



ISSN : 1875-4120  
Issue : Vol. 17, Issue 2  
Published : February 2020

#### Terms & Conditions

Registered TDM users are authorised to download and print one copy of the articles in the TDM Website for personal, non-commercial use provided all printouts clearly include the name of the author and of TDM. The work so downloaded must not be modified. **Copies downloaded must not be further circulated.** Each individual wishing to download a copy must first register with the website.

All other use including copying, distribution, retransmission or modification of the information or materials contained herein without the express written consent of TDM is strictly prohibited. Should the user contravene these conditions TDM reserve the right to send a bill for the unauthorised use to the person or persons engaging in such unauthorised use. The bill will charge to the unauthorised user a sum which takes into account the copyright fee and administrative costs of identifying and pursuing the unauthorised user.

For more information about the Terms & Conditions visit [www.transnational-dispute-management.com](http://www.transnational-dispute-management.com)

© Copyright TDM 2020  
TDM Cover v8.0

# Transnational Dispute Management

[www.transnational-dispute-management.com](http://www.transnational-dispute-management.com)

## Summary of Young-OGEMID Symposium No. 7: "Empirical Legal Research - How-To and Why-To (March 2018)" by V. Hristova

### About TDM

**TDM** (Transnational Dispute Management): Focusing on recent developments in the area of Investment arbitration and Dispute Management, regulation, treaties, judicial and arbitral cases, voluntary guidelines, tax and contracting.

Visit [www.transnational-dispute-management.com](http://www.transnational-dispute-management.com) for full Terms & Conditions and subscription rates.

### Open to all to read and to contribute

TDM has become the hub of a global professional and academic network. Therefore we invite all those with an interest in Investment arbitration and Dispute Management to contribute. We are looking mainly for short comments on recent developments of broad interest. We would like where possible for such comments to be backed-up by provision of in-depth notes and articles (which we will be published in our 'knowledge bank') and primary legal and regulatory materials.

If you would like to participate in this global network please contact us at [info@transnational-dispute-management.com](mailto:info@transnational-dispute-management.com): we are ready to publish relevant and quality contributions with name, photo, and brief biographical description - but we will also accept anonymous ones where there is a good reason. We do not expect contributors to produce long academic articles (though we publish a select number of academic studies either as an advance version or an TDM-focused republication), but rather concise comments from the author's professional 'workshop'.

**TDM** is linked to **OGEMID**, the principal internet information & discussion forum in the area of oil, gas, energy, mining, infrastructure and investment disputes founded by Professor Thomas Wälde.

## Summary of Young-OGEMID Symposium No. 7: “Empirical Legal Research - How-To and Why-To (March 2018)”

By *Velislava Hristova*\*

### **General Summary**

*Young-OGEMID’s (YO) seventh virtual symposium advised junior practitioners, academics and law students on how to and why to do empirical research, identifying the problems and possibilities associated with such analysis. As YO moderator **Prof. S.I. Strong** stated when introducing the event, “Ever wonder why lawyers do empirical research and whether it actually provides useful data?” During the discussion, academics and practitioners from around the world discussed the pros and cons of a variety of methodologies and provided tips for aspiring empiricists and critical readers.*

*The seventh YO virtual symposium featured the following speakers:*

- 1) **Prof. Stavros Brekoulakis** – Professor of International Arbitration and Commercial Law, School of Law, Queen Mary University of London, United Kingdom;<sup>1</sup>
- 2) **Prof. Jennifer K. Robbennolt** – Associate Dean for Research, Alice Curtis Campbell Professor of Law and Professor of Psychology, and Co-Director of the Illinois Program on Law, Behavior and Social Science, Illinois College of Law, University of Illinois, USA;<sup>2</sup>
- 3) **Ms Maryam Salehijam** – PhD Researcher, Transnational Law Centre of the University of Ghent, Belgium;<sup>3</sup>

---

\* **Velislava Hristova** is a Bulgarian qualified lawyer specialising in International Arbitration.

