Leaving the Doors Open or Keeping Them Closed? The Impact of Transparency on the Authority of Peer Reviews in International Organizations

Valentina Carraro and Hortense Jongen

Although transparency is frequently employed to enhance the legitimacy of public organizations, several scholars point to its potentially negative implications. This study analyzes the impact of transparency on the authority of peer reviews in international organizations. Authority, here conceived as rooted in legitimacy beliefs, is crucial for peer reviews to produce effects. This research is based on results from an online survey and forty-three interviews with actors involved in two United Nations peer reviews: the Universal Periodic Review in human rights and the Implementation Review Mechanism in the fight against corruption. The article shows that transparency positively affects the perceived development of pressure, yet negatively influences mutual learning and appears to be unable to ensure equal treatment of states. KEYWORDS: authority, transparency, United Nations, peer reviews.

FOR OVER A CENTURY, TRANSPARENCY HAS BEEN EMPLOYED TO ENHANCE THE LEGITIMACY of public organizations. Transparency allows interested actors to scrutinize the performance of public officials, holding them accountable for their actions. However, transparency may also compromise confidentiality, inhibiting open discussions. For these reasons, there is disagreement as to whether transparency should be pursued to enhance the legitimacy of public organizations, or whether it should be avoided because of its possible negative consequences such as diplomatic posturing.¹ Such conflict becomes particularly pronounced in international procedures that aim to stimulate dialogue among states.

Peer reviews among states are an example of such procedures. They are increasingly utilized instruments of global governance, employed for instance by the Organisation for Economic Co-operation and Development (OECD), the European Union, and the African Union. In peer reviews, states share information on their compliance with international standards in a policy area, and this information is subsequently assessed by other states.² Their outcome is usually a set of recommendations on how states could improve their performance. Peer reviews are soft governance instruments: they neither provide rewards for good performance, nor sanctions for noncompliance. Some scholars are therefore pessimistic about their potential to generate reform.³ Others sketch a more promising

picture, pointing to the observable impact that some peer reviews have had in states such as a reorientation of the policy debate.⁴ We contend that peer reviews can produce such results, but only if participating members consider them to bear authority. *Authority* is defined as a form of power rooted in a shared belief in the appropriateness and legitimacy of an actor or institution.⁵ When the intended audience collectively recognizes the authority of a peer review, they are more likely to seriously consider following the review's recommendations.

In this article, we investigate the transparency of peer reviews as a factor that potentially contributes to authority. We investigate whether transparency contributes to decrease or increase the authority of peer reviews in the eyes of involved actors, or whether it plays no role. For peer reviews to be authoritative, should we leave the doors to the meeting rooms open, or keep them closed?

To answer this question, we selected two peer reviews within the United Nations that strongly differ regarding their level of transparency while holding other variables relatively constant: the Universal Periodic Review (UPR) of human rights and the Implementation Review Mechanism (IRM) of the UN Convention against Corruption (UNCAC). These reviews are relatively similar as both are recent UN peer reviews employed in politically sensitive policy fields, where governments are generally hesitant to disclose information.⁶ Additionally, both were created to promote learning in a nonconfrontational environment. It is therefore striking that, while the UPR provides full informational disclosure to states and the wider public, the IRM grants a higher level of confidentiality. Our analysis is based on original survey data and forty-three interviews with officials involved in these reviews.

The article is structured as follows. We first engage in a conceptual discussion of authority and transparency and their possible connection. Subsequently, we discuss our methodology and outline the main transparency provisions of the UPR and the IRM. Next, we assess the authority beliefs of participants, before studying the influence of transparency on such beliefs.

Authority and Transparency of Peer Reviews

Authority

The literature abounds with different conceptualizations of authority in global governance.⁷ We deviate from more traditional conceptions of authority as the possession of (delegated) competences or decisionmaking powers (i.e., formal-legal authority). Formal-legal authority is inapplicable to global governance actors that have been delegated limited competences or that differ minimally in these competences, the latter of which is the case for peer reviews. Instead, we draw on a relational conception of authority.⁸ *Authority*, in this understanding, concerns a social relation between the actor or institution bearing authority on the one hand (in this case a peer review), and the authority-followers on the

other (i.e., the actors whose behavior the peer review seeks to influence).⁹ A peer review possesses authority only when the intended audience collectively recognizes its legitimacy; put differently “authority is conferred” by the authority-followers.¹⁰

The aforementioned discussion raises the question of how this abstract concept can be studied empirically. In answer, we draw on the following definition of *authority* offered by Bruce Cronin and Ian Hurd: “An institution acquires authority when its power is believed to be legitimate. Authority requires legitimacy and is therefore a product of the shared beliefs about the appropriateness of the organization’s proceduralism, mission and capabilities.”¹¹ Thus, authority is rooted in a shared belief in an actor’s or institution’s legitimacy vis-à-vis its mission, proceduralism, and capabilities.¹² We operationalize beliefs pertaining to these dimensions as follows (Table 1). First, *mission* concerns the legitimacy of the objectives or purpose of an actor or institution, relating to three aspects:

1. the appropriateness of the international organization (IO) hosting the reviews (the UN);
2. the appropriateness of using peer review to assess states’ performance in a policy area (human rights and anticorruption); and
3. the appropriateness of the standards used to assess states’ performance.

Table 1 Operationalization of Authority

Dimension	Operationalization: Perceptions of ...
Mission	<ul style="list-style-type: none"> – The international organization hosting the peer review – The policy field – Standards of assessment
Proceduralism	<ul style="list-style-type: none"> – The uniform application of assessment standards
Capabilities	<ul style="list-style-type: none"> – Peer pressure – Public pressure – Mutual learning – Accurate overview of reviewed states’ performance – Practically feasible recommendations

Second, the *proceduralism* dimension refers to the “procedural legitimacy” of a peer review, meaning legitimacy based on the correct and consistent application of rules. We assess this by investigating the extent to which the rules are perceived to be applied uniformly. The final dimension, *capabilities*, is concerned with the perceived performance of the reviews.¹³ Specifically, we focus on five types of results, selected on the basis of existing literature on peer reviews¹⁴ and a range of exploratory interviews:¹⁵ (1) to generate peer pressure; (2) to generate public pressure; (3) to facilitate mutual learning; (4) to present an accu-

rate overview of the reviewed states' performance; and (5) to deliver practically feasible recommendations.

