Meetings that make a difference: Making grant review panels work – online and onsite

Peer review panels assessing applications for research grants constitute a key redistributive and quality negotiation mechanism in the research system, but their added value is contested. Building on extensive observations of grant review panels and interviews with panellists in a broad variety of funding organisations, this study explores the value of grant panel review in terms of its scope for improving information sharing and processing, and whether this is affected by moving deliberations online. Adopting a systems view of the assessment process, we find that individual panellists' characteristics and the organizational set-up of the funding competition as well as how these factors are mediated by the group dynamics of panel discussions, determine whether panel review adds value to proposal assessment. While panel deliberation has potential intrinsic value by integrating diverse expertise, improving error detection in individual assessments and contributing to a more uniform understanding of review criteria and scoring scales, deliberation also has potential extrinsic value, affecting how individual assessments are carried out. The prospect of panel deliberation constitutes a potent accountability mechanism which leads panellists, motivated by reputational concerns, to assess proposals more systematically and thoroughly. While moving deliberation online is associated with certain limitations to information sharing and processing, these limitations can largely be overcome through organisational measures. However, evaluative complexity is challenging in an online environment, and and unconventional research is thus best assessed in onsite or standing rather than ad-hoc online panels.

1. Introduction

Peer review panels assessing applications for research grants constitute a key redistributive and quality negotiation mechanism in the research system. However, their added value is contested, with some studies indicating that the impact of panel deliberations is negligible compared to a simple arithmetic averaging of individual reviews (Fogelholm et al., 2012; Pina et al., 2015). With the onset of the Covid-epidemic, review panels went through a dramatic organizational change, switching from onsite, face-to-face deliberation to online deliberation. There are indications that this mode of operation is becoming the new normal also post Covid, but the impact of this change is poorly understood. A limited number of studies have analysed whether panel review has added value, and there are also contributions assessing how online panel meetings compare to onsite meetings. However, existing literature on these subjects has predominantly been of a quantitative nature, focusing on analysis of pre and post panel scores to ascertain the added value of panel deliberations and the effect of moving deliberations online. More qualitatively oriented studies of how grant review panels operate are needed, but such knowledge is surprisingly scarce (Demicheli & Di Pietrantonj, 2007; Grimaldo et al., 2018; Guthrie et al., 2018; Wood & Wessely, 2003).

While a number of studies have attempted to open the black box of panel review processes, these have predominantly been based on secondary sources such as analysis of reviewers' written comments and surveys and interviews with panellists. While providing important insight, such sources risk giving an incomplete view of the process, with written assessments giving a curated view of the assessments performed, and interviews about such assessments being subject to individual perception biases and social desirability bias. There have thus been consistent calls for observational studies of grant review panels' functioning (Guthrie et al., 2018; Lamont et al., 2006a; Laudel, 2006b; Mansilla, 2006; Sorrell et al., 2018; van Arensbergen et al., 2014), but very few such studies exist due to the confidential nature of review panels' deliberations. Furthermore, the majority of studies have

been carried out within single funding agencies, focusing on single funding instruments (Hug, 2022; Huutoniemi, 2012; Reinhart & Schendzielorz, 2021), thus limiting our understanding of how variability in organisational context impacts the added value of panel deliberation in grant review processes and the effect of a move to onsite deliberation.

This study aims to address these limitations in our understanding of grant review panels functioning. Drawing on extensive observations of grant panels and interviews with panellists, carried out over a 2,5 year period from the onset of the pandemic until its abatement, **we explore whether grant panel deliberation adds value to the grant review process and how a move from onsite to online deliberations affects this potential value.** Specifically, we seek to ascertain if panel deliberation can deliver on the potential associated with group decision making of improved information sharing and processing. The data set includes observations and interviews with panellists in nine different funding competitions in five different research funding organisations at national, Nordic and EU-level, encompassing 68 panels and 88 interviews. The quantity and variety that the empirical material covers is unparalleled by any previous qualitative studies on the subject, thus addressing criticism that research in the field has been characterised by small sample sizes with corresponding negative effects on robustness of results (Reinhart & Schendzielorz, 2021).

Adapting a systems view of how panel deliberation adds value to the assessment process, we aim to overcome the limitations associated with the simple input-output understandings of how evaluation outcomes are produced that characterizes much of the existing literature. We argue that while grant panel review has the potential to enlarge the pool of knowledge brough to bear in the assessment of grant applications and ensure more thorough processing of applications, individual panellists' characteristics, the organizational set-up of the funding competition and how these factors are mediated by the group dynamics of panel discussions, will determine to what extent this potential is brought to fruition.

We find that panel characteristics which encourage accountability - both at the individual and group stage of assessment, are key to ensuring that panel review can deliver on its potential of ensuring improved information sharing and processing. Accountability here refers to an explicit expectation that one will be called upon to justify one's assessment to others (Lerner & Tetlock, 2003). Such accountability measures are especially important in an online environment, due to the impediments to interaction that the online format entails. Appropriately set up, online panels provide a viable alternative to onsite discussions for customary assessments and can even serve to reduce the unwanted impact of certain interaction dynamics on outcomes, such as gaming and minority conformity. However, online discussions carried out in ad-hoc panels are ill suited for evaluatively complex assessments due to their limited ability to build trust among participants. Onsite or standing online panels are thus a preferable option for the assessment of unconventional research.

In the following section, we present an overview of the literature relevant to our research questions. We then present the data and methods employed to answer these questions. Next, we describe the key panellist and organisational characteristics as well as interaction dynamics that are relevant to understanding the potential added value of a grant panel deliberation and this value is affected by a move online. Finally, we discuss the policy implications of our findings.

2. Literature review

2.1 The added value of grant panel peer review

Surveys indicate that reviewers see clear added value in grant panel meetings, with most estimating that the quality and effectiveness of panel discussion is high and that discussions influence the outcome of the review (Stephen A Gallo et al., 2020). Relatedly, the literature on social psychology points to several positive effects of group decision making, primarily in terms of improved information sharing and processing. For one, groups can draw on a larger pool of knowledge as each member has unique knowledge they can contribute to the group. In a group context, members are also able to process the information they possess more thoroughly through discussion, asking questions, considering alternative options, weighing arguments against one another, etc. Through such information processing, the group is better positioned than individuals to identify errors. (Forsyth, 2014).

The main alternative to bringing experts together in groups to evaluate applications is to combine individual expert reviews, averaging their scores without any preceding discussion. Some funders practice this approach and some voices in the literature on peer review advocate it. The main argument is that it is more efficient and that panel meetings have limited effect as mean reviewer scores prior to panel meetings are similar to the panel consensus score (Fogelholm et al., 2012; Obrecht et al., 2007). Others, while overall agreeing that panel deliberation can be eliminated, find that there are subsets of proposals where panel discussions have an effect, predominantly those with considerable disagreement between individual reviewers (Martin et al., 2010; Pina et al., 2015; Thorngate et al., 2010). There are furthermore indications that complex proposals, involving multidisciplinary and inter-sectorial research groups, require a more elaborate review procedure (Pina et al., 2015).

Although individuals coming together in groups to perform a task possess the abilities and expertise required to complete an assigned task, they may fail to coordinate their efforts in a productive way, hindering them in reaping the full rewards of group decision making. An example of this is the common group dynamic termed shared information bias which describes a tendency for groups to focus more on the information possessed by all members (shared information) than on the information possessed only by an individual member (unshared information) (Jhangiani et al., 2014) . This will naturally hamper groups' ability to capitalise on the richness of knowledge they possess. Normative conformity is another well-known group dynamic that might limit information sharing in a group context. This is the tendency to express or supress opinions and to behave in ways aimed at encouraging acceptance in the group (Jhangiani et al., 2014).

2.2 Moving meetings online can have negative implications for information sharing and processing

Most of the literature exploring the effect of moving panel meetings online find little difference between onsite and online deliberation. A number of studies analysing scoring patterns pre and post discussion in a teleconferencing/online format vs onsite format find limited differences in scoring patterns, concluding that meeting medium does not influence the fairness of the review process (Gallo et al., 2013; Obrecht et al., 2007; Pier et al., 2015; Pina et al., 2015; Recio-Saucedo et al., 2022; Vo et al., 2016; Vo & Trocki, 2015). Others find small differences, noting reduced score shifts in teleconferences as compared to face-to-face meetings (Carpenter et al., 2015). Reduced engagement among teleconference reviewers was pointed to as a possible explanation.

Many of the studies exploring scoring patterns find that discussion time tends to be more limited in online discussions (Carpenter et al., 2015; Pier et al., 2015; Pina et al., 2015), potentially lending

support to the hypothesis of reduced engagement in an online setting. Further supporting indications of reduced engagement in an online context, a recent observational study of online grant panels found a reduction in spontaneous interjections and more formal 'turn-taking' deliberation which implicitly invited agreement only. In addition, the perspectives of new panel members proved difficult to integrate, and there were persistent challenges related to 'at home' distractions. (Derrick, unpublished).

When panellists are interviewed about the virtues and drawbacks of virtual panels, they point to similar issues, suggesting in-person meetings are superior in terms of thoroughness of discussion, greater ease of speaking up and reduced scope for multi-tasking or distractions (Pier et al., 2015). Compared to onsite meetings, online meetings are perceived to take a greater toll on attention and requiring the investment of more effort in communication and listening (Stephen A Gallo et al., 2020).

