Fears of peers? Explaining peer and public shaming in global governance

Valentina Carraro*, Thomas Conzelmann and Hortense Jongen

Abstract
This article conducts a comparative analysis of peer and public pressure in peer reviews among states. Arguing that such pressure is one increasingly important form of shaming in global politics, we seek to understand the extent to which five different peer reviews exert peer and public pressure and how possible variation among them can be explained. Our findings are based on responses to an original survey and semi-structured interviews among participants in the reviews. We find that peer and public pressure exist to different degrees in the peer reviews under study. Such differences cannot be explained by the policy area under review or the international organization in which peer reviews are organized. Likewise, the expertise of the actors involved in a peer review or perceptions of the legitimacy of peer review as a monitoring instrument do not explain the variation. Instead, we find that institutional factors and the acceptance of peer and public pressure among the participants in a peer review offer the best explanations.

Keywords
Anti-corruption policy, human rights, international authority, naming and shaming, peer review, trade and economic policy

Introduction
In an interview with the Financial Times in 2011, Mark Pieth, then Chairman of the Working Group on Bribery (WGB) in the Organisation for Economic Cooperation and Development (OECD), issued a warning to the UK. The WGB’s peer review, which monitors states’ performance in combating foreign bribery, had exposed substantial shortcomings and implementation delays in the UK’s legislation. In a public statement, Pieth eventually warned that the WGB ‘would consider “robust options” should the legislation be held up further, including the blacklisting of UK exporters’ (Boxell and Rigby, 2011).
This instance of peer and public pressure, mobilized by the WGB and other states, ultimately led the UK to speed up implementation of its Bribery Act.

In another episode, Thailand was faced with civil society pressure after it appeared in front of the Universal Periodic Review (UPR) in May 2016. The UPR is a peer review organized by the United Nations (UN) Human Rights Council to improve the global observance of human rights in all member states. In the UPR, states make recommendations to each other to address deficits in their human rights performance, which the reviewed state can either accept or simply ‘take note of’, which is a euphemism for disagreeing. Following the UPR meeting for Thailand in May 2016, Thai civil society began to publicly pressure the government to accept more recommendations than the country had initially intended. In the end, the government accepted several additional recommendations during the final adoption of its UPR report in September 2016 and committed to a national dialogue concerning the implementation of outstanding recommendations (UPR-info, 2016).

As the two episodes indicate, peer reviews among states can both ‘name’ and ‘shame’ transgressors. They can generate pressure by the ‘peers’, that is, the delegates and experts from other states, as in the case of the WGB, or can create public pressure, as shown in the UPR example (also see Nance, 2015; Tanaka, 2008; Terman and Voeten, 2017). The increasing use of peer reviews as a tool for monitoring international agreements means that naming and shaming through peer reviews may be observed more frequently in the future, which makes the study of this instrument relevant. Not all peer reviews, however, seem capable of exerting peer and public pressure on transgressors (Abebe, 2009; Greene and Boehm, 2012); strong political dynamics enter the field (Carraro, 2017b; Gutterman and Lohaus, 2018; Terman and Voeten, 2017), and some researchers even portray peer reviews as generally incapable of generating effects on domestic policy (Schäfer, 2006).

This article discusses the potential of peer reviews among states as naming and shaming instruments. While some authors have looked at individual cases (Abebe, 2009; Greene and Boehm, 2012; Nance, 2015; Terman and Voeten, 2017), we present the first comparative analysis of naming and shaming in peer reviews. Based on an analysis of five peer reviews in different international organizations (IOs) and policy fields, we discuss the extent to which they exert peer and public pressure. We research two peer reviews in the OECD, namely the WGB and the Economic and Development Review Committee (EDRC); two UN peer reviews, namely the UPR of human rights and the Implementation Review Mechanism (IRM) of the UN Convention against Corruption (UNCAC); and the Trade Policy Review Mechanism (TPRM) of the World Trade Organization (WTO).

The subsequent section reviews the literature on naming and shaming and on peer reviews in IOs. It establishes that the peer and public pressure exerted by peer reviews can be seen as instances of naming and shaming. Next, hypotheses are formulated about factors that may affect the exertion of peer and public pressure in peer reviews. Subsequently, we discuss our data, which include 85 semi-structured interviews, online documents and 375 responses to an original survey distributed to IO staff, diplomats and domestic civil servants participating in the five peer reviews.

We find that peer and public pressure are exerted to varying degrees in these mechanisms: the WGB and UPR are best capable of organizing peer and public pressure on
reviewed states, whereas the IRM is lagging behind, and the EDRC and TPRM are intermediate cases. The extent to which peer and public pressure take place does not systematically vary according to the policy field and the IO in which the peer reviews are organized. Rather, the institutional design of a peer review, and in particular the specificity of its recommendations, its transparency and the existence of follow-up monitoring provide convincing explanations for the variation in the reviews’ ability to generate peer and public pressure. Similarly, legitimacy perceptions of the acceptability of peer and public pressure in peer reviews further explains the reviews’ ability to generate pressure on states.

**Naming, shaming and peer reviews**

Naming and shaming is a social process that brings together three different kinds of actors: the *agents* of shaming (who shames), the *targets* of shaming (those being shamed) and the *audience* (which amplifies the social pressure on the target if it agrees with the shaming exercise). Naming and shaming thus depend on the audience’s disapproval of the target’s behaviour, and the audience’s support for exerting pressure on the target. Social opprobrium plays the key role, and not material sanctioning. Specifically, we define naming as the process of classifying certain behaviour as falling inside or outside of certain behavioural expectations. Shaming means to publicly denounce an actor and its behaviour, in the expectation that the social discomfort of being reprimanded pushes states towards compliance (Franklin, 2015: 44; Keck and Sikkink, 1998). Within the abovementioned definition, there can be empirical variation in the agents (states, IOs, specific bodies within IOs), the targets (states, firms, individuals) and the audience (the global or domestic public, financial institutions, international organizations or a smaller ‘in camera’ audience if a transgression is discussed among peers behind closed doors).

Peer reviews among states have potential for both naming and shaming. They build on the regular assessment of information on the policy performance and compliance of states by the IO secretariat and other states (the ‘peers’). Most peer reviews end with some praise for the reviewed member, but also recommendations to address certain policy shortcomings, thus ‘naming’ behaviour that falls outside acceptable standards. In the next step, the community of reviewing states may use ‘shaming’ to target states that fall behind expectations and make these states heed the recommendations received. This can be done in a smaller circle of peers by demanding that recommendations are addressed until a certain deadline, by revisiting review recommendations during the next review cycle or by not allowing laggards to move on to the next review phase. Some reviews combine this peer pressure with public pressure, exerted by publishing review documents and press releases online, or by organizing public events to present review outcomes. Pressure on laggards is enhanced if specific countries are singled out as poor performers, or ‘blacklisted’ (Nance, 2015; Sharman, 2009). Being assessed and possibly reprimanded by the peers or the public therefore constitutes a form of naming and shaming (also see Greene and Boehm, 2012; Terman and Voeten, 2017).