Thus far, the literature has been primarily concerned with the domestic impact and use of peer reviews.¹⁶ In principle, this is an important research interest: the peer reviews' soft nature raises questions of whether and how these instruments can engender policy reform. However, the approach we suggest to study the significance of peer reviews by focusing on their authority presents several advantages compared to an analysis of their effectiveness in inducing state compliance. The first advantage is methodological. An assessment of compliance would make it difficult to isolate the influence of peer reviews from confounding factors. Compliance with a review's recommendations might be caused by many factors unrelated to the peer review such as unilateral pressure by another state. Likewise, noncompliance with a peer review's recommendations does not necessarily imply that the review lacks authority; some states want to implement reform, but are plagued by other concerns that impede compliance such as political or budget constraints.¹⁷ Our focus on authority, however, does not mean that the issue of substantive compliance is irrelevant. Rather, we conceive of authority as a necessary though insufficient precondition for a peer review to engender compliance. Authority is necessary for a peer review to have an *independent* effect on member states. The instrument is devoid of sanctioning tools and, instead, seeks to engender effects through peer and public pressure, policy learning, and socialization processes. The ability of a peer review to successfully carry out these functions and to induce its member states to obey is contingent on its authority.

Transparency

Transparency is frequently invoked as a remedy to enhance an organization's credibility.¹⁸ Appeals for increased transparency are issued to national institutions and international organizations alike, with the argument that intergovernmental bargaining should be exposed to no less public monitoring than political debates at the national level.¹⁹ Transparency is therefore often employed in international bodies to enhance the legitimacy of organizations by allowing citizens to evaluate the performances of public officials, holding them accountable for their actions.²⁰ *Transparency*, broadly defined as the "availability of regime-relevant information,"²¹ can come in different forms and serve different purposes. Ronald B. Mitchell divides transparency into two macro-categories: transparency *for* governance, which aims to alter the behavior of actors by providing them with information on their conduct and its consequences; and transparency *of* governance, which concerns the possibility for the public to "observe the actions either of 'regulators' to whom they have delegated power or of other powerful actors of society."²² We focus on the latter form of transparency.

Even within the same IO, peer reviews largely vary in their transparency of governance; namely, their level of openness toward broader audiences. Applied

to peer reviews, we interpret *transparency of governance* as the ability of actors other than the reviewed state—specifically, the other states in the peer review and nonstate actors (NSAs)—to access review-related information. This can mean making review-related documents available, granting access to civil society, web-casting the event, or discussing review reports with other states in plenary. In this article, we study the extent to which a peer review's transparency of governance plays a role when it comes to the authority beliefs of directly involved actors. To this aim, we elaborated a set of expectations that we submit to empirical scrutiny. The following four dimensions of transparency of governance are considered: (1) public availability of country reports; (2) plenary discussions of individual country reports; (3) input by NSAs; and (4) possibility for NSAs to attend or follow plenary sessions.²³

First, concerning the *mission* of a peer review, we do not expect transparency to play any role. This is because the appropriateness of conducting a peer review in a specific policy field or organization relates to the existence of the review itself, rather than to its transparency provisions.

Second, we foresee transparency to play a substantial role regarding a peer review's *proceduralism*; specifically, the perceived fairness and consistency of rule application. We expect that transparency—in particular, the discussion of review reports in plenary and the possibility for NSAs to attend or follow review sessions—positively affects the proceduralism of a peer review: actors might be more prone to follow the rules since they know they are being scrutinized, and unequal treatment of states could be exposed. Consequently, transparency is likely to lead to positive views on the proceduralism of a peer review.

Finally, we expect that transparency strongly influences authority perceptions of the *capabilities* of peer reviews, yet this influence might take different directions depending on the review goal. Transparency is likely to facilitate the perceived achievement of peer and public pressure. Peer pressure can be successfully exerted only if the relevant information is available to the peers, by publishing review reports online and discussing these reports in plenary sessions. Likewise, public pressure is contingent on the public availability of this information to a broader audience, via the publication of review reports and the possibility for NSAs to attend or follow reviews. In contrast, we expect plenary discussions of review reports and the openness of these discussions to NSAs to be an obstacle to mutual learning, as a review conducted in a confidential environment might better serve this goal. Besides, we expect transparency to positively contribute to the peer reviews' ability to deliver accurate review reports: peer reviews that are more open to information submitted by NSAs might be better able to sketch an accurate picture of states' performances than reviews that rely on information provided only by states. Finally, we do not anticipate transparency to play a role on the ability of the reviews to deliver practically feasible recommendations, as this relates to the quality of recommendations.

Methods and Case Selection

Methods

We studied the effect of transparency on the authority of the UPR and the IRM. Data were collected through an online survey and forty-three semistructured interviews. The survey was employed as the most effective method to systematically collect information on participants' perceptions, whereas interviews served to contextualize survey findings.

The survey was distributed via e-mail between July and December 2015. The IRM survey was distributed to all UN Secretariat officials involved in the peer review (twenty-seven officials), one diplomat per member state with permanent representation in Vienna—where the review takes place—and for whom contact information could be retrieved (eighty diplomats), and one national expert per member state for whom contact information could be retrieved (ninety-eight experts).²⁴ The UPR survey was distributed to all state delegates involved in the peer review belonging to countries with a permanent representation in Geneva—where the review takes place—and for whom contact details could be retrieved (157 state delegates). The survey response rate was 38.7 percent and we collected 140 observations.

The survey data include the responses to a battery of questions, probing respondents' perceptions of the authority of the two reviews. Most questions were assessed on a scale of 1 to 4,²⁵ in which a score of 1 indicates that respondents viewed a certain aspect of the review as very inappropriate (or considered the review completely unable to carry out a specific function) and a score of 4 signifies that this aspect was recognized as very appropriate (or the review was completely able to perform this function). The analyses mainly consist of descriptive statistics and *t* tests, comparing the mean scores of the IRM and the UPR on survey items to identify possible statistically significant differences among them.

