

Legitimacy Judgment Formation as Deliberation: Evidence from Experiments on Corporate Tax Avoidance

ABSTRACT

Research has conceptualized organizational legitimacy as a multi-level process comprising validity and propriety. Propriety refers to an individual evaluator's judgment that an organization or organizational practice is appropriate for its social context, whereas validity denotes an institutionalized, collective-level judgment of appropriateness within that context. Scholars have explored how propriety judgments form and change and focused on how propriety interacts with validity. However, research has applied a monological conception of legitimacy, ignoring the fact that propriety judgments are shaped in communication and social interaction processes. In response, we advance a dialogical conception and study the formation of propriety judgments in deliberations about corporate tax avoidance. Results of two experiments featuring direct deliberations with real representatives of a Big Four tax advisory firm and its critics and indirect deliberations via videos on social media confirm that deliberation affects propriety and offer important insights regarding the role of validity in this process. In particular, we explore the conditions under which deliberation leads to the (de-)legitimation of organizations and their practices, and we analyze the impact of peers, experts, and social media commentators in the formation of propriety judgments. We then discuss implications for research on legitimacy and outline practical implications for business firms.

INTRODUCTION

Since the seminal contribution by Meyer and Rowan (1977), research on organizational legitimacy has become a key feature of management and organization studies, particularly of institutional theory (Deephouse, Bundy, Tost, & Suchman, 2017). Research has focused on the legitimacy process at the collective level and maintained that organizational legitimacy emerges as an irreducible whole from social interaction and is thus largely “independent from particular observers” (Suchman, 1995: 576). However, scholars have also acknowledged that legitimacy “resides in the eye of the beholder” and thus involves the judgments of individuals (Ashforth & Gibbs, 1990, p. 177). In response, research has sought to integrate both viewpoints and suggested that legitimacy construal is a multilevel process of social judgment formation in which different components of legitimacy interact (Johnson, Dowd, & Ridgeway, 2006; Tost, 2011).

Advancing this line of research, Bitektine and Haack (2015) defined “validity” and “propriety” as the collective- and individual-level judgments in the legitimacy process. *Validity* refers to an institutionalized, collective-level judgment and describes the extent to which a group as a whole considers the nature or activities of a legitimacy object to be appropriate for its social context (Suchman, 1995). In turn, *propriety* is an individual-level judgment and denotes an individual evaluator’s endorsement of a legitimacy object and her assessment that the essence, qualities, or actions of that object are appropriate for its social context (Tost, 2011). The individual evaluator’s propriety judgment about a legitimacy object is shaped by his or her validity belief, which refers to the evaluator’s perception of whether a focal organization or practice is considered appropriate by others in the evaluator’s reference group, independent of the focal evaluator’s private endorsement of that object (Johnson et al., 2006; Suddaby, Bitektine, & Haack, 2017). Following these theoretical advancements, empirical research has started to investigate changes in propriety judgments (Finch, Deephouse, & Varella, 2015; Jahn, Eichhorn, & Brühl, 2017; Lamertz & Bhave,

2017), the promotion of propriety and validity in communication (Gauthier & Kappen, 2017), and cross-level interactions between validity and propriety (Haack & Sieweke, 2018; Huy, Corley, & Kraatz, 2014).

Recognizing that legitimacy can be described as a process of social judgment formation operating at multiple levels *including* the individual level has been a critical step in extending insights from previous research on organizational legitimacy as a collective-level phenomenon. A major problem with this novel stream of research, however, is that its conception of legitimacy is *monological* (Palazzo & Scherer, 2006; Scherer & Palazzo, 2007); that is, it takes the view that individuals form propriety judgments in isolation from others and operate as a “mental Robinson Crusoe” (Zerubavel, 1997, p. 7). The monological conception of legitimacy does not take sufficiently into account that propriety judgments are shaped in communication and social interaction processes. Moreover, it seems particularly ill-equipped to account for the ramifications of public debate on propriety judgments about controversial organizations and practices that evolve around “grand challenges” or corporate scandals (Ferraro, Etzion, & Gehman, 2015; George, Howard-Grenville, Joshi, & Tihanyi, 2016).

Due to the complexity and “wicked” character of these legitimacy issues, individual evaluators are often unable to judge the appropriateness of corporate behavior but depend on the validation and social instruction of others (Bitektine & Haack, 2015). However, in a globalized world the environment has become heterogeneous, making it difficult for evaluators to determine the appropriate source of validity (Scherer, Palazzo, & Seidl, 2013). In addition, the rise of social media as a heterogeneous and co-produced environment changes how propriety judgments about organizations are produced and disseminated in the public domain and challenges traditional assumptions of legitimacy judgment formation (Castelló, Etter, & Nielsen, 2016; Etter, Ravasi, & Colleoni, 2017). Specifically, while research has suggested that business firms defend or maintain

their legitimacy by actively engaging in public discourses vis-à-vis civil society (Palazzo & Scherer, 2006), it remains unclear whether and how such engagement affects the formation of propriety judgments. As face-to-face discussion is often not available to reach mass audiences and business firms instead increasingly engage in social media communication, it is important to explore the role of social media in the production of deliberative effects (Halpern & Gibbs, 2013).

We thus believe that future legitimacy research would benefit from an enhanced framework, which acknowledges that propriety judgments are actively shaped through *deliberation*, either face-to-face or on social media. In this paper, we accordingly advance a *dialogical* conception of legitimacy, a perspective that models legitimacy as a process of politics (Kostova, Roth, & Dacin, 2008) and moral reasoning (Scherer & Palazzo, 2007; Scherer et al., 2013), which is affected by communicative interaction (e.g., Gauthier & Kappen, 2017; Maguire & Hardy, 2009; Suddaby & Greenwood, 2005). Our specific research objective is to theorize and test whether deliberation, understood as “debate and discussion aimed at producing reasonable, well-informed opinions” (Chambers, 2003: 309), affects individual evaluators’ propriety judgments, contributing to the legitimation or delegitimation of organizations and their practices. Thereby we look at how validity cues, such as opinions stemming from experts, other deliberation participants and social media commentators, mediate the influence of deliberation on propriety.

Addressing this objective requires a fundamental shift in how we theorize about change in propriety judgments and how we test it. We suggest that experimentally testing the impact of deliberations is essential for further developing legitimacy research, as it facilitates “the empirical investigation of the negotiations and struggles underlying both the production of prevailing judgments and silencing of marginalized voices in legitimacy judgment formation” (Suddaby et al., 2017, p. 470). Drawing on the context of corporate tax avoidance (Dharmapala, 2014; Hanlon & Heitzman, 2010) and the facilitating role of the Big Four tax advisory firms (Addison & Mueller,

2015; Sikka & Willmott, 2013), we carry out two experimental studies to examine the formation of propriety judgments through deliberation, contrasting direct deliberation in a face-to-face setting with indirect deliberation via video on social media. Corporate tax avoidance, which refers to the pursuit of transactions and structures in order to reduce a business firm's tax responsibility in a manner that is contrary to the spirit of the law (Prebble & Prebble, 2010), has been described as a grand challenge (United Nations, 2015) and a matter of public concern (e.g., BBC, 2012; Hodge, 2016; Zucman, 2017). The context of corporate tax avoidance thus seems ideal for our theory development, as it involves a controversial yet ambiguous topic of high relevance (Addison & Mueller, 2015; Sikka & Willmott, 2013). The results of our experiments confirm that both direct and indirect deliberation have a significant impact on the formation of propriety judgments and may offer a valid (de-)legitimation strategy for organizational actors. Our findings also offer important insights regarding the question of under what conditions a given validity source (i.e., organizational representatives, other deliberation participants, or social media commentators) gains predominance in the judgment formation process and overrides the influence of validity cues stemming from other sources.

We make three main contributions to the field. First, we show the need and advance the notion of a dialogical perspective in legitimacy research that complements monological conceptions in extant research. Second, to our knowledge, our study is the first large-scale empirical test in management research to provide evidence for the notion that legitimacy is established and maintained through public deliberation (Palazzo & Scherer, 2006). In particular, we explore under what conditions deliberation leads to the legitimation or delegitimation of an organization and a controversial practice to which it contributes. Third, a narrow set of research methods has constrained theory development in legitimacy (Suddaby et al., 2017) and deliberation research (Bächtiger, Dryzek, Mansbridge, & Warren, 2018). Our use of deliberation experiments

under controlled conditions addresses this concern by allowing examination of cause-and-effect relationships in the formation of propriety judgments. Furthermore, we add a generalizable legitimacy scale to the literature that allows to measure propriety judgments about a business practice or organization.

THEORETICAL BACKGROUND

Legitimacy as a Multi-level Process

Organizational legitimacy is widely recognized as important for organizational growth and survival, given that a lack of legitimacy may lead constituents and resource holders to withhold material or ideational support (Aldrich & Fiol, 1994; Deephouse, 1999; Zimmerman & Zeitz, 2002). However, legitimacy has been applied to a plethora of theoretical and empirical contexts, which have added considerable surplus meaning to the construct (Deephouse et al., 2017; Suddaby et al., 2017). In response, Bitektine and Haack (2015) re-conceptualized the legitimacy construct as an intricate social process consisting of judgments at both the collective level (*validity*) and the individual level (*propriety*). Drawing on previous research, they stressed that the interaction between propriety and validity is complex. On the one hand, validity is the result of individual propriety judgments that institutionalized over time into a collective judgment of appropriateness (Bitektine & Haack, 2015; Tolbert & Zucker, 1996; Tost, 2011). On the other hand, validity shapes the formation of propriety judgments (Bitektine & Haack, 2015). Specifically, propriety judgments are likely to be influenced by whether an individual believes that the legitimacy object is perceived as appropriate by others in a collective of individuals—which is referred to as *validity belief* (Suddaby et al., 2017). To form such a validity belief, and ultimately a propriety judgment, individuals may attend to different sources of validity. We define a *validity source* as an actor or a group of actors signaling perceived appropriateness of a legitimacy object for a specific collective of individuals to which they belong.

Although it is widely acknowledged that there can be many “sources of legitimacy” or “legitimizing audiences” that apply different standards to assess organizations and their behavior (Aldrich & Fiol, 1994; Deephouse & Suchman, 2008; Kostova & Roth, 2002), the questions of why different evaluators view a legitimacy object differently and why the assessment criteria of legitimacy change over time (Huy et al., 2014) have remained largely underexplored. While in highly institutionalized settings the choice of the standard may be guided by the perceived validity of a specific source (Bitektine & Haack, 2015; Tost, 2011), exemplifying that validity beliefs influence or even determine propriety beliefs (Massey, Freeman, & Zelditch, 1997; Walker, Rogers, & Zelditch, 1988), we do not know how evaluators form propriety judgments when confronted with “multiple validities,” as for example in situations of institutional change or institutional complexity (Bitektine & Haack, 2015, p. 59). To tackle this question, we turn to research on deliberation to advance a *dialogical* conception of legitimacy judgment formation.

Legitimation-as-Deliberation

The roles of discourse, rhetoric, narratives, frames, and other communicative means in the process of legitimation have been highlighted since early research in this area (Berger & Luckmann, 1966), and various scholars have empirically assessed communicative legitimation strategies at both the organizational (e.g., Elsbach, 1994; Vaara & Tienari, 2008) and institutional levels (e.g., Green, Li, & Nohria, 2009; Hoffman, 1999; R. E. Meyer & Höllerer, 2010; Suddaby & Greenwood, 2005). Building upon these ideas, Palazzo and Scherer (2006) suggested that multinational corporations gain legitimacy in heterogeneous normative environments by actively participating in public deliberation vis-à-vis actors of civil society. Palazzo and Scherer (2006) describe this process as the politicization of the firm and develop a political conception of private actors, calling for a paradigm shift in the analysis of multinational corporations (Scherer & Palazzo, 2007, 2011;

Scherer, Rasche, Palazzo, & Spicer, 2016). Importantly, their legitimization-as-deliberation perspective emphasizes a *dialogical* conception of legitimation.

In practice, deliberations can take the form of “stakeholder dialogues” in which firms’ representatives directly engage in dialogues with critics and/or representatives of stakeholders to address issues of public concern (e.g., Ferraro & Beunza, 2018; Roloff, 2008). In addition, deliberations take place as debates in the public realm, mediated through traditional news media, such as newspapers, TV channels, broadcast channels, or news opinion websites (Habermas, 1989). More recently, social media platforms, such as Twitter, Facebook, and YouTube, offer a forum where experts, firm and civil society representatives, and laypeople interact. Within these deliberative processes, actors discuss issues of public concern, such as the acceptability of a corporation’s behavior, or explore who is responsible for the externalities of corporate strategies (Fung, 2003; Schouten, Leroy, & Glasbergen, 2012). These processes create and disseminate a variety of opinions which individual evaluators use as “validity cues” to render their own propriety judgments (Tost, 2011). Thus, in addition to allowing firms to learn about and handle various stakeholder demands, deliberation can be assumed to be a mechanism that helps evaluators to decide which validity cues to draw on when assessing organizational legitimacy.

HYPOTHESES

Based on these considerations, we develop a research model that blends the largely cognitive multi-level perspective of legitimacy (Bitektine & Haack, 2015) with the communicative perspective of “legitimation-as deliberation” (Palazzo & Scherer, 2006). This integration allows us to develop a dialogical conception of legitimacy judgment formation, a framework that satisfies the increasingly dynamic nature of firm-stakeholder relationships in the age of social media (Barnett, Henriques, & Husted, 2018; Etter et al., 2017) and grand challenges (Ferraro et al., 2015). In the following sections, we develop hypotheses regarding the impact of deliberation on the

interplay between validity and propriety. Specifically, we theorize on how competing validity sources (experts, other deliberation participants, and social media commentators) affect the formation of propriety judgments in face-to-face deliberations (Hypotheses 1-2) and in indirect deliberations via video on social media (Hypotheses 3-4).

The Effects of Experts

In line with deliberative theory, we suspect that the effect of deliberation on propriety judgments depends on the relative cogency of the exchanged arguments in favor vs. against the object of legitimacy (e.g., Habermas, 1996; Manin, 1987). In a deliberative setting, the “unforced force of the better argument” should prevail (Habermas, 1996: 306). Yet the quality of an argument is not “objectively” given as a sole property of the argument or the person making the argument, though it may be influenced by all of these factors (Habermas, 1984; Rehg, 2013). The assessment of an argument as good or bad by an individual evaluator is instead the result of deliberation as a social process, in which the pros and cons of an argument are considered together with its proponents and opponents. In this process, as we hypothesize, individuals not only attend to the content of an argument but may also (automatically or unknowingly) draw from different sources of validity, signaling the approval of a legitimacy object among a specific collective, in order to evaluate arguments and reach their propriety judgments.

A major source of validity that is prevalent in public debates are experts. The influence of experts on opinion formation in general is well-known and documented (Das, Kamruzzaman, & Karmakar, 2018; Moussaïd, Kämmer, Analytis, & Neth, 2013; Page, Shapiro, & Dempsey, 1987). Also in the context of deliberation, which involves moral reasoning and debate, the arguments of experts, in comparison to the arguments of laypersons, can be expected to be particularly effective regarding their impact on propriety judgments. For two reasons: First, their status as authorities and knowledge actors grants their arguments increased salience and thus attention (Mitchell, Agle,

& Wood, 1997; Parent & Deephouse, 2007). Second, because they possess expert knowledge, their arguments can be expected to be of high quality and to be more developed than those of laypersons (Bächtiger, 2017; Fung, 2006). Their role and expertise make experts a high credibility source, enabling them, according to the persuasion literature, to produce stronger attitude changes than low credibility sources (Briñol & Petty, 2009; Fishbein & Ajzen, 1975).

Taken together, it can be assumed that in a deliberative setting featuring experts and laypersons, the arguments of experts are more salient and in the center of attention. As a result, the perceived persuasiveness of experts is expected to drive the effects of deliberation on propriety.

Hypothesis 1. Experts (primary source of validity) determine the direction and size of the effect of deliberation on propriety via the (relative) cogency of their arguments.

Hypothesis 1a. If an expert or the expert with more convincing arguments argues in favor of the legitimacy subject, propriety will increase.

Hypothesis 1b. If an expert or the expert with more convincing arguments argues against the legitimacy subject, propriety will decrease.

Hypothesis 1c. The size of the effect of an expert on propriety is contingent on the (relative) cogency of the expert's arguments.

The Effects of Other Deliberation Participants

Another source of validity, which may affect propriety judgments, is the majority opinion of the deliberation group. Since deliberation participants are all in the same situation and share the same information and experience, they are likely to psychologically identify as an in-group *or peer group* (Terry & Hogg, 1996). Research on social influence has found that individual evaluators adjust their opinion towards the opinion of their peers (e.g., Moussaïd et al., 2013), either due to conformity preferences and to maintain a positive self-concept (Baddeley, 2013; Cialdini & Goldstein, 2004) or to reduce cognitive dissonance (Festinger, 1962; Groeber, Lorenz, &

Schweitzer, 2014). Importantly, the *perceived* rather than the actual majority opinion affects judgments and decision making (Taylor, 1982). The other members of a deliberation group may thus influence propriety judgments directly via individual evaluators' validity beliefs (i.e., their beliefs about the propriety judgments of others in the group; Bitektine & Haack, 2015).