Panellist surveys confirm that the online discussion format entails reduced focus and engagement. A survey of grant reviewers for the US National Institutes of Health found that reviewers have shorter attention spans and lower engagement during video grant-review meetings. Compared with in-person grant panels, 46% of respondents said that they paid less attention during the video meetings, and 51% said that their engagement was worse (Chawla, 2021). Similarly, in a survey among grant reviewers at the Research Council of Norway, 30% estimated that online discussions were less thorough than onsite discussions, and there was a clear preference for shorter meetings in the online format compared to an onsite format (NIFU policy brief om online review meetings 2023). A survey among US reviewers in the fields of physics and natural sciences similarly found that panellists consider onsite discussions more thorough, pointing to better conditions for interpreting body language, building rapport, listening actively, voicing disagreement and sustaining attention (Stephen A. Gallo et al., 2020). Reviewers furthermore point out that the online format entails loss of informal interaction (Gallo et al., 2013; Pier et al., 2015) + NIFU policy brief om online review meetings 2023). However, there are indications that such interaction has the flipside that it allows for gaming and behaviour that violates important codes of conduct (Coveney et al., 2017).

However, there are also positive aspects associated with a move to online grant panels, including reduced costs and reduced time required for meetings (REF) and greater ease in composing more diverse panels as it enables recruitment of members that for lack of time or resources do not have the possibility to travel and/or be absent from home (Davis et al., 2020; Meadmore et al., 2020). More diverse panels have been shown to improve information sharing as when opinions diverge, a greater variety of perspectives – unshared information – is introduced into the discussion (Olbrecht & Bornmann, 2010). People are also more motivated to engage in deep and systematic information processing as they become less confident in the accuracy of their own opinions (De Dreu et al., 2011).

2.3 Organisational measures can improve information sharing and processing

Accountability

While much of the literature on the added value of grant panel review and the effect of moving panel review online predominantly focus on panel deliberations, panellists point out that the most important work is that done individually, when panellists read and rate applications prior to the panel meeting (Huutoniemi, 2012). This is corroborated by work showing that individual panellists' proposal scores has a strong anchoring effect on panel deliberations (Roumbanis, 2017) (Oxley in press), as the average of panellist's scores constitute an "anchor" from which the final result will

rarely deviate markedly. This furthermore aligns with findings that mean reviewer scores prior to panel meetings are similar to the panel consensus scores (Fogelholm et al., 2012; Martin et al., 2010; Obrecht et al., 2007; Pina et al., 2015; Thorngate et al., 2010).

However, evidence from behavioural decision research shows that the quality of individual reviews cannot be taken for granted. Individuals have a tendency to jump to conclusions on the basis of insufficient evidence, do not engage in deep thinking and use a variety of heuristics and shortcuts to simplify decisions (Lerner & Tetlock, 1999; Scholten et al., 2007). Making individuals accountable for their judgements appear to counteract such tendencies, with predecisional accountability to an audience with unknown views shown to stimulate effortful, self-critical thought (Lerner & Tetlock, 1999). As they want to avoid appearing foolish in front of an audience, people prepare themselves more effortfully, decreasing their susceptibility to a number of biases that arise from lack of critical attention to one's decision processes and failure to use all relevant information (Lerner & Tetlock, 2003).

Like individuals, groups also engage in more or less deliberate information processing and in the same way as for individuals, accountability can influence groups' information processing (De Dreu, Nijstad, et al., 2008; Scholten et al., 2007). In particular, accountability serves to increase group members epistemic motivation – "their willingness to expend effort to achieve a thorough, rich, and accurate understanding of the world, including the group task or decision problem at hand" (De Dreu, Baas, et al., 2008, p. 23). Groups characterized by high epistemic motivation engage in more systematic and thorough search for and processing of information. As such they are unlikely to be influenced by inaccurate decision heuristics and reasoning errors. They are also more open to minority dissent and are characterised by more egalitarian and participative interaction patterns, stimulating the dissemination of unshared information (De Dreu, Nijstad, et al., 2008). Accountability also serves to improve groups' information processing by reducing the anchoring effect (Lerner & Tetlock, 2003).

In a panel review context, common accountability measures are to require all or a subset of panellists to submit their scores and/or comments to the proposals under review prior to the panel meeting. Often these will then in turn be shared with their fellow panellists. A 2022 survey among the Research Council of Norway's panellists shows they deem such sharing of written individual reviews to have considerable potential to improve panel meetings in an online setting. (NIFU policy brief). Research in social psychology corroborates that the collection of individual judgements prior to discussion can serve to enhance information sharing by attenuating groups' tendency for informational cascades. This is a dynamic whereby an initial positive or negative assessment affects the info that is shared, with info that goes against the initial statement tending to be suppressed (Kahneman et al., 2020; Kahneman et al., 2021).

Moderation style

There are indications that the group leader has an important role to play in fostering group-level epistemic motivation (De Dreu, Nijstad, et al., 2008). The leader can foster norms of open discussion and actively solicit unshared information from group members (Jhangiani et al., 2014) thus controlling the group's tendency to focus on shared information (Larson et al., 1994). The chair should not be involved in reviewing proposals, but should be independent in order to avoid efforts at gaming the system in favour of certain proposals (Coveney et al., 2017), and enabling him/her to ensure all panellists get the chance to express their opinions, playing the role of devil's advocate in cases where panels are too eager to reach consensus quickly (van Arensbergen et al., 2014).

This chair role is in line with the expectations reviewers have of a well-functioning chair. Interview and survey studies find that reviewers estimate that a strong chair can mitigate biases and ensure engagement and inclusion, optimally leveraging panel expertise (Coveney et al., 2017; Stephen A Gallo et al., 2020; Mow, 2010). However, many report experiences of ineffective chairs, suggesting chairs should receive training in how to lead and facilitate a discussion appropriately (Stephen A Gallo et al., 2020).

Discussion style and time

There are indications that the shared information bias is particularly pronounced in unstructured, free-flowing discussions (Stasser & Titus, 1985). A discussion protocol where panellists discuss review criteria one by one rather than having a general discussion can thus result in improved information sharing.

There are numerous contributions indicating that time pressure accentuates the shared information bias. Unshared information has been shown to become more prevalent in group discussion over time (Larson et al., 1994), while time pressure has the effect of increasing groups' desire for uniformity of opinion (Kruglanski et al., 1993) and reduce people's epistemic motivation (Kelly & Karau, 1999; Kelly & Loving, 2004). Time pressure has also been shown to affect information processing, with groups under stress tending to narrow attention onto more vital task features and engage in more simplified, heuristic information processing (Kaplan et al., 1993; Karau & Kelly, 1992). Similarly, epistemic motivation is reduced when people become fatigued (Kruglanski & Webster, 1996) and there are indications that panel members' ability to interact effectively will diminish as they tire after considering many applications in succession (Thornley et al., 2002).

3. Methodology

3.1 Data

Nine different funding competitions in five different research funding organisations at national, Nordic and EU-level are included in the data analysis, encompassing observations of 68 panels and interviews with 88 panellists. Observations were carried out at regular intervals over a two-and-a-half-year period, some of which in the same funding organisations. This longitudinal data allowed us to monitor the progressive adjustments made to align the assessment process to the characteristics of an online context. A project carried out be the author in 2016 for the Research Council of Norway, comparing its grant review process with EU-level review processes, provided a comprehensive view of onsite assessment practices prior to the move to online deliberations. The 2016 project entailed extensive observations of panels within the same RCN funding programme and EU funding agencies for which data for the current study was collected, thus providing a point of reference for how the move to online deliberations affected review processes in these organisations.

The calls were selected as they constitute a maximum variation sample with respect to the key organisational dimensions which the literature indicates is of relevance to the added value of panel deliberation: Accountability measures in the form of rre- and post-discussion evaluative documentation required, discussion model, moderation model and discussion time:

Organisation	***Call***	***Panels observed / total no. of panels***	***Panel characteristics***
NordForsk	2020 call in Nordic Programme for Interd. Research	1/1	- Assigned reviewers provide general evaluative documentation pre and post discussion. Non-assigned reviewers not required to assess proposals - Very limited discussion time (12 min) - Mix of unstructured discussions/ discussion structured according to the main review criteria - Modest moderation, ensured by one of the panellists
Research Council of Norway	- 2021 call for Large-scale Interd. Researcher Projects - 2021 call for Researcher Projects	20/40	- Assigned reviewers provide general evaluative documentation pre and post discussion. Unassigned reviewers provide scores pre-discussion. - Limited discussion time (20- 30 min) - Mix of unstructured discussions/ discussion structured according to the main review criteria - Modest moderation, ensured by agency staff
Research Council of Norway	2021 call for Researcher Project for Technological Convergence	5/8	
Dam	2021 call for Research	6/6	- All provide general evaluative documentation pre discussion - Limited discussion time (20 min) - Mix of unstructured discussions/ discussion structured according to main review criteria - Modest moderation, ensured by agency staff
CINEA	2021 Horizon Europe call in the fields of transport, climate and energy	24/24	- All provide detailed evaluative documentation pre discussion and detailed post-discussion documentation developed cooperatively - Long discussion time (3 hours) - Highly structured discussions, according to sub-criteria - Active moderation, ensured by agency staff
CINEA	2022 Horizon Europe call on climate adaptation	6/6	
CINEA	2022 Horizon Europe call on climate sciences	3/3	
Europe's Rail	2023 EU-level call on rail	3/3	- All provide detailed evaluative documentation pre discussion and detailed post-discussion documentation developed cooperatively - Long discussion time (1,5 hours) - Highly structured discussions, according to sub-criteria - Active moderation, ensured by agency staff

While the cases selected allow us to explore our core question of whether panel deliberation has added value in organisational contexts that vary in analytically interesting ways, the calls are sufficiently similar with respect to how the review process is organised to enable comparison. For all calls observed except the NordForsk call, the main evaluation criteria were excellence, impact and implementation. Proposals were judged by a panel of experts consisting of 3-7 members that first read and scored proposals individually before meeting to discuss the proposals. Panels were composed in an ad-hoc manner with a view to ensuring that panel expertise matched the proposals under review. Discussions were moderated by personnel from the funding agency. NordForsk was an exception also here, as one of the panel members were designated chair.