Next to the survey, we interviewed thirty-three state delegates, who represented member states from all five UN regional groupings. In addition, ten interviews were conducted with UN Secretariat members of the two peer reviews. During interviews, emphasis was placed on why respondents considered a particular dimension (in)appropriate.

Case Selection

The monitoring activities of the UN reflect the growing popularity of peer reviews. Next to the UPR and the IRM, many peer reviews can be found in policy fields from environmental protection to competition policy. The UN peer reviews take many different forms: some are voluntary (e.g., competition policy²⁶), others subject all states to review (e.g., corruption control, human rights); some discuss national reports in plenary (e.g., environmental policy, human rights), others hold more general meetings (e.g., anticorruption).

To study the relationship between transparency and authority, we selected two intergovernmental peer reviews that show similarities in their format and functioning, but differ in terms of their transparency, along the four dimensions identified above. As for similarities, we aimed for peer reviews that have been in place for a relatively similar time. Moreover, we included only compulsory peer reviews as opposed to voluntary instruments, and peer reviews that are carried out regularly as opposed to one-off exercises.

Table 2 shows that the IRM is an obvious case study of a peer review with limited transparency. It is one of the few UN peer reviews for which the answer to the aforementioned four criteria is completely or partly negative. In contrast, the UPR is among the few mechanisms that are extremely transparent. Notwithstanding these differences, the two cases show many similarities. They have been operating for a relatively similar time (the UPR since 2007, the IRM since 2010), they hold regular meetings (quadrennially for the UPR, biennially for the IRM), and the meetings are compulsory for the 183 states parties to the UNCAC and all 193 UN member states in the case of the UPR.

Table 2 The Transparency of the UPR and the UNCAC Peer Reviews

	UPR	IRM
Publication of all review reports	Yes	Optional (except for executive summaries)
Individual country reports are discussed in plenary	Yes	No
NSAs can provide input in the review process	Yes	Optional
NSAs and the public can attend plenary sessions	Yes	No

Note: UPR, Universal Periodic Review; UNCAC, UN Convention Against Corruption; IRM, Implementation Review Mechanism; NSAs, nonstate actors.

Elucidating Table 2, the first dimension pertains to the public availability of review reports. The UPR reviews are based on three reports, one compiled by the reviewed state and two drafted by the UN Secretariat, which are available online. Similarly, the UPR output report is published online. In the IRM, three reports are relevant: the reviewed state's responses to a self-assessment checklist, the final evaluation report (containing country-specific recommendations), and an executive summary of the final report. Apart from the executive summary, the online publication of these reports is optional.

The second dimension relates to the discussion of country reports in plenary. In the UPR, country reports are discussed during the so-called interactive dialogue, which is attended by diplomats and other interested stakeholders. During this dialogue, the reviewed state presents its report and the diplomats issue their recommendations. In contrast, in the IRM there is no plenary discussion of individual reports. The reviewing team, which consists of experts from two other states assisted by the Secretariat, writes the evaluation report and formu-

lates recommendations in consultation with the reviewed state. The assessment is based on the reviewed state's responses to the self-assessment checklist and, if it allows, on information collected during a country visit. Experts other than those carrying out the review cannot issue recommendations and have no insight into their peers' performance. This is because in plenary sessions only thematic and synthesis reports are presented and discussed.

The third dimension concerns the input by NSAs in the review. In the UPR, NSAs can submit information on countries under review to the UN Secretariat, which compiles the information received in one report. In the IRM, NSAs can be requested to provide information during the country visit; however, only if the reviewed state allows this. Moreover, the reviewed state can prevent the reviewing team from consulting certain NSAs.

The fourth dimension pertains to NSAs' attendance at plenary sessions. In the UPR, interested actors may attend the interactive dialogue as part of the audience. In addition, these dialogues are webcasted and thus available to the wider public. In the IRM, NSAs cannot attend plenary sessions, which are likewise not webcasted.²⁷

The Authority of the UPR and the IRM

Mission

The mission of a peer review relates to the perceived appropriateness of using peer reviews in a particular policy field and in a specific IO. Accordingly, respondents were asked: "How appropriate or inappropriate do you find:

1. that a peer review is used to assess states' performance in the field of [human rights/corruption]?
2. that the UN is used as a framework to organize the peer review?
3. the [anticorruption/human rights] standards that are used to assess states' performance?"

Response options to the three survey items were converted into the following numerical scores: 1 = very inappropriate, 2 = inappropriate, 3 = appropriate, 4 = very appropriate.

The findings (Table 3) indicate that the use of peer review in anticorruption and human rights is largely deemed appropriate; no statistically significant differences between the IRM ($M = 3.37$) and the UPR ($M = 3.27$) were found to exist. Officials in both peer reviews expressed appreciation for the peer-to-peer nature of the instruments (Interviews UPR 2, 6, 7, 8, 9, 11, 12, 13, 20, 21; Interviews IRM 8, 11, 13, 19, 21). Interestingly, as to the appropriateness of the IO, the analyses reveal that the UN was perceived as somewhat more appropriate in the UPR ($M = 3.56$) than in the IRM ($M = 3.45$), but these differences are

not statistically significant. Also in the IRM, the UN was deemed largely appropriate, in line with the interview findings. Finally, the analyses of the standards of assessment reveal that these were perceived as somewhat, though not statistically significantly, more appropriate in the UPR ($M = 3.38$) than in the IRM ($M = 3.25$). Likewise, hardly any issues were raised regarding this aspect in the interviews.