If an expert joins the discussion, multiple types of validity sources (expert statements and group opinion) become available. To explain which source gains predominance and how they interact in the formation of propriety judgments, we mainly focus on two aspects: *source salience*, which refers to the extent to which a validity source stands out relative to others and draws attention, and *source identification*, which refers to the extent to which an individual identifies with the reference group represented by validity source. We assume that the more salient a validity source is, the more focal and relevant are the corresponding validity cues in the formation of propriety judgments (Fishbein & Ajzen, 1975). However, if a validity source does not represent a close reference group of an evaluator, that source and its cues may require validation by another, closer and thus more trusted source of validity. In line with the idea that social influence is stronger in situations of ambiguity or uncertainty (Cialdini, 1993; Cialdini & Goldstein, 2004), we assume that the less an evaluator identifies with a source of validity and/or the more controversial its validity cues (i.e., the more likely the cues are not in line with or contradict the opinion of an evaluators reference group of choice), the more evaluators will attend to alternative sources for both validation and further cues of validity.

Specifically, we suggest a model of validation where validity sources (e.g., expert or peers) work as "validation layers" that are hierarchically structured along the degree of identification of the evaluator with the reference group represented by a validity source. We posit that an evaluator pays attention to how lower-ordered layers of validation (i.e., validity sources representing less close reference groups) are observed and evaluated by higher-ordered layers of validation (i.e.,

validity sources representing more close reference groups). In turn, the reactions stemming from a higher-ordered layer of validation to validity cues stemming from lower-ordered layers may then validate or invalidate these cues, affecting the formation of propriety judgments. The more controversial a source of validity or its cues are (i.e., the less likely the signaled opinions align with the opinion of an evaluators reference group of choice), the more an evaluator will rely on this process of validation.

As elaborated above, expert statements are more salient than those of other deliberation participants. Experts are thus assumed to be the primary source of validity in a deliberative setting and the effects of deliberation are assumed to depend on their (relative) persuasiveness. However, the less evaluators identify with an expert or the more controversial his or her opinion, the more likely they will attend to the other deliberation participants (representing a closer reference group) for both validation and further validity cues. The persuasiveness of an expert can thus become a function of peer reactions during the deliberation.

Hypothesis 2. Other deliberation participants (secondary source of validity) affect propriety judgments directly as well as indirectly by validating signals of experts.

Hypothesis 2a. Other deliberation participants directly affect propriety judgments in the direction of the perceived majority opinion.

Hypotheses 2b. The (relative) persuasiveness of experts depends on the perceived majority opinion of other deliberation participants.

Hypothesis 2c. The relevance of other deliberation participants as secondary source of validity and layer of validation increases the more controversial an expert as primary source of validity is.

Scalability of Deliberative Effects via Video and the Effects of Social Media Commentators

In the following, we discuss a potential additional layer of validity—that of social media. Since direct, face-to-face deliberation with all stakeholders, sitting at the same table in small groups, is an exception rather than the rule in a modern society, it is important to explore the scalability of deliberative effects through social media. Social media have become a battleground where NGOs, corporations, and other actors fight to shape public opinion in the pursuit of social change. Indeed, social media is changing how judgments about organizations are produced and disseminated in the public domain, and these dynamics require research (Etter et al., 2017). Today, the phenomenon of sharing videos and initiating discussions over social media platforms such as Facebook, Twitter and YouTube seems ideally suited for exploring the scalability of deliberative effects in social media, because such platforms make it possible not only to reach a large variety of stakeholders (Hanna, Rohm, & Crittenden, 2011) but also to show actual footage of direct deliberations. Although viewers do not participate directly in the discussion, they are able to follow and evaluate the arguments and views expressed in the discussion. Therefore, this indirect participation (i.e., the observation of the deliberations of others) is likely to influence the viewers' judgments. Consequently, we expect to observe the same deliberative effects for indirect deliberation via video as for direct deliberation (see Hypotheses 1–2).

Hypothesis 3. The effects of direct deliberation on propriety translate to indirect deliberation via video.

Moreover, social media platforms offer users the possibility of participative and collaborative communication by means of one-to-many messages (Walther, Carr, et al., 2010; Walther, DeAndrea, Kim, & Anthony, 2010). The video-platform YouTube, for example, allows its users to express and share their own opinions in the comment section below a specific video. By writing, reading or liking comments, they are able to discuss the video content and also get an

idea of what other people think and of overall tendencies in opinion, creating a sense of social consensus (Shi, Messaris, & Cappella, 2014). Thus, social media comments also constitute a source of validity, which can be very powerful in general.

Social media users are likely to perceive each other as peers (Lee & Chun, 2016). The (relative) anonymity of users on such sites even fosters social identification with others on the basis of depersonalization and assumed similarity, leading to greater potential social influence by group members (Reicher, Spears, & Postmes, 1995). Moreover, social media users share perspective by experiencing content (e.g., a video) in the same way. Given that they share or access the same content on these platforms, they also are likely to share interests (e.g., because they follow the same site or channel posting a video). Social media users as peers are thus said to exhibit “optimal heterophily” (Rogers & Shoemaker, 1971; Walther, Carr, et al., 2010), which means being similar in most respects but superior in a component of credibility (e.g., experience or expertise), thus giving a source maximal persuasive ability (Rogers & Shoemaker, 1971). The notion that individual evaluators identify with anonymous users and are influenced by such sources has been supported by a number of studies in online environments (e.g., Edwards, Edwards, Qing, & Wahl, 2007; Sundar & Nass, 2001; Wang, Walther, Pingree, & Hawkins, 2008).

Indirect deliberation via social media (i.e., watching a video of direct deliberation on social media along with comments of other users) requires theorizing on the interplay between three different sources of validity: experts, other deliberation participants, and social media users. Importantly, participants in indirect deliberation via social media see the users on social media rather than the members of the deliberation group on video as peers, because they share the same experience as mere observers of the discussion (Lee & Chung, 2016). Therefore, social media comments are likely to have a bigger impact on the formation of propriety judgments than the statements of the deliberation group. It follows that, according to our model of validation, the social

media users represent the top layer of validation (i.e., the validity source representing the closest reference group). Thus, it can be expected that evaluators pay close attention to how social media users evaluate the statements of experts as well the reactions of the deliberation group members. In contrast to reactions of direct deliberation participants, the social media comments do not offer live feedback to a deliberation statement, but rather express an evaluation of the discussion overall. For this reason—and in line with past research showing that people’s perceptions vary according to social media comments’ valence (e.g., Park & Lee, 2007; Shi et al., 2014; Sung & Lee, 2014; Walther, Carr, et al., 2010) and that online users can affect attitudes more strongly than experts (e.g., Wang et al., 2008)—we expect that the deliberative effects on propriety judgments are affected by social media comments in the direction of the expressed opinions.

Hypothesis 4. Social media commentators affect propriety judgments in the direction of the valence of expressed opinions.

STUDY OVERVIEW

We conducted two deliberation experiments in order to test our hypotheses. Study 1 investigates how firm representatives (who defend a controversial business or practice) and NGO representatives (who criticize a controversial business or practice) entering direct, face-to-face deliberation affect propriety judgments about a controversial business practice (i.e., corporate tax avoidance) and organization (i.e., a Big Four firm), addressing Hypotheses 1a–1c. In this study, we also examine how the interaction between experts and other deliberation participants (representing different sources of validity) affect the formation of propriety judgments, addressing Hypotheses 2a–2c. For this purpose, we had participants deliberate either with a firm representative, with an NGO representative, with both, with no representative or not at all. The treatment manipulations thus consisted of whether deliberation took place and in what group compositions. While employing common features of face-to-face deliberation studies (Carpini, Cook, & Jacobs, 2004;

Fishkin, 2009; Fishkin & Luskin, 2005; Ryfe, 2005; Thompson, 2008), the experiment applies random treatment assignment to enable causal inference about the deliberative effects—an asset that is usually lacking in deliberation research (Karpowitz & Mendelberg, 2011). Study 2 examines whether indirect deliberation via watching videos on social media (YouTube) is an effective means to scale the effects of direct deliberation and how comments on social media affect propriety judgments, addressing Hypotheses 3–4. To compare the effects of different sources of validity (representatives, other deliberation participants, social media commentators), we showed participants videos of the statements of representatives alone, together with the reactions of other discussion members and/or positively or negatively valenced social media comments about the Big Four firm and the tax avoidance practice. The treatments varied along three dimensions: video content (deliberation in firm/NGO/mixed condition), video cut (full version/only expert statements), and social media comments (positive/negative/none). Figure 1 shows an overview of the different combinations of validity layers that we examine in the course of our two experiments.

[Insert Figure 1 about here]

STUDY CONTEXT: CORPORATE TAX AVOIDANCE

Corporate tax avoidance and the facilitating role of the “Big Four” (the four largest auditing and tax accounting firms: Deloitte, EY, PwC, and KPMG) form the context of our studies. These firms are important players in the tax avoidance field and are key actors in the global dissemination of the corporate tax avoidance practice. Corporate tax avoidance – or “tax optimization,” the term used by tax accountants – can be defined as the pursuit of transactions and structures in order to reduce a business firm’s tax responsibility in a manner that is contrary to the spirit of the law (Prebble & Prebble, 2010). These practices involve transfer pricing, profit shifting, or rulings that are negotiated between the corporations and tax authorities (OECD, 2013). The context of tax avoidance is ideal for our theory-building on the role of deliberation in legitimacy judgment

formation. The practice of corporate tax avoidance has increasingly come under public scrutiny following widely publicized media reports on several multinational corporations paying hardly any taxes (e.g., BBC, 2012; Hodge, 2016). While the news media in Europe and the US largely focused on tax-avoiding MNCs, such as Apple, Google, and Starbucks, several other news outlets, especially in the UK and Australia, critically discussed the role of tax accounting firms in tax avoidance (e.g., Hudson, Chavkin, & Mos, 2014; Syal, 2015; Toynbee, 2013). Such outlets have referred to the Big Four as facilitators, enablers, or even “masterminds” of tax avoidance, portraying them as the architects of tax schemes that cost governments and taxpayers more than US \$1 trillion a year (e.g., West, 2016), including US \$160 billion in the US alone (Zucman, 2015). Thus, the topic of tax avoidance is clearly controversial and of high relevance (Addison & Mueller, 2015; Dowling, 2014; Sikka & Willmott, 2010, 2013). Furthermore, since the role of the Big Four in tax avoidance has been scarcely addressed by the media in Switzerland – where our experiments were conducted –, the salience is not too high, and participants can be expected not to hold fully informed or firmly determined positions *ex ante*, a critical feature for successful deliberations (Dryzek & Pickering, 2017; Farrar et al., 2010). Although the topic is fairly complex and allows for various viewpoints, the key issues and arguments of critics and advocates of corporate tax optimization and the role of the Big Four can easily be understood and discussed, making this a debatable and ambiguous topic.

STUDY 1: DIRECT DELIBERATION WITH REPRESENTATIVES

The experiment was conducted on two consecutive afternoons at the end of January 2017 at a public university in Switzerland (anonymized for review purposes). A total of 293 participants (43.7% female, 94.5% students, mean age of 25.3 years) took part in the study. The experimental sessions lasted 105 minutes, and participants were paid the equivalent of 80 USD.

Methods

Study Procedure. Participants were recruited via email, online marketplaces and social media forums. The participants could sign up for one of ten time slots (experimental sessions). Before signing up, participants learned about the topic of the study (corporate tax avoidance and the role of the Big Four), that they would receive information material about the topic at the beginning of the study, and that they would subsequently be asked to discuss the topic in small groups.

The experiment consisted of two parts. In the first part, all participants of an experimental session were welcomed and instructed in the same room. They received information material on the topic and filled out a first questionnaire after reading the material. To ensure that participants read the information material and understood some key aspects before taking the questionnaire, they had to answer two control questions correctly before they were given the questionnaire. The information material was prepared in collaboration with Big Four firms and NGOs and aimed to inform the participants on the matter in a neutral way. It contained an introduction to the topic, a balanced number of arguments in favor of and against corporate tax optimization, a summary of the main points of criticism about the role of the Big Four, respective counterarguments by proponents of the Big Four, and the following three open questions for the group discussion in the second part of the experiment: (1) “Is corporate tax optimization problematic or desirable (with regard to the economy/society/government)?” (2) “Who is responsible for the consequences of tax optimization (e.g., consultants, clients or governments)?” and (3) “How do you evaluate the criticism of the Big Four (e.g., conflicts of interest between auditing and consulting, conflicts of interest when advising law makers, power imbalance of consultants vs. regulators, operating within a grey area of legality, issues of transparency, call for splitting up the Big Four)?” The questions

were chosen to ensure that aspects pertinent to the debate on the legitimacy of tax optimization and the role of the Big Four would be addressed in the deliberation.

During the second part, participants discussed these questions in small groups of seven in separate rooms for 45 minutes. Afterward, they answered a second questionnaire. Depending on the treatment, the discussion groups either were or were not joined by representatives of the Big Four and/or of its critics (i.e., a critical NGO). The discussion groups and the treatments were randomly assigned by letting participants pick cards with imprinted room numbers (“room cards”) at the beginning of the experiment, determining in which room they would participate in the second part of the experiment. Participants were not aware that the different discussion groups had different treatment conditions. The experiment was carefully planned such that participants also did not learn about the potential presence of experts in the discussions, so as to prevent experimenter demand effects or potential frustration if some participants learned that others’ discussions included experts who were not present in their own group. At the end of the experiment, participants learned about a voluntary follow-up survey a week later and were instructed not to speak to participants of other discussion groups about the study until they had taken this survey. The follow-up survey was mainly intended to test for longevity of potential treatment effects. After the discussions and completion of the second questionnaire, the participants received their payment.

Treatment Groups. There was one control group and four different treatment groups (i.e., five rooms) in each session that took place simultaneously: *control condition*, *firm condition*, *NGO condition*, *mixed condition*, and *wait condition*. The first four conditions (discussion treatments) only varied in terms of the participation of firm and NGO representatives as external experts during the discussion. The number of participants in the discussion groups was fixed at seven. For efficiency reasons we chose a group size at the lower end of the spectrum compared to other deliberation studies (cf. Mansbridge, Hartz-Karp, Amengual, & Gastil, 2006). In the “control

condition,” the group of seven participants was not joined by an expert. In the “firm condition,” the seven participants were joined by a representative of the Big Four (a corporate tax specialist working for a Big Four firm that collaborated with us for the experimental study). In the “NGO condition,” the seven participants were joined by a representative of the critics of the Big Four (a member of a Swiss NGO that is critical of tax avoidance). In the “mixed condition,” the participants were joined by both a representative of the Big Four and a representative of the NGO. The “wait condition” was a treatment without discussion. Participants assigned to this condition simply waited for 45 minutes in a separate room (instead of discussing the topic for 45 minutes in small groups) and then filled out the second questionnaire. This condition allowed us to check whether the passage of time alone would lead to systematic changes in opinions on the topic (e.g., by allowing the participants to think more about the topic), and it also fulfilled a pragmatic purpose. Since we wanted to ensure that all groups in the four discussion treatments would consist of seven participants, we recruited an additional three participants per session to account for potential no-shows (i.e., 31 instead of 28 participants total). If more than the needed 28 participants in a session showed up, we added the appropriate number of additional room cards corresponding to the wait condition when assigning treatments.¹

Form of Discussion. Each of the four discussion groups in a session was joined by a neutral moderator who managed the process of discussion but who neither knew the goals of the experiment nor took part in terms of content. The moderators were research assistants at the department of business administration of our home university and were instructed to ensure that the rules of the discussion were adhered to. The rules of discussion, based on those applied in

¹ In one session, there were less than 28 participants, such that there were only two participants engaged in discussion in the control condition. We deleted this group from the analysis and recruited new participants to repeat the control condition of this session a week later. All effects are robust when repeating the analyses with the full sample or without the extra session we conducted a week later.

deliberative polling (Fishkin & Luskin, 2005), were intended to make the conditions of the discussion approximate the “ideal speech situation” (see Habermas, 1984, 1993), so as to support the positive effects of deliberation. The discussion was led as an essentially free conversation among the participants, in which the moderator managed the time, opened and closed the discussion questions included in the information material, and made sure that no one participant either dominated the conversation or refrained from talking entirely.