However, the extent to which specific peer reviews are able to mobilize shame is an empirical question. Not all peer reviews are purposely designed to pressure and shame states, or may not be used as such. Our analysis demonstrates that the peer reviews under
study exert peer and public pressure to different degrees. How can such variation be understood?

**Hypotheses**

In studying naming and shaming, many international relations scholars have focused on the effects of naming and shaming on targets and their motivation to give in to pressure. Rationalists and liberal scholars point out that the target might succumb to pressure in order to maintain its reputation and not to forego specific benefits (DeMeritt, 2012: 602–603; Krain, 2012; Murdie and Davis, 2012). Constructivists focus on the signalling function of shaming and socialization processes and point out that successful shaming depends on the aspiration of the target to be accepted by the community of peers (Risse and Sikkink, 1999: 15). Another strand of scholarship focuses on the act of naming and shaming, and the strategic considerations of the agent. This literature has, for instance, focused on the decisions by IOs and non-state actors to address specific transgressions and specific targets (Lebovic and Voeten, 2006, 2009; Murdie and Urpelainen, 2015). An implicit assumption in this scholarship is that not only are shamers unitary rational actors that select the most promising or worthy targets for shaming efforts, but that they are also able to shame.

This last issue is the one on which we focus in the next steps. We assume that the capability and readiness to exert peer and public pressure depends on specific conditions. We hypothesize that such conditions may be located on three distinct levels: firstly, the contexts provided by policy fields and the respective IOs in which the peer review is organized; secondly, specific institutional design features of the reviews that make the exertion of pressure possible; and, thirdly, the extent to which the practice of exerting pressure and the expertise of reviewers is seen as appropriate. We discuss each of these conditions below.

**Organizational and policy contexts**

While there is some literature on shaming by single IOs (Hafner-Burton, 2008; Krain, 2012; Lebovic and Voeten, 2009; Nance, 2015), there is little theoretical reflection on which IOs are more likely to shame. One recent contribution (Squatrito et al., 2017) argues that IOs with large memberships may shame more frequently, simply because the potential targets of shaming and the potential violations are more numerous than in smaller IOs. A constructivist argument pointing in the same direction is that larger IOs are less likely to create shared identities and feelings of trust and solidarity between member states. Such shared identities are, however, important to socialize states into rule-conforming behaviour (Checkel, 2001). Shaming, therefore, is one of the options that larger organizations have to push recalcitrant members to comply. As observed by Johnston (2001: 502–503), the benefits of being famed as a leader or the costs of being shamed as a laggard increase with group size, making the strategy of shaming particularly effective in larger organizations. The distinction between public pressure and peer pressure made above is relevant in this context: IOs with a smaller membership may more effectively use ‘in camera’ (peer) pressure to socialize members and to bring them in line with common
standards, while larger IOs will more frequently resort to public pressure. We would therefore expect public pressure to be more prevalent in reviews with larger memberships, that is, in the peer reviews housed by the WTO and the UN. In turn, shaming should be less prevalent in the OECD reviews.

A further contextual factor is the policy field in which reviews are organized. Naming and shaming can only work if the norms and rules that states are expected to comply with are widely accepted (Pawson, 2002); otherwise, shaming targets may, in some cases, challenge and even transform a dominant moral discourse (Adler-Nissen, 2014). As concerns the three policy areas under review in this article, there are fairly limited differences in terms of norm acceptance. Franklin (2015: 45) observes that human rights norms are still fairly broadly accepted, while there are also attempts by some states to question the universality of human rights and to prevent intrusive human rights monitoring (Inboden and Chen, 2012; Carraro, 2017a: 21–22). Similarly, Gutterman and Lohaus (2018) find that the global anti-corruption norm ‘appears robust in terms of public acceptance, international treaty ratification, and institutionalization’, but also observe that less economically developed states in particular engage in ‘applicatory contestation’ (pp. 251, 256). There is broad acceptance of liberal economic norms and a firm institutionalization of IOs that foster free trade and liberalization (Simmons et al., 2006), while the plethora of cases in front of the WTO Dispute Settlement Body also show a considerable degree of contestation over norm application. The limited differences between the three policy areas under research lead us to expect no strong divergences in the existence of peer and public pressure among the three policy fields. In any case, differences between reviews in the same policy field should be small.

Institutional design

A second set of hypotheses relates to institutional features of peer reviews that may facilitate exerting pressure on transgressors. We loosely follow the discussion in the rational legalization literature (Abbott et al., 2000; Koremenos et al., 2001) by distinguishing specific design features of the reviews, but with two important modifications. Firstly, we are interested in the effects of institutional provisions, not the reasons why they have been designed in a specific way. Secondly, we neither assume that (formal) institutional design features fully determine participant interaction within reviews, nor that state actors use institutions to their full potential. As argued below, appropriateness perceptions of peer and public pressure may inhibit the extent to which they are exerted, even if all institutional conditions are in place. Likewise, appropriateness perceptions may facilitate pressure even under adverse institutional circumstances (also see Wendt, 2001).

We focus on the following three institutional aspects (also see Pawson, 2002). Firstly, how specific or unspecific are recommendations to the reviewed state? We expect that the exertion of peer and public pressure is facilitated if transgressions are clearly defined and recommendations are clearly formulated. In their attempt to exert peer and public pressure, peer reviews cannot risk ambiguity in the conduct they require.3 We assess this measure qualitatively by looking at the recommendations that emanate from the review exercises. Secondly, how transparent are the peer reviews to the outside world? We
assess transparency by looking at the public availability of review documents as well as the openness of plenary meetings. Review documents such as country reports and recommendations may only be shared among state delegates, in which case we expect peer pressure to dominate. Review documents can also be published more widely online. Furthermore, in some peer reviews the publication of all review documents is voluntary, whereas in others it is (partially) mandatory. Transparency can also be increased by web-casting review sessions, as in the UPR. We expect that transparent reviews will attract more public attention, and will be more likely to trigger public pressure (see Carraro and Jongen, 2018). Thirdly, is there a possibility during reviews to assess whether states have implemented recommendations from the previous round? Such follow-up monitoring offers opportunities to criticize noncompliant behaviour, and is often delegated to the secretariats of the peer reviews. Due to its largely technical nature, follow-up monitoring is, however, more relevant for peer than for public pressure, with the exception of the public denouncement of states in cases of persistent non-implementation of recommendations. Empirically, we distinguish between formalized follow-up procedures, in which states are required to report on progress made, and informal ad hoc practices in which previous review results may be brought up, depending on the initiative of individual member states. Some peer reviews lack a system for follow-up monitoring altogether. We hypothesize that follow-up monitoring primarily facilitates peer pressure, but may also have some effects on public pressure.