The NordForsk call further differed from the others in that it had a different criteria set. This call also included an additional evaluation step where all applications were assessed by two individual external experts which applicants could then respond to in a written rebuttal, before individual assessment and discussion of applications by panel members. The panel was larger than in the other calls, consisting of 11 experts.

The Foundation Dam call differed from the others in that it used standing panels and did not require them to reach consensus on assessments. Rather, panellists presented their respective assessments of the proposals under review and had a brief discussion, after which all panellists had the opportunity to adjust (or not) their original scores based on the information that came up in discussions. The final panel score for each proposal was calculated based on the individual scores set.

INSERT figure showing the main assessment process with the Dam and NordFOrsk variations included.

In the Research Council of Norway's call for Researcher Projects and Large-scale Interdisciplinary Researcher Projects, the large number of panels active in the assessment process (40) prohibited observations of all panels involved. However, care was taken to observe a broad variety of panels, covering the broad research fields of life science, physical sciences and engineering, and social sciences and humanities to enable exploration of commonalities and differences across fields.

A minimum of two experts per panel were interviewed in the RCN calls and in the Foundation Dam, while all panel members bar two which declined to participate were interviewed in the NordForsk call. At EU-level, data collection restrictions imposed by the organizing authority meant only very limited interview data could be collected.

3.2 Data collection and analysis

The study is primarily based on observations of grant panels at work. This is complemented with semi-structured interviews with panellists and document analysis, primarily of call-level information and summary tables of panellists' pre-panel and post-panel scores. As all data collection was done during the Covid epidemic, both panel meetings and interviews were carried out on online platforms.

The observation technique used was an overt (the observed know about the observation), non-participant, naturalistic observation (observing the spontaneous behaviour of participants in natural surroundings) (Angrosino, 2016). That meetings took place online is likely to have reduced the effect of being observed on the observed (Hawthorne effect) compared to onsite observation, as the observer switched off both camera and microphone after the initial introduction. Confirming this, interviewees testified that they tended to forget that an observer was present as discussion progressed. Panel deliberations were recorded through detailed field notes primarily aimed at capturing the dialogue between participants as fully as possible. In addition, more subtle aspects that

added meaning to the exchange was recorded, such as if something was said in a joking manner, an irritated tone, etc.

Interviews with panel members were carried out shortly after panel meetings took place to reduce recall errors. They lasted an average of one hour and were recorded and transcribed verbatim. Care was taken to avoid leading questions, and to adopt a non-judgmental approach of active listening and probing. Interviews were focused on three main lines of inquiry: panellists' individual assessments, their views on the panel discussions, and their assessment of the organisation of the evaluation process. The interview guide was adjusted continuously, guided by ongoing interpretation of data and the emergence of key categories. For example, while interviews initially focused primarily on panel discussions and organisational measures, adjustments were subsequently made to improve understanding of assessments in the individual evaluation stage, as the positive effect that the prospect of panel deliberation had on the thoroughness of individual reviews became increasingly clear.

Interviews took a semi-structured form and were tailored to each interviewee based on a review of panel field notes and studies of pre- and post-discussion scores. Linking generic interview questions to specific situations and discussions in the panel and through such "anchored interviewing" (REF) aided participant recall and also uncovered differences between interviewees' general reasoning regarding their reviewing methods, and their reasoning with regards to specific review situations. Furthermore, different experts' interpretations of the same situations could be collected, allowing for an analysis of similarities and dissimilarities in perspectives, as well as an analysis of how participants' views of panel situations differed from the observers' own interpretation. Triangulating observation and interview data thus guarded against uncritically adopting participants' view and aided in critically questioning own emerging interpretations of the data. Furthermore, it enabled the identification of social desirability bias in interviewees' responses, as it allowed for contrasting people's panel behaviour and their post-meeting accounts.

Interview transcripts and field notes were analysed using Nvivo. A preliminary list of core coding categories was developed based on the interview guide, the preliminary literature review and the analytical memos developed in the early stage of the data gathering process. In the first stage, a very rudimentary coding structure was developed, divided into the three overall categories of panellist characteristics, organisational characteristics and interaction dynamics. Dedicated codes captured interview and observation comments pertaining specifically to panellists' assessments of the added value of panel meetings and how the move to online deliberations affected panel functioning. The three broad coding categorises were subdivided into more specific codes based on a mix of a deductive an inductive approach. For example, the code organisational characteristics was initially subdivided into the the key organisational dimensions which the literature indicates is of relevance to the added value of panel deliberation: Pre- and post-discussion evaluative documentation required, discussion model, moderation model and discussion time. Based on the constant interplay between data collection and analysis it was reworked and adjusted iteratively as codes were carefully compared with each other and with data. Redundant codes were deleted and relevant codes further elaborated and grouped together to form more comprehensive codes or decomposed into sub-codes to capture important nuances.

Concepts identified and the emerging associations among them were recorded and developed through analytical memos to guide the continuous analysis phase. The potential explanatory power of various theories in relation to the data was systematically assessed as the research process

unfolded. As system theory emerged as the theory holding greatest explanatory power, we used the input-process-output (IPO) model as a structuring device to draw out the relationships between concepts. This model constitutes the dominant system theoretical perspective on group performance in the psychology literature on groups (Forsyth, 2014), and is well adapt at capturing the complexity of interacting elements which make up the grant review system. In line with systems theory, we find that the component parts of the grant review system can best be understood in the context of their relationships with each other rather than in isolation (Wilkinson, 2011).

4. Results

Here we present the key panellist, panel and online review characteristics as well as the main group dynamics of importance to ascertaining the added value of panel deliberation. Quotes are provided to illustrate how we moved from observations and interviews to key concepts. Quotes from interviews carried out in Norwegian, Swedish and Danish are our translations. A broad variety of representative quotes illustrating the most important panellist, call and online review characteristics as well as group dynamics are provided in dedicated tables in annex to enable us substantiate these key concepts thoroughly. Quotes used in the running text are limited to one illustrative quote. Interview quotes dominate over observation quotes, as observation quotes often necessitate an understanding of the overall discussion context to grasp their true meaning.

Generally, in trying to capture the complexity of the evaluation system and how its constituent components interact, we have had to forgo detailed discussion and analysis of its individual elements.

4.1 Panellists' characteristics

When panellists describe how they go about carrying out their individual assessments, they frequently describe how their desire to make a good impression on their fellow panellists motivates them to carry out more thorough assessments:

I once woke up at 4 in the morning and went through all my grades and notes one more time before a panel meeting /…/ It means something to you, there are people who sit on the committees, and I know some of them, so I want them to think that I make smart judgments, that I think in a smart fashion. I don't want to sit there with lots of smart people and just say stupid things. So it's a little scary. (Interview Dam)

This contrasts with evaluations where they are merely requested to supply written assessments, with no requirements to substantiate these assessments to colleagues:

"I do the occasional individual evaluation, but that does feel much more like just work, in a way that feels more administrative. This morning I got an automatic email updating me on the status of one of these applications. /…/ I have no memory of it. But in the RCN one I will remember the proposals, because you are that much deeper into it because it involves, you know, dialogical process and you're exposed in front of colleagues and then you just take it more seriously." (Interview RCN Fripro)

A clear pattern can be discerned whereby increased levels of accountability leads to more thorough assessments. While the mere prospect of having to defend their assessment to peers leads panellists to carry out more thorough assessments, adding additional accountability measures tends to lead to further improvement in information processing. For example, RCN apply a practice of assigning two to three panel members a special responsibility for composing written comments to proposals prior to panel discussion and composing feedback to applicants post discussion. The remaining panellists are also expected to read and score all proposals, but are only required to submit scores, no

comments, prior to the panel meeting and are not involved in composing feedback to applicants. RCN panellists reported that when they were not assigned a special responsibility, this generally led them to carry out a more cursory reading of proposals, using a number of decision shortcuts rather than processing information systematically:

You are kind of reading through it probably much quicker than if you are first or second assessor. I'm trying to capture the idea and then of course it is also emotional. Do you like the idea or do you do not? Do you understand the idea or not? And you're probably making judgments rather on the non-technical things. So is the proposal put together in a proper way? (RCN FRIPRO interview).