Table 3 Quantitative Assessment of the Mission Dimension

Survey Item (Scale of 1 to 4)	Mean: IRM	Mean: UPR	<i>t</i> Statistic	Respondent <i>N</i>
1. Policy field	3.37 (.55)	3.27 (.61)	0.97	140
2. IO	3.45 (.58)	3.56 (.65)	-1.08	140
3. Standards of assessment	3.25 (.62)	3.38 (.61)	-1.20	139

Notes: IRM, Implementation Review Mechanism; UPR, Universal Periodic Review; IO, international organization.

t-test (two-tailed); standard deviations are in brackets; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

Proceduralism

Proceduralism concerns fair and consistent rule application. To study perceptions of equal treatment, the survey asked respondents: “How would you assess the extent to which standards of assessment are uniformly applied across reviews?” This could be 1 = far too low, 2 = too low, 3 = just right, 2 = too high, 1 = far too high. As the allocation of numerical scores reveals, the answer option indicating the highest degree of authority was the middle category, which was assigned a score of 3. This is because the answer category “just right” indicates that the application of standards of assessment was perceived as appropriate. The answer options “too high” and “too low” indicate lower levels of authority and, hence, were allocated a score of 2. The lowest level of authority relates to answer options “far too low” and “far too high,” receiving a value of 1. The reasoning for such categorization is this: if standards of assessment were perceived to be applied uniformly to a (too) low extent, this indicates that the reviews treated some states unjustly as compared to others. In contrast, if the application of standards of assessment was perceived as excessively uniform, this indicates that the review did not sufficiently take into account country specificities.

The survey findings (Table 4) reveal statistically significant variation between the IRM ($M = 2.53$) and the UPR ($M = 2.26$; $p < 0.05$), with an eta-squared value of 0.04. Figure 1 shows that 59.1 percent of respondents deemed the degree to which the rules were applied uniformly “just right” compared to only 37.0 percent in the UPR. In the latter, 56.6 percent considered this to be “too low” or even “far too low” as opposed to only 38.7 percent in the IRM.

Table 4 Quantitative Assessment of Uniform Rule Application

Survey Item (Scale of 1 to 3)	Mean: IRM	Mean: UPR	<i>t</i> Statistic	Respondent <i>N</i>
Uniform rule application	2.53 (.61)	2.26 (.65)	2.42*	134

Notes: IRM, Implementation Review Mechanism; UPR, Universal Periodic Review.
t-test (two-tailed); standard deviations are in brackets; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

Figure 1 Assessment of Uniform Rule Application

Note: IRM, Implementation Review Mechanism; UPR, Universal Periodic Review.

The interviews elucidated the observed variation. Even though the rules of procedure of the UPR grant equal treatment to states, they leave delegates substantial room for maneuver to engage in political horse-trading. Political power and alliances with other states are reportedly among the main reasons why some states are more gently treated than others. Specifically, officials reported that bilateral relations strongly come to the fore in determining the content of recommendations: the more allies that the state under review manages to summon, the more lenient its review will be (Interviews UPR 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19). These dynamics were perceived to undermine equal treatment in the UPR, and affect the proceduralism of the mechanism.

Officials were more positive about equal treatment in the IRM. Many interviewees indicated that the review exercises were rather technical and objective, and the type of political bias that was observed in the UPR evaluations was reportedly largely absent (Interviews IRM 10, 13, 18, 22). However, plenary sessions

were overall perceived as more political than the evaluation exercises (Interviews IRM 10, 16, 41, 21, 22). As opposed to the UPR, no country evaluations are discussed in the IRM plenary. Instead, plenary meetings primarily focus on procedural matters related to the peer review, such as the budget and civil society involvement, which can be highly contentious.

Capabilities

The two peer reviews' capabilities concerned their perceived ability to generate meaningful outcomes. The survey asked respondents: "Generally speaking, to what extent do you believe that the [respective peer review] successfully,

1. Exerts state-to-state (peer) pressure
2. Exerts public pressure
3. Triggers mutual learning
4. Provides an accurate overview of reviewed states' performance
5. Provides practically feasible recommendations to states?"

Answer options were: 1 = not at all, 2 = to some extent, 3 = to a large extent, 4 = completely. Respondents who indicated not to know the answer to the question were treated as item nonresponse. The results are presented in Table 5.

Table 5 Quantitative Assessment of the Capabilities of the UPR and the IRM

Survey Item (Scale of 1 to 4)	Mean: IRM	Mean: UPR	<i>t</i> Statistic	Respondent <i>N</i>
1. Peer pressure	2.30 (.79)	2.64 (.61)	-2.59*	133
2. Public pressure	2.10 (.73)	2.61 (.68)	-3.91***	134
3. Mutual learning	2.84 (.78)	2.52 (.72)	2.32*	135
4. Accurate reports	2.62 (.78)	2.59 (.77)	0.237	133
5. Practically feasible recommendations	2.74 (.75)	2.63 (.65)	0.857	135

Notes: UPR, Universal Periodic Review; IRM, Implementation Review Mechanism.
t-test (two-tailed); standard deviations are presented in brackets; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

Table 5 shows that the two peer reviews exhibited variation in their perceived ability to successfully execute several functions. First, the UPR was believed to be considerably better at organizing peer ($M = 2.64$) and public pressure on states ($M = 2.61$) than the IRM ($M = 2.30$; $p < 0.05$; and $M = 2.10$; $p < 0.001$, respectively). The effect size was an eta-squared value of 0.05 for peer pressure and 0.10 for public pressure. Interviews indicated that in the UPR peer pressure has a bilateral dimension, as recommendations are delivered by one governmental representative to another (Interviews UPR 1, 2, 10, 11, 12, 13, 15, 16, 18). Plenary sessions of the IRM are characterized by a diplomatic atmosphere; state delegates do not directly ask one another questions about their perfor-

mance, let alone criticize each other. Hence, peer pressure does not materialize (Interviews IRM 3, 10, 21).

Second, concerning public pressure, NSAs reportedly hold governments accountable for the recommendations they accepted during the UPR. Consequently, when a country review takes place, governmental officials reportedly feel the need to show their audiences that they have already acted on many of the recommendations received in the previous review cycle (Interviews UPR 3, 10, 11, 12, 13, 16, 18). In the IRM, public pressure was overall felt to be low. Only two interviewed officials mentioned that the mechanism generated some public awareness in member states, mainly through the media and nongovernmental organizations (NGOs) such as Transparency International (Interviews IRM 10, 15).

Third, the two peer reviews exposed statistically significant variation in their perceived ability to trigger mutual learning among states. This time, the IRM ($M = 2.84$) performed considerably better than the UPR ($M = 2.52$; $p < 0.05$), though the eta-squared value only amounted to 0.04. IRM delegates reported that they had learned a lot from the evaluation exercise; one of the delegates described it as “a whole new learning experience” (Interview IRM 18). As to the UPR, interviewees did not mention mutual learning as one of the major results achieved by the UPR.