**Legitimacy perceptions**

Even if institutional preconditions for naming and shaming are in place, the exertion of peer and public pressure may not be socially accepted. As pointed out by Pagani and Wellen, ‘these methods are appropriate and produce positive results only when the “rules of the game” are clear and the countries accept them’ (2008: 263). Further, the shamer needs to have ‘established (legal and moral) authority’ (Pawson, 2002: 225). Hafner-Burton suggests that naming and shaming in the human rights area was ‘unproductive’ during the early 2000s, as ‘NGOs [non-governmental organizations] and the media lack authority over states and the UNCHR [the former UN Commission on Human Rights], packed full of despots, lacks legitimacy’ (2008: 691). To cover this dimension, we discuss two elements. Firstly, we research perceptions of the legitimacy of exerting peer and public pressure. Some scholars have warned that peer reviews might degenerate into a ‘condemnatory system of oversight’ (Abebe, 2009: 3; also see Comley, 2008: 122–124). We expect peer and public pressure to be inhibited if they are not widely deemed legitimate. Secondly, we research how the expertise of the IO staff and state representatives involved in peer reviews is assessed by participants. Higher levels of perceived expertise are expected to positively contribute to the exertion of peer and public pressure.

Table 1 gives an overview of the factors that we hypothesize to be conducive to peer or public pressure. The peer reviews under study are used to provide an explorative assessment of the relevance of each factor for the observed outcome. The fact that we study a limited number of peer reviews does not allow a true empirical test of our hypotheses. Still, the discussion provides evidence for the plausibility of some of the presumed causal links. Moreover, our research design does not consider the effects of peer and public pressure on
domestic policy – that is, whether specific instances of naming and shaming have actually led to a behavioural change. Such an endeavour would require a much more encompassing study looking at domestic policy change in different legislations.

**Table 1. Potential explanatory factors.**

| Categories                                      | Factors                                                                 |
|------------------------------------------------|--------------------------------------------------------------------------|
| Organizational and policy contexts             | Large size of the IO (*facilitates public pressure*)                     |
|                                                | Small size of the IO (*facilitates peer pressure*)                       |
|                                                | Broadly accepted norms that allow the mobilization of shame against transgressors (*facilitates peer and public pressure*) |
|                                                | Recommendations are specific (*facilitates peer and public pressure*)     |
| Institutional opportunities for peer and public pressure | High transparency of the review procedure and its outcomes to outsiders (*facilitates public pressure*) |
|                                                | Formalized possibilities for follow-up exist (*facilitates peer and public pressure*) |
| Legitimacy perceptions pertaining to reviewers and the exertion of pressure | The exertion of peer and public pressure is perceived as appropriate (*facilitates peer and public pressure*) |
|                                                | The IO secretariat and member state officials are perceived to hold the necessary expertise (*facilitates peer and public pressure*) |

IO: international organization.

**Peer and public pressure in peer reviews**

Our empirical analysis is based on data collected by means of a web-based survey with 375 distinct observations and 85 semi-structured interviews. As in the survey, interviews targeted the officials who are directly involved in the reviewing mechanisms, namely secretariat officials, state delegates and national experts in the case of the IRM and WGB. Survey and interview findings allow us to understand the extent to which peer and public pressure are exerted in the peer reviews and to research legitimacy perceptions. Interviews helped to reconstruct the causal mechanisms linking naming and shaming and the explanatory factors discussed above.

We assess the extent to which peer and public pressure exist in the five reviews under scrutiny through a number of survey questions. Participant perceptions of whether peer and public pressure is exerted offer more relevant information and are more feasible to research than whether alleged norm violations have been taken up by the media or in NGO reports (see Hafner-Burton, 2008; Murdie and Davis, 2012; Murdie and Urpelainen, 2015). On the one hand, peer pressure happens during plenary sessions, which are not open to the public in our cases, except for the UPR. On the other hand, public pressure triggered by peer reviews is usually exerted at the domestic (as opposed to the transnational) level. It is practically unfeasible to assess the extent to which local media or NGOs in all member states that participate in the review are taking up review recommendations.
We asked survey respondents to what extent they believe that peer and public pressure is exerted in the peer review in which they participate. Answer options were as follows: 1 = not at all; 2 = to some extent; 3 = to a large extent; 4 = completely. ‘I do not know’ answers treated as item non-response. The analyses find clear differences between the five mechanisms: the peer review in which respondents participate has a statistically significant effect on their assessment of peer and public pressure ($\eta^2 = 0.10$ for peer pressure and 0.13 for public pressure, $p < 0.001$). The WGB is perceived as best able to organize peer pressure (Mean value ($M$) = 2.92) (Table 2), showing statistically significant differences with the IRM and the EDRC in the pairwise comparisons ($p < 0.001$ for both cases). Many interviewees in the WGB extensively discussed peer pressure (Interviews CO6 2, 3, 5, 6, 25, 26, 27, 28). Some state delegates reportedly hold each other accountable for the progress their countries have made on implementing the Anti-Bribery Convention and exert pressure on underperforming states. Also, the UPR is perceived as successful in generating peer pressure ($M = 2.64$), performing significantly better than the IRM ($p < 0.05$). The recommendations received by states under review are generally perceived as politically binding, because they were issued by a fellow government (Interviews HR 1, 2, 3, 4, 8, 22, 26, 28, 29, 30, 31, 32, 34, 38, 39). The TPRM ($M = 2.59$) lags somewhat behind, while the IRM ($M = 2.30$) and the EDRC ($M = 2.34$) are viewed as least able to organize peer pressure. Interviewees indicated that states are not critically questioned on their performance in IRM plenary sessions (Interviews CO 1, 4, 13, 26), and that the EDRC can be better understood as a framework to stimulate open discussion and learning (Interviews ET 4, 6, 11).

The UPR and the WGB are perceived as best able to generate public pressure ($M = 2.61$ for both reviews; Table 2). Civil society actors reportedly play a crucial role in generating public pressure: they are directly involved in the country reviews and hold governments accountable for the recommendations that they have accepted (Interviews HR 3, 10, 11, 12, 13, 28, 30). In the WGB, civil society is not present when review reports are discussed and adopted; however, interviewees mentioned that review reports at times receive attention by the media and NGOs, such as Transparency International (Interviews

---

**Table 2.** The perceived ability of the peer reviews to generate peer and public pressure.$^7$

|                   | Total | WGB | IRM | UPR | EDRC | TPRM |
|-------------------|-------|-----|-----|-----|------|------|
| Peer pressure***  | 2.54  | 2.92| 2.30| 2.64| 2.34 | 2.59 |
| Public pressure***| 2.27  | 2.61| 2.10| 2.61| 2.29 | 1.87 |

**Survey item:** To what extent do you believe that the [peer review] successfully...
- Peer pressure: ... exerts state-to-state (peer) pressure
- Public pressure: ... exerts public pressure

**Answer options:** 1 = not at all; 2 = to some extent; 3 = to a large extent; 4 = completely. ‘I do not know’ answers treated as item non-response.

**Note:** One-way analysis of variance (Games–Howell post hoc test).

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

WGB: Working Group on Bribery; IRM: Implementation Review Mechanism; UPR: Universal Periodic Review; EDRC: Economic and Development Review Committee; TPRM: Trade Policy Review Mechanism.
Both peer reviews differ significantly from the IRM\(^8\) and from the TPRM \((p < 0.001)\). The IRM \((M = 2.10)\) and the TPRM \((M = 1.87)\) are commonly perceived as the least capable of organizing public pressure. In the IRM, several officials mentioned that they had barely observed any public pressure in their countries. Others reported on some instances in which the media or NGOs expressed interest in the review outcomes (Interviews CO 1, 13, 18). Finally, the EDRC represents a middle case \((M = 2.29)\), performing better than the TPRM \((p < 0.05)\).