<sup>1</sup> **Prof. Stavros Brekoulakis** is a Professor in International Arbitration at Queen Mary University of London and an associate member of 3 Verulam Buildings. His work includes the leading publications on Third Parties in International Commercial Arbitration, Arbitrability, the ICCA-Queen Mary Report on Third Party Funding and numerous publications in leading legal journals and reviews. He is a member of the LCIA Court, ICC Task Force on Emergency Arbitrator, a Member of the ICC Commission on Arbitration, an assistant Rapporteur in the International Law Association Committee on International Commercial Arbitration, the General Editor of the Journal of International Dispute Settlement, the Editor-in-Chief of the (CIArb’s) International Journal of Arbitration, Mediation and Dispute Management and co-editor of the Kluwer’s International Arbitration Law Library series.

<sup>2</sup> **Prof. Jennifer Robbennolt**, J.D., Ph.D, is the Associate Dean for Research, the Alice Curtis Campbell Professor of Law and Professor of Psychology, and Co-Director of the Illinois Program on Law, Behavior and Social Science at the University of Illinois, College of Law. Professor Robbennolt’s research integrates psychological theory and empirical methods into the study of law and legal institutions, particularly in the areas of torts and dispute resolution. Her book, *Empirical Methods in Law* (written with colleagues Bob Lawless and Tom Ulen), is designed to give lawyers and law students an accessible introduction to empirical research methods and statistics and their use in legal contexts. She is co-author of several additional books, including the influential casebook *Dispute Resolution and Lawyers*, *The Psychology of Tort Law* (with Valerie Hans), and *Psychology for Lawyers: Understanding the Human Factors in Negotiation, Litigation, and Decision Making* (with Jean Sternlight).

<sup>3</sup> **Ms Maryam Salehijam** is a Canadian/Iranian BOD funded PhD researcher at the Transnational Law Center of the University of Ghent. She focuses on the legal issues pertaining to the use of alternative dispute resolution (ADR) in commercial disputes. Her specific research questions address the validity and enforceability of ADR agreements, the obligations of the parties to an ADR agreement, and whether such legal questions can be regulated at the international level. Maryam was awarded her LL.M. in International Laws and LL.B. in European Law by Maastricht University (exchanges in Glasgow and McGill University). She is also a member of YMI, CEPANI40, DIS40, YIAC, LCIA, and OGEMID. Furthermore, due to her international background,

- 4) **Prof. Wolfgang Alschner** – Professor of Law, Common Law Section of the University of Ottawa, Canada;<sup>4</sup> and
- 5) **Dr. Suha Jubran Ballan** – Research Associate, Law and Social Science Faculty, SOAS, University of London, United Kingdom.<sup>5</sup>

### **Speaker 1: Overview of the Emergence of Empirical Studies in Arbitration and General Observations on the Use of Social Science in Arbitration**

Speaking first, **Prof. Brekoulakis** began the symposium with a brief overview of the emergence of empirical studies in arbitration stating that judicial decision-making, encompassing both state court and arbitral decision-making in this term, has traditionally been a subject of empirical and interdisciplinary research. He pointed out that at first empirical studies focused only on state court-decision making (giving as examples the studies *The Supreme Court and the Attitudinal Model* (CUP, 1993) and *The Supreme Court and The Attitudinal Model Revisited* (CUP, 2002) by Jeffrey A. Segal and Harold J. Spaeth) but in the last fifteen years there was an increase in the scholars who have taken an interest in private decision-making and in particular in international arbitration.

The first studies on arbitral decision-making examined the way international arbitrators take decisions and whether they tend to favour certain groups of parties. At first, the scholars focused on commercial arbitration concerning consumer or public policy disputes, but then switched to investment arbitration and disputes concerning public international law and sovereign states. **Prof. Brekoulakis** noted that like the studies on national judiciaries, most of the studies on arbitral decision-making intended to assess the bias scientifically.

**Prof. Brekoulakis** then made some general observations on the appropriateness and limitations of the use of social science in law and particularly in arbitration.

