (Breunig et al., 2020; Giger et al., 2020). The downside of this strategy is that it is not able to preclude all types of deception while at the same time results in loss of experimental control and increased funding needs; that is, increased societal costs. With regard to the types of deception involved, asking confederates to send out emails avoids identity deception but still involves motivation and activity deception. Politicians continue to be unaware that they are part of a scientific study and about the motivations of this study (Landgrave, 2020). It furthermore remains untested whether this particular form of deception is indeed able to prevent a backlash; for example, when it is combined with a de-brief option. With regard to the issue of experimental control, recruiting confederates increases the odds of non-intended variation in email texts and also drop-out incidents among confederates who did not send out their emails. This introduces unmeasured variation into the data. If this is random measurement error, this makes it harder to detect a causal effect. Systematic measurement error poses an even bigger problem. As the content is likely to correlate with our intended treatment, this effectively violates the other core assumption behind experiments, namely the assumption of excludability (Gerber and Green, 2012: 39–43). The bottom line here is that involving confederates poses risk for the quality of our causal inferences. This is not to say that designs in which distinct types of confederates are recruited for distinct purposes might not result in positive trade-offs (Landgrave, 2020; Wiener, 2020). This is just to highlight the challenges and to contradict portraying the use of confederates in CSFEs as a magic cure.

A final manifestly contentious strategy, judged by the courses of action adopted in published work, is to minimize the number of observations. While this helps to minimize potential negative externalities – that is, the danger of distorting political processes and also the risk of a backlash – it puts statistical power and thus opportunities for strong causal inferences at risk. This is in line with King et al.’s (1994) classic advice to maximize the number of testable implications of a hypothesis.

The solution here is straightforward: researchers should draw a large enough sample to reach conventional levels of statistical power (using informed estimates of realistic and relevant effect sizes) but not go beyond that. Conducting a power analysis prior to designing the experiment is crucial to identify the optimal number of cases in this regard, and is often not reported in published studies.

How to take national context into account: CSFEs in comparative research

While the use of CSFEs has become more common in political science, also beyond the US, cross-national comparative studies remain a rare species (one exception is Magni and Ponce de Leon, 2020). This is unfortunate since they can facilitate answers about how political institutions moderate interactions between representatives and voters and also help to test for the robustness of assumed direct citizen-level effects. In this vein, they deepen our understanding of the implications of institutional design and provide input to debates about possible reforms.

The ethical challenges of comparative CSFEs could be one reason for their scarcity. Our own experiences in designing a cross-national comparative CSFE show that reconciling the costs and benefits of CSFEs – and hence, reaching a sufficient trade-off – may depend on the national context in which the experiment is conducted. In other words, the advice that we provide in this article needs to be critically explored in view of distinct national contexts.

Context affects the trade-offs discussed above in various ways. First, the costs of CSFEs are lower when political elites are known to be generally supportive of them. This touches upon Scott Desposato's (2018) advice to practice 'empirical ethics'; that is, to ask potential subjects about their views towards experiments and to advance CSFEs only in contexts that are supportive in this regard. Desposato argues that this eases tensions between researchers and experimental subjects, and, while this may or may not be the case, widespread approval of CSFEs could reduce the backlash that they will elicit. We need to acknowledge, however, that empirical agreement is able to define the severity of challenges and risks regarding a potential backlash rather than setting backdrops that allow us to evaluate whether CSFEs are ethical; that is, whether they can be publicly justified (see e.g. Whitfield, 2019).

Second, country-specific institutional factors are likely to structure how the costs and benefits of CSFEs are weighted by political elites and mass publics, and thus how they can be justified. For example, strong freedom of information regimes ought to facilitate permissive attitudes and improve the odds for justification and agreement compared to weaker frameworks. Furthermore, professionalized parliaments lower the time and effort that are taken up by the experiment *in relative terms* since the treatment will be one of many thousands received by politicians and since staff members will be available to handle incoming emails. Arguably, it is also more important to audit democratic norms in a professionalized setting, since legislators can be held to higher standards of responsiveness in such a context (Malesky, 2016).

Third, besides context-dependent cost and benefits, comparative CSFEs are challenging in terms of treatment design. On the one hand, treatments should be as similar as possible across contexts to strengthen the comparative analysis. On the other hand, we have argued in favor of sending realistic emails to decrease risk of exposure and increase external validity of the experiment. However, realistic emails may vary across context for reasons related to the theoretically relevant variables. For instance, in cases that involve electoral systems with single-member districts a realistic email is conventionally sent by a district constituent and is often related to district-level concerns or service requests. In contrast, in proportional electoral systems that involve multi-member districts realistic emails are less likely to include information about the district and may rather relate to policy-related

issues. Comparative studies thus require additional considerations and trade-offs regarding the design of treatments. These need to be informed by the research question and contextual factors that constrain what is deemed to be a realistic email.