We conclude that the WGB takes the lead on peer pressure, followed by the UPR and the TPRM. In contrast, the EDRC and especially the IRM are perceived as less capable of organizing peer pressure. In terms of public pressure, the UPR and the WGB are perceived to be best able to achieve it and the EDRC is a middle case, while both the IRM and the TPRM are lagging behind.

**Understanding peer and public pressure**

Returning to our hypotheses, we find that organizational and policy context do not have a strong impact on these scores. The two corruption cases (WGB, IRM) and the two economics and trade cases (EDRC, TPRM) show statistically significant divergences for the existence of peer and public pressure despite covering similar policy fields. Likewise, the peer reviews organized in the OECD (the WGB and the EDRC) and the UN (IRM and UPR) show strongly divergent results. Hence, the next section looks for alternative explanations for these differences.

**Institutional opportunities**

As discussed above, we expected three institutional features to facilitate the exertion of peer and public pressure: (a) the specificity of recommendations; (b) the transparency of reviews; and (c) possibilities for follow-up.

**Specificity of recommendations.** Most of the WGB and EDRC recommendations are very specific, clearly setting out expectations and shortcomings. Many IRM country review reports also set out recommendations for improvement, but not in all cases. In the UPR, the specificity of recommendations largely varies depending on the state delivering them. Recommendations vary from extremely general or rather action-oriented.

**Transparency to the outside world.** The UPR is definitely the most transparent peer review among our cases (Carraro and Jongen, 2018). All review-related documents are available on the UN website, review sessions are webcast and interested individuals are allowed to attend as members of the public. Likewise, the TPRM is relatively transparent. While no webcasts are available, all documents pertaining to the reviews, including meeting minutes, are published online. The WGB and the EDRC are in-between cases. On the one hand, they are much less transparent than the UPR, as plenary sessions take place in an *in camera* setting. Neither can civil society organizations, the media or the public attend these sessions, nor are minutes of meetings made public. On the other hand, all country review reports and the outcome documents are publicly
available on the OECD website and are complemented with press statements. The OECD actively seeks to draw attention to these reports (OECD Website, n.d.) and, in the case of the EDRC, organizes high-profile launching events in national capitals (Interviews ET 1, 2). For the IRM, only the executive summaries of the reviews are available online, but it is not mandatory for states to publish the full country reports. As in the other reviews, plenary sessions cannot be attended by officials other than UN staff members and state delegates. Transparency to the outside world thus is comparatively low and there exist fewer opportunities for public pressure in the IRM than in the other cases. These institutional features correspond with the strong scores for public pressure for the UPR and the weak scores for the IRM. The three reviews that show only limited transparency (EDRC, TPRM and WGB), however, strongly diverge on the public pressure scores.

Follow-up monitoring. The WGB has a well-developed system for follow-up monitoring (Jongen, 2018). The review process consists of several phases. Each review phase focuses on a different stage, starting with an assessment of the adequacy of domestic legislation to implement the OECD Anti-Bribery Convention, to the effective application of the Convention and ultimately to its enforcement in practice. States cannot proceed to the next review phase unless the other members of the Working Group deem its performance under the previous phase satisfactory. In addition, delegates are expected to update their peers on their progress in implementing recommendations. In the EDRC, it is common to return to previous review exercises. In fact, this has been made a formal requirement recently (Interviews ET 2, 4), although not in a similarly sophisticated manner as in the WGB. In the UPR, there is no specific system for follow-up, which is left to states’ discretion. Some reviewed states are very open in highlighting the progress made in implementing the recommendations received, but they are under no obligation to discuss these points. Similarly, some reviewing states in the UPR explicitly ask questions to the reviewed regarding their compliance with previous recommendations, but this is equally voluntary. A similar system exists in the TPRM. No mechanism for follow-up monitoring exists in the IRM. In summary, we can identify considerable institutional differences between the five peer reviews (Table 3).

We thus find that the institutional opportunities for peer pressure correspond fairly closely with the factual existence of peer and public pressure in the five reviews. The fairly low degree of peer and public pressure in the IRM corresponds to the very limited institutional opportunities in place to exert such pressure. The absence of a plenary discussion was mentioned as one reason why it is much harder to organize peer pressure in the IRM when compared to the WGB (Interviews CO 4, 13, 30; see also Jongen, 2018). Vice versa, the high peer pressure in the WGB is in line with the very specific recommendations it issues, and with the advanced system for follow-up monitoring, which is recognized to enhance peer accountability and peer pressure (Interviews CO 6, 7, 8). The UPR and the TPRM come in second and third for peer pressure, which is in line with their often broad recommendations and limited follow-up activities. The fact that both still show a relatively high degree of peer pressure seems to be linked to the stronger emphasis on state-to-state recommendations (Interviews HR 1, 2, 10, 11, 12, 13, 24, 26, 28, 36; ET 16, 21, 22, 23, 25). In both the UPR and the TPRM,
questions, demands and recommendations are made by individual states, while the
chair only offers a more general summary of the discussion. The EDRC’s low score on
peer pressure contradicts the frequently specific recommendations and its formal sys-
tem for follow-up. One element in understanding this contradiction is that EDRC rec-
ommendations must be negotiated with the reviewed state, which makes them
consensual and not in need of further enhancement through peer pressure (Interviews
ET 4, 8, 10).

Legitimacy perceptions pertaining to the shamer and shaming

To refine our explanatory model, we study perceptions of the appropriateness of exerting
peer and public pressure and of the expertise of the reviewers as possible explanations
for differences in peer and public pressure between the reviews. We hypothesize that
peer and public pressure are unlikely to be exerted if they are not widely accepted.
Similarly, if the expertise of reviewers is deemed to be low, this may undermine the
legitimacy of the review.

The legitimacy of peer and public pressure. To study the perceived legitimacy of exerting
peer and public pressure, we requested respondents to indicate on a scale of 1–10 whether
they consider peer and public pressure a valuable contribution of a peer review. A score
of 1 indicates that this is not at all valuable, whereas a score of 10 implies it is seen as
extremely valuable. Participants were requested to answer this question without looking
at the specific peer review in which they were involved.