In the Dam evaluation, an experimental intervention was set up designed to test the effect of strengthened accountability measures on individual information processing and collective information sharing and processing. Members of four out of eight panels were required to disclose their assessment scores to their fellow panellists, while the other half of panels were explicitly prohibited from doing so. Ultimately, the effect of this intervention on individual information processing could not be confidently ascertained as panellists received the information later than intended, meaning the majority had already carried out their assessments upon receipt of the information. Even so, observations and interviews indicate that such strengthening of accountability measures leads panellists to carry our more thorough information processing:

I did the assessments quite quickly this year because I had a lot of other things to do. I put those scores down a bit hastily, so I felt that they were a bit poorly thought through. Had I known that the scores were going to be shared, I might have spent a little more time on it before I just clicked my assessments into the form, but I will make a note of that for next year. (Panel meeting Dam).

4.2 Panel characteristics

A number of panel characteristics have an impact on the degree to which panels engage in thorough information sharing and processing. These include discussion model, moderation model and discussion time. Furthermore, the form of pre- and post- discussion evaluative documentation required and how such documentation is used to structure discussions has an important impact.

Structured discussions are observed to lead to improved information sharing and processing. RCN panels which only carried out a general discussion of proposals had shorter and less thorough discussions than those which structured the discussion according to the evaluation criteria. The same tendency could be observed in the NordForsk panel, as different constellations of panellists varied in the degree to which they structured their deliberation according to the review criteria. At EU-level all discussions were consistently structured strictly according to the criteria – at sub-criteria level, and these were also the longest and most through discussions observed.

An experimental intervention in the Foundation Dam, designed to test the effect of a structured vs unstructured discussion format, lends support to the importance of a structured discussion. Analysis of this intervention showed more information sharing and more systematic information processing in the panels applying a structured discussion format. An unforeseen experimental alteration whereby one of the groups which initially practiced a structured discussion format switched to an unstructured discussion format in the course of discussions, had the observable effect of reducing information sharing and processing. This suggests that the differences observed in this respect between the other groups practicing structured vs unstructured discussion was not down to the characteristics of the panel members, but due to the discussion format.

Active moderation is also observed to be important in order to encourage information sharing and processing, serving to ensure that dominant and non-dominant members alike contribute to discussions, and that all relevant elements of assessments are discussed:

Everybody was, you know everybody not only had the opportunity to contribute, but everybody was asked. That meant the discussion went on a bit longer, but I think there's something different about saying. Does anybody want to say anything? To saying, Okay, Carol, what's your view? (FIRPRO interview)

Active moderation is particularly important in an online environment where it is more difficult to rely on body language for picking up on dissent and consent. This means that the moderators' job tends to be more taxing online than onsite. In the words of one EU-level moderator:

"The difficult time here with remote evaluations is that you cannot have facial expressions and see the body language, while in the meeting you can easily see this through nodding of heads etc. Now I have to go round the group and ask all of them individually what they think."

Interviews and observations indicate that it is important that the moderator is independent, in terms of not being involved in assessing applications. The risk is otherwise that applications favoured by the chair are given preferential treatment in terms of discussion time or scoring:

"I: When I was chair I would have a list with the five best proposals in the order I wanted and I would have them in my back pocket and three days later I would see if I had actually succeeded in getting what I wanted before I left home. I never showed anybody.
Q: And most of the time did you succeed? How did you succeed?
I: I succeed because I badly wanted to. (FRIPRO interview)

Time pressure was overall observed to influence panels' propensity for information sharing and systematic information processing. Here, there was considerable variation among the calls observed, ranging from the NordForsk panel with an average discussion time per proposal of 12 minutes, to the Horizon Europe evaluation model with an average discussion time of 3 hours per proposal. Unsurprisingly, the general pattern was that longer discussion times lead to more extensive information sharing and more systematic information processing. Very limited discussion time was observed to lead to a dynamic whereby discussions constituted little more than score bargaining as this example from a NordForsk panel debate illustrates:

Pi: how much time do we have left now for the other applications? We need to be faster.

Pa: Yes we should speed up. For the next one, I am happy to mark down my six on interdisciplinarity as it was really difficult for me to get my head around it. I am happy to put it as a 4. It was really messy.

U: I agree. It was difficult to asses, messy.

P: Scientific quality was very unclear. I favour a 4.

M: I agree.

P. Overall for the same reason it cannot be above 4. M OK ?

M: That is fine.

Pat: OK – moving on. How about that? Fast ha?

Finally, the requirements for pre- and post-discussion evaluative documentation were observed to be key in determining the degree to which panellists engaged in systematic information processing at both individual and panel level, and also influenced information sharing in panels.

As discussed, in the RCN context, the designation of assigned assessors affected the thoroughness of individual information processing, with assigned assessors carrying out more thorough reviews than non-assigned. This in turn had consequences for information sharing and processing in panel discussions, where assigned assessors were consistently observed to take prime responsibility for sharing information on their assessments and the remaining panel members taking a more pared back role. In terms of information processing, the general pattern observed was that primary weight was attributed to assigned assessors' information, even in cases where panel members had appropriate expertise and had carried out a thorough review:

There were one or two situations where I was not one of the first assessors, I was just on the panel but I understood very well the project and I was very convinced with the project. But in order to be able to persuade the other people, especially the number one and number two assessors, it was very difficult. (RCN FRIPRO interview)

Consequently, when requirements for pre-discussions evaluative documentation are not applied uniformly to all panel members this potentially restricts the pool of knowledge the panel can draw on.

The manner in which pre-discussion evaluative documentation was used to structure panel discussion was also observed to influence information sharing and processing. As discussed, in the experimental intervention in the Dam call, panellists in four out of eight panels were required to disclose their assessment scores to their fellow panellists, while the other half of panels were explicitly prohibited from doing so. This was observed to result in improved information sharing in the panels allowing discussion of scores, primarily through reducing the effect of informational cascades whereby info that goes against initial statements tends to be suppressed. This dynamic was counteracted when scores were shared as it prompted a need to justify scores that went against the majority opinion. However, when scores were not shared, there were numerous occasions where the scores given (and which the observer had access to) indicated that panellists' views were not aligned with the majority opinion, but where they nonetheless indicated overall agreement with previous speakers.

Similarly, differing manners of using pre-discussion evaluative documentation to structure panel discussion was observed to influence information sharing and processing in the RCN FRIPRO call. While some panels merely displayed panellists individual scores on screen during discussions, some also displayed panellists' comments substantiating the scores. This latter approach was observed to aid both information sharing and processing, acting as a focusing device for panellist to probe each other's assessments more thoroughly and helping to integrate these:

It gives you two ways of following the argument, I mean you can listen to the person who introduces it or you can just read it. So it gives you more options, so that's that's good. And of course having that visualised sometimes helps you seeing where the big arguments are and where that might differ from your own arguments and while you're discussing, it is still on the screen, so you get all the facts and all the positives and negatives of the proposals written on the screen so you can revisit that as often as you like and think about it. While if it is just done verbally, you only get it once. (FRIPRO interview)

At EU-level, this approach was taken one step further, as panellists' individual assessment comments were integrated into a draft panel assessment report prior to panel discussions to function as a discussion aid. Where the draft report revealed differences of opinion in the individual assessments, discussions were pursued to achieve consensus. This meant that information from more timid and more dominant members as well as minority and majority opinions alike were brought into the discussion in equal measure, ensuring comprehensive information sharing. The use of pre-discussion evaluative documentation as a structuring device for discussions thus functions as something of an equalizer in terms of information sharing.

Some EU-level panels extended the requirements for pre-discussion preparations further, by requiring panellists to carry out a pre-discussion through written comments in the online evaluation platform. Similarly, this measure appeared to have an equaliser effect, reducing the influence of minority/majority dynamics and power plays on discussion outcomes and largely eliminating the effect of time pressure.

Furthermore, requirements for collective development of post-panel documentation of evaluative outcomes was observed to encourage systematic information processing at panel level. All panels except the Dam panels were required to provide a unified reasoning for their panel scores in the form of explanatory comments to be shared with applicants. Dam panels were not required to reach consensus and thus not required to provide such unified comments. This was observed to lead to reduced information processing in the Dam panels compared to the other panels observed. Interaction would often merely consist of a series of monologues whereby panellists presented their primary observations of the application under review, but with no discussion about these observations. An extreme example of the limitations to information processing that this approach entailed was one instance observed where one of the panellist mistakenly presented the assessment of a different proposal than that under discussion, and this was not immediately picked up:

*Expert1: *presentation of assessment*

*Expert2: *presentation of assessment*

Expert3: The positive aspect was the ultimate output in the form of a calculator than can have potential clinical value. The weak part is that the PhD student's tasks are not sufficently described.

*Expert4: I agree with what is being said. *Presentation of assessment.*

Expert2: To expert3 – are you sure you are not commenting on the wrong application now? There was no calculator in this project?

*Expert3: It is a good thing that you are paying attention. I was commenting on the wrong applicaiton. On this one i have noted that *description of shortcomings.*

Moderator: Any other comments? No? Then you can adjust your individual scores if you see a need.

Naturally, the reduced information processing in the Dam panels compared to the other panels observed cannot exclusively be explained by the fact that these panels did not have to provide unified feedback to applicants. The fact that these panels did not have to reach consensus also sets them apart from the other panels and contributes to explaining the reduced information processing observed. However, differences in requirements for post-panel evaluative documentation in the other panels observed underlines that this is an important panel characteristic affecting information processing.