All three points just stressed result in a fourth and final point, namely the need to aim at local ethics reviews in comparative CSFEs to heed differences in country context (see Desposato, 2016: Part II). However, this also results in significant challenges. As it currently stands, in European contexts, researchers need to submit their applications for ethical approval to university or department-level ethics boards that may apply different yardsticks contingent upon place and dominant professional orientation. This is different from the American context, where a federal regulatory regime guides the deliberations of university-based ethics boards across the country. In Europe, some national professional associations – for example, in the fields of psychology and sociology – have taken on the task of working towards common guidelines (Gell-Redman et al., 2018; Grose, 2014; Mendez and Grose, 2018). This might harmonize ethics audits within disciplines within one country. But this is hardly able to increase the consistency of ethic audits across national and professional contexts in Europe.

The segmentation of audit processes in European contexts raises special challenges to comparative CSFEs seeking local review. This is because they might face different decisions that involve different requirements within one and the same CSFE. This is what we experienced in our own research, particularly with regard to de-briefing. For example, one of the national ethics audits that our experiment was subject to asked for a de-brief with opt-in option. In line with our previous considerations, this resulted in the decision to withdraw this particular CSFE since it would have not been able to avoid potentially significant social desirability biases. Our experiences in a second of our four national experiments further corroborate the challenges comparative CSFEs face under local review and the risk involved. The decision taken in his case to run the experiment after the ethics review board required a de-brief with opt-out option resulted in a major backlash. Overall, our advice is to pay close attention to this issue in comparative CSFEs. Differences in experimental design between countries would compromise good science and also could result in non-voluntary exposure and resulting professional backlash.

Conclusion

In this article, we have discussed the ethical implications of CSFEs with political elites from a cost-benefit perspective. Advancing from a synthesis of published work on the issue we have argued that, despite the costs that CSFEs involve, it would be misguided to abandon them given their potential to further good governance. This article developed a rationalistic framework to navigate the trade-off between the costs and benefits of CSFEs. It particularly stressed some of the most challenging and contentious issues in this regard and also reflected upon the role of context in comparative designs. By following the outlined guidelines and by taking into account how context affects the trade-offs, we can increase the odds that CSFEs present a minimal burden to legislators, minimize externalities and provide valuable knowledge to society as a whole. Such a transparent and deliberate approach is of special importance in view of the legitimate concerns that CSFEs face and the resulting questions about their value and scientific integrity. We hope our article will inspire ethically sound field experiments in comparative legislative studies research and promote further discussion in this regard.

Acknowledgments

Previous versions of this article were presented in a workshop at Goethe-University Frankfurt on 2 and 3 May 2019 and at the ECPR General Conference in September 2019 in Wroclaw. We thank participants of these

meetings for most valuable comments, in particular Julie Sevenans who served as a discussant on the Wroclaw panel. We also are indebted to three anonymous reviewers for their most valuable and constructive comments.

Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This work was supported by grants from the Deutsche Forschungsgemeinschaft (Grant Number: ZI 608/8-1) and the Nederlandse Organisatie voor Wetenschappelijk (Grant Number: 464.18.108).

ORCID iDs

Thomas Zittel  <https://orcid.org/0000-0003-2964-6617>

Tom Louwerse  <https://orcid.org/0000-0003-4131-2724>

Notes

1. For a list of reviewed CSFEs, see Appendix 1.
2. This is commonplace but merits special emphasis due to its critical nature. We also are aware of one specific transgression, where individual-level data have been exposed in non-anonymized ways.

References

- Auspurg, Katrin, Andreas Schneck and Thomas Hinz (2018) Closed Doors Everywhere? A meta-analyses of field experiments on ethnic discrimination in rental housing markets. *Journal of Ethnic and Migration Studies* 45(1): 95–114.
- Bol, Damien, Thomas Gschwend, Thomas Zittel and Steffen Zittlau (2021) The Importance of Personal Vote Intentions for the Responsiveness of Legislators: A field experiment. *European Journal of Political Research* 60(2): 455–473.
- Breunig, Christian, Emiliano Grossman and Miriam Hanni (2020) Responsiveness and Democratic Accountability: Observational evidence from an experiment in a mixed-member proportional system. *Legislative Studies Quarterly*. Epub ahead of print 30 December 2020. DOI: 10.1111/lsq.12326.
- Broockman, David E (2013) Black Politicians are More Intrinsically Motivated to Advance Blacks' Interests: A field experiment manipulating political incentives. *American Journal of Political Science* 57(3): 521–536.
- Butler, Daniel M (2014) *Representing the Advantaged: How Politicians Reinforce Inequality*. New York, NY: Cambridge University Press.
- Butler, Daniel M and David E Broockman (2011) Do Politicians Racially Discriminate Against Constituents? A field experiment on state legislators. *American Journal of Political Science* 55(3): 463–477.
- Carlson, Elizabeth (2020) Field Experiments and Behavioral Theories: Science and ethics. *PS: Political Science & Politics* 53(1): 89–93.
- Costa, Mia (2017) How Responsive are Political Elites? A meta-analysis of experiments on public officials. *Journal of Experimental Political Science* 4: 241–254.
- Desposato, Scott (2016) *Ethics and Experiments. Problems and Solutions for Social Scientists and Policy Professionals*. New York, NY: Routledge.
- Desposato, Scott (2018) Subjects and Scholars' Views on the Ethics of Political Science Field Experiments. *Perspective on Politics* 16(3): 739–750.
- Dinesen, Peter Thisted, Malte Dahl and Mikkel Schiøler (2021) When are Legislators Responsive to Ethnic Minorities? Testing the role of electoral incentives and candidate selection for mitigating ethnocentric responsiveness. *American Political Science Review* 115(2): 450–466.
- Esaiasson, Peter, Mikael Gilljam and Mikael Persson (2017) Responsiveness Beyond Policy Satisfaction: Does it matter to citizens? *Comparative Political Studies* 50(6): 739–765.