There exists no significant main effect of the peer review in which respondents par-
ticipate on their perceptions of the added value of peer pressure (Table 4), and also the
pairwise comparisons⁹ reveal no statistically significant differences between the reviews.
As for public pressure, the peer reviews do have a significant main effect on perceptions

| Specificity of recommendations | WGB | IRM | UPR | EDRC | TPRM |
|-------------------------------|-----|-----|-----|------|------|
| Very specific                 | Variation across cases | Large variation across cases, more often general than specific | Variation across cases, more often specific than general | Large variation across cases, more often general than specific |

| Transparency | WGB | IRM | UPR | EDRC | TPRM |
|--------------|-----|-----|-----|------|------|
| Medium transparency | Low transparency | High transparency | Medium transparency | Medium transparency |

| Follow-up | WGB | IRM | UPR | EDRC | TPRM |
|-----------|-----|-----|-----|------|------|
| Formal system of follow-up | No follow-up | No formal follow-up, but follow-up may take place | Formal system of follow-up | No formal follow-up, but follow-up may take place |

WGB: Working Group on Bribery; IRM: Implementation Review Mechanism; UPR: Universal Periodic Review; EDRC: Economic and Development Review Committee; TPRM: Trade Policy Review Mechanism.
The effect size is $\eta^2 = 0.06$. Participants in both the WGB ($M = 6.75$) and the UPR ($M = 6.29$) generally appreciate the exertion of public pressure. Perceptions of WGB respondents differ significantly from those involved in the IRM ($M = 5.75$; $p < 0.05$). TPRM respondents appreciate public pressure the least ($M = 4.83$). The scores on the TPRM differ significantly from all other peer reviews: WGB ($p < 0.001$), IRM ($p < 0.05$), UPR ($p < 0.01$), and EDRC ($p < 0.01$).

### Table 4. Views on the extent to which peer and public pressure are valued functions of a peer review.

|         | Total          | WGB  | IRM  | UPR  | EDRC | TPRM |
|---------|----------------|------|------|------|------|------|
| Mean SD | Mean SD        | Mean SD | Mean SD | Mean SD | Mean SD | Mean SD |
| Peer pressure | 6.50 2.40 | 7.10 1.89 | 6.22 2.75 | 6.24 2.86 | 6.38 1.98 | 6.57 2.34 |
| Public pressure | 5.89 2.61 | 6.75 2.13 | 5.75 2.63 | 6.29 2.94 | 6.13 2.37 | 4.83 2.61 |

Survey item: Generally speaking, what would you see as a valuable contribution of a peer review?
- That state-to-state (peer) pressure is exerted
- That public pressure is exerted

Answer options: Scale from 1 (not at all valuable) to 10 (extremely valuable).

Note: One-way analysis of variance (least significant difference post hoc test).

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

WGB: Working Group on Bribery; IRM: Implementation Review Mechanism; UPR: Universal Periodic Review; EDRC: Economic and Development Review Committee; TPRM: Trade Policy Review Mechanism.

of the added value of public pressure ($p < 0.001$). The effect size is $\eta^2 = 0.06$. Participants in both the WGB ($M = 6.75$) and the UPR ($M = 6.29$) generally appreciate the exertion of public pressure. Perceptions of WGB respondents differ significantly from those involved in the IRM ($M = 5.75$; $p < 0.05$). TPRM respondents appreciate public pressure the least ($M = 4.83$). The scores on the TPRM differ significantly from all other peer reviews: WGB ($p < 0.001$), IRM ($p < 0.05$), UPR ($p < 0.01$), and EDRC ($p < 0.01$).

**Perceptions of the expertise of reviewers.** To probe into the perceived expertise of the officials involved in the reviews, we asked respondents to assess the expertise of both staff members of the IO secretariat and member state officials involved in the review, on a scale from 1 (very low degree of expertise) to 4 (very high degree). Generally speaking, the expertise of these actors is assessed as high to very high across all reviews. The peer review in which respondents participate does not have a statistically significant main effect on perceptions of the expertise of the actors involved in the review (Table 5).\(^{10}\) Several differences catch one’s attention. The expertise of the secretariat members involved in the TPRM is assessed to be the highest ($M = 3.62$), and the pairwise comparisons\(^{11}\) reveal statistically significant differences with the UPR and the EDRC ($p < 0.05$). Likewise, the expertise of the WGB secretariat is assessed to be higher than that of the UPR ($p < 0.05$). As for the perceived expertise of the member state officials, the UPR is viewed most negatively ($M = 2.88$), differing significantly from the EDRC ($p < 0.05$).

Linking the findings of this section to the findings on the degree of peer and public pressure in the five peer reviews, two results stand out. Regarding the perceived expertise of member state officials and secretariat staff members, statistically significant differences between the peer reviews were found in some cases. They do, however, not correspond with the degree to which peer and public pressure are experienced to exist in the reviews. Perceptions of the extent to which peer and public pressure are valued functions of a peer review correspond more closely with the actual existence of peer and public pressure. Perhaps unsurprisingly, respondents involved in the peer review in which peer and public pressure
Table 5. Perceptions of the expertise of secretariat members and member state officials.

|                 | Total | WGB | IRM | UPR | EDRC | TPRM |
|-----------------|-------|-----|-----|-----|------|------|
|                 | Mean  | SD  | Mean| SD  | Mean | SD  |
| Secretariat     | 3.49  | .61 | 3.56| .62 | 3.53 | .61 |
| Member state    | 3.05  | .65 | 3.09| .72 | 3.01 | .62 |
| officials       | 2.88  | .63 | 3.17| .62 | 3.03 | .62 |

*Survey item: Based on your experiences with the [peer review], how would you in general assess…*
- the expertise of the staff members of the [IO] secretariat involved in the review
- the expertise of the (other) member state officials involved in the review

*Answer options: 1 = very low; 2 = low; 3 = high; 4 = very high. ‘I do not know answers’ treated as item non-response.*

*Note: One-way analysis of variance (least significant difference post hoc test).*

*p < 0.05; **p < 0.01; ***p < 0.001.*

WGB: Working Group on Bribery; IRM: Implementation Review Mechanism; UPR: Universal Periodic Review; EDRC: Economic and Development Review Committee; TPRM: Trade Policy Review Mechanism; IO: international organization.

Pressure are most appreciated (the WGB), also perceive the WGB as best able to organize this. IRM respondents, who value peer and public pressure less, also perceive this review as largely incapable of generating pressure. For the EDRC and the TPRM, the appreciation of peer and public pressure corresponds with the respective scores for the existence of pressure in these reviews. More surprising, however, is the observation that peer pressure is overall not perceived as a valuable function among UPR respondents, but that this peer review is nevertheless largely able to exert peer pressure.

**Discussion and conclusion**

Peer reviews are an increasingly important instrument for exerting pressure on states. The fact that states’ policy performance is critically evaluated and assessed, that recommendations are delivered by peers and the publication of these recommendations open ample possibilities for exerting pressure on laggards, and thus for naming and shaming. To what extent these opportunities are de facto used was unknown thus far. Based on original survey data and interviews, we found that the WGB, and to a lesser extent the UPR, are overall best capable of organizing peer and public pressure. The UPR is especially strong in public pressure, while the WGB shows very high scores on peer pressure. The TPRM comes close to the UPR in terms of its ability to generate peer pressure, but is overall perceived as the least capable of organizing public pressure. The EDRC and the IRM show comparable scores on peer pressure, which are lower than for the other reviews. The EDRC outperforms the IRM in terms of its ability to organize public pressure.