To incorporate the firm and NGO representatives in the treatments in which they were present, they were asked to make an introductory statement of two minutes for each question. If two experts were present, both made an introductory statement for each question in alternating order (between questions and sessions). Since the participants’ discussions in the different treatment groups were held simultaneously, two representatives of the Big Four and two NGO representatives needed to be available. The representatives varied between the two days of the study for both parties, such that a total of four representatives of the Big Four and four representatives of its critics participated in the study. To ensure that the content was identical within a group of representatives, they were asked to prepare common introductory statements and to discuss the questions listed in the information material with their team ahead of time so as to agree on a joint line of argumentation. To support a coherent perception of representatives of each team, the moderators introduced them to the discussion groups using prepared descriptions that were scripted to be similar. Since individual differences between representatives of either team cannot be ruled out, the representatives switched treatments between sessions to counterbalance potential effects resulting from such individual differences. Moderators also switched treatments between sessions to counterbalance potentially occurring interdependencies. Between the two days of the study, the different rooms were also assigned to different treatments. Table B1 in the appendix shows the assignment of representatives and moderators to treatments.

The moderators closed the discussions after 45 minutes, asked any experts present to leave the room, and asked participants to fill out the second questionnaire. All discussions were videotaped, enabling us to make sure everything had gone as planned.

Dependent Variables. Our main dependent variable is propriety judgment, defined as an individual evaluator's endorsement of a legitimacy object and her assessment that the essence, qualities, or actions of that object are appropriate for its social context (Tost, 2011). In order to assess propriety judgments about the legitimacy of the Big Four and of corporate tax avoidance separately, we constructed our own legitimacy scale. Although different measures were already existing in the literature at that time, they were context-dependent and thus not suitable for the context of tax avoidance (e.g., Chung, Berger, & DeCoster, 2016; Elsbach, 1994; Foreman & Whetten, 2002; Jahn et al., 2017). More recently, Alexiou and Wiggins (2018) as well as Bitektine, Hill, Song, and Vandenberghe (2018) published more generalizable scales to measure organizational legitimacy. However, their items are either difficult to apply to measure the legitimacy of an organization and one of its practices separately, as we intend to do, or they mix statements relating to propriety and validity (e.g., "The general public would approve of this organization's policies and procedures", Alexiou & Wiggins, 2018, p. 11).

Similar to Alexiou and Wiggins (2018) we based our scale on the three dimensions of legitimacy as identified by Suchman (1995): *pragmatic legitimacy* (is it beneficial?), *cognitive legitimacy* (is it taken for granted?), and *moral legitimacy* (is it the right thing to do?). Yet we made sure to only include statements relating to propriety, statements that can easily be applied to measure either a business organization or one of its practices, and we further divided the legitimacy dimensions into subdimensions to account for related key components, which also loaded on separate factors in exploratory factor analyses.

The details regarding the scale construction (formulation and selection of items, exploratory and confirmatory factor analyses, etc.) can be found in Appendix A. Table 1 shows the final set of items, the Cronbach's alphas for the (sub)dimensions as well as the factor loadings of a confirmatory factor analysis, for which we conducted an online survey with 175 students. All items were rated on an 11-point Likert scale (-5=totally disagree, 0=neither agree nor disagree, +5=totally agree). While our scale was found to significantly correlate with the recently published scales measuring legitimacy of Alexiou and Wiggins (2018: $r=.80$, $p<.001$) and Bitektine et al. (2018: $r=.61$, $p<.001$), our scale showed a higher reliability (higher Cronbach's alpha within as well as over the different legitimacy dimensions; our scale: $\alpha=.93$; Alexiou & Wiggins, 2018: $\alpha=.83$; Bitektine et al., 2018: $\alpha=.56$) and a better overall goodness of fit (our scale: $\chi^2_{75}=119.90$, $p<.01$; RMSEA=.058; CFI=.976; SRMR=.038; Alexiou & Wiggins, 2018: $\chi^2_{51}=83.52$, $p<.01$; RMSEA=.060; CFI=.958; SRMR=.065; Bitektine et al., 2018: $\chi^2_8=7.2$, $p=.51$; RMSEA=0; CFI=1; SRMR=.044).

[Insert Table 1 about here]

In addition to comparing the changes in the above-described legitimacy scale before and after the discussions, we also wanted a more explicit measure of changes in participants' propriety judgments. Therefore, in the second questionnaire (after the discussion), we asked participants to indicate whether their judgments regarding tax optimization or the Big Four had changed during the discussion and to what extent, again using an 11-point scale (-5=worse, 0=same, +5=improved).

Controls. The participants in both the mixed and single treatment conditions rated the representatives after the deliberation regarding their power to convince ("The representative was convincing") and their expertise ("The representative was competent"), again indicating their

agreement on an 11-point Likert scale. To create a general measure for the persuasiveness of representatives, we averaged the ratings of these two items.

As our hypotheses for the mixed conditions crucially depend on the relative cogency of the representatives' arguments, we also explicitly asked the participants in the mixed condition after the discussion, "Whose arguments were more convincing, those of the employee of [name of the company] (Big Four) or those of the member of [name of the NGO] (critics)?" This question was rated on an 11-point scale (0=NGO, 5=both equally, 10=Big Four).

The initial position of participants is likely to be an important factor in the potential effects of deliberation, for both effect size and direction. In order to be able to control for participants' attitudes, which we define as their pre-existing opinions about the focal legitimacy object, we asked participants to indicate their attitude towards the Big Four and corporate tax optimization on an 11-point scale (-5=contra, 0=neutral, +5=pro). We also asked participants to rate the position of other members of the discussion group (after the discussion), using the same scale, to be able to control for the majority opinion of the group as perceived by the individual evaluator (a potential validity cue).

To control for other individual characteristics, we assessed demographic measures, the participants' left/right political orientation (as usual for studies in political science; e.g., Isernia & Fishkin, 2014), and other individual characteristics (e.g., the Big Five personality traits, cf. Gerlitz & Schupp, 2005); the latter are not listed here in detail, as they did not prove to be important for the focus of our study.

Design and Analysis. To check for treatment effects on the main dependent variable (i.e., our legitimacy scale), the study uses a difference-in-differences design, comparing the within-subjects observed changes in legitimacy ratings between the first and the second questionnaire between participants of different treatments.

For comparisons between treatments, we treat each discussion group and each participant in the wait condition (who did not interact) as one independent observation. To test our hypotheses, we divide participants in the mixed condition into three subgroups, depending on whose arguments the participants indicated to be more convincing after the discussion (control measure for relative performance of representatives). Participants who indicated that the firm representative was more convincing form the subgroup “mixed advantage firm”; participants who indicated that the NGO representative was more convincing form the subgroup “mixed advantage NGO”; and participants who indicated both representatives were equally convincing form the subgroup “mixed tie.”

Results and Discussion

Treatment Effects on Legitimacy Judgments. Table 2 summarizes the mean values of propriety judgments about the Big Four (cf. organization) and corporate tax optimization (cf. practice) measured in the questionnaire before (cf. Q1) and after (cf. Q2) the discussions (as well as in the follow-up survey; cf. Q3)² for the different treatments. The table also displays the mean changes in propriety judgments, given by the differences in the legitimacy ratings (cf. implicit change) and the explicitly indicated changes in opinion in the second questionnaire (cf. explicit change), and reports significance levels regarding whether the means are different from zero. In the mixed condition, 18 participants (26%) thought the firm representative was more convincing (cf. “mixed adv. firm”), 40 participants (57%) thought the NGO representative was more convincing (cf. “mixed adv. NGO”), and 12 participants (17%) perceived both as equally convincing (cf. “mixed tie”). Before turning to the testing of our hypotheses, it should be noted that the discussions in the control condition tended to decrease propriety and led to a significant

² As not all participants took the follow-up survey and not all effects reached statistical significance, we omit the follow-up survey from further analyses. However, the results of the follow-up survey show the same patterns as the findings for the main study, indicating longevity of effects.

decrease in the propriety of the Big Four. As we do not observe significant changes in propriety judgments for participants in the wait condition, it seems that deliberation – as opposed to the mere passage of time (in which participants could think about the topic) – indeed affects legitimacy judgment formation.

[Insert Table 2 about here]

In order to test our hypotheses, we compared the treatment effects on changes in propriety judgments to the control condition (i.e., baseline). Figure 2 shows these marginal effects of the different treatment groups for our two measures of changes in propriety (implicit measure via legitimacy scale and explicit measure via direct question about change in opinion). According to Hypotheses 1a and 1b, we expected a positive marginal effect in the firm condition (i.e., we expected the change in propriety to be more positive in comparison to the control condition) and a negative marginal effect in the NGO condition (i.e., we expected the change in propriety to be more negative in comparison to the control condition). In line with these hypotheses, the presence of a firm representative in the firm condition tends to lead not only to more positive changes in propriety compared to the control condition (i.e., baseline) but also to an overall increase in propriety (cf. Table 2). Analogously, the presence of an NGO representative in the NGO condition not only tends to lead to more negative changes in propriety compared to the control condition but also to an overall decrease in propriety (cf. Table 2). The pattern in the mixed condition, as expected by Hypotheses 1a–1b, critically hinged on which representative was perceived as more convincing. In line with Hypothesis 1a, the effects of the mixed condition do not significantly differ from those of the firm condition for the participants who perceived the arguments of the firm representative to be relatively more convincing (“mixed adv. firm”). Analogously, the effects of the NGO condition are in line with the observed changes for participants in the mixed condition who perceived the NGO representative to be more convincing (“mixed adv. NGO”), supporting Hypothesis 1b. For

the participants who perceived both representatives to be equally convincing (“mixed tie”), the results did not significantly differ from the control condition, overall supporting Hypothesis 1.

[Insert Figure 2 about here]

Interestingly, the effects of the two types of representatives differ with regard to the type of legitimacy object. That is, firm representatives had a relatively stronger impact on the perception of the organization than on that of the practice, whereas the reverse was the case for the NGO representatives. As can be seen in Figure 2, firm representatives or their relative performance in mainly affected the propriety judgments about the organization, whereas they had less of an effect on the propriety judgments about the practice. The NGO representatives, in contrast, affected the propriety judgments about the practice more strongly than those about the organization. The relative importance of representatives is also evident in the mixed condition as a whole, where the marginal effect on the organization tends to be positive and the marginal effect on the practice tends to be negative.

Validity Cues and Their Effect on Legitimacy Judgment Formation. To investigate how the different sources of validity (experts and peers) affect the formation of propriety judgments, we regressed individual changes in propriety for the different treatments on a participant’s perception of the position of the other discussion members (cf. effect of peers) and the participant’s rating of the persuasiveness of the representative(s) (cf. effect of expert), controlling for the individual’s initial position towards the organization and the practice, political orientation and session (discussion group). As will be argued below, political orientation and individuals’ initial position are crucial control variables. Table 3 shows the results. For the control condition, there is a highly significant positive relationship between the perceived position of others towards the organization or the practice and changes in the corresponding legitimacy measure. For example, the more approving (disapproving) an individual thinks the position of others is towards the

organization or practice, the more positive (negative) will be the changes in propriety judgments. Thus, in the absence of experts, peers seem to function as the primary source of validity and affect propriety judgments directly, confirming Hypothesis 2a. For treatments featuring representatives, the perceived persuasiveness of the representatives clearly drives the hypothesized deliberative effects, whereas the relevance of peer evaluations for propriety judgments only shows in a few exceptions in this case. This again confirms Hypotheses 1, according to which the organizational representative becomes the primary source of validity, and Hypothesis 1c, according to which the size of the effect of deliberation depends on the (relative) cogency of the representative's arguments. More concretely, the relevance of peer evaluations in the presence of a representative only shows for changes in propriety of the Big Four and only when a firm representative is present. According to our theory, this implies that the expert's validity cues are not trusted by the participants in this case and need to be validated by attending to peer reactions. In line with Hypothesis 2c, peer reactions as a secondary source of validity are more relevant in the presence of a firm (vs. NGO) representative, as firm representatives were apparently a more controversial source of validity than NGO representatives.

[Insert Table 3 about here]

As assumed by Hypothesis 2, the relevance of peers thereby not only shows via affecting propriety judgments directly, but also by affecting the evaluation of the expert's persuasiveness. Table 4 displays the results of regressing the persuasiveness of representatives on the perceived position of other deliberation members, controlling for the session (i.e., deliberation group), the participant's political orientation and initial position towards the organization and practice (see next section). In line with Hypothesis 2b, the persuasiveness of the firm representative and the performance of the firm representative against the NGO representative in the mixed condition are functions of the perceived majority opinion (i.e., how individuals perceive the position of other

discussion members towards the Big Four). The persuasiveness of the NGO representative, on the other hand, is not contingent on peer reactions. This again confirms Hypothesis 2c, given that representatives of a publically criticized firm appear to be a more controversial source of validity than representatives of an NGO.

[Insert Table 4 about here]

Robustness Checks. Since the participants in the mixed condition were divided into subgroups depending on which representative they indicated to be more convincing, the results for these subgroups may be driven by individual differences that are exogenously given rather than created by deliberation. This means, for example, that participants who perceived the firm representative as more convincing could differ from other participants, which may explain the observed differences in perceived legitimacy.

To control for this, we first assessed whether any variables strongly predict whether a participant will find the representative of the firm or the representative of the NGO more convincing. Corresponding logistic regressions, shown in Table B2 in the appendix, capture the two best predictors in our dataset: the participants' left-right political orientation and initial position towards the Big Four. Accordingly, participants are significantly more likely to perceive the firm (NGO) representative as more convincing if their political orientation is more rightist (leftist) and their initial position toward the Big Four is more favorable (critical).

When controlling for those individual differences, the treatment effects for the mixed condition subgroups (mixed adv. firm/NGO) remain significant, as captured by Table B3 in the appendix. Hence, which representative is perceived as more convincing as well as the corresponding marginal effects cannot be explained by exogenous differences in individuals alone; rather, these variables are endogenous and therefore depend on the deliberation itself.

STUDY 2: INDIRECT DELIBERATION THROUGH SOCIAL MEDIA

Our second experiment was conducted online during August and September 2018 with a total of 472 participants (53.6% female, 97% students, mean age of 23 years). The experiment lasted about 90 minutes, and participants were paid the equivalent of 45 USD.

Methods

Study Procedure. Participants for the online survey were again recruited via email, online marketplaces and social media forums. The survey was set up as follows. First, the participants were required to enter some personal information, along with their postal address to receive the payment. Requiring a Swiss residence allowed us to ensure that all participants were embedded in the same economic and political context. Next, participants were asked to read the information material from Study 1. To induce reading the material, a timer was set, and participants were only able to continue after viewing the material for a minimum of 10 minutes. To further encourage reading the material, participants were told upfront that they would have to answer two control questions correctly after reading the material before they would be allowed to continue. After correctly answering the questions, participants were directed to the first questionnaire, indicating their initial propriety judgments about corporate tax avoidance and the Big Four company. Next, participants were shown one of several videos of the small group discussions of Study 1 on YouTube, which varied according to the treatment conditions (see below). Watching the video was again encouraged by a timer of corresponding length, which had to expire before the participants were allowed to continue. After the video, participants were directed to the second questionnaire. Finally, participants were asked to join the conversation on YouTube and to write at least one comment below the video. For this part, the participants were directed to a YouTube link covering the same video material, but featuring different settings regarding the comment section, such that an open discussion among participants was possible.

Treatment Groups. Participants were randomly assigned to one of 13 treatment groups corresponding to 13 YouTube videos. Treatment groups (i.e., videos) varied along three dimensions: (1) taped discussion group and condition of Study 1 (firm/NGO/mixed condition), (2) video cut (full version/only expert statements), and (3) YouTube comments (positive/negative/none). These dimensions are discussed further below.

(1) The video material of the discussions in Study 1 formed the basis for the YouTube videos in Study 2. We showed participants a discussion in either the firm (“F”), NGO (“N”) or mixed (“MIX”) condition of Study 1. As we recorded all sessions in Study 1, we were able to select from among all of the discussion group sessions. To investigate whether indirect deliberation via video can replicate the effects of direct deliberation, we wanted to pick discussion groups where the representatives performed best and the effects were the clearest. Thus, for both the firm and NGO condition, we selected the group where the representative was rated the most persuasive and the effects on legitimacy judgment formation were the strongest (see Table B4 in the appendix). For the mixed condition, we chose the discussion that included the same representatives, so as to prevent potential biases relating to individual differences between experts. In this particular group, the firm representative happened to be perceived as more persuasive, and the discussion led, on average, to an increase in propriety (for more details, see Table B4).