- Gell-Redman, Micah, Neil Visalvanich, Charles Crabtree and Christopher J Fariss (2018) It's All About Race: How state legislators respond to immigrant constituents. *Political Research Quarterly* 71(3): 517–531.
- Gerber, Alan S and Donald P Green (2012) *Field Experiments. Design, Analysis, and Interpretation*. New York, NY: WW Norton.
- Giger, Nathalie, Simon Lanz and Catherine de Vries (2020) The Motivational Basis of Constituency Work: How intrinsic and extrinsic motivations interact. *Political Science Research and Methods* 8(3): 493–508.
- Grose, Christian R (2014) Field Experimental Work on Political Institutions. *Annual Review of Political Science* 17: 355–370.
- Grose, Christian R (2016) Field Experiments on Elected Public Officials: Ethical obligations and requirements. In Scott Desposato (ed.) *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*. New York, NY: Routledge, 227–238.
- Habel, Philip and Sarah Birch (2019) A Field Experiment on the Effects of Ethnicity and Socioeconomic Status on the Quality of Representation. *Legislative Studies Quarterly* 44(3): 389–420.
- Israel, Mark (2015) *Research Ethics for Social Scientists*. London: SAGE.
- Johnson, Jeremy B (2018) Protecting the Community: Lessons from the Montana flyer project. *PS: Political Science & Politics* 51(3): 615–619.
- King, Gary, Robert O Keohane and Sidney Verba (1994) *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- Landgrave, Michelangelo (2020) Can We Reduce Deception in Elite Field Experiments? Evidence from a field experiment with state legislative offices. *State Politics & Policy Quarterly* 20(4): 489–507.
- Lensveld-Mulders, Gerty (2008) Surveying Sensitive Topics. In Edith D de Leeuw, Joop J Hox and Don A Dillman (eds) *International Handbook of Survey Methodology*. New York, NY: Lawrence Erlbaum Associates, 461–478.
- McClendon, Gwyneth (2012) Ethics of Using Public Officials as Field Experiment Subjects. *Newsletter of the APSA Experiments Section* 3(1): 13–20.
- Magni, Gabriele and Zoila Ponce de Leon (2020) Women Want an Answer! Field experiments on elected officials and gender bias. *Journal of Experimental Political Science*. Epub ahead of print 10 August 2020. DOI: 10.1017/xps.2020.23.
- Malesky, Edmund J (2016) Manipulating Elites. In Scott Desposato (ed.) *Ethics and Experiments. Problems and Solutions for Social Scientists and Policy Professionals*. New York, NY: Routledge, 217–226.
- Mendez, Matthew S and Christian R Grose (2018) Doubling Down: Inequality in responsiveness and the policy preferences of elected officials. *Legislative Studies Quarterly* 43(3): 457–491.
- Naurin, Elin and Patrik Öhberg (2019) Ethics in Elite Experiments: A perspective of officials and voters. *British Journal of Political Science* 51(2): 890–898.
- Rhinehart, Sarina (2020) Mentoring the Next Generation of Women Candidates: A field experiment of state legislators. *American Politics Research* 48(4): 492–505.
- Riach, Peter A and Judith Rich (2004) Deceptive Field Experiments of Discrimination: Are they ethical? *Kyklos* 57(3): 457–470.
- Thomsen, Danielle M and Bailey K Sanders (2020) Gender Differences in Legislator Responsiveness. *Perspectives on Politics* 18(4): 1017–1030.
- Whitfield, Gregory (2019) TRENDs: Toward a separate ethics of political field experiments. *Political Research Quarterly* 72(3): 527–538.
- Wiener, Elizabeth (2020) Getting a High Heel in the Door: An experiment on state legislator responsiveness to women's issue lobbying. *Political Research Quarterly*. Epub ahead of print 9 July 2020. DOI: 10.1177/1065912920939186.
- World Medical Association (2008) Declaration of Helsinki: Ethical principles for medical research involving human subjects. <https://www.wma.net/what-we-do/medical-ethics/declaration-of-helsinki/doh-oct2008/>
- Zimmermann, Brigitte (2016) Information and Power. Ethical considerations of political information experiments. In Scott Desposato (ed.) *Ethics and Experiments. Problems and Solutions for Social Scientists and Policy Professionals*. New York, NY: Routledge, 183–197.