We find that neither the policy area under review nor the IO that hosts the peer review exercise are of relevance in explaining these findings. Cross-case comparisons show that the specificity of recommendations, the transparency of the peer reviews and systems for follow-up monitoring offer plausible explanations for some of the observed
variation in peer and public pressure. The low scores for the IRM correspond to the limited institutional structures it has in place to exert pressure (i.e., low transparency and no system for follow-up monitoring). The WGB, which is equipped with a formal system for follow-up monitoring, has a medium level of transparency and formulates very specific review recommendations, is commonly perceived as very able to organize public and especially peer pressure. The exertion of peer pressure can be linked to the closed setting in which the WGB reviews happen (Interviews CO 8, 14, 33). Vice versa, the high degree of public pressure found in the UPR is in line with its detailed transparency provisions.

Rather difficult to explain is the UPR’s and the TPRM’s ability to organize peer pressure. Despite their lack of a system for follow-up monitoring and rather generic recommendations, these two reviews are perceived as quite able to generate peer pressure. One possibility to interpret this divergence is the fact that both reviews centre around bilateral exchanges, in which individual countries make individual recommendations to the reviewed member, without the necessity to have these recommendations endorsed by the entire peer group. Due to this bilateral nature, review recommendations are creating peer pressure (Interviews HR 1, 2, 10, 11, 12, 13, 24, 26, 28, 36; Interviews ET 16, 21, 22, 23, 25; Carraro 2017b; Conzelmann 2008). EDRC recommendations are issued by the entire review body, but the fact that the reviewed country has to agree to the recommendations may explain why peer pressure is fairly limited in the EDRC (Interviews ET 2, 9). The study of legitimacy perceptions offers further explanations for variations in the peer reviews’ ability to generate peer and public pressure. The degree to which peer and public pressure are valued functions of a peer review is in general closely related to the reviews’ ability to generate such peer and public pressure. It is difficult to determine, though, in which direction the causal vectors run. Are peer reviews purposely designed to induce peer and public pressure, because these processes are deemed legitimate? Or are these processes deemed legitimate, because the peer reviews are able to generate them? These questions were outside the scope of the present article, but merit further reflection and investigation. The UPR, however, shows that strong peer pressure can be exerted even in the presence of a comparatively low appreciation for this practice among (some) participants.

It was beyond the scope of this contribution to consider how far peer reviews generate domestic policy change. With the exception of the anti-corruption peer reviews, peer pressure is often targeted at diplomats and only sometimes at high-level representatives of reviewed states. Public pressure might be strong, but it is only one element influencing policy change, alongside financial constraints and political expediency. Systematic research into the significance of peer and public pressure for domestic policy change will offer a more complete picture of the effectiveness of peer reviews as naming and shaming instruments in global governance.

Acknowledgements

For helpful comments on earlier versions of this article, the authors would like to thank Giovanni Mantilla, Christian Kreuder-Sonnen, the anonymous reviewer and the organizers and participants of the 2016 ‘Shaming in World Politics’ workshop at Stockholm University. All authors contributed equally to this article.
Funding
This work was supported by the Netherlands Organisation for Scientific Research (NWO; grant number 452-11-016).

Notes
1. See Tanaka (2008). Apart from the peer reviews covered in this study, the instrument is used in other policy fields by the OECD and the UN as well as by the International Monetary Fund (IMF), the Council of Europe, the European Union (EU), the African Union and the Financial Stability Board.
2. One example of such an approach is the OECD WGB discussed below.
3. This institutional design feature corresponds to what Abbott and co-authors coin the ‘precision dimension’ of the legalization of an institution (2000: 401).
4. Information on the sampling strategy for the survey and response rates is available in Appendix 1.
5. Significance levels were computed with the Games–Howell post hoc test, since our data shows unequal variances across groups. The Levene test yields $p \leq 0.05$ for both peer and public pressure.
6. Interviews dealing with the UPR are coded as HR, interviews on the corruption cases (IRM and WGB) are coded as CO and those conducted for the economics and trade cases (EDRC and TPRM) are coded as ET.
7. The tables indicate whether there is a statistically significant main effect of the independent variable (i.e., the peer review in which respondents participate) on the dependent variable. This is shown under the first column. The results of the pairwise comparisons are discussed in the text.
8. For the WGB: $p < 0.001$; UPR: $p < 0.01$.
9. The Levene test yields values of $p < 0.05$ for peer pressure and $p > 0.05$ for public pressure. We therefore used the Games–Howell post hoc test for peer pressure and the least significant difference (LSD) post hoc test for public pressure.
10. The effect sizes are $\eta^2 = 0.04$ for the expertise of IO secretariat staff; $\eta^2 = 0.02$ for member state officials.
11. Calculated with the LSD post hoc test. The Levene test yields $p > 0.05$ both for staff members and member state officials.
12. This question was not asked to secretariat staff members.
13. The insertion ‘other’ was used for respondents that are themselves member state representatives.
14. Even though the year 2009 falls outside of our research period (2014–2017), this list was used because no more recent official list of EDRC delegates was available. Since the survey asked respondents to indicate the time in which they were involved in the respective review, respondents who terminated their involvement in the EDRC before 2014 were removed from the sample.
15. Member states with no permanent mission in Geneva were removed from the sample, as these countries do not regularly participate in review meetings.
16. The national or governmental experts are selected by the governments of the States Parties. Many of them work for the Ministry of Justice, Ministry of Foreign Affairs or Anti-Corruption Agencies, but may also be judges or university professors. These experts are responsible for conducting the peer review and in many instances attend plenary sessions.
17. It was decided to give respondents who were sampled for more than one peer review the opportunity to fill out the survey twice. In particular, these individuals have extensive experience with peer reviews and are able to assess them from a comparative perspective.
Preventing them from filling out the survey more than once would lead to a bias in the study against individuals who are most actively involved in the peer reviews and would result in a loss of valuable information.

**ORCID iDs**

Valentina Carraro [https://orcid.org/0000-0002-4739-3281](https://orcid.org/0000-0002-4739-3281)

Thomas Conzelmann [https://orcid.org/0000-0003-2019-9443](https://orcid.org/0000-0003-2019-9443)

**References**

Abbott KW, Keohane RO, Moravcsik A, et al. (2000) The concept of legalization. *International Organization* 54 (3): 401–419.

Abebe AM (2009) Of shaming and bargaining: African states and the universal periodic review of the United Nations Human Rights Council. *Human Rights Law Review* 9(1): 1–35.

Adler-Nissen R (2014) Stigma management in international relations: Transgressive identities, norms, and order in international society. *International Organization* 68(1): 143–176.

Boxell J and Rigby E (2011) Exports warning as bribery law is delayed. *Financial Times*, 31 January. Available at: [https://www.ft.com/content/07b244dc-2d3e-11e0-9b0f-00144feab49a](https://www.ft.com/content/07b244dc-2d3e-11e0-9b0f-00144feab49a) (accessed 12 October 2017).

Carraro V (2017a) *A Double-Edged Sword: The Effects of Politicization on the Authority of the UN Universal Periodic Review and Treaty Bodies*. Maastricht University: unpublished PhD thesis.