Some panels (the majority of RCN panels, the NordForsk panel) did not develop comments substantiating the scores until several days after panel discussions were finalised. Assigned assessors were then required to hand in feedback based on their own written pre-panel comments and notes taken based on the panel discussion. However, a limited number of RCN panels collectively agreed feedback to applicants in the course of discussion, and this was observed to act as a focusing device, encouraging improved information sharing and processing:

"It [agreeing comments] can be a bit tedious, but at the same time I think that it was transparent, so if you had any objections you had your time, you had your moment to give that objection. And people did. People took it seriously. So while it was a bit more time consuming, I think it's probably a fairer way of doing it, because everybody can contribute to that." (FRIPRO interview)

EU level-panels similarly collectively agreed feedback in the course of discussion. Adding to this accountability measure was a system of quality control whereby designated quality controllers checked all feedback forms and send them back to panels for revision if found to be sub-standard. This was observed to encourage very systematic information sharing and processing, with panellists and moderators diligently ensuring that all necessary sub-criteria were comprehensively addressed in order to avoid having their feedback forms returned to them from the quality controllers with new rounds of discussions required as a result.

4.3 Online review characteristics

The large majority of panellists believe that carrying out meetings online detract somewhat from their added value. They primarily point to that the online format impairs ability to sustain focus over time and to pick up body language as well as reduces scope for spontaneous dialogue.

Overall, interviews and observations show that online meetings are more tiring, making it harder to concentrate:

Mod: Today we only have one meeting which is better I think, otherwise it is very tiring

V: Yes looking at the screen all day is very tiring. I had to have a facial massage from one of my youngest yesterday ha ha

J: Yes, it is surprising how tiring it is to just sit in front of a screen and discuss. (Panel discussion EASME 2021)

Adding to this difficulty is the fact that a home setting involves multiple distractions that would not be present in an onsite setting:

I think to be honest, the other thing was when the panel was going on, I was having some plumbing problems, so weird water was coming in the kitchen, and I had to move offices, so I was not as attentive as I might have been when I wasn't having some major domestic disaster. (RCN FRIPPRO interview)

Panellists also report that it is more challenging to block whole days in the agenda when you are not physically away from your normal work-context, resulting in panellist moving in an out of meetings, disrupting the flow of discussions and reducing the pool of knowledge that the panel can draw on:

People dialled in and out and I had to all kinds of stuff in parallel by myself, for example, I had to excuse myself for a PhD thesis on the second day which I had to attend. All these things would not be

possible if you were there in person, and I think this helps also overall the dynamics, but also the interpersonal relationships in the panel. (RCN FRIPRO interview)

Overall, the online format impair participants' ability to pick up body language, making it harder to intervene in a discussion online, thus resulting in less engagement and information sharing in an online environment compared to onsite:

"it's harder to...two or three people discussing something, it's harder to jump in, because they don't see you... normally you just look around and yeah, it makes it a bit more natural... possibly that online environment just enables... potentially it's a little easier to step back from that interactive process isn't it, and potentially if you're in the room, you might pick up on non-verbal quick cues when you look around and think, oh, this person, they look a little unsure about that, and bring them back in, where, when you're doing it, I mean, you got the screen, so you can obviously look at people's faces, but it's not the same."(FRIPRO interview).

While less information tends to be shared in an online environment, the processing of the information that *is* shared also tends to be impaired. This is because the online format tends to result in a discussion model characterised by more formal turn taking, making it harder to contribute in a timely manner to the questions under discussion, thereby raising the bar for intervening even higher:

The kind of like going back and forth on an argument also goes better in real life than online. When you have to kind of get in a queue before you're allowed to speak, which I think... yeah, so the whole interaction is slowed down a little bit, and less good I think, or less deep as well. (RCN FRIPRO interview)

Taken together, these issues lead to a discussion framework that reduces information sharing and processing compared to onsite deliberation. The NordForsk call provides an interesting case in this respect as it was carried out in two stages, including many of the same people in the initial "culling" round and in the second finalist round. However, while the first round of panel deliberations was carried out onsite, the second round was carried out online due to the Covid pandemic. Panellists unprompted frequently referred to the higher quality assessments achieved in the onsite environment in terms of information sharing and processing:

Yeah, so I think this time round the dynamics were just really toned down. Yeah, I mean the interaction was very functional and minimal so people didn't want to step in /…/ The nature of online means you can't just speak. You have to kind of raise your hand up. So then you feel like you have to really have to have something to say that's important. (Interview NordForsk)

However, while online discussions overall compare negatively with onsite discussions both in terms of information sharing and processing, there has been a clear positive development over the 2,5 years of observations. While all interviewees in the start of the pandemic tended to underline that online discussion would have to be a strictly temporary solution, solely acceptable due to the challenges imposed by the pandemic, both panellists and funding agency staff have over time become increasingly positive to the online medium. Those interviewed in the later stages of observations frequently pointed out the positive aspects associated with online meetings such as increased flexibility, ability to recruit a more diverse set of panellists and a reduced carbon footprint associated with the evaluation.

This increasingly positive attitude to online assessments is likely due on the one hand to the participants themselves having become more accustomed to working in an online environment, meaning technical barriers, etc have become reduced at the user-side. On the other hand, funders at

their end have ensured technical infrastructure upgrades and improved working methods. At the start of the pandemic, the traditional manner of organising panel assessment was simply transposed from an onsite to an online context without much alteration, and technical challenges were rife. As funders have increasingly adopted discussion formats more tailored to the online context, and technical issues causing frustration at the start have largely been eliminated, the negative aspects associated with online panel meetings have been reduced:

"Panellist: Earlier meetings were exhausting as you had to draft everything online and the connectivity was really bad. We really haven't had many technical problems this time round, and it is much better the way we do it now, where much more of the work is done offline. It is much less tiring".
Moderator: "Yes, they have upgraded the server capacity substantially now, so it is good that connectivity has generally been good."
(CINEA2022 panel meeting).

At the EU-level in particular, substantial modifications of the assessment process have been carried out. The key change has been increased focus on pre-panel preparations, with panellists' pre-discussion evaluative documentation used in a more strategic manner to ensure shorter, more focused discussions as this quote from a CINEA2021 moderator illustrates:

"In virtual meetings it is much more tiring to have long meetings […] It was therefore decided that rapporteurs should develop a high quality first draft of the feedback report and that the focus of meetings should only be on the points were there was disagreement"

4.4 Panel interaction

Panellists almost uniformly believe that panel discussion leads to higher quality assessments than what can be achieved by simply averaging individual assessments. Out of the 88 interviews conducted, only one believed arithmetic averaging constituted a viable option to the conduct of panel meetings. Panellists primarily point to the importance of integrating different perspectives on the applications under review, ensuring that the most appropriate expertise has primary say in the faith of a proposal, as well as the potential to identify and correct mistakes and biases in individual assessments. Some also points to that discussions afford panellists with an opportunity to calibrate their understanding of the scoring scale, the review criteria and the call text and receive guidance from the funding agency in this respect.

Observations confirm that overall, these are indeed the main factors which constitute the potential added value of panel interaction. However, the degree to which this potential is brought to fruition will largely depend on the panel characteristics discussed in the above: Requirements for pre- and post-discussion evaluative documentation and their use in discussions, discussion structuring, discussion time and moderation style.

++++ Quotes and additional detail on these added value of panel meetings, including self-checks

The characteristics of the online review format affects the added value of panel interaction in a number of ways, some with positive and some with negative effects. There are indications that online discussions are more equitable, making it harder for more dominant participants to monopolise discussions. Both self-confessed shy and dominant experts pointed to this effect. Participants estimated that this can on the one hand be down to the reduced effect of body language in an online setting, which might serve to accentuate the profiling of participants into more shy and more dominant members. On the other hand, it might be down to the more formal turn-taking

required by an online format, which makes it more difficult to monopolise discussions:

"I found it [online meetings] easier, because I'm not a particularly dominant sort of person in a meeting, my default is to be quiet, not speak, and you know, to step back rather than being in everyone's face. And in certain meetings I've found it easier to talk then, because you can put your hand up virtually. "(RCN FRIPRO interview)

"The other interesting thing is that there is an adaptation that was needed, in particular for a person like myself [who is dominant] […] During the central evaluation, when you could read the people's attitudes towards your argument, it was sometimes easier to sway the room because you would talk to the person who you'd see immediately would be your ally. (CINEA2021 interview)

Panellists consistently point to the online format's limited ability to nurture trust and understanding among participants due to limited possibility for informal interaction. While the discussion itself offers some opportunity to progressively become acquainted, the scope for doing so online is much more limited than onsite, where participants will connect and form ties through more informal settings such as common coffee breaks, lunch and dinner. The opportunity for these kinds of informal exchanges is not easily re-created online:

To trust among each other's - we don't have the time among us to know each other. When the discussion is not so smooth, it is more difficult to assess which expertise to trust. The interdisciplinary team spirit is more difficult to establish online. (EASME2021 interview)

Due to the online medium's limited ability to establish trust, it is clearly easier to carry out online discussions with people you already know. This was a general pattern across observations, whereby experienced panellists participating in panel constellations where they had cooperated with many of the panellists in previous evaluations reporting less problems with the online format than newcomers to the evaluation. The NordForsk evaluation provides an illustrative example of this effect, where panellists reported greater ease of communicating with panellists they knew from the first round of onsite evaluations compared to those that were recruited specifically just for the second round:

So the people on this panel, the people who were on the last panel who I got to know I felt much easier having a conversation with them. So a lot of my projects were with A and M. M I don't know. A I know fairly well from the first round /…/ so we had a way of working together and I would bow to his expertise and ideas and he might bow to mine in others. But with M it's like I don't really know.. there is no connection. There's nothing other than a black screen. It's hard to have that honest and genuine interaction then. (NordForsk interview)

The interviews point to an important negative consequence of the online medium's failure to foster trust among participants. The online context is ill-suited to handle evaluative complexity compared to an onsite context. While the online format has become progressively more well-functioning over the 2,5 years of panel observation, with overall improved information sharing and processing as a result, this positive development primarily applies to more customary and uncontested assessments. Out-of the ordinary assessments, such as the evaluation of highly novel, highly interdisciplinary or highly transdisciplinary proposals as some of the observed calls specifically solicited, consistently proved challenging to achieve in an online format.