On the two remaining functions, the IRM and the UPR did not exhibit significant variation. The IRM ($M = 2.62$) was perceived as marginally, but not statistically significantly, better able to provide an accurate overview of states’ performances than the UPR ($M = 2.59$). In the UPR, while most recommendations are judged to be well informed, others are considered to be politically motivated or stemming from a lack of knowledge of the country. Consequently, it often is not possible to identify the level of human rights standards in a country from the output report (Interviews UPR 1, 2, 3, 4, 6, 8, 12). In the IRM, considering that a substantial number of country evaluation reports are not published online and none are discussed in plenary, it is more difficult for interviewed officials to comment on the accuracy of other countries’ evaluation reports. However, when asked to assess the accuracy of their own country’s review reports, none of the interviewees reported any inaccuracies. Differences with regard to the practical feasibility of recommendations were even smaller ($M = 2.74$ for the IRM and $M = 2.63$ for the UPR). In the IRM, interviewees expressed appreciation for the peer review’s flexibility in terms of its recommendations. The review reports only indicate what needs to be achieved, leaving the choice of how to achieve this to the reviewed state (Interviews IRM 7, 13, 19). In the UPR, respondents particularly appreciated the fact that recommendations are realistic; namely, they recommend only what is feasible for that country to achieve (Interviews UPR 3, 6, 7, 18).

The Transparency Effect: More Authority, Less Authority, or No Effect at All?

The previous section showed that clear differences exist regarding perceptions of the UPR's and the IRM's proceduralism and capabilities. Based on interviews, this section examines the extent to which this variation can be explained by the reviews' different transparency provisions.

Mission

We did not observe substantial differences in perceptions of the mission of the two mechanisms. This fits with our prior expectations, as we anticipated that transparency would play no role with relation to this dimension.

Proceduralism

We expected transparency to positively affect the proceduralism of peer reviews since we anticipated that transparency would lead involved actors to follow the rules more closely, thus positively influencing fairness perceptions. Our empirical findings are therefore highly surprising: the UPR was perceived to fare rather poorly in terms of consistent rule application while the IRM scored much better.

In the UPR, transparency alone appeared unable to ensure equal treatment of states—as far as the involved actors' perceptions were concerned. Even though UPR review sessions are highly transparent—as state reports are discussed in a plenary session that can be attended by the wider public—states negotiate behind closed doors when deciding on the recommendations to be issued. Thus, the UPR's transparency provisions cannot prevent political bias to emerge and, as a consequence, the mechanism was perceived as unfair.²⁸

Equally surprising is our finding that the IRM's lack of transparency has not negatively affected its perceived fairness. According to directly involved officials, the confidential setting seemingly did not stimulate politically motivated behavior. One explanation for this is the involvement of technical experts in the IRM as opposed to the diplomats in the UPR. The involvement of technical experts creates the impression that the IRM is an objective, and nonpolitical, exercise (Interviews IRM 10, 13, 18, 22). Trust in the expertise and objectivity of the evaluators appeared to be high, even when discussions take place behind closed doors.²⁹ These findings tie in with observations on the OECD peer reviews which, like the IRM, mostly bring together technical experts. Kenneth Abbott and Duncan Snidal, for instance, describe the technical experts (prosecutors) involved in the OECD Working Group on Bribery as a “nascent epistemic community,” in which shared norms and mutual trust develop among officials with a similar professional background.³⁰

Capabilities

Finally, the two reviews displayed interesting variation in their capabilities. The UPR's ability to generate peer and public pressure was deemed superior to that of

the IRM, which, however, scored better in its perceived ability to trigger mutual learning. In this regard, we expected transparency to play a strong, yet ambivalent, role.

Confirming our expectations, we argue, first, that transparency has a strong impact on the mechanisms' perceived ability to trigger peer pressure. In the UPR, peer pressure arises because recommendations are strongly politically charged. It is difficult for countries to ignore recommendations by another state: once a recommendation is accepted, it takes the form of a bilateral commitment with the reviewing country. As put by an interviewee, "It is so difficult to reject a recommendation because after each recommendation in the parenthesis you have the name of the country that made the recommendation" (Interview UPR 15). Although this is caused by the highly political nature of the mechanism, the fact that reports are available to all states and are discussed during plenary sessions allows states to be informed about the human rights situation in the reviewed country and to monitor its progress. Additionally, and in contrast to the IRM, all UN states have the opportunity to issue recommendations to states under review during plenary sessions. Transparency is thus a facilitating condition for peer pressure to develop. In the IRM, we observed a more limited ability to generate peer pressure for two reasons. One, naming and shaming is often frowned on by some officials, even when the objects of criticism are unresponsive states (personal observations³¹). Two, the IRM has few institutional structures in place to exert pressure. The plenary sessions hardly offer any opportunities to discuss country reports and ask critical questions. As mentioned earlier, these sessions devote considerable time to issues related to the management of the peer review rather than the substantive content of reviews. Whenever substantive issues are discussed, country reports are presented as anonymized and aggregated data, making it difficult to identify laggards and leaders, let alone to criticize or commend them. Hence, the lack of peer pressure in the IRM can, to a large extent, be attributed to its limited transparency to the peers.

Second, transparency seems to have a considerable impact on the reviews' perceived ability to trigger public pressure. In the UPR, the fact that all review-related documents are available online and that review sessions can be attended by NSAs and are webcasted facilitates the emergence of public pressure. Conversely, the IRM's opaque nature limits opportunities for public pressure. NSAs cannot attend plenary sessions and reports are not published online. As Fabrizio Pagani, an OECD official, also confirms: the "impact will be greatest when the outcome of the peer review is made available to the public" and the media.³² However, evaluation reports are, in any case, technical and written in inaccessible legal language. It therefore remains doubtful that, if the IRM were more transparent, it would generate more public pressure. As one delegate mentioned, "If you are an activist you really have to scrutinize these reports in detail to get something out of it" (Interview IRM 10). Transparency, in terms of publishing reports, may be necessary for public pressure to develop, but it is not sufficient to do so.