Carraro V (2017b) The United Nations Treaty Bodies and Universal Periodic Review: Advancing Human Rights by Preventing Politicization? *Human Rights Quarterly* 39(4): 943–970.

Carraro V and Jongen H (2018) Leaving the doors open or keeping them closed? The impact of transparency on the authority of peer reviews in international organisations. *Global Governance* 24(4): forthcoming.

Checkel JT (2001) Why comply? Social learning and european identity change. *International Organization* 55(3): 553–588.

Conzelmann T (2008) Beyond the carrot and the stick: State reporting procedures in the WTO and the OECD. In: Joachim J, Reinalda B and Verbeek B (eds) *International Organizations and Implementation*. London and New York: Routledge, pp. 35–47.

DeMeritt JHR (2012) International organizations and government killing: Does naming and shaming save lives? *International Interactions* 38(5): 597–621.

Franklin JC (2015) Human rights naming and shaming: International and domestic processes. In: Friman HR (ed.) *The Politics of Leverage in International Relations: Name, Shame, and Sanction*. Basingstoke: Palgrave Macmillan, pp. 43–60.

Greene EF and Boehm JL (2012) The limits of name-and-shame in international financial regulation. *Cornell Law Review* 97: 1083–1139.

Gutterman E and Lohaus M (2018) What is the “anti-corruption” norm in global politics? Norm robustness and contestation in the transnational governance of corruption. In: Kubbe I and Engelbert A (eds) *Corruption and Norms: Why Informal Rules Matter*. Cham: Springer, pp. 241–268.

Hafner-Burton EM (2008) Sticks and stones: Naming and shaming the human rights enforcement problem. *International Organization* 62(4): 689–716.

Inboden RS and Chen TC (2012) China’s response to international normative pressure: The case of human rights. *The International Spectator* 47(2): 45–57.

Johnston AI (2001) Treating international institutions as social environments. *International Studies Quarterly* 45(4): 487–515.
Jongen H (2018) The authority of peer reviews among states in the global governance of corruption. *Review of International Political Economy*. Epub ahead of print 23 November. DOI: 10.1080/09692290.2018.1512891.

Keck ME and Sikkink K (1998) *Activists beyond Borders: Advocacy Networks in International Politics*. Ithaca, NY: Cornell University Press.

Koremenos B, Lipson C and Snidal D (2001) The rational design of international institutions. *International Organization* 55(4): 761–799.

Krain M (2012) J’accuse! Does naming and shaming perpetrators reduce the severity of genocides or politicides? *International Studies Quarterly* 56(3): 574–589.

Lebovic JH and Voeten E (2006) The politics of shame: The condemnation of country human rights practices in the UNCHR. *International Studies Quarterly* 50(4): 861–888.

Lebovic JH and Voeten E (2009) The cost of shame: International organizations and foreign aid in the punishing of human rights violators. *Journal of Peace Research* 46(1): 79–97.

Murdie AM and Davis DR (2012) Shaming and blaming: Using events data to assess the impact of human rights INGOs. *International Studies Quarterly* 56(1): 1–16.

Murdie AM and Urpelainen J (2015) Why pick on us? Environmental INGOs and state shaming as a strategic substitute. *Political Studies* 63(2): 353–372.

Nance MT (2015) Naming and shaming in financial regulation: Explaining variation in the Financial Action Task Force on Money Laundering. In: Friman HR (ed.) *The Politics of Leverage in International Relations: Name, Shame, and Sanction*. Basingstoke: Palgrave Macmillan, pp. 123–142.

Pagani F and Wellen U (2008) The OECD peer review mechanism: Concept and function. In: Tanaka K (ed.) *Shaping Policy Reform and Peer Review in Southeast Asia: Integrating Economies amid Diversity*. Paris: OECD, pp. 261–277.

Pawson R (2002) Evidence and policy naming and shaming. *Policy Studies* 23(3–4): 211–230.

Risse T and Sikkink K (1999) The socialization of international human rights norms into domestic practices: Introduction. In: Risse T, Ropp SC and Sikkink K (eds) *The Power of Human Rights: International Norms and Domestic Change* (Cambridge Studies in International Relations). Cambridge: Cambridge University Press, pp. 1–38.

Schäfer A (2006) Resolving deadlock: Why international organisations introduce soft law. *European Law Journal* 12(2): 194–208.

Sharman JC (2009) The bark is the bite: International organizations and blacklisting. *Review of International Political Economy* 16(4): 573–596.

Simmons BA, Dobbin F and Garrett G (2006) Introduction: The international diffusion of liberalism. *International Organization* 60(4): 781–810.

Squatrito T, Lundgren M and Sommerer T (2017) Naming and Shaming by International Organizations: Mapping Condemnatory Speech Acts Across 27 IOs, 1980–2015. Unpublished manuscript.

Tanaka K (ed.) (2008) *Shaping Policy Reform and Peer Review in Southeast Asia: Integrating Economies amid Diversity*. Paris: OECD.

Terman R and Voeten E (2017) The relational politics of shame: Evidence from the universal periodic review. *The Review of International Organizations* 5(1): 33.

UPR-info (2016) Thailand: CSOs lobby with government to accept pending recommendations. Available at: https://www.upr-info.org/en/news/thailand-csos-lobby-with-government-to-accept-pending-recommendations (accessed 12 October 2017).

Wendt A (2001) Driving with the rearview mirror: On the rational science of institutional design. *International Organization* 55(4): 1019–1049.
Author biographies

Valentina Carraro is a visiting researcher at the Ludwig Boltzmann Institute of Human Rights in Vienna, and a postdoctoral researcher at the Faculty of Arts and Social Sciences of Maastricht University. Since 2013, her research has been dealing with soft governance mechanisms in the field of human rights.

Thomas Conzelmann is Professor of International Relations at Maastricht University. He has published on the role of non-binding regulation and implementation mechanisms at the global level, such as peer reviews among states and business self-regulation.

Hortense Jongen is postdoctoral researcher at the School of Global Studies of the University of Gothenburg. She holds a PhD from Maastricht University (2017). Her PhD research dealt with the instrument of peer review among states in the global fight against corruption.

Appendix 1

Respondent selection for the survey

This annex explains the way in which potential survey respondents were identified in each of the five peer reviews under study. We decided to focus the survey on the perceptions of individuals who are directly involved in the preparation and actual conduct of the reviews, rather than the broader circle of individuals, institutions and organizations confronted with the outcomes of the reviews. These individuals were expected to have a sufficient level of expertise and familiarity with the peer review mechanisms. We defined the target population as all respondents out of the total population for which we managed to obtain valid personal office email addresses.

The survey for the OECD EDRC was distributed to all EDRC counsellors (member state representatives) identified through two publicly available lists of EDRC member state delegates dating from 200914 and 2016. On the secretariat side, the survey was distributed to all officials working in the Country Studies Branch of the OECD Economics Department at the time the survey was sent out, as well as to all secretariat officials who were mentioned as contributors in the prefaces of any of the OECD Economic Surveys published between 2014 and 2016. Hence, the total population comprises 190 persons. We retrieved valid personal email addresses of 146 out of these 190 respondents, yielding a coverage rate of 76.8%.