(2) To test whether and how seeing the reactions and statements of other deliberation members matters compared to only seeing statements of the expert, we prepared two different video versions for the deliberations in the firm and NGO condition. Accordingly, we prepared a long (“L”) and a short (“S”) video version of the deliberations. The long video version showed the full discussion, while the short video version showed only the statements of the expert, including the opening statements to the three official discussion questions as well as their answers to questions and comments of other deliberation members. All reactions of deliberation members were edited

out. However, to clarify the questions or statements to which the experts were referring in the short video versions, the relevant question or comment was presented in the video as text right before the statement as well as during the statement. In the short video versions, we also blurred all parts of the video except for the expert, to make sure there were no cues regarding how other deliberation members reacted to the expert's statements (e.g., facial expressions). For both video versions, we beeped out the names of all deliberation members as well as the names of the organizations of the representatives (to protect the identities of these people in case of unauthorized distribution of the videos). However, survey participants were informed upfront whom the experts were representing and for whom they worked so that the participants would have the same knowledge as the participants in the first experiment.

(3) The videos were shown via the social media platform YouTube. To make sure only the participants would view the videos, they were “not listed” and were only accessible via the URL provided to participants. Below the videos, there were either no comments or comments exclusively in favor of (“positive”) or exclusively against (“negative”) the organization and practice. In the positive and negative conditions, there were nine comments regarding the organization and the practice (18 in total). The participants learned that the comments stemmed from former participants but were reposted as quotes by the study administrator to ensure anonymity. How the comments were gathered and selected is described below. The order of comments was fixed across treatments (the comments regarding tax optimization were presented before the comments regarding the Big Four, matching the sequence of the discussion). The video settings on YouTube ensured that the comment section looked the same for all survey participants within a treatment and that potential comments of other survey participants were not visible.

Based on the discussion selected from the firm condition and NGO condition, there were five treatments each. We presented some participants with the full-length video with no comments

(FL, NL) and some participants with the short video with no comments (FS, NS) to test whether and how seeing the reactions and statements of other deliberation members matter compared to only seeing statements of the expert. To test whether the comments may enhance the effects of the expert, we also presented the short video version along with positive comments in the case of the firm representative (FS+) or negative comments in the case of the NGO representative (NS-). Analogously, to test whether the comments may dampen or even reverse effects, some participants were shown a short video version with negative comments in the case of the firm representative (FS-) and positive comments in the case of the NGO representative (NS+). Finally, to identify the effectiveness of the social media comments when counteracting both the expert's statements and the reactions of other members of the deliberation group, we also presented a long video version along with negative comments for the firm representative (FL-) or positive comments for the NGO representative (NL+). For the mixed condition, we only used a full-length video version that was presented either with positive comments (MIXL+), with negative comments (MIXL-) or without (MIXL) comments. The purpose here was to find out how effective social media comments are when presented with a full and balanced discussion, integrating experts of both sides as well as "neutral" participants. Table 5 gives an overview of the treatment groups.

[Insert Table 5 about here]

Social Media Comments. The comments that were posted as quotes below the videos in the YouTube comment section by the study administrator were gathered from two sources: a first wave of data collection for Study 2 and the follow-up survey from Study 1. Trying to gather positively or negatively valenced comments as authentically as possible, we ran the FL and NL treatment as described above (with at least 60 participants each). That is, participants saw the full-length deliberation either with the firm representative (in advance to generate positively valenced comments) or with the NGO representative (in advance to generate negatively valenced comments)

and were free to discuss the topic via YouTube comments (below separate videos) afterwards. In addition, we asked participants in the follow-up study of Study 1 to write a comment regarding tax optimization and a comment regarding the Big Four and let them decide whether or not they wanted those comments to be published on the website of our department. If so, we added them to the pool of comments to choose from, as the participants made the choice to share them (similar as when posting comments on social media). As a result, we selected nine comments that expressed a clear position in favor of or against the Big Four and nine comments that were in favor of or against tax optimization (36 in total).

Dependent Variables. We used the same dependent variables as in Study 1. Thus, to assess implicit changes in propriety judgments about the organization and practice, we again employed the legitimacy scale (cf. Table 1) before and after watching the YouTube videos (treatments). To assess explicit changes, we again asked survey participants after watching the YouTube videos whether and to what extent their opinion regarding tax optimization or the Big Four had changed as a result of the discussion as a whole, using the same 11-point scale (-5=worse, 0=same, +5=improved).

Controls. We applied the same control variables as in Study 1. Additionally, in order to assess how participants themselves rated the relevance of different sources of validity, we asked them how important the statements of the present representative(s), the reactions of the deliberation group and the social media comments were for them in forming an opinion on tax avoidance or the Big Four; again using an 11-point Likert scale to indicate their agreement.

Results and Discussion

Figure 3 shows the treatment effects on the implicit and explicit measures of changes in propriety of the organization and the practice. The first thing that should be noted about these results is that the general treatment effects of the first study are replicated when participants only

watched a video of the deliberation rather than took part in the deliberation. That is, watching the video with the firm (NGO) representative led to an increase (a decrease) in propriety. Thus, indirect deliberation via video seems to work, confirming Hypothesis 3. Interestingly, the effects were stronger when participants were only shown the statements of the representatives instead of the whole discussion (cf. FL vs. FS; NL vs. NS). The reason for this might be that participants who watched the short version of the video only heard expert statements in favor of or against the organization or the practice in a concise manner, whereas participants that watched the full version also heard more moderate or also opposing opinions. Regarding Hypothesis 4 on the impact of the comments on social media, the results are striking. The comments not only boosted the effects of experts if they were aligned (cf. FS+, NS-) but also were able to completely reverse the representative's effect if they were contradictory (cf. FS-, FL-, NS+, NL+). Moreover, for the mixed condition, the position of the comments dictated the direction of the effect (cf. MIXL+ vs. MIXL-). In line with the idea that individuals attend more to peer reactions in presence of controversial sources, the effects of social media comments seem slightly more pronounced in treatments featuring a firm (vs. NGO) representative.

[Insert Figure 3 about here]

Additionally, we looked at whether the social media comments also affected the open deliberations on social media at the end of the experiment. Watching a video with positive (negative) comments indeed tended to increase the share of positively (negatively) valenced comments about the Big Four or tax optimization posted by participants (see Figure C1 the appendix). In sum, watching a video along with substantial comments (positive or negative) not only influenced legitimacy judgment formation but even set the tone for the open online discussions.

Intriguingly, this influence of social media comments seems to be largely unconscious to the individual. Our results suggest that participants generally underestimate the relevance and

power of social media comments in forming their judgments. When they were asked how important the different validity sources (expert statements, discussion group, comments) were to them in forming an opinion regarding the Big Four or tax optimization, they rated the comments least important overall, as shown in Table 6.³ Comparing the effects on the implicit vs. explicit measure, the participants seem to particularly underestimate the effects of social media comments in favor of the organization or practice (cf. treatments with positive comments in Figure 3).

[Insert Table 6 about here]

DISCUSSION

Drawing on the multi-level perspective of legitimacy (Bitektine & Haack, 2015) and the legitimization-as-deliberation-thesis (Palazzo & Scherer, 2006), we develop a dialogical conception of legitimacy judgment formation and investigate whether and under what conditions entering deliberation may serve as a (de-)legitimation strategy for organizations. For this purpose, we conducted two experiments with real firm and NGO representatives in the context of corporate tax avoidance. The first study confirms the deliberation-as-legitimation-thesis by showing that entering direct, face-to-face deliberation offers an effective strategy for legitimation or delegitimation. Our findings provide empirical evidence that direction and size of deliberation effects are contingent on the persuasiveness of the representatives' arguments through which the propriety judgments of individuals are influenced. The persuasiveness of firm representatives, however, is a function of the perceived opinion of other discussion members, indicating that the validity cues stemming from firm representatives were less trusted than those of NGO representatives and needed to be validated by attending to peer reactions. In the second study we found that indirect deliberation via watching

³ Note that in line with our observations in Study 1, the importance of the firm representative was, on average, rated higher for forming a judgment about the organization than the practice and vice versa for the NGO representative. The differences were, however, not significant.

videos on social media (YouTube) is an effective means to scale the effects of direct deliberation. Furthermore, our results highlight that comments on social media were especially effective in influencing propriety judgments, to the extent that they were even able to override and reverse the effects of the arguments from representatives and other discussion members. We now elaborate the study's main contributions to theory, discuss its practical limitations, and highlight several avenues for future research.

Theoretical Contributions

We make three main contributions. First, we advance recent literature on how individual evaluators render their propriety judgments (Bitektine, 2011; Finch et al., 2015; Jahn et al., 2017; Tost, 2011). This literature has complemented legitimacy research, because individual judgments formation is related to the social construction of legitimacy at the collective level (Bitektine & Haack, 2015; Johnson et al., 2006). However, due to its “monological” conception this perspective has ignored the role of communication and argumentation in the legitimation process (Palazzo & Scherer, 2006). Consequently, the studies on individual judgment formation have overlooked the fact that legitimacy is inherently political (Kostova et al., 2008) and grounded in the collective understanding of issues (Hoffman, 1999; R. E. Meyer & Höllerer, 2010), which is affected by rhetoric, narratives, discourse, and other communicative means (Maguire & Hardy, 2009; Suddaby & Greenwood, 2005). These dynamics are particularly important in the context of social media, which exposes evaluators to a multiplicity of judgments and enable evaluators to collectively engage in the *co-production* of judgments on an ongoing basis (Castello, Etter, & Nielsen, 2016; Etter et al., 2017; Glozer, Caruana, & Hibbert, 2018).

Our study allows us to see legitimacy judgment formation in a new light, as the outcome of a deliberation process among evaluators and other actors discussing a controversial organization or practice. The conception of legitimacy envisioned here is thus *dialogical* and corresponds to the

legitimation-as-deliberation-perspective developed in business ethics (Palazzo & Scherer, 2006; Scherer & Palazzo, 2007; Scherer et al., 2013, 2016). A dialogical conception demands a fundamental shift in how we theorize and test legitimacy judgments towards an understanding that legitimacy is grounded in both cognition and social interaction (Cornelissen, Durand, Fiss, Lammers, & Vaara, 2015; Hofer & Green, 2016). Accordingly, we posit that legitimacy judgment formation in deliberations depends on the perceived quality of arguments, which is itself subject to the deliberative process and can be influenced by various sources of validity (e.g., experts, other deliberation members, social media commentators).

To explain the influence of different validity sources we distinguish between direct effects on propriety, regulated by source salience, and indirect effects via validation by other sources, whereby validity sources function as “validation layers” ordered by source identification. Specifically, we posit that individuals pay attention to how more proximate, more trusted, sources (i.e., higher ordered layers such as peers) evaluate less proximate, less trusted, sources (i.e., lower ordered layers such as experts), especially if the latter seem controversial. Our results indicate that peer validation by other deliberation participants in direct deliberation, such as in the first study, or by social media users in indirect deliberation, such as in the second study, was more important in presence of firm representatives. The firm representatives we argue, in our case, to be more controversial due to representing the criticized organization and practice, than NGO representatives, thus requiring more validation.

Second, our study is the first to provide a large-scale empirical test about the influence of public deliberation on judgment formation about organizational legitimacy (Palazzo & Scherer, 2006; Suchman, 1995). Extant management literature on deliberations is either conceptual (e.g., Palazzo & Scherer, 2006; Scherer & Palazzo, 2007) or relies on qualitative case studies on topics such as stakeholder dialogues (Dentoni, Bitzer, & Schouten, 2018; Roloff, 2008; Schouten et al.,

2012), ethical sense-making (Reinecke & Ansari, 2015), or shareholder engagement (Ferraro & Beunza, 2018), that do not systematically explore the influence on legitimacy. Our quantitative study examines the conditions under which deliberation leads to the legitimation or delegitimation of an organization and a practice and provides evidence for the significance of the relative cogency of the arguments put forward by representatives of the organization and its critics. Further, we show that in deliberative settings it is not just the arguments, but also the source of the arguments and conditions such as source salience or identification that influence the individual judgment process.

Our study also goes beyond the empirical political sciences that have explored in field experiments the effects of deliberations among ‘mini publics’ (Goodin & Dryzek, 2006) on collective decision-making about public issues such as local security, renewable energy projects, or waste management (Dryzek, 1990; Fast, 2013; Fishkin, 2009). These studies provide evidence under which conditions deliberations lead to better collective decisions (Ryfe, 2005; Thompson, 2008). However, despite their general interest in legitimacy as an outcome of deliberations in the public realm (Fung, 2006; Thompson, 2008) political scientists have not studied the effect of deliberations on legitimacy judgments about organizational entities and the practices with which they are involved. Further, we take account of the significance of social media as a newly emerging deliberation platform that complements the traditional news media (such as newspapers, radio broadcasts, or TV) and enriches the public sphere in which deliberations about societal issues take place (Habermas, 1989; Roberts, 2004).

Third, we make a methodological contribution. Most empirical research on deliberation struggles to make causal claims about deliberative effects, because it is missing random treatment assignment (Bächtiger, 2017; Karpowitz & Mendelberg, 2011; for exceptions see, e.g., Baccaro, Bächtiger, & Deville, 2016; Farrar, Green, Green, Nickerson, & Shewfelt, 2009), or display design

flaws such as self-selection biases, varying group sizes or deliberation times, that potentially distort the analysis of causal relationships. In addition, some studies do not use moderators or other means to safeguard the general rules of deliberation (e.g., Druckman & Nelson, 2003; Karpowitz, Mendelberg, & Shaker, 2012; Mendelberg, Karpowitz, & Goedert, 2014). As deliberation is not just mere communication among a group of people, it is unclear what the corresponding results contribute to deliberation research. Our experimental design accounts for these potential flaws by not only implementing random treatment assignment but also controlling for potential confounds and paying attention to an epistemically valid setting for deliberation.

Organizational scholars have recently argued that a too narrow set of empirical research methods has constrained theory development in legitimacy research (Suddaby et al., 2017). Specifically, empirical research has thus far neglected one of the central tenets of legitimacy theory—that legitimacy “represents a *relationship* with an audience, rather than being a *possession* of the organization” (Suchman, 1995: 594, emphasis in the original). Whereas scholars argue that legitimacy should be understood as a perceptual phenomenon (Suddaby et al., 2017), empirical insight into how exactly the legitimacy perceptions and judgments of individual evaluators form and change is largely lacking (Haack & Sieweke, 2018; Jahn et al., 2017, p. 20). Our experimental study allows legitimacy researchers to grasp legitimacy as a perception that “resides in the eye of the beholder” (Ashforth & Gibbs, 1990, p. 177) while acknowledging that this perception is molded in social interaction. It thus contributes to examining legitimacy at the individual level as an explanandum and empirical attribute in its own right. To further promote legitimacy research at the individual level, we developed a scale to measure propriety judgments that can be used to measure either the legitimacy of a business practice or organization across different contexts.

Practical Implications

Our study also offers several practical implications for organizations entering deliberations. Specifically, whereas entering deliberations discussing the legitimacy of a criticized business firm is fairly risk-free for the firm itself and may even make evaluators' propriety judgments more favourable, organizations should be mindful of entering deliberations discussing the legitimacy of a controversial business practice, as our study results provide evidence that entering deliberations may lead to the practice's delegitimation rather than its legitimation. For NGOs, representing the societal critics of business firms and their activities, the pattern is reversed. For them, entering deliberation seems to be risk-free or beneficial in terms of legitimacy of the practice as deliberation has been shown to affect evaluators' propriety judgments about the practice negatively, yet risky in terms of the legitimacy of the criticized firm, as results have shown that deliberation may lead to the legitimation rather than delegitimation of the firm. It follows that it would be beneficial for firms to stay out of deliberations in the event that only a practice in which they are engaging, and not their organization in particular, is publicly criticized. If, however, the name of a firm is picked up and criticized publically, a firm should proactively engage in deliberation and foster communication with relevant audiences to prevent delegitimation.

According to the results of our second experiment, the positive or negative effects on propriety judgments that occur when deliberating directly and face-to-face with organizational representatives can also be achieved when watching a video of the deliberation. Since it is not feasible to engage in direct deliberation with all stakeholders, this form of indirect deliberation can be a good way to scale the effects of direct deliberation and affect the judgments of broader audiences. While sharing such videos over social media would allow reaching large parts of society, our results indicate several risks for business firms. Our study has shown that comments on social media can be a very powerful source of validity that can completely reverse the intended effects of

deliberation. Therefore, diligently monitoring online communication and evaluators' social comments is an important first step in the management of organizational legitimacy.

Although we only looked at the effects of exclusively positive or exclusively negative comments, former research suggests that even a single negative comment can already affect corporate reputation negatively (Park & Lee, 2007). Although limiting users' ability to set comments in the first place may seem like a safeguarding option, such strategies can also backfire (Walther, DeAndrea, et al., 2010). Naturally, the here outlined risks for business firms provide opportunities for NGOs to more effectively criticize and delegitimize business firms and controversial practices.