Third, transparency seems to be an obstacle to mutual learning. As shown above, the IRM is perceived to perform better than the UPR in this regard. The answer to the question of whether this can be attributed to the IRM's limited transparency is two-pronged. There is little evidence that the closed doors of the review's plenary sessions explain the observed variation. As discussed above, the information disclosed during plenaries is rather generic and individual country reports are not discussed. Closed doors to plenary sessions do not seem to stimulate a frank debate among delegates on sensitive issues and, hence, seem to play a limited role in organizing learning. Yet some interviewees indicated that the nonpunitive nature of the peer review stimulates openness and mutual learning; for instance, during desk reviews and country visits. Because information cannot be made publicly available without their consent, states might be more forthcoming in sharing information with the peers and, as such, foster mutual learning (Interview IRM 9). In this line of reasoning, it seems plausible that the fact that NSAs can attend review sessions of the UPR—either in person or via a webcast—stimulates diplomatic posturing. Our findings are thus partially comparable to those of Jane Cowan and Julie Billaud, who reported that several involved actors perceived the UPR as a potentially humiliating exam, where “bad students” are put in the spotlight. Rather than a forum for learning, the UPR thus appears to be a setting where diplomatic posturing and political maneuvering take place.³³ While Markku Lehtonen argues that exposing countries to public criticism and triggering learning do not necessarily have to be mutually exclusive goals, our findings suggest—in line with arguments by Jeffrey Checkel and Andrew Moravcsik—that learning is more likely to take place in protected closed-door settings.³⁴

Fourth, against our expectations, we observed that the UPR (which offers NSA input in the review process) was not perceived as better at formulating accurate review reports than the IRM. In the IRM, interviewed officials indicated that the evaluation exercises may go less in-depth due to the nonobligatory nature of country visits and the possible exclusion of NGOs from providing input in the evaluations (Interviews IRM 2, 3, 12). Thus, this transparency provision seems to account for some concerns about the IRM's ability to formulate accurate reports. Nevertheless, the UPR does not perform better than the IRM because recommendations are perceived as strongly politically motivated and often stemming from a lack of knowledge of country situations.

Finally, with regard to the perceived ability to deliver practically feasible recommendations, we did not expect transparency to play any role. Our empirical results confirm this expectation.

Conclusion: Time to Leave the Door Ajar?

This article tells the story of two peer reviews designed with a similar goal in mind: to evaluate the performance of states in a highly sensitive field. Nonethe-

less, completely different choices were made regarding their transparency. We studied the extent to which the transparency of the UPR and the IRM influences their authority. Notably, we sought to understand the extent to which the observed differences in two authority dimensions, proceduralism and capabilities, can be attributed to the different transparency provisions of these mechanisms.

Our findings on proceduralism were highly surprising. Though the two peer reviews, as anticipated, exhibited variation on this dimension, the pattern is the opposite of what we expected. The transparent UPR fares more poorly in terms of (perceived) equal treatment of states than the more opaque IRM. Thus, transparency appears unable to prevent political bias in the UPR. Likewise, lack of transparency in the IRM did not evoke perceptions of unequal treatment. This finding is particularly striking, considering that the UPR was established with the explicit aim of ensuring equal treatment of states, and that transparency was established as one of the tools to achieve this goal. We explained these differences by focusing on the political nature of the UPR vis-à-vis the more technical-oriented IRM reviews. What matters for the development of a politically unbiased review is not so much the level of transparency of the meetings, but rather the type of actors who are present at these meetings and who carry out the reviews. The involvement of technical experts might create the impressions of “depoliticized” and “objective knowledge” and be conducive to the formation of an epistemic community where mutual trust and shared norms develop.³⁵

As to capabilities, our findings fall more in line with expectations. First, the UPR was perceived as comparably better at organizing peer and public pressure than the IRM. Access to information about states’ performances and plenary discussions on country reviews appear crucial for peer and public pressure to develop, as perceived by respondents. In the IRM, delegates have insufficient insight into other states’ performance to exert peer pressure. Likewise, the paucity of publicly available information on countries’ performances inhibits opportunities for public pressure to develop. Second, the more confidential setting of the IRM appears to stimulate learning. This, however, mostly occurs in the closed setting of the country reviews (e.g., the country visits and desk reviews), not in plenary sessions. In contrast, we argue that transparency leads some states to withhold sensitive information on their human rights performance, and triggers diplomatic posturing. This offers a plausible explanation for the UPR’s limited perceived ability to trigger learning, falling in line with arguments by Checkel and Moravcsik.³⁶

Returning to our initial question, and based on the findings from our study, should we conclude that, for the development of peer review authority, the doors to peer reviews should be left open or closed? This calls for a qualified answer, and depends on the goals of the review and the transparency dimension at hand. In our two cases, transparency strongly contributes to perceptions of peer and public pressure, yet negatively affects mutual learning. Additionally, transparency appears unable to ensure equal treatment of states.

Returning to the four dimensions of transparency identified above—namely, the publication of review reports, plenary discussion of country reports, input by NSAs, and possibilities for the public to attend review sessions—we noticed that transparency indeed plays a role in the authority of peer reviews, but that not all transparency dimensions are equally relevant. First, the public availability of country reports cannot guarantee a fair review, yet it does not appear to be harmful either. However, it plays a strong role when it comes to facilitating peer and public pressure. Second, and similarly, discussing country reports in plenary sessions is valuable in triggering peer and public pressure, as perceived by respondents. Third, we cannot conclude that allowing NSAs input in the review process necessarily improves the perceived quality of review reports, yet it does not decrease it either. Finally, allowing NSAs to attend or follow review sessions is successful in creating public pressure, although it does not necessarily prevent the emergence of political bias.

These findings should be placed within the context of the UN, which brings together many states in a political setting. Transparency may have a different effect on the authority of peer reviews conducted within different organizational or policy contexts, which remain areas for further research.