Such calls require more extensive support from the funding agency, in terms of clearly communicating the intentions of the call and supporting panellist in the appropriate operationalisation of call requirements and evaluation criteria. Similarly, panellist need to calibrate

their respective understandings of the call and criteria and the appropriate manner of assessing unconventional proposals. This sort of dialogue and support is more easily achieved in an onsite context and is greatly aided by panellists' ability to establish a personal rapport to one another:

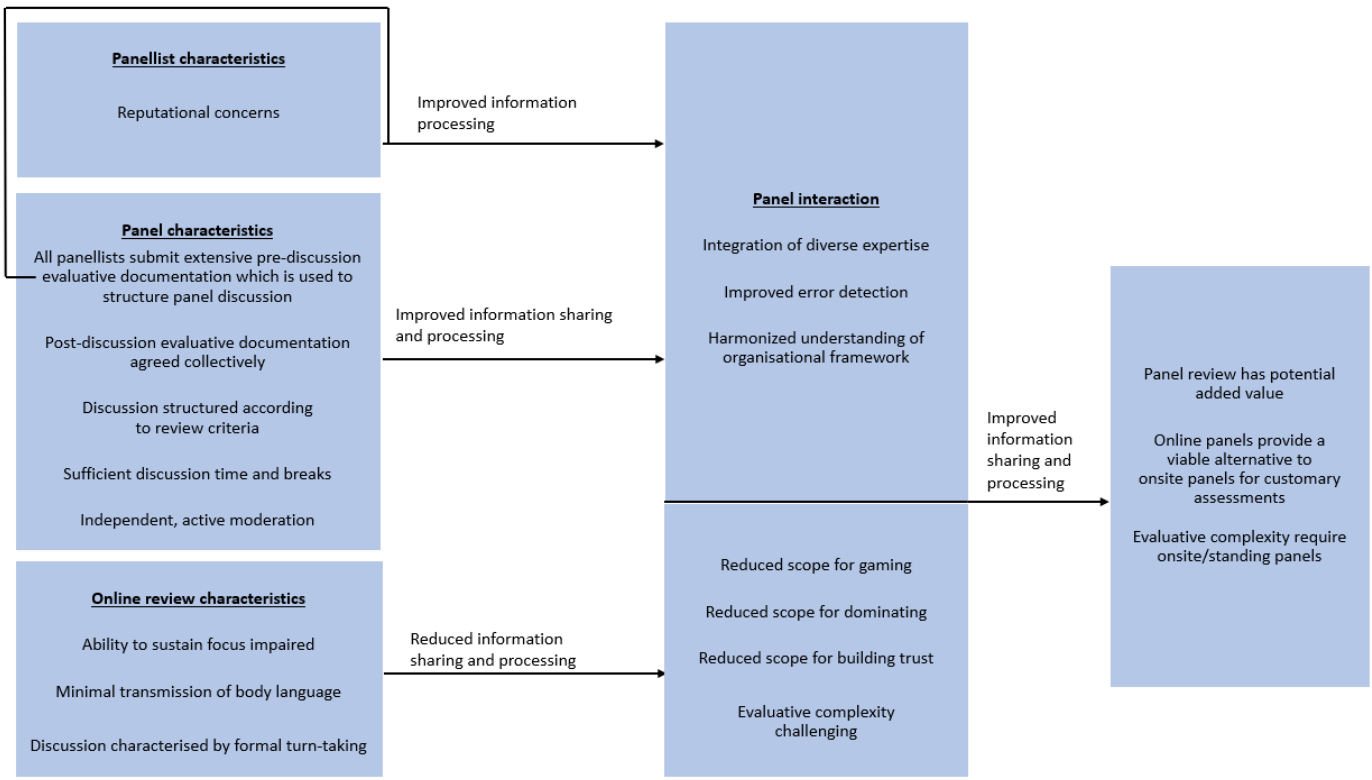
I think online discussions are good for things that are more or less routine, but especially for matters that require a lot of discussion, different points of view and considerations, I think they are less effective […] and in the end, it reflects... of course, because it's more difficult, then in the end it reflects also on the quality of the evaluations. (Interview RCN Teknogov)

While the lack of opportunity for informal networking might entail negative implications in the form of reduced ability to handle evaluative complexity, it might also have positive effects in the form of reduced potential for gaming the system and reduced pertinence of quid pro quo dynamics:

"It would help if we could be together in Brussel. I miss the informal discussion. It impacts the evaluation. In a skype call you always have to decide when to speak and not. It is harder to find out what is really important to some. Over a coffee you can really find out what is important to a person." (Joint Rail interview)

5. Discussion

Panel review has the potential to enlarge the pool of knowledge available in the assessment of grant applications and ensure more thorough and uniform processing of applications. Adopting a systems view of the review process, we find that individual panellists' characteristics and the organizational set-up of the funding competition as well as how these factors are mediated by the group dynamics of panel discussions, determine whether panel review can deliver on this potential added value. We furthermore find that while ad hoc online panels provide a viable alternative to onsite panels for customary assessments given certain key organisational characteristics are ensured, onsite or standing online panels are a preferable option for the assessment of unconventional research:



5.1 Panellist characteristics

In line with evidence from behavioural decision research, we find that individual panellists tend to process information in a superficial manner and use a variety of shortcuts to simplify decisions (Lerner & Tetlock, 1999; Scholten et al., 2007). Equally, we find that making panellists accountable for their assessments to colleagues counteract such tendencies (Lerner & Tetlock, 2003) as their reputational concerns lead them to carry out more thorough assessments. This points to a key finding resulting from our systems perspective that quantitative studies and qualitative studies exclusively focused on panel dynamics cannot capture: Panel discussion leads to higher quality individual assessments.

We extend our understanding of the accountability effect by showing that individuals' information processing is responsive to the strength of accountability measures. While previous contributions have compared accountable to non-accountable individuals (De Dreu, Nijstad, et al., 2008; Lerner & Tetlock, 2003; Scholten et al., 2007), we compare individuals that are subject to varying degrees of accountability. We find that exposing panellist to stronger peer scrutiny serves to further improve information processing. Panellists required to disclose their scores to their fellow panellists process information more thoroughly than those not required to disclose scores, and panellist required to submit written comments substantiating their scores process information more systematically than those merely required to submit scores.

5.2 Organisational characteristics

We thus find that requirements for evaluative documentation constitute a key organisational characteristic affecting the value-added of panel deliberation. In addition to affecting the thoroughness of individual panellists' information processing, we align with the social psychology literature in finding that accountability measures also affect collective information sharing and processing in panel discussions (De Dreu, Nijstad, et al., 2008).

On the one hand, using pre-discussion evaluative documentation actively as a basis for discussions, as is systematically done in the EU evaluations observed, attenuates the effect of the shared information basis as it ensures relevant individual contributions are brought into the discussion irrespective of these individuals' propensity for conformity. On the other hand, requirements for post-discussion evaluative documentation act as a focusing device for discussions, enabling more systematic and effective information processing by concentrating panels' attention on those elements of the assessment most in need of deliberation.

We extend our understanding of the group-level effect of accountability measures by showing that it is – in the same manner as individual information processing – responsive to the strength of accountability measures. Groups required to collectively agree scores carry out more thorough and systematic information processing than groups which are not required to agree scores, while groups required to collectively agree comments to substantiate their agreed scores in turn process information more systematically and share more information than groups which are merely required to agree scores.

Requirements for pre- and post-discussion evaluative documentation are particularly important in an online context. Generally we align with previous working in finding that an online assessment model is more taxing on panellists than onsite discussions, posing greater challenges in maintaining concentration over time (Stephen A Gallo et al., 2020; Pier et al., 2015), and involving more

distractions (Derrick, in press (Pier et al., 2015). We furthermore echo previous work in finding that spontaneous interaction is reduced, with formal turn-taking being the main norm (Derrick, in press).

We extend our understanding of online peer review by identifying pre- and post-discussion evaluative documentation as they key organisational measure that attenuates the negative effects on information sharing and processing associated with the online medium. Basing discussions on panellists' individual written pre-discussion evaluative documentation facilitates broad information sharing in a context where the barrier to sharing information orally are higher. Similarly, the focusing effect of having to agree post-discussion evaluative documentation is particularly important in an online context where distractions are rife and the impediments to dialogue are higher due to the need for more formal turn-taking and loss of body language.