For the WTO TPRM, we followed a similar approach as in both the UN cases discussed below. The survey was sent to one delegate in the permanent missions of each WTO member state in Geneva who was identified as the person whose portfolio covered the meetings of the Trade Policy Review Body (TPRB) in 2016.15 In the first step, the web presence of member state missions in Geneva were screened to identify those representatives who were mentioned as covering trade matters in their portfolio. If more than one person was identified in this way, the person with the highest professional rank was selected. In cases where this information was not available through the web pages, the permanent missions of member states in Geneva were contacted through email or phone, requesting to identify the staff member of the permanent mission who regularly covered the meetings of the TPRB in 2016. Out of the 155 WTO member states with permanent missions in Geneva, we managed to identify the representative who
regularly attends the meetings of the TPRB in 135 cases. In 19 cases, no valid personal mail addresses could be retrieved, yielding a sample of 116 member state representatives. In addition, the survey was sent to all 18 staff officials working in the Trade Policy Review Division in 2016, the department within the WTO secretariat that is responsible for the conduct of the Trade Policy Reviews. This created a target population of 134 (116 + 18) potential respondents out of a total population of 173 (155 + 18), implying a coverage rate of 77.5%.

In the OECD WGB, 10 members of the OECD Anti-Corruption Division, one chairman of the Working Group and 183 national experts who acted as delegates to the WGB meetings or as focal points for the country reviews were involved during the chosen timeframe. Of these 194 officials, 192 could be contacted, which implies a coverage rate of 98.9%. The high coverage can be explained by the availability of an overview of the involved officials in the WGB in 2015.

The UNCAC IRM is a more complex case. Contrary to the WGB and the Group of States against Corruption (GRECO), the IRM actively involves three types of actors: IO secretariat members, national anti-corruption experts and diplomats. For a considerable number of countries, diplomats are the only representatives to attend these sessions. However, the sheer size of the IRM makes it practically unfeasible to identify all the involved actors during the chosen timeframe. Thus, we aimed to include the 27 secretariat members of the UN Office on Drugs and Crime (UNODC) that are involved in the IRM, as well as one national expert for each member state, and one diplomat for each member state that has diplomatic representation in Vienna.

The selection model for diplomats for the IRM case was as follows.

- During attendance of the resumed fifth session of the Implementation Review Group (IRG), held in October 2014, and the sixth session of the IRG, held in June 2015, many contacts were established with diplomats and national experts. As it can be said with absolute certainty that these diplomats were present at the IRG sessions, first priority was given to these officials compared to other officials mentioned on the attendance list. If contact was established with more than one diplomat per country, priority was given to the diplomat who had attended the most meetings during the timeframe of June 2014–June 2015. If several diplomats had attended an equal number of meetings, the diplomat who appeared first on the attendance list after the ambassador was contacted. In total, 14 diplomats were selected following this strategy.
- The next step consisted of contacting the 95 remaining embassies by email, asking them to provide the names and email addresses of the responsible diplomats for the UNCAC file. In those cases where no reply was given, embassies were directly contacted by phone and asked for the relevant contact details. In total, 48 diplomats were selected following this strategy.
- Of the remaining 47 state parties with diplomatic representation in Vienna, 18 could be contacted through an internet search. Again, priority was given to the diplomat who had attended the most meetings during the chosen timeframe. In the case of an equal number of meetings, the diplomat who appeared first on the attendance list after the ambassador was contacted.
The selection model for national civil servants involved in the IRM was as follows.

- Priority was given to officials with whom personal contact was established, for instance, during the resumed fifth session of the IRG, during the sixth session of the IRG and during the GRECO 68 plenary session. If contact was established with more than one expert per country, priority was given to the experts who had attended the most meetings during the timeframe of June 2014–June 2015. If several experts had attended an equal number of meetings, the expert who appeared first on the attendance list was contacted. In total, the contact details of 31 experts could be retrieved.

- Secondly, the email addresses of 64 national experts could be retrieved by means of an internet search. Priority was given to the expert who had attended the most meetings during the identified timeframe, as indicated on the attendance list. If the contact details of the experts listed on the attendance list could not be retrieved, the contact details of national coordinators were located.

- The contact details of three national experts were collected by contacting embassies and national administrations.

Out of 170 member states, 98 national experts could be included in the sample. Out of 109 states with diplomatic representation in Vienna, 80 diplomats could be contacted. The total coverage rate of this study, which also includes the secretariat members, therefore amounts to 67.0%.

Adding up the sampled officials for all three anti-corruption peer reviews brings the total sample size to 558 officials. Considering that several officials were sampled for more than one anti-corruption peer review, 532 distinct individuals were contacted. If respondents filled out the survey for two peer reviews, their responses regarding each peer review were included in the study as separate entries.17

In the case of the UN UPR, the survey was targeted to all reviewees and reviewers involved in the UPR in 2015. More specifically, the target population consisted of all state delegates involved in the mechanism at the time the survey was sent out, meaning the human rights delegates in charge of the UPR portfolio belonging to UN member states with a mission in Geneva (177 state delegates). The survey was sent to all the individuals belonging to the target population and for whom a valid email address could be retrieved (157 state delegates). This yielded a coverage rate of 88.7%. The survey was intended also for staff of the Office of the High Commissioner for Human Rights, which is responsible for the conduct of the UPR. However, the Office of the United Nations High Commissioner for Human Rights (OHCHR) secretariat declined permission to circulate the survey among its staff members. This group could therefore not be included in the survey.

A.1 shows an overview of the total and target population for each of the individual surveys, as well as coverage and response rates. For most surveys a response rate between 40% and 50% was achieved, in one case about 35% and in one case above 60%.

The response rate was calculated by including each respondent who started the survey. When distributing the survey, all possible efforts were made to ensure a high
response rate, including an invitation email outlining the goals of the survey and the confidentiality precautions, as well as three reminder emails. Outreach activities were organized at the Council of Europe and the UN Human Rights Council to raise awareness about the project and motivate respondents to participate in our research.

Table A.1  Overview of total populations, target populations and response rates per case study.

| Case study                                                      | Total population | Target population | Coverage rate | Collected responses | Response rate |
|---------------------------------------------------------------|------------------|-------------------|---------------|---------------------|---------------|
| OECD Economic and Development Review Committee                | 190              | 146               | 76.8%         | 62                  | 42.5%         |
| WTO Trade Policy Review Mechanism                             | 173              | 134               | 77.5%         | 83                  | 61.9%         |
| OECD Working Group on Bribery                                 | 194              | 192               | 98.9%         | 77                  | 40.1%         |
| UNCAC Implementation Review Mechanism                         | 306              | 205               | 67.0%         | 99                  | 48.3%         |
| UN Universal Periodic Review                                  | 177              | 157               | 88.7%         | 54                  | 34.4%         |

OECD: Organisation for Economic Cooperation and Development; WTO: World Trade Organization; UNCAC: United Nations Convention against Corruption; UN: United Nations.