IRM Interviews (conducted by Hortense Jongen)

1. UN Secretariat official, phone, December 2013
2. Member state official, Western European and Others Group (WEOG), national capital, January 2014
3. Member state official, WEOG, national capital, January 2014
4. Member state official, African Group, Vienna, July 2014
5. UN Secretariat official, Vienna, July 2014
6. Member state official, WEOG, Vienna, July 2014
7. Member state official, Eastern European Group (EEG), phone, May 2014
8. UN Secretariat official, Vienna, July 2014
9. Member state official, Asia-Pacific Group, Vienna, July 2014
10. Member state official, WEOG, Vienna, July 2014
11. Member state official, EEG, phone, May 2014
12. Member state official, Latin American and Caribbean Group (GRU-LAC), Vienna, June 2015
13. Member state official, GRULAC, Vienna, June 2015
14. Member state official, GRULAC, Vienna, June 2015
15. Member state official, African Group, Vienna, June 2015
16. Member state official, African Group, Vienna, June 2015
17. Member state official, Asia-Pacific Group, Vienna, June 2015
18. Member state official, GRULAC, Vienna, June 2015
19. Member state official, Asia-Pacific Group, Vienna, June 2015

20. Member state official, EEG, Strasbourg, June 2015
21. Member state official, WEOG, national capital, August 2015
22. Member state official, WEOG, phone, October 2015

UPR Interviews (conducted by Valentina Carraro)

1. Member state official, GRULAC, Geneva, February 2014
2. Member state official, GRULAC, Geneva, February 2014
3. Member state official, WEOG, Geneva, February 2014
4. Member state official, WEOG, Geneva, February 2014
5. Member state official, WEOG, Geneva, February 2014
6. Member state official, EEG, Geneva, October 2014
7. Member state official, EEG, Geneva, October 2014
8. Member state official, WEOG, Geneva, October 2014
9. Member state official, EEG, Geneva, October 2014
10. Member state official, WEOG, Geneva, October 2014
11. Member state official, WEOG, Geneva, June 2015
12. Member state official, Asia-Pacific Group, Geneva, June 2015
13. Member state official, WEOG, Geneva, June 2015
14. Member state official, EEG, Geneva, June 2015
15. UN Secretariat official, Geneva, February 2014
16. UN Secretariat official, Geneva, October 2014
17. UN Secretariat official, Geneva, October 2014
18. UN Secretariat official, Geneva, October 2014
19. UN Secretariat official, Geneva, October 2014
20. UN Secretariat official, Geneva, October 2014
21. UN Secretariat official, Geneva, October 2014

Notes

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. Hortense Jongen is postdoctoral researcher at the School of Global Studies of the University of Gothenburg. Her PhD dealt with the instrument of peer review among states in the global fight against corruption.

The authors were equal contributors to this article. The research performed in this article was part of the project “No Carrots, No Sticks: How Do Peer Reviews Among States Acquire Authority in Global Governance?” led by Professor Thomas Conzelmann and funded by the Netherlands Organisation for Scientific Research (NWO) (grant number 452-11-016).

We are very grateful to Thomas Conzelmann, Sophie Vanhoonacker, Giselle Bosse, Andreea Năstase, Aneta Spendzharova, and the participants in The Quest for Legitimacy in World Politics—International Organisations’ Self-Legitimations workshop (ECPR Joint

Sessions 2015) for their comments on earlier drafts. We also express our appreciation to Ian Lovering and Lea Smidt, who offered invaluable research assistance.

1. David Heald, "Fiscal Transparency: Concepts, Measurement and UK Practice," *Public Administration* 81, no. 4 (2003): 723–759; Christopher Hood, "What Happens when Transparency Meets Blame-avoidance?" *Public Management Review* 9, no. 2 (2007): 191–210; Christopher Hood, "Accountability and Transparency: Siamese Twins, Matching Parts, Awkward Couple?" *West European Politics* 33, no. 5 (2010): 989–1009; Christopher Hood and David Heald, *Transparency: The Key to Better Governance?* (Oxford University Press, 2006); Ronald B. Mitchell, "Sources of Transparency: Information Systems in International Regimes," *International Studies Quarterly* 42, no. 1 (1998): 109–130; Ronald B. Mitchell, "Transparency for Governance: The Mechanisms and Effectiveness of Disclosure-based and Education-based Transparency Policies," *Ecological Economics* 70, no. 11 (2011): 1882–1890; Onora O. Neill, *A Question of Trust* (Cambridge: Cambridge University Press, 2002); David Stasavage, "Open-door or Closed-door? Transparency in Domestic and International Bargaining," *International Organization* 58, no. 4 (2004): 667–703.

2. Fabrizio Pagani, "Peer Review as a Tool for Cooperation and Change," *African Security Review* 11, no. 4 (2002): 15–24.

3. Armin Schäfer, "A New Form of Governance? Comparing the Open Method of Coordination to Multilateral Surveillance by the IMF and the OECD," *Journal of European Public Policy* 13, no. 1 (2006): 70–88.

4. Martin Heidenreich and Jonathan Zeitlin, eds., *Changing European Employment and Welfare Regimes: The Influence of the Open Method of Coordination on National Reforms* (London: Routledge, 2009).

5. Bruce Cronin and Ian Hurd, "Introduction," in Bruce Cronin and Ian Hurd, eds., *The UN Security Council and the Politics of International Authority* (London: Routledge, 2008), pp. 3–22.

6. Human rights protection is a policy field that frequently raises sovereignty concerns. See Xinyuan Dai, "Information Systems in Treaty Regimes," *World Politics* 54, no. 4 (2002): 405–446. Similarly, corruption is receiving growing attention, with high-level politicians being forced to resign over corruption scandals. For a comparative assessment of the authority of the IRM to two other anticorruption peer reviews, see Hortense Jongen, "The Authority of Peer Reviews in the Global Governance of Corruption," *Review of International Political Economy* (forthcoming).