Other panel characteristics that affect information sharing and processing in panel discussions are discussion style, moderation style and discussion time. We confirm previous work in social psychology which show structured discussion to facilitate discussion of unshared information (Stasser & Titus, 1985) and that active moderation serves a similar purpose (De Dreu, Nijstad, et al., 2008; Jhangiani et al., 2014; Larson et al., 1994). We align with research on peer review in finding that the moderator should not be involved in reviewing proposals (Coveney et al., 2017). The risk is otherwise that proposals favoured by the moderator are given preferential treatment in terms of more discussion time or elevated scores. We extend our understanding of the moderator function by showing that its importance is responsive to evaluation context. The need for active moderation is accentuated in an online environment, primarily due to the loss of body language as a discussion enabler.

We align with research in social psychology in finding that discussion time affects information sharing (Kelly & Loving, 2004; Kruglanski et al., 1993; Larson et al., 1994) and information processing (Kaplan et al., 1993; Karau & Kelly, 1992). We extend our understanding of this general mechanism by showing how it plays out in the concrete context of panel discussions. Panels under time pressure tend to forego substantial examination and exchange of information on the proposals under review, concentrating instead on their primary task of agreeing scores. They do so in a manner characterised by bargaining rather than information processing.

We furthermore confirm that fatigue has negative implications for information sharing (Kruglanski & Webster, 1996; Thornley et al., 2002) with proposals discussed at the end of a long work session subject to decreased information sharing and less thorough information processing. This suggests breaks should be taken frequently to avoid that applications assessed late in a work session are subject to differential treatment compared to applications assessed earlier. Furthermore, the total length of the panel meeting should be restricted to counteract fatigue towards the end of the meeting. This is particularly important for online meetings which are more taxing on participants.

5.4 Panel discussion

Contrary to much of the quantitative literature on the added value of panel meetings (Fogelholm et al., 2012; Martin et al., 2010; Obrecht et al., 2007; Pina et al., 2015; Thorngate et al., 2010), we find that panel deliberation has added value given certain key organisational characteristics are ensured. Through discussion, panellists potentially integrate their respective expertise, facilitating a more comprehensive and thorough review of applications by enabling the weighting of different pieces of information against each other. Deliberation also enables improved identification of errors in the assessments carried out through the twin mechanisms of peer assessment and improved self-

assessment. Finally, deliberation contributes to harmonizing reviewers understanding of the organisational framework, reducing differential treatment of applications due to reviewer variations in scoring and understanding of the review criteria and call text.

Our longitudinal, cross-organisational empirical material provides new insight into the key organisational characteristics that facilitate this added value: Requirements for extensive pre- and post-discussion evaluative documentation, structured discussion, generous discussion time and an independent, active moderator. These organisational characteristics will serve to improve information sharing and processing irrespective of whether deliberations are organised online or onsite, but they are especially important in an online environment due to the impediments to interaction that the online format entails.

While the online discussion model is associated with certain drawbacks, positive effects are also evident. Online discussions reduce the unwanted impact of certain interaction dynamics on outcomes, such as gaming and minority conformity. However, we find that online discussions are ill suited for evaluatively complex assessments due to their limited ability to build trust among participants. Therefore, calls soliciting unconventional research, such as highly novel and highly interdisciplinary proposals should preferably be carried out onsite. Alternatively, standing online panels which meet regularly should be established specifically for such purposes, as this will allow for the required trust-building among participants that is important for handling evaluative complexity. As previous work, we find that online interaction is eased if panellists know each other beforehand (Derrick, in press), indicating that standing panels are overall a preferable option to ad-hoc panels as online rather than onsite panel deliberation becomes the rule rather than the exception.

6. Conclusion

In conclusion we find that panel deliberation has clear added value given certain organisational characteristics are ensured. A key finding is that much of the previous work exploring the added value of panel deliberation has – due to its exclusive focus on how deliberation affects output in the form of scores – failed to recognise deliberation's important effect on a key input in the review process, the individual assessments. The prospect of having to defend their assessments to peers constitutes a potent accountability mechanism which leads panellists to review proposals more systematically and thoroughly. Eliminating panel deliberation and replacing it with a system of averaging individual reviews as previous work has suggested (Fogelholm et al., 2012; Obrecht et al., 2007), might thus constitute a small efficiency gain but risks a substantial quality loss in assessment processes.

Our results illustrate the value of going beyond the simple input-output understandings of how evaluation outcomes are produced that characterizes much of the existing literature, applying a systems perspective on grant panel peer review. Such system understanding is essential for undertaking effective interventions in peer review (Oxley, in press). The risk is otherwise that adjustments in one part of the system will produce unforeseen and unwanted effects in other parts, i.e. eliminating panel deliberation has negative repercussions for individual assessments.

We find that online panels provide a viable alternative to onsite panels for customary assessments. However, a number of drawbacks are evident. To minimize these and capitalize on the strong sides of online deliberation, an informed systems understanding of how online panels can best be organised is needed. We see signs that the necessary adjustments are already being carried out. Going forward, the challenge is to ensure that also unconventional research is assessed in an appropriate manner in the new era of online panel review.

Annexes

(To be added)

References

Coveney, J., Herbert, D. L., Hill, K., Mow, K. E., Graves, N., & Barnett, A. (2017). 'Are you siding with a personality or the grant proposal?': observations on how peer review panels function. *Research Integrity and Peer Review*, *2*(1), 1-14.

De Dreu, C. K., Baas, M., & Nijstad, B. A. (2008). Hedonic tone and activation level in the mood-creativity link: toward a dual pathway to creativity model. *Journal of personality and social psychology*, *94*(5), 739.

De Dreu, C. K., Nijstad, B. A., & Van Knippenberg, D. (2008). Motivated information processing in group judgment and decision making. *Personality and Social Psychology Review*, *12*(1), 22-49.

Fogelholm, M., Leppinen, S., Auvinen, A., Raitanen, J., Nuutinen, A., & Väänänen, K. (2012). Panel discussion does not improve reliability of peer review for medical research grant proposals. *J Clin Epidemiol*, *65*(1), 47-52. https://doi.org/10.1016/j.jclinepi.2011.05.001

Gallo, S. A., Schmaling, K. B., Thompson, L. A., & Glisson, S. R. (2020). Grant reviewer perceptions of the quality, effectiveness, and influence of panel discussion. *Research Integrity and Peer Review*, *5*(1), 1-9.

Huutoniemi, K. (2012). Communicating and compromising on disciplinary expertise in the peer review of research proposals. *Social Studies of Science*, *42*(6), 897-921. https://doi.org/10.1177/0306312712458478

Jhangiani, R., Tarry, H., & Stangor, C. (2014). Principles of social psychology-1st international edition.

Kameda, T. (1991). Procedural influence in small-group decision making: Deliberation style and assigned decision rule. *Journal of personality and social psychology*, *61*(2), 245.

Kaplan, M. F., Wanshula, L. T., & Zanna, M. P. (1993). Time pressure and information integration in social judgment: The effect of need for structure. *Time pressure and stress in human judgment and decision making*, 255-267.

Karau, S. J., & Kelly, J. R. (1992). The effects of time scarcity and time abundance on group performance quality and interaction process. *Journal of Experimental Social Psychology*, *28*(6), 542-571.

Kelly, J. R., & Loving, T. J. (2004). Time pressure and group performance: Exploring underlying processes in the attentional focus model. *Journal of Experimental Social Psychology*, *40*(2), 185-198.

Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind:" Seizing" and" freezing.". *Psychological review*, *103*(2), 263.

Kruglanski, A. W., Webster, D. M., & Klem, A. (1993). Motivated resistance and openness to persuasion in the presence or absence of prior information. *Journal of personality and social psychology*, *65*(5), 861.

Larson, J. R., Foster-Fishman, P. G., & Keys, C. B. (1994). Discussion of shared and unshared information in decision-making groups. *Journal of personality and social psychology*, *67*(3), 446.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological bulletin*, *125*(2), 255.

Lerner, J. S., & Tetlock, P. E. (2003). 13 Bridging Individual, Interpersonal, and Institutional Approaches to Judgment and Decision Making: The Impact of Accountability on Cognitive Bias. *Emerging perspectives on judgment and decision research*, 431.
Martin, M. R., Kopstein, A., & Janice, J. M. (2010). An analysis of preliminary and post-discussion priority scores for grant applications peer reviewed by the Center for Scientific Review at the NIH. *PLoS One*, *5*(11), e13526. https://doi.org/10.1371/journal.pone.0013526
Obrecht, M., Tibelius, K., & D'Aloisio, G. (2007). Examining the value added by committee discussion in the review of applications for research awards. *Research Evaluation*, *16*(2), 79-91.
Pier, E. L., Raclaw, J., Nathan, M. J., Kaatz, A., Carnes, M., & Ford, C. E. (2015). Studying the Study Section: How Group Decision Making in Person and via Videoconferencing Affects the Grant Peer Review Process. WCER Working Paper No. 2015-6. *Wisconsin Center for Education Research*.
Pina, D. G., Hren, D., & Marušić, A. (2015). Peer review evaluation process of Marie Curie actions under EU's seventh framework programme for research. *PLoS One*, *10*(6), e0130753.
Scholten, L., Van Knippenberg, D., Nijstad, B. A., & De Dreu, C. K. (2007). Motivated information processing and group decision-making: Effects of process accountability on information processing and decision quality. *Journal of Experimental Social Psychology*, *43*(4), 539-552.
Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of personality and social psychology*, *48*(6), 1467.
Thorngate, W., Dawes, R. M., & Foddy, M. (2010). *Judging merit*. Psychology Press.
Thornley, R., Spence, M. W., Taylor, M., & Magnan, J. (2002). New decision tool to evaluate award selection process. *Journal of Research Administration*, *33*(2/3), 49-58.