Appendix I. List of published correspondence study field experiments with legislators reviewed with regard to ethical considerations in chronological order

- Bol, Damien, Thomas Gschwend, Thomas Zittel and Steffen Zittlau (2021) The Importance of Personal Vote Intentions for the Responsiveness of Legislators: A field experiment. *European Journal of Political Research* 60(2): 455–473.
- Breunig, Christian, Emiliano Grossman and Miriam Hanni (2020) Responsiveness and Democratic Accountability: Observational evidence from an experiment in a mixed-member proportional system. *Legislative Studies Quarterly*. Epub ahead of print 30 December 2020. DOI: 10.1111/lsq.12326.
- Broockman, David E (2013) Black Politicians are More Intrinsically Motivated to Advance Blacks' Interests: A field experiment manipulating political incentives. *American Journal of Political Science* 57(3): 521–536.
- Butler, Daniel M and David E Broockman (2011) Do Politicians Racially Discriminate Against Constituents? A field experiment on state legislators. *American Journal of Political Science* 55(3): 463–477.
- Butler, Daniel M and David W Nickerson (2011) Can Learning Constituency Opinion Affect How Legislators Vote? Results from a field experiment. *Quarterly Journal of Political Science* 6(1): 55–83.
- Butler, Daniel M, Christopher F Karpowitz and Jeremy C Pope (2012) Field Experiment on Legislators' Home Style: Service versus policy. *Journal of Politics* 74(2): 474–486.
- Dinesen, Peter Thisted, Malte Dahl and Mikkel Schiøler (2021) When are Legislators Responsive to Ethnic Minorities? Testing the role of electoral incentives and candidate selection for mitigating ethnocentric responsiveness. *American Political Science Review* 115(2): 450–466.
- Gell-Redman, Micah, Neil Visalvanich, Charles Crabtree and Christopher J Fariss (2018) It's All About Race: How state legislators respond to immigrant constituents. *Political Research Quarterly* 71(3): 517–531.
- Giger, Nathalie, Simon Lanz and Catherine de Vries (2020) The Motivational Basis of Constituency Work: How intrinsic and extrinsic motivations interact. *Political Science Research and Methods* 8(3): 493–508.
- Habel, Philip and Sarah Birch (2019) A Field Experiment on the Effects of Ethnicity and Socioeconomic Status on the Quality of Representation. *Legislative Studies Quarterly* 44(3): 389–420.
- Kalla, Joshua, Frances Rosenbluth and Dawn Langan Teele (2018) Are You My Mentor? A field experiment on gender, ethnicity, and political self-starters. *Journal of Politics* 80(1): 337–341.
- Landgrave, Michelangelo (2020) Can We Reduce Deception in Elite Field Experiments? Evidence from a field experiment with state legislative offices. *State Politics & Policy Quarterly* 20(4): 489–507.
- Magni, Gabriele and Zoila Ponce de Leon (2020) Women Want an Answer! Field experiments on elected officials and gender bias. *Journal of Experimental Political Science*. Epub ahead of print 10 August 2020. DOI: 10.1017/xps.2020.23.
- Malesky, Edmund, Paul Schuler and Anh Tran (2012) The Adverse Effects of Sunshine: A field experiment on legislative transparency in an authoritarian assembly. *American Political Science Review* 106(4): 762–786.
- Mendez, Matthew S and Christian R Grose (2018) Doubling Down: Inequality in responsiveness and the policy preferences of elected officials. *Legislative Studies Quarterly* 43(3): 457–491.
- Rhinehart, Sarina (2020) Mentoring the Next Generation of Women Candidates: A field experiment of state legislators. *American Politics Research* 48(4): 492–505.
- Thomsen, Danielle M and Bailey K Sanders (2020) Gender Differences in Legislator Responsiveness. *Perspectives on Politics* 18(4): 1017–1030.
- Wiener, Elizabeth (2020) Getting a High Heel in the Door: An experiment on state legislator responsiveness to women's issue lobbying. *Political Research Quarterly*. Epub ahead of print 9 July 2020. DOI: 10.1177/1065912920939186.