7. Michael Barnett and Martha Finnemore, *Rules for the World: International Organizations in Global Politics* (New York: Cornell University Press, 2004); Rodney B. Hall and Thomas J. Biersteker, eds., *The Emergence of Private Authority in Global Governance*, vol. 85, Cambridge Studies in International Relations (Cambridge: Cambridge University Press, 2002); Ian Hurd, "Legitimacy and Authority in International Politics," *International Organization* 53, no. 2 (1999): 379–401; David Lake, "Rightful Rules: Authority, Order and the Foundations of Global Governance," *International Studies Quarterly* 54, no. 3 (2010): 587–613; James N. Rosenau, "Governing the Ungovernable: The Challenge of a Global Disaggregation of Authority," *Regulation and Governance* 1, no. 1 (2007): 88–97; Michael Zürn, Martin Binder, and Matthias Ecker-Ehrhardt, "International Authority and Its Politicization," *International Theory* 4, no. 1 (2012): 69–106.

8. Barnett and Finnemore, *Rules for the World*; David Lake, "Relational Authority and Legitimacy in International Relations," *American Behavioral Scientist* 53, no. 3 (2009): 331–353; Lake, "Rightful Rules."

9. Thomas Conzelmann and Hortense Jongen, "The Power of the Peers: Assessing the Authority of Peer Reviews in Curbing Corruption," paper presented at Workshop International Authority, WZB Berlin, 10–11 December 2015.

10. Barnett and Finnemore, *Rules for the World*, p. 20.

11. Cronin and Hurd, "Introduction," p. 12.

12. Ibid.; Wayne Sandholtz, "Creating Authority by the Council: The International Criminal Tribunals," in Bruce Cronin and Ian Hurd, eds., *The UN Security Council and the Politics of International Authority* (London: Routledge, 2008), pp. 131–153.

13. Cronin and Hurd, "Introduction"; Sandholtz, "Creating Authority by the Council."

14. Jean-H. Guilmette, "Peer Pressure Power: Development Cooperation and Networks—Making Use of Methods and Know-how from the Organisation for Economic Co-Operation and Development (OECD) and the International Research Development Centre (IRDC)" (Ottawa: OECD, 2004), <https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/25923/119954.pdf?sequence=1>; Mohamad Ikhsan, "Economic Reform Under a Democratic Transition," in Kensuke Tanaka, ed., *Shaping Policy Reform and Peer Review in Southeast Asia* (Paris: OECD, 2008), pp. 177–198; Markku Lehtonen, "OECD Environmental Performance Review Programme: Accountability (f)or Learning?" *Evaluation* 11, no. 2 (2005): 169–188; Pagani, "Peer Review as a Tool for Cooperation and Change"; Theodor Rathgeber, *The HRC Universal Periodic Review: A Preliminary Assessment*, Dialogue on Globalization Briefing Papers No. 6 (Berlin: Friedrich-Ebert-Stiftung, 2008).

15. For details, see Conzelmann and Jongen, "The Power of the Peers."

16. For example, Klaus Armingeon, "OECD and National Welfare State Development," in Klaus Armingeon and Michelle Beyeler, eds., *The OECD and European Welfare States* (Cheltenham: Edward Elgar, 2004), pp. 226–241; Bernard Casey and Michael Gold, "Peer Review of Labour Market Programmes in the European Union: What Can Countries Really Learn from One Another?" *Journal of European Public Policy* 12, no. 1 (2005): 23–43; Rik de Ruiters, "EU Soft Law and the Functioning of Representative Democracy: The Use of Methods of Open Co-ordination by Dutch and British Parliamentarians," *Journal of European Public Policy* 17, no. 6 (2010): 874–890; Mariely López-Santana, "The Domestic Implications of European Soft Law: Framing and Transmitting Change in Employment Policy," *Journal of European Public Policy* 13, no. 4 (2006): 481–499.

17. Conzelmann and Jongen, "The Power of the Peers"; Ian Hurd, "Theories and Tests of International Authority," in Bruce Cronin and Ian Hurd, eds., *The UN Security Council and the Politics of International Authority* (London: Routledge, 2008), pp. 23–39.

18. See note 1.

19. Stasavage, "Open-door or Closed-door?"

20. Heald, "Fiscal Transparency"; Hood, "What Happens when Transparency Meets Blame-avoidance?"; Hood, "Accountability and Transparency"; Stasavage, "Open-door or Closed-door?"

21. Mitchell, "Sources of Transparency," p. 110.

22. Mitchell, "Transparency for Governance," p. 1882.

23. Thomas Conzelmann, "Peering at the Peers: How Do Peer Reviews Among States

Take Shape in Four International Organizations?" European Consortium for Political Research Joint Sessions, Mainz, 11–16 March 2013.

24. The national experts are responsible for conducting the peer review and often attend plenary sessions.

25. One survey item was assessed on a scale of 1 to 3.

26. For further information, see the official website of the United Nations Conference on Trade and Development: <http://unctad.org/en/Pages/DITC/CompetitionLaw/Voluntary-Peer-Review-of-Competition-Law-and-Policy.aspx>, accessed 7 June 2017.

27. Except for meetings of the Conference of the States Parties.

28. Valentina Carraro, "The United Nations Treaty Bodies and Universal Periodic Review: Advancing Human Rights by Preventing Politicization?" *Human Rights Quarterly* 39, no. 4 (2017): 943–970.

29. Hortense Jongen, "The Authority of Peer Reviews in the Global Governance of Corruption."

30. Kenneth Abbott and Duncan Snidal, "Values and Interests: International Legalization in the Fight Against Corruption," *Journal of Legal Studies* 31, no. S1 (2002): 166.

31. During the resumed fifth and the sixth session of the Implementation Review Group.

32. Pagani, "Peer Review as a Tool for Cooperation and Change," p. 16.

33. Jane Cowan and Julie Billaud, "Between Learning and Schooling: The Politics of Human Rights Monitoring at the Universal Periodic Review," *Third World Quarterly* 36, no. 6 (2015): 1175–1190.

34. Jeffrey T. Checkel and Andrew Moravcsik, "A Constructivist Research Program in EU Studies?" *European Union Politics* 219, no. 2 (2001): 219–249; Lehtonen, "OECD Environmental Performance Review Programme."

35. Barnett and Finnemore, *Rules for the World*, p. 24; Peter Haas, "Introduction: Epistemic Communities and International Policy Coordination," *International Organization* 46, no. 1 (1992): 1–35.

36. Checkel and Moravcsik, "A Constructivist Research Program in EU Studies?"