Limitations and Future Research

This study generates several opportunities for future research. First, it would be important to elucidate the socio-psychological mechanisms underlying the impact of deliberation on legitimacy judgment formation. Dual-process theories of judgment and decision-making (Kahneman, 2011), social cognition (Chaiken & Trope, 1999) and persuasion (Petty & Cacioppo, 1986) argue that human behaviour is influenced by mental operations that can be distinguished in "System 1" processes, which are quick, associative, and intuitive, and "System 2" processes, which are rather slow, rule-governed, and deliberate. While previous legitimacy research drew on this research and distinguished an "active" and "passive" mode of legitimacy judgment formation (Bitektine, 2011; Bitektine & Haack, 2015; Tost, 2011), also highlighting that evaluators often derive their assessments intuitively (Haack, Pfarrer, & Scherer, 2014). As our results suggest that even within deliberative settings individuals draw from heuristics in an automatic fashion, future work could examine further the joint relevance of System 1 and System 2 processes in shaping deliberation outcomes and explore in more depth the intuitive and emotional underpinnings of legitimacy judgment formation (Gauthier & Kappen, 2017; Huy et al., 2014).

Second, given that our sample consisting predominantly of students, we need to explore the generalizability to other populations. The use of student samples is often criticized for lacking external validity (Aguinis & Bradley, 2014; Bitektine, Lucas, & Schilke, 2018; Druckman & Kam, 2009; Sears, 1986). However, the justification of this critique depends on the goal of the particular study (Bitektine, Lucas, et al., 2018; Druckman & Kam, 2009). The focus of our study was primarily to test and further advance a theory, namely to examine whether deliberation affects legitimacy judgment formation and to generate insights regarding the role of validity in this process, rather than to generalize an already established relationship. Achieving control and internal validity has thus been the focus of this study. However, student samples do not necessarily undermine generalizability to other contexts and populations. If there are sample differences in variables that influence the size and direction of a treatment effect, the effect might only be misestimated if those moderators are not included in the theoretical model and/or if the sample shows virtually no variance on the moderators (Druckman & Kam, 2009). In our case, the treatment effects were robust when controlling for the moderating variables (cf. Table B3 in the appendix). As we are not aware of a moderator for which we did not control that may have substantial influence on the relationship between deliberation and legitimation, we are confident that the observed effects of deliberation would also generalize to other samples.

Finally, a limitation of our study is that our conclusions are derived from deliberations in the context of corporate tax avoidance. Thus, it may be possible that the findings are bounded by the special characteristics and contingencies of our context. While the domain to which the relationships and processes we theorized may apply is potentially large (Bitektine, Lucas, et al., 2018), a major task for future research is to explore whether and how our findings can be generalized to other settings and translate to other grand challenges in global governance, such as climate change, poverty alleviation, and anti-corruption.

REFERENCES

- Addison, S., & Mueller, F. (2015). The dark side of professions: the big four and tax avoidance. *Accounting, Auditing & Accountability Journal*, 28(8), 1263–1290.
- Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, 17(4), 351–371.
- Aldrich, H. E., & Fiol, C. M. (1994). Fools rush in? The institutional context of industry creation. *The Academy of Management Review*, 19(4), 645–670.
- Alexiou, K., & Wiggins, J. (2018). Measuring individual legitimacy perceptions: Scale development and validation. *Strategic Organization*.
- Ashforth, B. E., & Gibbs, B. W. (1990). The double-edge of organizational legitimation. *Organization Science*, 1(2), 177–194.
- Baccaro, L., Bächtiger, A., & Deville, M. (2016). Small differences that matter: The impact of discussion modalities on deliberative outcomes. *British Journal of Political Science*, 46(3), 551–566.
- Bächtiger, A. (2017). A preface to studying deliberation empirically. In A. Bächtiger, J. S. Dryzek, J. Mansbridge, & M. E. Warren (Eds.), *The Oxford Handbook of Deliberative Democracy* (forthcoming). Oxford: Oxford University Press.
- Bächtiger, A., Dryzek, J. S., Mansbridge, J. J., & Warren, M. (Eds.). (2018). *The Oxford handbook of deliberative democracy* (First edition). Oxford, United Kingdom ; New York: Oxford University Press.
- Baddeley, M. (2013). Herding, social influence and expert opinion. *Journal of Economic Methodology*, 20(1), 35–44.

- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16(1), 74–94.
- Barnett, M., Henriques, I., & Husted, B. (2018). The rise and stall of stakeholder influence: How the digital age limits social control. *Academy of Management Perspectives*.
- BBC. (2012, November 12). Starbucks, Google and Amazon grilled over tax avoidance. *BBC News*. Retrieved from <http://www.bbc.com/news/business-20288077>
- Berger, P. L., & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*. Garden City, NY: Doubleday.
- Bitektine, A. (2011). Toward a Theory of Social Judgments of Organizations: The Case of Legitimacy, Reputation, and Status. *Academy of Management Review*, 36(1), 151–179.
- Bitektine, A., & Haack, P. (2015). The “Macro” and the “Micro” of Legitimacy: Toward a Multilevel Theory of the Legitimacy Process. *Academy of Management Review*, 40(1), 49–75.
- Bitektine, A., Hill, K., Song, F., & Vandenberghe, C. (2018). Organizational legitimacy, reputation and status: Insights from micro-level measurement. *Academy of Management Discoveries*.
- Bitektine, A., Lucas, J., & Schilke, O. (2018). Institutions under a microscope: experimental methods in institutional theory. In A. Bryman & D. Buchanan (Eds.), *Unconventional Methodology in Organization and Management Research*. Oxford: Oxford University Press.
- Briñol, P., & Petty, R. E. (2009). Chapter 2 Persuasion: Insights from the Self-Validation Hypothesis. In *Advances in Experimental Social Psychology* (Vol. 41, pp. 69–118). Academic Press.
- Carpini, M. X. D., Cook, F. L., & Jacobs, L. R. (2004). Public deliberation, discursive participation, and citizen engagement: A review of the empirical literature. *Annual Review of Political Science*, 7(1), 315–344.

- Castelló, I., Etter, M., & Nielsen, F. Å. (2016). Strategies of Legitimacy Through Social Media: The Networked Strategy. *Journal of Management Studies*, 53(3), 402–432.
- Castello, I., Etter, M., & Nielsen, F. Å. (2016). Strategies of Legitimacy Through Social Media: The Networked Strategy. *Journal of Management Studies*, 53(3).
- Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. New York: Guilford Press.
- Chambers, S. (2003). Deliberative Democratic Theory. *Annual Review of Political Science*, 6(1), 307–326.
- Chung, J. Y., Berger, B. K., & DeCoster, J. (2016). Developing Measurement Scales of Organizational and Issue Legitimacy: A Case of Direct-to-Consumer Advertising in the Pharmaceutical Industry. *Journal of Business Ethics*, 137(2), 405–413.
- Cialdini, R. B. (1993). *Influence: Science and practice* (3rd Edition). New York, NY, US: HarperCollins College Publishers.
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: compliance and conformity. *Annual Review of Psychology*, 55, 591–621.
- Cornelissen, J. P., Durand, R., Fiss, P. C., Lammers, J. C., & Vaara, E. (2015). Putting communication front and center in institutional theory and analysis. *Academy of Management Review*, 40(1), 10–27.
- Das, R., Kamruzzaman, J., & Karmakar, G. (2018). Modelling majority and expert influences on opinion formation in online social networks. *World Wide Web*, 21(3), 663–685.
- Deephouse, D. L. (1999). To be different, or to be the same? It's a question (and theory) of strategic balance. *Strategic Management Journal*, 20(2), 147–166.

- Deephouse, D. L., Bundy, J., Tost, L., & Suchman, M. (2017). Organizational legitimacy: Six key questions. In R. Greenwood, C. Oliver, T. Lawrence, & R. Meyer (Eds.), *The SAGE Handbook of Organizational Institutionalism* (2nd ed.). Thousand Oaks CA: Sage.
- Deephouse, D. L., & Suchman, M. (2008). Legitimacy in organizational institutionalism. In R. Greenwood, C. Oliver, R. Suddaby, & K. Sahlin (Eds.), *The SAGE Handbook of Organizational Institutionalism* (pp. 49–77). London: Sage.
- Dentoni, D., Bitzer, V., & Schouten, G. (2018). Harnessing Wicked Problems in Multi-stakeholder Partnerships. *Journal of Business Ethics*, *150*(2), 333–356.
- Dharmapala, D. (2014). What Do We Know about Base Erosion and Profit Shifting? A Review of the Empirical Literature. *Fiscal Studies*, *35*(4), 421–448.
- Dowling, G. R. (2014). The Curious Case of Corporate Tax Avoidance: Is it Socially Irresponsible? *Journal of Business Ethics*, *124*(1), 173–184.
- Druckman, J. N., & Kam, C. (2009). *Students as experimental participants: A defense of the “narrow data base”* (SSRN Scholarly Paper No. ID 1498843). Rochester, NY: Social Science Research Network.
- Druckman, J. N., & Nelson, K. R. (2003). Framing and deliberation: How citizens’ conversations limit elite influence. *American Journal of Political Science*, *47*(4), 729–745.
- Dryzek, J. S. (1990). *Discursive democracy: Politics, policy, and political science*. Cambridge, Mass.: Cambridge University Press.
- Dryzek, J. S., & Pickering, J. (2017). Deliberation as a catalyst for reflexive environmental governance. *Ecological Economics*, *131*, 353–360.
- Edwards, C., Edwards, A., Qing, Q., & Wahl, S. T. (2007). The Influence of Computer-Mediated Word-of-Mouth Communication on Student Perceptions of Instructors and Attitudes Toward Learning Course Content. *Communication Education*, *56*(3), 255–277.

- Elsbach, K. D. (1994). Managing organizational legitimacy in the California cattle industry: The construction and effectiveness of verbal accounts. *Administrative Science Quarterly*, 39(1), 57–88.
- Etter, M., Ravasi, D., & Colleoni, E. (2017). Social media and the formation of organizational reputation. *Academy of Management Review*, amr.2014.0280.
- Farrar, C., Fishkin, J. S., Green, D. P., List, C., Luskin, R. C., & Levy Paluck, E. (2010). Disaggregating Deliberation's Effects: An Experiment within a Deliberative Poll. *British Journal of Political Science*, 40(02), 333–347.
- Farrar, C., Green, D. P., Green, J. E., Nickerson, D. W., & Shewfelt, S. (2009). Does discussion group composition affect policy preferences? Results from three randomized experiments. *Political Psychology*, 30(4), 615–647.
- Fast, S. (2013). A Habermasian analysis of local renewable energy deliberations. *Journal of Rural Studies*, 30, 86–98.
- Ferraro, F., & Beunza, D. (2018). Creating Common Ground: A Communicative Action Model of Dialogue in Shareholder Engagement. *Organization Science*, 1–21.
- Ferraro, F., Etzion, D., & Gehman, J. (2015). Tackling Grand Challenges Pragmatically: Robust Action Revisited. *Organization Studies*, 36(3), 363–390.
- Festinger, L. (1962). *A Theory of Cognitive Dissonance*. Stanford University Press.
- Finch, D., Deephouse, D., & Varella, P. (2015). Examining an Individual's Legitimacy Judgment Using the Value–Attitude System: The Role of Environmental and Economic Values and Source Credibility. *Journal of Business Ethics*, 127(2), 265–281.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behaviour: An introduction to theory and research*. Reading, MA: Addison-Wesley.

- Fishkin, J. S. (2009). *When the People Speak: Deliberative Democracy and Public Consultation*. Oxford: Oxford University Press.
- Fishkin, J. S., & Luskin, R. C. (2005). Experimenting with a Democratic Ideal: Deliberative Polling and Public Opinion. *Acta Politica*, 40(3), 284–298.
- Foreman, P., & Whetten, D. A. (2002). Members' Identification with Multiple-Identity Organizations. *Organization Science*, 13(6), 618–635.
- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39.
- Fung, A. (2003). Deliberative democracy and international labor standards. *Governance*, 16(1), 51–71.
- Fung, A. (2006). Varieties of Participation in Complex Governance. *Public Administration Review*, 66(s1), 66–75.
- Gauthier, J., & Kappen, J. A. (2017). Corporate rhetorical strategies in the legitimation of genetically modified foods. *Journal of Communication Management*, 21(3), 218–235.
- George, G., Howard-Grenville, J., Joshi, A., & Tihanyi, L. (2016). Understanding and Tackling Societal Grand Challenges through Management Research. *Academy of Management Journal*, 59(6), 1880–1895.
- Gerlitz, J.-Y., & Schupp, J. (2005). *Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP*. DIW Research Notes, Berlin.
- Glozer, S., Caruana, R., & Hibbert, S. A. (2018). The Never-Ending Story: Discursive Legitimation in Social Media Dialogue. *Organization Studies*, 017084061775100.
- Goodin, R. E., & Dryzek, J. S. (2006). Deliberative Impacts: The Macro-Political Uptake of Mini-Publics. *Politics & Society*, 34(2), 219–244.

- Green, S. E., Li, Y., & Nohria, N. (2009). Suspended in Self-Spun Webs of Significance: A Rhetorical Model of Institutionalization and Institutionally Embedded Agency. *The Academy of Management Journal*, 52(1), 11–36.
- Groeber, P., Lorenz, J., & Schweitzer, F. (2014). Dissonance Minimization as a Microfoundation of Social Influence in Models of Opinion Formation. *The Journal of Mathematical Sociology*, 38(3), 147–174.
- Haack, P., McKinley, W., Schilke, O., & Zucker, L. G. (2016). The theory-method gap in organizational legitimacy research: A critical review, synthesis, and directions for future research. Presented at the American Sociological Association Annual Meeting, Seattle, WA.
- Haack, P., Pfarrer, M. D., & Scherer, A. G. (2014). Legitimacy-as-Feeling: How Affect Leads to Vertical Legitimacy Spillovers in Transnational Governance. *Journal of Management Studies*, 51(4), 634–666.
- Haack, P., & Sieweke, J. (2018). The Legitimacy of Inequality: Integrating the Perspectives of System Justification and Social Judgment. *Journal of Management Studies*, 55(3), 486–516.
- Habermas, J. (1984). *The theory of communicative action. Trans. Thomas Mccarthy. Vol. 1: Reason and the rationalization of society*. Boston: Beacon.
- Habermas, J. (1989). *The Structural transformation of the public sphere: An inquiry into a category of bourgeois society*. Cambridge: Polity.
- Habermas, J. (1993). *Justification and application: Remarks on discourse ethics*. Cambridge, MA: The Mit Press.
- Habermas, J. (1996). *Between facts and norms*. Cambridge, MA: MIT Press.
- Halpern, D., & Gibbs, J. (2013). Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior*, 29(3), 1159–1168.

- Hanlon, M., & Heitzman, S. (2010). A review of tax research. *Journal of Accounting and Economics*, 50(2–3), 127–178.
- Hanna, R., Rohm, A., & Crittenden, V. (2011). We're all connected: The power of the social media ecosystem. *Business Horizons*, 54, 265–273.
- Hinkin, T. R. (1998). A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires. *Organizational Research Methods*, 1(1), 104–121.
- Hodge, M. (2016). *Called to account: How corporate bad behaviour and government waste combine to cost us millions*. London: Little, Brown Book Group.
- Hoefer, R. L., & Green, S. E. (2016). A Rhetorical Model of Institutional Decision Making: The Role of Rhetoric in the Formation and Change of Legitimacy Judgments. *Academy of Management Review*, 41(1), 130–150.
- Hoffman, A. J. (1999). Institutional evolution and change: Environmentalism and the U.S. chemical industry. *The Academy of Management Journal*, 42(4), 351–371.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Hudson, M., Chavkin, S., & Mos, B. (2014, November 6). Big four audit firms behind global profit shifting. *The Sydney Morning Herald*. Retrieved from <http://www.smh.com.au/business/big-four-audit-firms-behind-global-profit-shifting-20141106-11i08q.html>
- Huy, Q. N., Corley, K. G., & Kraatz, M. S. (2014). From support to mutiny: Shifting legitimacy judgments and emotional reactions impacting the implementation of radical change. *Academy of Management Journal*, 57(6), 1650–1680.