Angrosino, M. V. (2016). *Naturalistic observation*. Routledge.
Carpenter, A. S., Sullivan, J. H., Deshmukh, A., Glisson, S. R., & Gallo, S. A. (2015). A retrospective analysis of the effect of discussion in teleconference and face-to-face scientific peer-review panels. *BMJ Open*, *5*(9), e009138.
Chawla, D. S. (2021). Zoom fatigue saps grant reviewers' attention. *Nature*, *590*(7844), 172-172.
Coveney, J., Herbert, D. L., Hill, K., Mow, K. E., Graves, N., & Barnett, A. (2017). 'Are you siding with a personality or the grant proposal?': observations on how peer review panels function. *Research Integrity and Peer Review*, *2*(1), 1-14.
Davis, M. S. L. E., Conner, T. R., Miller-Bains, K., & Shapard, L. (2020). What makes an effective grants peer reviewer? An exploratory study of the necessary skills. *PLoS One*, *15*(5), e0232327. https://doi.org/10.1371/journal.pone.0232327
De Dreu, C. K., Baas, M., & Nijstad, B. A. (2008). Hedonic tone and activation level in the mood-creativity link: toward a dual pathway to creativity model. *Journal of personality and social psychology*, *94*(5), 739.
De Dreu, C. K., Nijstad, B. A., Bechtoldt, M. N., & Baas, M. (2011). Group creativity and innovation: A motivated information processing perspective. *Psychology of Aesthetics, Creativity, and the Arts*, *5*(1), 81.

De Dreu, C. K., Nijstad, B. A., & Van Knippenberg, D. (2008). Motivated information processing in group judgment and decision making. *Personality and Social Psychology Review*, *12*(1), 22-49.
Fogelholm, M., Leppinen, S., Auvinen, A., Raitanen, J., Nuutinen, A., & Väänänen, K. (2012). Panel discussion does not improve reliability of peer review for medical research grant proposals. *J Clin Epidemiol*, *65*(1), 47-52. https://doi.org/10.1016/j.jclinepi.2011.05.001
Forsyth, D. R. (2014). *Group dynamics* (6th ed., International ed. ed.). Wadsworth Cengage learning.
Gallo, S. A., Carpenter, A. S., & Glisson, S. R. (2013). Teleconference versus face-to-face scientific peer review of grant application: effects on review outcomes. *PLoS One*, *8*(8), e71693.
Gallo, S. A., Schmaling, K. B., Thompson, L. A., & Glisson, S. R. (2020). Grant reviewer perceptions of the quality, effectiveness, and influence of panel discussion. *Research Integrity and Peer Review*, *5*(1), 1-9.
Gallo, S. A., Schmaling, K. B., Thompson, L. A., & Glisson, S. R. (2020). Grant reviewer perceptions of the quality, effectiveness, and influence of panel discussion. *Research Integrity and Peer Review*, *5*(1), 7. https://doi.org/10.1186/s41073-020-00093-0
Hug, S. E. (2022). Towards theorizing peer review. *Quantitative Science Studies*, *3*(3), 815-831.
Huutoniemi, K. (2012). Communicating and compromising on disciplinary expertise in the peer review of research proposals. *Social Studies of Science*, *42*(6), 897-921. https://doi.org/10.1177/0306312712458478
Jhangiani, R., Tarry, H., & Stangor, C. (2014). Principles of social psychology-1st international edition.
[Record #422 is using a reference type undefined in this output style.]
Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little, Brown.
Kaplan, M. F., Wanshula, L. T., & Zanna, M. P. (1993). Time pressure and information integration in social judgment: The effect of need for structure. *Time pressure and stress in human judgment and decision making*, 255-267.
Karau, S. J., & Kelly, J. R. (1992). The effects of time scarcity and time abundance on group performance quality and interaction process. *Journal of Experimental Social Psychology*, *28*(6), 542-571.
Kelly, J. R., & Karau, S. J. (1999). Group decision making: The effects of initial preferences and time pressure. *Personality and Social Psychology Bulletin*, *25*(11), 1342-1354.
Kelly, J. R., & Loving, T. J. (2004). Time pressure and group performance: Exploring underlying processes in the attentional focus model. *Journal of Experimental Social Psychology*, *40*(2), 185-198.
Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind:" Seizing" and" freezing.". *Psychological review*, *103*(2), 263.
Kruglanski, A. W., Webster, D. M., & Klem, A. (1993). Motivated resistance and openness to persuasion in the presence or absence of prior information. *Journal of personality and social psychology*, *65*(5), 861.
Larson, J. R., Foster-Fishman, P. G., & Keys, C. B. (1994). Discussion of shared and unshared information in decision-making groups. *Journal of personality and social psychology*, *67*(3), 446.
Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological bulletin*, *125*(2), 255.

Lerner, J. S., & Tetlock, P. E. (2003). 13 Bridging Individual, Interpersonal, and Institutional Approaches to Judgment and Decision Making: The Impact of Accountability on Cognitive Bias. *Emerging perspectives on judgment and decision research*, 431.

Martin, M. R., Kopstein, A., & Janice, J. M. (2010). An analysis of preliminary and post-discussion priority scores for grant applications peer reviewed by the Center for Scientific Review at the NIH. *PLoS One*, *5*(11), e13526. https://doi.org/10.1371/journal.pone.0013526

Meadmore, K., Fackrell, K., Recio-Saucedo, A., Bull, A., Fraser, S. D. S., & Blatch-Jones, A. (2020). Decision-making approaches used by UK and international health funding organisations for allocating research funds: A survey of current practice. *PLoS One*, *15*(11), e0239757. https://doi.org/10.1371/journal.pone.0239757

Mow, K. (2010). *Inside the black box: research grant funding and peer review in Australian research councils*. Lambert Academic Publishing.

Obrecht, M., Tibelius, K., & D'Aloisio, G. (2007). Examining the value added by committee discussion in the review of applications for research awards. *Research Evaluation*, *16*(2), 79-91.

Olbrecht, M., & Bornmann, L. (2010). Panel peer review of grant applications: what do we know from research in social psychology on judgment and decision-making in groups? *Research Evaluation*, *19*(4), 293-304. https://doi.org/10.3152/095820210X12809191250762

Pier, E. L., Raclaw, J., Nathan, M. J., Kaatz, A., Carnes, M., & Ford, C. E. (2015). Studying the Study Section: How Group Decision Making in Person and via Videoconferencing Affects the Grant Peer Review Process. WCER Working Paper No. 2015-6. *Wisconsin Center for Education Research*.

Pina, D. G., Hren, D., & Marušić, A. (2015). Peer review evaluation process of Marie Curie actions under EU's seventh framework programme for research. *PLoS One*, *10*(6), e0130753.

Recio-Saucedo, A., Crane, K., Meadmore, K., Fackrell, K., Church, H., Fraser, S., & Blatch-Jones, A. (2022). What works for peer review and decision-making in research funding: a realist synthesis. *Research Integrity and Peer Review*, *7*(1), 1-28.

Reinhart, M., & Schendzielorz, C. (2021). Peer Review Procedures as Practice, Decision, and Governance–Preliminaries to Theories of Peer Review.

Roumbanis, L. (2017). Academic judgments under uncertainty: A study of collective anchoring effects in Swedish Research Council panel groups. *Social Studies of Science*, *47*(1), 95-116. https://doi.org/10.1177/0306312716659789

Scholten, L., Van Knippenberg, D., Nijstad, B. A., & De Dreu, C. K. (2007). Motivated information processing and group decision-making: Effects of process accountability on information processing and decision quality. *Journal of Experimental Social Psychology*, *43*(4), 539-552.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of personality and social psychology*, *48*(6), 1467.

Thorngate, W., Dawes, R. M., & Foddy, M. (2010). *Judging merit*. Psychology Press.

Thornley, R., Spence, M. W., Taylor, M., & Magnan, J. (2002). New decision tool to evaluate award selection process. *Journal of Research Administration*, *33*(2/3), 49-58.

van Arensbergen, P., van der Weijden, I., & van den Besselaar, P. (2014). The selection of talent as a group process. A literature review on the social dynamics of decision making in grant panels. *Research Evaluation*, *23*(4), 298-311. https://doi.org/10.1093/reseval/rvu017

Vo, N. M., Quiggle, G. M., & Wadhwani, K. (2016). Comparative outcomes of face-to-face and virtual review meetings. *International Journal of Surgery Open*, *4*, 38-41.
Vo, N. M., & Trocki, R. (2015). Virtual and Peer Reviews of Grant Applications at the Agency for Healthcare Research and Quality. *South Med J*, *108*(10), 622-626.
Wilkinson, L. A. (2011). Systems Theory. In S. Goldstein & J. A. Naglieri (Eds.), *Encyclopedia of Child Behavior and Development* (pp. 1466-1468). Springer US. https://doi.org/10.1007/978-0-387-79061-9_941