- Isernia, P., & Fishkin, J. S. (2014). The EuroPolis deliberative poll. *European Union Politics*, 15(3), 311–327.
- Jahn, J., Eichhorn, M., & Brühl, R. (2017). How do individuals judge organizational legitimacy? Effects of attributed motives and credibility on organizational legitimacy. *Business & Society*, 0007650317717959.
- Johnson, C., Dowd, T. J., & Ridgeway, C. L. (2006). Legitimacy as a Social Process. *Annual Review of Sociology*, 32(1), 53–78.
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Lane.
- Karpowitz, C. F., & Mendelberg, T. (2011). An experimental approach to citizen deliberation. *Cambridge Handbook of Experimental Political Science*, 258–272.
- Karpowitz, C. F., Mendelberg, T., & Shaker, L. (2012). Gender inequality in deliberative participation. *American Political Science Review*, 106(3), 533–547.
- Kostova, T., & Roth, K. (2002). Adoption of an Organizational Practice by Subsidiaries of Multinational Corporations: Institutional and Relational Effects. *The Academy of Management Journal*, 45(1), 215–233.
- Kostova, T., Roth, K., & Dacin, M. T. (2008). Institutional theory in the study of multinational corporations: A critique and new directions. *Academy of Management Review*, 33(4), 994–1006.
- Lamertz, K., & Bhave, D. P. (2017). Employee perceptions of organisational legitimacy as impersonal bases of organisational trustworthiness and trust. *Journal of Trust Research*, 7(2), 129–149.
- Lee, M. J., & Chun, J. W. (2016). Reading others' comments and public opinion poll results on social media: Social judgment and spiral of empowerment. *Computers in Human Behavior*, 65, 479–487.

- Maguire, S., & Hardy, C. (2009). Discourse and deinstitutionalization: The decline of DDT. *Academy of Management Journal*, 52(1), 148–178.
- Manin, B. (1987). On Legitimacy and Political Deliberation. *Political Theory*, 15(3), 338–368.
- Mansbridge, J., Hartz-Karp, J., Amengual, M., & Gastil, J. (2006). Norms of deliberation: An inductive study. *Journal of Public Deliberation*, 2(1).
- Massey, K., Freeman, S., & Zelditch, M. (1997). Status, Power, and Accounts. *Social Psychology Quarterly*, 60(3), 238–251.
- Mendelberg, T., Karpowitz, C. F., & Goedert, N. (2014). Does descriptive representation facilitate women's distinctive voice? How gender composition and decision rules affect deliberation. *American Journal of Political Science*, 58(2), 291–306.
- Meyer, J. W., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, 83(2), 340–363.
- Meyer, R. E., & Höllerer, M. A. (2010). Meaning structures in a contested issue field: A topographic map of shareholder value in Austria. *Academy of Management Journal*, 53(6), 1241–1262.
- Mitchell, R. K., Agle, B. R., & Wood, D. J. (1997). Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts. *The Academy of Management Review*, 22(4), 853–886.
- Moussaïd, M., Kämmer, J. E., Analytis, P. P., & Neth, H. (2013). Social Influence and the Collective Dynamics of Opinion Formation. *PLOS ONE*, 8(11), e78433.
- OECD. (2013). *Addressing base erosion and profit shifting*. Paris: OECD Publishing.
- Page, B. I., Shapiro, R. Y., & Dempsey, G. R. (1987). What Moves Public Opinion? *The American Political Science Review*, 81(1), 23–43.

- Palazzo, G., & Scherer, A. G. (2006). Corporate Legitimacy as Deliberation: A Communicative Framework. *Journal of Business Ethics*, 66(1), 71–88.
- Parent, M. M., & Deephouse, D. L. (2007). A case study of stakeholder identification and prioritization by managers. *Journal of Business Ethics*, 75(1), 1–23.
- Park, N., & Lee, K. M. (2007). Effects of online news forum on corporate reputation. *Public Relations Review*, 33(3), 346–348.
- Petty, R., & Cacioppo, J. (1986). *Communication and Persuasion Central and Peripheral Routes to Attitude Change*. New York, NY: Springer.
- Prebble, Z., & Prebble, J. (2010). The morality of tax avoidance. *Creighton Law Review*, 43(3), 693–745.
- Rehg, W. (2013). Rhetoric, cogency, and the radically social character of persuasion: Habermas's argumentation theory revisited. *Philosophy & Rhetoric*, 46(4), 465–492.
- Reicher, S. D., Spears, R., & Postmes, T. (1995). A Social Identity Model of Deindividuation Phenomena. *European Review of Social Psychology*, 6(1), 161–198.
- Reinecke, J., & Ansari, S. (2015). What Is a “Fair” Price? Ethics as Sensemaking. *Organization Science*, 26(3), 867–888.
- Roberts, N. (2004). Public deliberation in an age of direct citizen participation. *The American Review of Public Administration*, 34(4), 315–353.
- Rogers, E. M., & Shoemaker, F. F. (1971). *Communication of innovations: A cross-cultural approach*. New York, NY, US: Free Press.
- Roloff, J. (2008). Learning from Multi-Stakeholder Networks: Issue-Focussed Stakeholder Management. *Journal of Business Ethics*, 82(1), 233–250.
- Ryfe, D. M. (2005). Does Deliberative Democracy Work? *Annual Review of Political Science*, 8(1), 49–71.

- Scherer, A. G., & Palazzo, G. (2007). Toward a political conception of corporate responsibility: Business and society seen from a Habermasian perspective. *The Academy of Management Review*, 32(4), 1096–1120.
- Scherer, A. G., & Palazzo, G. (2011). The New Political Role of Business in a Globalized World: A Review of a New Perspective on CSR and its Implications for the Firm, Governance, and Democracy. *Journal of Management Studies*, 48(4), 899–931.
- Scherer, A. G., Palazzo, G., & Seidl, D. (2013). Managing Legitimacy in Complex and Heterogeneous Environments: Sustainable Development in a Globalized World: Sustainable Development in a Globalized World. *Journal of Management Studies*, 50(2), 259–284.
- Scherer, A. G., Rasche, A., Palazzo, G., & Spicer, A. (2016). Managing for Political Corporate Social Responsibility: New Challenges and Directions for PCSR 2.0. *Journal of Management Studies*, 53(3), 273–298.
- Schouten, G., Leroy, P., & Glasbergen, P. (2012). On the deliberative capacity of private multi-stakeholder governance: The Roundtables on Responsible Soy and Sustainable Palm Oil. *Ecological Economics*, 83(Supplement C), 42–50.
- Sears, D. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, 51, 515–530.
- Shi, R., Messaris, P., & Cappella, J. N. (2014). Effects of Online Comments on Smokers' Perception of Antismoking Public Service Announcements. *Journal of Computer-Mediated Communication*, 19(4), 975–990.
- Sikka, P., & Willmott, H. (2010). The dark side of transfer pricing: Its role in tax avoidance and wealth retentiveness. *Critical Perspectives on Accounting*, 21(4), 342–356.

- Sikka, P., & Willmott, H. (2013). The tax avoidance industry: accountancy firms on the make. *Critical Perspectives on International Business*, 9(4), 415–443.
- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *The Academy of Management Review*, 20(3), 571–610.
- Suddaby, R., Bitektine, A., & Haack, P. (2017). Legitimacy. *Academy of Management Annals*, 11(1), 451–478.
- Suddaby, R., & Greenwood, R. (2005). Rhetorical strategies of legitimacy. *Administrative Science Quarterly*, 50(1), 35–67.
- Sundar, S. S., & Nass, C. (2001). Conceptualizing Sources in Online News. *Journal of Communication*, 51, 52–72.
- Sung, H. K., & Lee, M. (2014). Do Online Comments Influence the Public's Attitudes Toward an Organization? Effects of Online Comments Based on Individuals' Prior Attitudes. *The Journal of Psychology*, 149, 141217142405003.
- Syal, R. (2015, March 26). Prosecute financial advisers who assist in tax avoidance, say MPs. *The Guardian*. Retrieved from <https://www.theguardian.com/business/2015/mar/26/prosecute-financial-advisers-tax-avoidance-mps>
- Taylor, D. G. (1982). Pluralistic Ignorance and the Spiral of Silence: A Formal Analysis. *The Public Opinion Quarterly*, 46(3), 311–335.
- Thompson, D. F. (2008). Deliberative Democratic Theory and Empirical Political Science. *Annual Review of Political Science*, 11(1), 497–520.
- Tolbert, P. S., & Zucker, L. G. (1996). The Institutionalization of Institutional Theory. In S. Clegg, C. Hardy, & W. Nord (Eds.), *The handbook of organization studies* (pp. 175–190). Thousand Oaks, CA: SAGE.

- Tost, L. P. (2011). An Integrative Model of Legitimacy Judgments. *Academy of Management Review*, 36(4), 686–710.
- Toynbee, P. (2013, February 1). Accountancy's Big Four are laughing all the way to the tax office. *The Guardian*. Retrieved from <https://www.theguardian.com/commentisfree/2013/feb/01/accountancy-big-four-laugh-tax-office>
- United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. Resolution adopted by the General Assembly on 25 September 2015, A/RES/70/1.
- Vaara, E., & Tienari, J. (2008). A Discursive Perspective on Legitimation Strategies in Multinational Corporations. *The Academy of Management Review*, 33(4), 985–993.
- Walker, H. A., Rogers, L., & Zelditch, M. (1988). Legitimacy and Collective Action: A Research Note. *Social Forces*, 67(1), 216–228.
- Walther, J. B., Carr, C., Choi, S., DeAndrea, D., Kim, J., Tong, S., & Van Der Heide, B. (2010). Interaction of interpersonal, peer, and media influence sources online: A research agenda for technology convergence. In Z. Papacharissi (Ed.), *A Networked Self: Identity, Community, and Culture on Social Network Sites* (1st ed., pp. 17–38). NY: Routledge.
- Walther, J. B., DeAndrea, D., Kim, J., & Anthony, J. C. (2010). The Influence of Online Comments on Perceptions of Antimarijuana Public Service Announcements on YouTube. *Human Communication Research*, 36(4), 469–492.
- Wang, Z., Walther, J. B., Pingree, S., & Hawkins, R. P. (2008). Health Information, Credibility, Homophily, and Influence via the Internet: Web Sites Versus Discussion Groups. *Health Communication*, 23(4), 358–368.
- West, M. (2016, July 11). Oligarchs of the Treasure Islands. Retrieved from <http://www.michaelwest.com.au/oligarchs-of-the-treasure-islands/>

Zerubavel, E. (1997). *Social mindscapes: An invitation to cognitive sociology*. Cambridge: Harvard University Press.

Zimmerman, M. A., & Zeitz, G. J. (2002). Beyond Survival: Achieving New Venture Growth by Building Legitimacy. *The Academy of Management Review*, 27(3), 414–431.

Zucman, G. (2015). *The hidden wealth of nations: The scourge of tax havens*. Chicago: Chicago University Press.

Zucman, G. (2017, November 10). How Corporations and the Wealthy Avoid Taxes (and How to Stop Them). *The New York Times*. Retrieved from <https://www.nytimes.com/interactive/2017/11/10/opinion/gabriel-zucman-paradise-papers-tax-evasion.html>

Table 1 – Legitimacy Scale

Organization (Big Four)		Practice (Corporate Tax Avoidance)	
Items	Factor Loadings	Items	Factor Loadings
Cognitive Legitimacy ($\alpha=.81$)		Cognitive Legitimacy ($\alpha=.79$)	
Factor 1: Taken-For-Grantedness ($\alpha=.90$)		Factor 1: Taken-For-Grantedness ($\alpha=.79$)	
1. The Big Four are needed.	.80	1. Tax optimization is needed.	.61
2. The Big Four are necessary corporations.	.79	2. Tax optimization is a necessary service.	.61
Factor 2: Comprehensibility / Understandability ($\alpha=.86$)		Factor 2: Comprehensibility / Understandability ($\alpha=.80$)	
3. I can understand the behavior of the Big Four.	.77	3. I can understand the use of tax optimization.	.76
4. The behavior of the Big Four is understandable.	.84	4. The use of tax optimization is understandable.	.81
Moral Legitimacy ($\alpha=.92$)		Moral Legitimacy ($\alpha=.87$)	
Factor 3: Moral Integrity ($\alpha=.90$)		Factor 3: Moral Integrity ($\alpha=.85$)	
8. The Big Four are acting morally acceptable.	.58	8. Tax optimization is morally acceptable.	.51
9. The Big Four are acting morally reprehensible.*	.71	9. Tax optimization is morally reprehensible.*	.68
Factor 4: Approval of Criticism ($\alpha=.90$)		Factor 4: Approval of Criticism ($\alpha=.85$)	
5. The criticism of the Big Four is justified.*	.91	5. The criticism of tax optimization is justified.*	.82
6. I agree with the critique of the Big Four.*	.64	6. I agree with the critique of tax optimization.*	.76
7. The behavior of the Big Four is problematic.*	.81	7. Tax optimization is problematic.*	.68
Pragmatic Legitimacy ($\alpha=.85$)		Pragmatic Legitimacy ($\alpha=.79$)	
Factor 5: Economy – Direct Benefits ($\alpha=.74$)		Factor 5: Economy – Direct Benefits ($\alpha=.68$)	
10. The economy benefits from the Big Four.	.66	10. The economy benefits from tax optimization.	.63
11. The Big Four hurt the economy.*	.70	11. Tax optimization hurts the economy.*	.66
Factor 6: Society – Indirect Benefits & Responsiveness ($\alpha=.88$)		Factor 6: Society – Indirect Benefits & Responsiveness ($\alpha=.82$)	
12. The Big Four benefit the society.	.78	12. Tax optimization benefits the society.	.78
13. The Big Four contribute positively to society.	.68	13. Tax optimization contributes positively to society.	.68
14. The behavior of the Big Four is compatible with public interests.	.71	14. Tax optimization is compatible with public interests.	.53
15. Public interests are taken into account by the Big Four.	.75	15. Public interests are taken into account when optimizing taxes.	.74

* Reverse coded items.

Note: Item loadings stem from a CFA using principal axis factoring with oblimin rotation. All cross loadings are below .2 for the organization and below 0.33 for the practice.

Table 2 – Summary of Legitimacy Judgments Across Treatments

Treatment	Subject of Legitimacy	Implicit Measure of Legitimacy (via Legitimacy Scale)							
		Before (Q1)	After (Q2)	Change (Q2-Q1)	Explicit Change (Q2)	Follow up (Q3)	Change (Q3-Q1)	N/n	
Control	Organization	-0.754***	-0.981***	-0.227**	-0.929***	70/10	-1.039***	-0.321**	55/10
	Practice	-0.639***	-0.698***	-0.0583	-0.129	70/10	-0.925***	-0.285**	55/10
Mixed tie	Organization	-0.0266	-0.382	-0.355	0	12/7	-0.516	-0.490*	12/7
	Practice	0.0648	-0.556	-0.620**	-0.250	12/7	-0.647	-0.712**	12/7
Firm	Organization	-0.748***	-0.366*	0.382***	-0.200	70/10	-0.356	0.336**	51/10
	Practice	-0.647***	-0.361*	0.285***	0.243	70/10	-0.440*	0.161	51/10
Mixed adv. firm	Organization	0.966**	1.289***	0.323	0.944***	18/8	0.880**	0.113	15/7
	Practice	0.968**	1.090***	0.121	0.722**	18/8	0.713	-0.158	15/7
NGO	Organization	-0.725***	-1.366***	-0.642***	-0.786***	70/10	-1.142***	-0.404***	58/10
	Practice	-0.799***	-1.467***	-0.668***	-0.614***	70/10	-1.376***	-0.662***	58/10
Mixed adv. NGO	Organization	-1.203***	-1.661***	-0.458***	-1.500***	40/10	-1.603***	-0.491***	36/9
	Practice	-1.000***	-1.507***	-0.507***	-1.175***	40/10	-1.591***	-0.544***	36/9
Wait	Organization	-0.121	-0.146	-0.0253	-0.182	11/11	-0.356	-0.241	8/8
	Practice	-0.338	-0.456	-0.117	0.0909	11/11	-0.380	0.0729	8/8

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ for t-Tests (H_0 : mean=0)

N: Total observations (i.e., participants) per treatment (subgroup)

n: Total independent observations (i.e., discussion groups) per treatment (subgroup)

Table 3 – Relevance of Peers and Experts as Sources of Validity

	Change in Legitimacy of the Big Four										Change in Legitimacy of Tax Optimization									
	(1) control	(2) firm	(3) firm	(4) firm	(5) NGO	(6) NGO	(7) NGO	(8) mixed	(9) mixed	(10) mixed	(1) control	(2) firm	(3) firm	(4) firm	(5) NGO	(6) NGO	(7) NGO	(8) mixed	(9) mixed	(10) mixed
Others' position ...																				
towards Big Four	0.128** (0.024)	0.157** (0.020)		0.0941 (0.147)	-0.0179 (0.812)		0.0246 (0.744)	0.186** (0.024)		0.133 (0.195)	0.134*** (0.008)	0.0834 (0.123)		0.0689 (0.199)	0.0202 (0.732)		0.0455 (0.450)	0.0691 (0.411)		0.0595 (0.477)
towards tax optimization																				
Persuasiveness of representatives:																				
Firm			0.208*** (0.000)	0.184*** (0.002)									0.101** (0.049)	0.0906* (0.077)						
NGO						-0.147** (0.027)	-0.153** (0.027)									-0.128** (0.047)	-0.137** (0.037)			
Firm vs. NGO								0.118** (0.010)	0.0944** (0.027)									0.0676 (0.163)	0.0647 (0.186)	
Controls:																				
Initial position ...																				
towards Big Four	-0.101* (0.083)	-0.0390 (0.554)	-0.0359 (0.561)	-0.0316 (0.607)	0.0566 (0.456)	0.00640 (0.931)	-0.000647 (0.993)	-0.0834 (0.231)	-0.0846 (0.217)	-0.0991 (0.238)	0.118** (0.044)	0.0822 (0.175)	0.0605 (0.289)	0.0829 (0.164)	0.165*** (0.006)	0.127* (0.080)	0.118 (0.108)	-0.0733 (0.327)	-0.0799 (0.281)	-0.0881 (0.242)
towards tax optimization	0.0800 (0.193)	0.00134 (0.983)	-0.0284 (0.646)	-0.0326 (0.595)	-0.0269 (0.703)	-0.0331 (0.623)	-0.0299 (0.661)	-0.0352 (0.533)	-0.0450 (0.423)	-0.0456 (0.478)	-0.0508 (0.449)	-0.0693 (0.259)	-0.0554 (0.332)	-0.0830 (0.174)	-0.201*** (0.003)	-0.208*** (0.002)	-0.210*** (0.002)	-0.0703 (0.252)	-0.0738 (0.225)	-0.0776 (0.206)
Political orientation	-0.0800 (0.157)	0.0472 (0.466)	0.0331 (0.587)	0.0331 (0.584)	0.0324 (0.581)	0.0160 (0.779)	0.0153 (0.790)	0.137** (0.035)	0.107 (0.103)	0.107 (0.184)	-0.146** (0.015)	-0.0405 (0.476)	-0.0527 (0.349)	-0.0488 (0.383)	0.116 (0.145)	0.101* (0.070)	0.101* (0.073)	0.125* (0.074)	0.105 (0.138)	0.103 (0.148)
Session	0.000921 (0.979)	0.0547 (0.182)	0.0400 (0.302)	0.0342 (0.375)	-0.00672 (0.871)	-0.000705 (0.985)	0.00427 (0.915)	-0.00203 (0.965)	0.0123 (0.779)	-0.0131 (0.713)	0.00109 (0.977)	-0.000914 (0.980)	-0.00608 (0.864)	-0.0125 (0.726)	0.0551 (0.199)	0.0542 (0.134)	0.0611 (0.103)	-0.00605 (0.910)	-0.000496 (0.992)	-0.0189 (0.726)
Constant	-0.149 (0.570)	0.258 (0.363)	-0.223 (0.396)	-0.0619 (0.826)	-0.532* (0.052)	-0.115 (0.702)	-0.0711 (0.830)	0.0804 (0.817)	-0.213 (0.459)	0.0875 (0.764)	-0.00484 (0.985)	0.351 (0.162)	0.0946 (0.695)	0.213 (0.408)	-0.759*** (0.001)	-0.447 (0.128)	-0.353 (0.269)	-0.221 (0.568)	-0.318 (0.307)	-0.154 (0.692)
N	69	68	68	68	68	68	68	69	69	69	69	68	68	68	68	68	68	69	69	69
R ²	0.127	0.143	0.244	0.270	0.018	0.092	0.094	0.164	0.183	0.217	0.247	0.064	0.087	0.112	0.174	0.224	0.231	0.078	0.097	0.104
adj. R ²	0.057	0.074	0.183	0.198	-0.061	0.019	0.005	0.097	0.119	0.141	0.187	-0.011	0.013	0.024	0.108	0.162	0.156	0.005	0.025	0.017

OLS regressions with robust standard errors where necessary (Breusch-Pagan test with $p < 0.05$)

p -values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4 – Peer Effects on Evaluation of Experts

	Persuasiveness of Firm Representative			Persuasiveness of NGO Representative			Persuasiveness of Firm vs. NGO Representative		
	(1) firm	(2) firm	(3) firm	(4) NGO	(5) NGO	(6) NGO	(7) mixed	(8) mixed	(9) mixed
Others' position towards Big Four	0.341** (0.016)		0.608** (0.014)	0.278 (0.250)		0.224 (0.413)	0.553** (0.012)		0.668** (0.011)
Others' position towards tax optimization		0.160 (0.230)	-0.302 (0.179)		0.184 (0.220)	0.0871 (0.619)		0.148 (0.497)	-0.204 (0.411)
Controls:									
Initial position towards Big Four	-0.0402 (0.770)	-0.00807 (0.957)	-0.124 (0.410)	-0.374* (0.060)	-0.336* (0.058)	-0.373* (0.056)	0.167 (0.368)	0.229 (0.239)	0.180 (0.335)
Initial position towards tax optimization	0.185 (0.177)	0.151 (0.321)	0.288* (0.067)	-0.0198 (0.883)	-0.0652 (0.640)	-0.0304 (0.822)	0.110 (0.467)	0.113 (0.475)	0.121 (0.425)
Political orientation	0.0771 (0.570)	0.0921 (0.513)	0.0576 (0.671)	-0.112 (0.506)	-0.109 (0.517)	-0.111 (0.508)	0.315* (0.068)	0.339* (0.061)	0.320* (0.065)
Session	0.112 (0.193)	0.128 (0.152)	0.119 (0.163)	0.0719 (0.296)	0.0440 (0.547)	0.0741 (0.280)	0.118 (0.347)	0.199 (0.154)	0.157 (0.244)
Constant	1.745*** (0.004)	1.522** (0.016)	1.649*** (0.007)	3.009*** (0.000)	2.957*** (0.000)	3.076*** (0.000)	-0.0755 (0.935)	-1.044 (0.299)	-0.366 (0.712)
<i>N</i>	68	68	68	68	68	68	69	69	69
<i>R</i> ²	0.234	0.178	0.256	0.313	0.296	0.318	0.379	0.318	0.386
adj. <i>R</i> ²	0.172	0.111	0.183	0.258	0.239	0.251	0.330	0.263	0.326

OLS regressions with robust standard errors where necessary (Breusch-Pagan test with $p < 0.05$)

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5 – Study 2 Treatment Groups

Discussion condition	Video version	YouTube comments	Treatment
Firm condition	long	none	FL
		negative	FL-
	short	none	FS
		positive negative	FS+ FS-
NGO condition	long	none	NL
		positive	NL+
	short	none	NS
		negative positive	NS- NS+
Mixed condition	long	none	MIXL
		positive	MIXL+
		negative	MIXL-

Table 6 – Relevance of Validity Sources for Opinion Formation

	Importance for Opinion Formation on Organization				Importance for Opinion Formation on Practice			
	Firm Expert	NGO Expert	Group	Comments	Firm Expert	NGO Expert	Group	Comments
mean	1.51	1.60	0.90	0.48	1.36	1.83	0.95	0.46
(sd)	2.52	2.24	2.33	2.54	2.39	2.17	2.32	2.57

Figure 1 – Study Overview

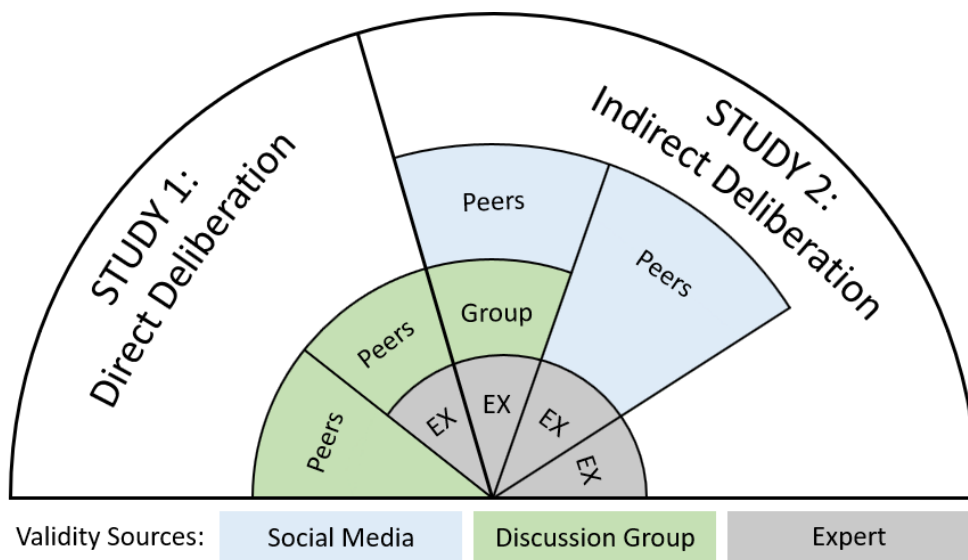


Figure 2 – Marginal Treatment Effects on Changes in Legitimacy

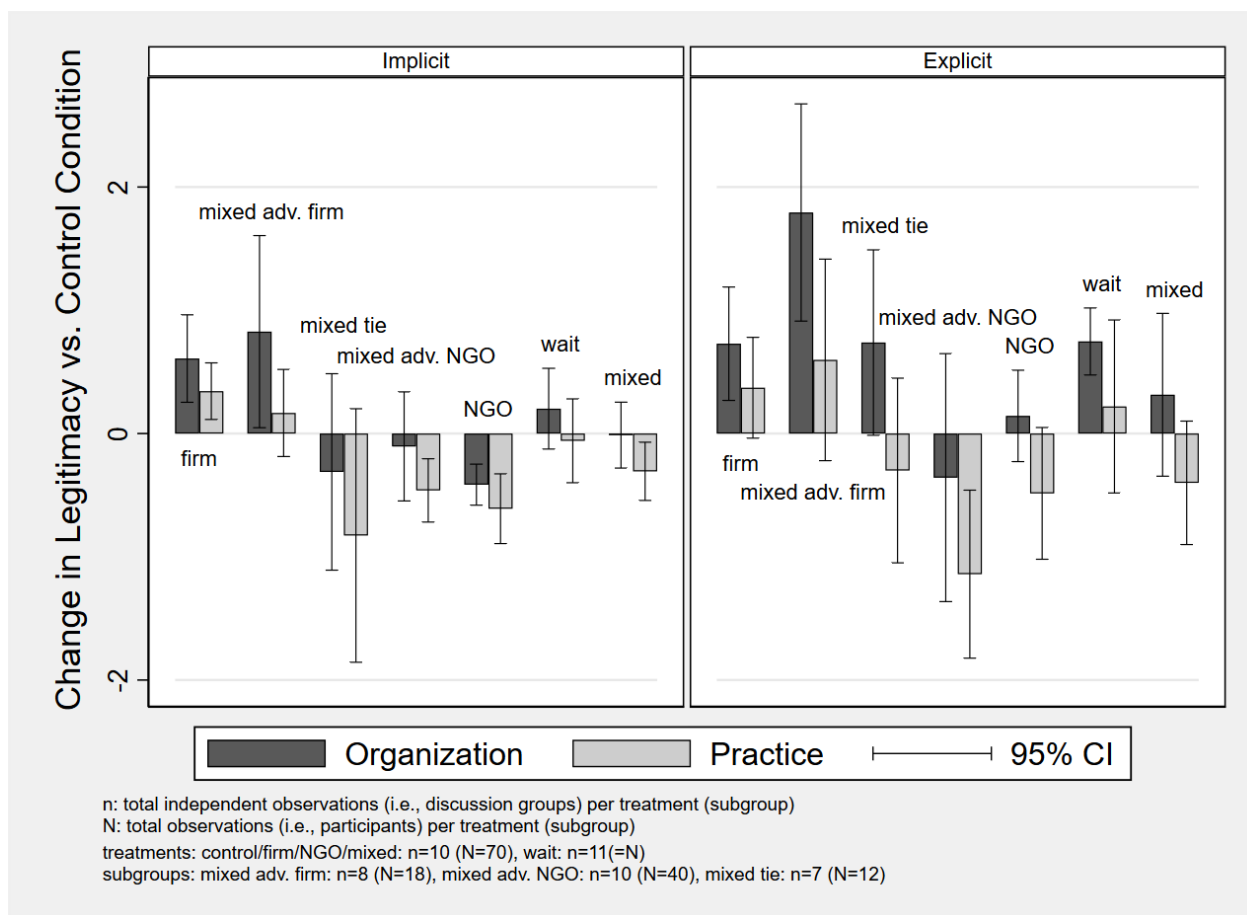
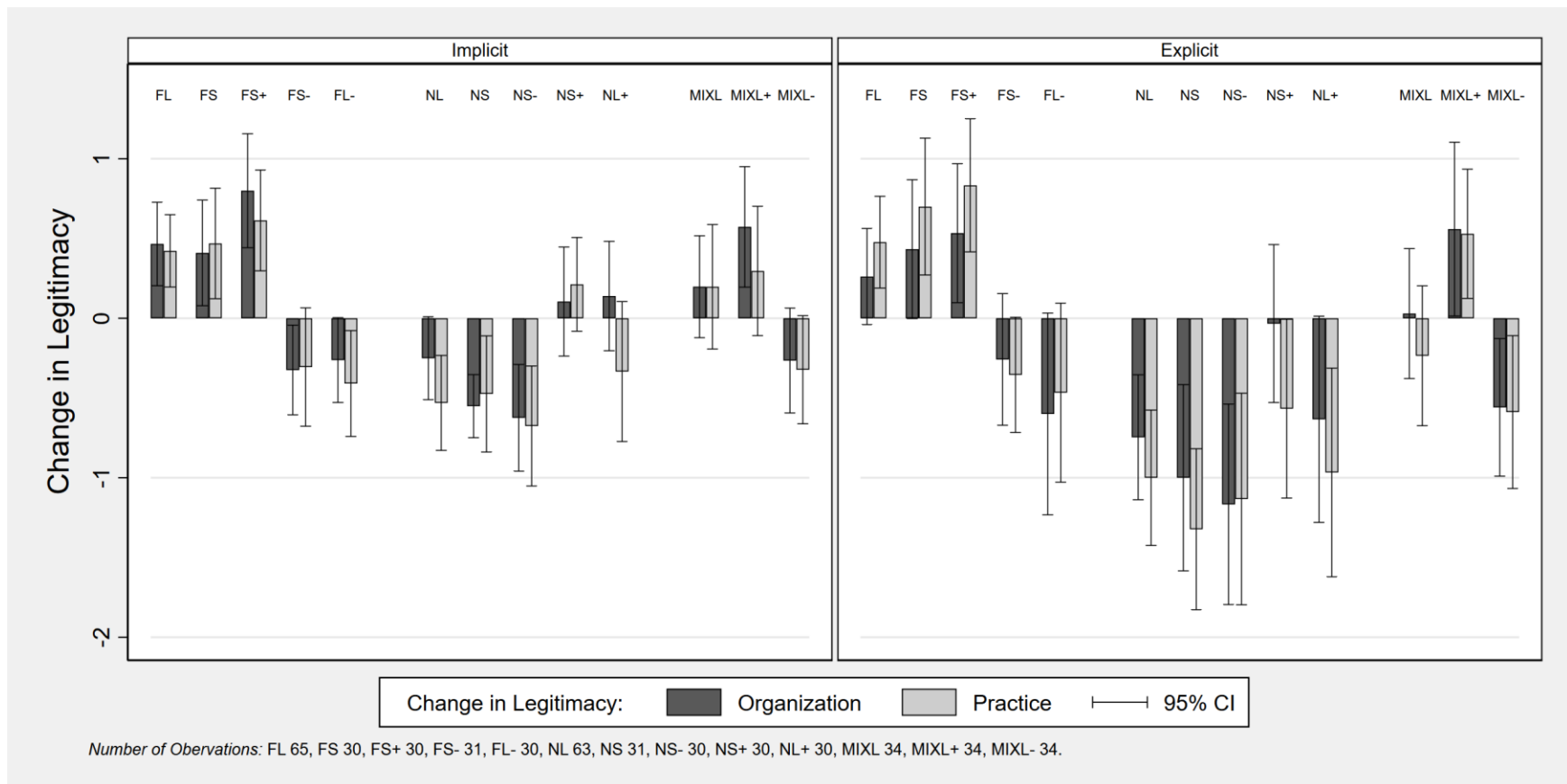


Figure 3 – Study 2 Treatment Effects



APPENDIX

Appendix A: Construction of Legitimacy Scale

To create our own scale to measure the propriety judgments about a publically criticized business practice (i.e., tax optimization) or organization (i.e., the Big Four), we largely followed the recommendations of Hinkin (1998).

Item Development and Initial Selection

First, we formulated a series of items in relation to the three dimensions of legitimacy described by Suchman (1995) and tested them. By looking closely at Suchman's (1995) original definition of the legitimacy dimensions (pragmatic, cognitive, and moral), we identified separate aspects underlying the cognitive, moral and pragmatic dimension. Accordingly, we have subdivided cognitive legitimacy into separate items describing "comprehensibility" or understanding (i.e., whether one can understand the use of the practice or the behavior of the organization), "knowledge" (i.e., whether one is familiar with the practice or the organization), and "taken-for-grantedness" (i.e., whether one generally sees the practice or organization as necessary). Within the moral dimension of legitimacy, we distinguished between "moral integrity" and approval of "criticism" about the legitimacy object. We divided the dimension of pragmatic legitimacy into subdimensions to cover the benefits or harm to different groups or actors of the organization or practice – in this case the economy, society, and the state or government – as well as the feature of being responsive to larger interests (i.e., "responsiveness"). Within the pragmatic subdimension of "economy," we further distinguished between more direct (cf. highly concentrated) and more aggregated (cf. weakly concentrated) benefits or harm.

The survey to pre-test the corresponding items was conducted between December 20th, 2016, and January 4th, 2017, with 32 participants (12 males and 20 females, mean age of 23.6 years), all but two of whom were students from Switzerland; thus, the sample had characteristics similar to that of the main experiment. From an initial set of 92 items (46 each for the practice and the organization), we chose at this early stage to keep 62 items (31 each). We based the item selection at this stage on two main criteria: (1) a high correlation of the item with an overall measure for legitimacy and (2) a high Cronbach's alpha of the items within a legitimacy (sub)dimension. The overall measure for legitimacy consisted of a compound measure of the participants' indicated agreement with the two items "I think [tax optimization/the Big Four's behavior] is legitimate and appropriate" and "I think [tax optimization/the Big Four's behavior] is illegitimate and

inappropriate.” All items were rated on an 11-point Likert scale (-5=totally disagree, 0=neither agree nor disagree, +5=totally agree). This was done to make sure the set of items and related dimensions seem indeed to be relevant for an overall judgment on whether an organization or practice is appropriate (cf. content validity) and to assure the items add to the internal consistency of a dimension. The resulting of items can be found in Table A1. Due to time constraints, we decided to include all these remaining items (Table A1) in our experiments to measure legitimacy and reduce the items and create our final scale afterwards based on further analyses and studies. Note that all results reported in the paper are robust to using a composite score of all recorded items.

Item Reduction and Scale Development

To develop a generalizable scale to assess propriety judgments for both a publically criticized business practice or organization, we decided to remove the items of the subdimension of pragmatic legitimacy relating to the benefits for the government due to a lack of generalizability. We also removed the items of the subdimension of cognitive legitimacy relating to general knowledge about the legitimacy object, as it was found to hardly correlate with other legitimacy dimensions across samples and seems hardly relevant for the overall judgment of whether a legitimacy object is perceived as proper (judging by the correlations with our overall compound measure).

To further reduce the number of items, we conducted exploratory factor analyses based on the samples of our two main experiments. Similar to Alexiou and Wiggins (2018) we used principal axis factoring with oblimin rotation and in line with usual recommendations we retained items with factor loadings greater than .4 and cross loadings less than .3. We also played with the number of factors to be retained to examine on how many different factors the items would load in accordance with the assumed dimensions of legitimacy. In the end, we only kept items that showed high loadings and low cross loadings across samples and across type of legitimacy object (i.e., items that worked for both, measuring the legitimacy of the practice as well as the organization). Those items were found to load on six different factors (cf. Table 1).

Scale Validation

To confirm and validate our scale we conducted an online survey in August 2019 with 175 subjects (48.6% female, mean age of 23.4 years, only one non-student). The subjects of the online survey were asked to read the same information material on the Big Four and global tax avoidance as the study participants of our two main experiments. Afterwards, they rated the legitimacy of the

Big Four as well as the legitimacy of tax optimization using our own scale and the legitimacy of the Big Four using the scales of Alexiou and Wiggins (2018), Bitektine et al. (2018), and Elsbach (1994) for comparison.

A confirmatory factor analysis confirmed the previously established structure of the scale. The resulting factors and factor loadings are captured by Table 1. The discriminant validity between the six dimensions in the scale was assured using the well-accepted criterion by Fornell and Larcker (1981). Accordingly, the average variance extracted for each factor exceeded the squared correlation of the factors.

Our scale significantly correlates with previous measures (Alexious & Wiggins, 2018: $r=.80$, $p<.001$; Bitektine et al., 2018: $r=.61$, $p<.001$; Elsbach, 1994: $r=.51$, $p<.001$), whereas it demonstrates better psychometric properties. Overall goodness of fit statistics suggest that the data fits our model well (our scale for the Big Four: $\chi^2_{75}=119.90$, $p<.01$; RMSEA=.058; CFI=.976; SRMR=.038; our scale for tax optimization: $\chi^2_{75}=146.67$, $p<.01$; RMSEA=.074; CFI=.947; SRMR=.055) and better than previous measures (Alexious & Wiggins, 2018: $\chi^2_{51}=83.52$, $p<.01$; RMSEA=.060; CFI=.958; SRMR=.065; Bitektine et al., 2018: $\chi^2_8=7.2$, $p=.51$; RMSEA=0; CFI=1; SRMR=.044; Elsbach, 1994: $\chi^2_{54}=256.46$, $p<.001$; RMSEA=.146; CFI=.537; SRMR=.135). Note that values of up to .08 for the root mean square error of approximation (RMSEA) and the standardized root-mean-square residual (SRMR) are considered acceptable, as well as values greater than .90 for the confirmatory fit index (CFI) (Bagozzi & Yi, 1988; Hu & Bentler, 1999). In addition, our scale showed a higher reliability within and across legitimacy dimensions than previous scales (our scale for the Big Four: $\alpha=.93$; Alexious & Wiggins, 2018: $\alpha=.83$; Bitektine et al., 2018: $\alpha=.56$; Elsbach, 1994: $\alpha=.69$). Taken together, these results confirm that our scale provides a valid measure and an improvement to previous measures.

Table A1 – Complete List of Assessed Items

Dimension of Legitimacy		Items		
		Organization (Big Four)	Practice (Corporate Tax Optimization)	
Cognitive Legitimacy	Taken-for-Grantedness	The Big Four are needed. The Big Four are necessary corporations. The criticism of the Big Four is justified. I agree with the criticism of the Big Four.	Tax optimization is needed. Tax optimization is a necessary service. The criticism of tax optimization is justified. I agree with the criticism of tax optimization.	
	Knowledge	I am familiar with the Big Four. I know a lot about the Big Four.	I am familiar with the service of tax optimization. I know a lot about tax optimization.	
	Comprehensibility / Understanding	I can understand the behavior of the Big Four. I have no understanding of the behavior of the Big Four. The behavior of the Big Four is understandable.	I can understand the use of tax optimization. I have no understanding of the use of tax optimization. The use of tax optimization is understandable.	
Moral Legitimacy	Moral Integrity	The Big Four's behavior is appropriate. The Big Four's behavior is desirable. The Big Four's behavior is morally acceptable. The Big Four's behaviour is morally reprehensible. The Big Four's behavior is inappropriate.	Tax optimization is appropriate. Tax optimization is desirable. Tax optimization is morally acceptable. Tax optimization is morally reprehensible. Tax optimization is inappropriate.	
	Approval of Criticism	The criticism of the Big Four is justified. I agree with the criticism of the Big Four. The Big Four's behavior is problematic.	The criticism of tax optimization is justified. I agree with the criticism of tax optimization. Tax optimization is problematic.	
Pragmatic Legitimacy	Economy	Direct Benefit	The Big Four are good employers. The Big Four are helping corporations.	Tax optimization creates good jobs. Tax optimization helps corporations.
		Aggregated Benefit	The economy benefits from the Big Four. The Big Four hurt the economy. The Big Four hurt competition. The Big Four are value-creating corporations.	The economy benefits from tax optimization. Tax optimization hurts the economy. Tax optimization hurts competition. Tax optimization is a value-creating service.
	Society	The Big Four benefit society. The Big Four contribute positively to society. The Big Four hurt society.	Tax optimization benefits society. Tax optimization contributes positively to society. Tax optimization hurts society.	
	State	The Big Four are useful for governments. States benefit from the Big Four. The Big Four hurt the state.	Tax optimization is useful for governments. States benefit from tax optimization. Tax optimization hurts the state.	
	Responsiveness	The behavior of the Big Four is compatible with public interests. Public interests are taken into account by the Big Four.	Tax optimization is compatible with public interests. Public interests are taken into account when optimizing taxes.	
Overall Legitimacy		I think the Big Four and their actions are legitimate and appropriate. I think the Big Four and their actions are illegitimate and inappropriate.	I think tax optimization is legitimate and appropriate. I think tax optimization is illegitimate and inappropriate.	

Appendix B: Additional Tables

Table B1 – Assignment of Representatives and Moderators to Treatments

Day 1 (30.01.2017)				
Treatment	Control	Firm	NGO	Mixed
Room	E-27	E-08	E-21	E-12
Session #1	M1	M2 F1	M3 N1	M4 F2 N2 f
Session #2	M4	M3 F2	M2 N2	M1 F1 N1 n
Session #3	M3	M4 F1	M1 N2	M2 F2 N1 f
Session #4	M2	M1 F2	M4 N1	M3 F1 N2 f
Session #5	M1	M2 F1	M3 N1	M4 F2 N2 n
Day 2 (31.01.2017)				
Room	E-12	E-21	E-08	E-27
Session #6	M3	M4 F4	M1 N4	M2 F3 N3 n
Session #7	M2	M1 F3	M4 N3	M3 F4 N4 f
Session #8	M4	M3 F3	M2 N4	M1 F4 N3 n
Session #9	M1	M2 F4	M3 N3	M4 F3 N4 n
Session #10	M3	M4 F4	M1 N4	M2 F3 N3 f

M1-M4: Moderators

F1-F4: Representatives of the Big Four (employees of a Big Four firm)

N1-N4: Representatives of critics of the Big Four (members of a Swiss NGO)

f or n: F (firm) or n (NGO) indicates who starts with introductory statement at the first question

Note: The representatives and moderators are arranged such that: each moderator encounters each representative at least once; each representative takes part in each treatment at least twice; each moderator switches treatments after each session and chairs each treatment at least once per day; the combinations in the first session of the day are repeated in the last session of the day to control for learning effects; the order of introductory statements is reversed between days; and the treatment-room combinations are rotated between days.

Table B2 – Logit Regressions:
% of Viewing Firm or NGO Representative as More (or Equally) Convincing

	(1)	(2)	(3)
	%mixed pro firm	%mixed pro NGO	%mixed tie
Political orientation (-5=left, +5=right)	0.634*** (0.005)	-0.308* (0.063)	-0.113 (0.523)
Position towards Big Four (-5=contra, +5=pro)	0.906*** (0.002)	-0.827*** (0.001)	0.102 (0.530)
Economics	-3.439** (0.011)	2.174** (0.033)	0.629 (0.453)
Session (discussion group)	0.148 (0.299)	-0.301** (0.021)	0.151 (0.185)
Constant	-0.492 (0.597)	0.310 (0.689)	-2.644*** (0.002)
<i>N</i>	70	70	70
pseudo <i>R</i> ²	0.503	0.398	0.059

p-values in parentheses

* *p* < 0.10, ** *p* < 0.05, *** *p* < 0.01

Table B3 – Robustness Check: Controlling for Exogenous Factors

	Change in Legitimacy of Big Four		Change in Legitimacy of Tax Optimization	
	(1) implicit	(2) explicit	(3) implicit	(4) explicit
Mixed tie	-0.352 (0.317)	0.679* (0.098)	-0.856* (0.061)	-0.381 (0.333)
Mixed adv. NGO	-0.0485 (0.836)	-0.230 (0.545)	-0.359** (0.031)	-0.817** (0.030)
Mixed adv. firm	0.772* (0.083)	1.628*** (0.000)	0.0103 (0.959)	0.0896 (0.829)
Firm	0.609*** (0.002)	0.741** (0.046)	0.359** (0.016)	0.424 (0.234)
NGO	-0.416*** (0.003)	0.153 (0.676)	-0.595*** (0.001)	-0.434 (0.224)
Wait	0.178 (0.289)	0.719** (0.048)	-0.0668 (0.696)	0.204 (0.558)
Position towards Big Four	0.0517 (0.217)	0.0928 (0.190)	0.0610* (0.099)	0.185*** (0.009)
Political orientation	-0.0222 (0.696)	-0.00618 (0.931)	0.0197 (0.640)	0.0764 (0.270)
Constant (control)	-0.192 (0.162)	-0.827*** (0.004)	0.0370 (0.771)	0.180 (0.508)
<i>N</i>	66	66	66	66

OLS regressions with robust standard errors where necessary (Breusch-Pagan test with *p* < 0.05)
p-values in parentheses, * *p* < 0.10, ** *p* < 0.05, *** *p* < 0.01

Table B4 – Selection of Videos for Study 2

treatment	session	firm representative				NGO representative				Change in Legitimacy				position of others towards	
		id	persuasive	competent	Ø	id	persuasive	competent	Ø	implicit		explicit		Organization	Practice
										Organization	Practice	Organization	Practice	Organization	Practice
firm condition	1	F1	0.29	1.43	0.86	0.29	0.89	-0.14	0.57	-1.71	-1.57
	2	F2	-0.43	2.00	0.79	0.43	0.00	-0.57	0.00	-0.71	-0.86
	3	F1	-0.29	2.71	1.21	0.50	0.30	-1.29	-0.71	-2.00	-2.00
	4	F2	-0.29	2.14	0.93	0.90	0.65	0.14	0.14	-0.57	-0.71
	5	F1	1.14	4.00	2.57	0.55	0.20	-0.43	-0.29	-1.00	-0.57
	6	F4	2.71	3.57	3.14	0.79	0.15	0.29	1.00	0.29	-0.57
	7	F3	1.43	4.00	2.71	1.00	0.92	0.57	0.57	-0.71	-0.57
	8	F3	-1.43	2.71	0.64	-0.48	0.20	-1.14	-0.29	-2.43	-2.29
	9	F4	1.29	2.86	2.07	0.55	0.42	0.14	0.43	0.86	1.14
	10	F4	1.86	2.71	2.29	1.56	-0.11	0.43	1.00	-0.43	-0.57
NGO condition	1	N1	2.29	4.29	3.29	-0.45	-1.05	-0.29	0.43	-1.00	-1.00
	2	N2	2.29	3.43	2.86	-0.46	-0.71	-0.71	-0.86	-2.71	-2.43
	3	N2	3.14	3.71	3.43	-0.63	-1.35	-1.14	-1.00	-2.57	-3.14
	4	N1	2.57	3.29	2.93	-0.32	-0.53	-0.57	-1.29	-1.00	-1.71
	5	N1	4.00	4.43	4.21	0.20	-0.31	-0.57	-0.86	-1.43	-1.71
	6	N4	3.43	4.14	3.79	-0.16	-0.16	-2.00	-1.00	-3.14	-3.29
	7	N3	2.57	3.00	2.79	-0.51	-0.24	-0.71	1.00	-2.86	-2.43
	8	N4	3.86	4.29	4.07	-1.10	-0.81	-0.71	-1.29	-2.71	-3.43
	9	N3	3.71	4.14	3.93	-0.21	-0.69	-1.00	-0.57	-3.29	-2.57
	10	N4	3.00	3.14	3.07	-0.22	-0.34	-0.14	-0.71	-3.14	-3.29
mixed condition	1	F2	-2.14	2.14	0.00	N2	3.00	3.71	3.36	-0.61	-0.54	-1.71	-0.43	-2.86	-3.00
	2	F1	-0.43	2.43	1.00	N1	3.29	3.29	3.29	-0.41	-0.62	-1.14	-1.86	-1.86	-2.00
	3	F2	-0.14	2.86	1.36	N1	3.57	4.43	4.00	-0.15	-0.02	-0.86	-0.86	-3.29	-3.00
	4	F1	1.57	3.86	2.71	N2	1.14	2.00	1.57	0.15	-0.30	0.14	0.14	-0.86	-1.00
	5	F2	1.00	2.86	1.93	N2	2.29	2.71	2.50	0.36	-0.45	-1.14	-0.29	-1.57	-1.71
	6	F3	2.86	4.29	3.57	N3	2.57	3.71	3.14	0.42	0.05	-0.14	-0.86	-0.71	-1.57
	7	F4	1.29	2.86	2.07	N4	3.14	3.86	3.50	-0.64	-0.65	-1.57	-0.86	-2.43	-1.86
	8	F4	-1.00	0.86	-0.07	N3	2.14	2.43	2.29	0.05	-0.09	0.00	-0.14	-1.43	-1.00
	9	F3	4.29	4.57	4.43	N4	2.00	3.29	2.64	0.31	0.07	1.29	0.71	0.57	1.57
	10	F3	1.14	2.57	1.86	N3	1.57	1.86	1.71	0.04	0.06	-1.00	-0.86	-0.57	-0.14

Note: Marked sessions were picked for showing in Study 2.

Appendix C: Additional Figures

Figure C1 – Treatment Effects on Social Media Discussions