First, **Prof. Brekoulakis** pointed out that it is difficult to design a method for direct observation and measurement of bias. The experimental methods, employed in a small number of studies as a technique to examine arbitral decision-making, provide findings that should be usually treated with caution and a certain degree of uncertainty. He confirmed the view of other scholars that it is extremely difficult to simulate in an artificial setting the complex incentives and motivations that drive the arbitrators' decisions or the parties' decisions when selecting their arbitrator.

---

she is a native speaker of English and Persian, intermediate in Spanish and French, and beginner in German and Dutch.

<sup>4</sup> **Prof. Wolfgang Alschner** is an empirical legal scholar specialized in international economic law and the computational analysis of law. He is a permanent faculty member of the Common Law Section of the University of Ottawa with cross-appointment to the Faculty of Engineering, School of Electrical Engineering and Computer Science. Prior to joining the University of Ottawa, Wolfgang worked for several years as an individual contractor for UNCTAD's Section on International Investment Agreements and as a research fellow at the Graduate Institute in Geneva and the World Trade Institute in Bern, Switzerland. He is also cofounder of the investment treaty analytics portal [www.mappinginvestmen\\_rea\\_es.com](http://www.mappinginvestmen_rea_es.com).

<sup>5</sup> **Suha Jubran-Ballan** is an associate researcher at the Faculty of Law and Social Sciences, University of London. In 2018 she has been a visiting researcher at the Dickson Poon School of Law, King's College London, and a postdoctoral fellow at the Political Science Department, Haifa University. Her research on investment treaty arbitration involves democratization and foreign investors, economic crises, institutionalism, and empirical research. She obtained her Ph.D. from the Buchmann Faculty of Law, Tel Aviv University under the supervision of Prof. Eyal Benvenisti. Her dissertation examined the judicial reasoning of investment treaty arbitration and identifies different patterns of judicial reasoning according to the institutional background of the arbitration panel.

Secondly, **Prof. Brekoulakis** explained that, because of the challenges, the direct assessment of bias poses, many scholars tried to assess bias in arbitral decision-making indirectly, relying on proxy-variables. Proxy-variables have been very popular amongst US scholars who have been able to claim general observations on judicial behaviour by establishing a strong correlation between certain characteristics or traits of some US judges and the way these judges voted on certain occasions. For example, several well-known studies have found that US Supreme Court judges that were appointed by a democratic government tended to vote more liberally, whereas US Supreme Court judges that were appointed by a republican government tended to vote more conservatively.

Similar proxy-variables have been used by scholars to explain arbitral decision-making. For example, the number of times an arbitrator has been appointed by a state or an investor has been used as a proxy-variable by scholars to make inferences on policy preferences and judicial attitude of arbitrators deciding investment treaty disputes. Another popular proxy-variable used in empirical studies on investor-state dispute settlement (“ISDS”) is demographics.

In conclusion, **Prof. Brekoulakis** invited the listserv members to share their views as to what kind of legal phenomena can be subject to social science methods, and whether scientifically measuring bias in decision-making can provide objective data. He also asked what the views of the listserv members on using proxies to measure bias in arbitral decision-making are.

#### *Q&A Session*

In the Q&A session that followed, participant **Mr Christian Campbell**, Director of the Center for International Legal Studies, pointed out that since 2015 the participants who serve as arbitrators and the students who participate as counsel in the Foreign Direct Investment International Arbitration Moot (“FDI Moot”) are surveyed about the issues raised by the case which are drawn from current controversial ISDS cases. **Mr Campbell** further noted that Prof. Anne van Aaken, and he discussed the possibility to use the FDI Moot case to empirically establish the connection between the rule of law/corruption indices and bias; however he has not managed to figure out yet how to do that without disrupting the competition. **Mr Campbell** then posed the question whether a respondent state being high or low on corruption index can be a proxy or one should look only at the characteristics of the adjudicators.

Moderator **Prof. Strong** shared two pieces from the news reflecting the role of empirical research in the law. First, she referred to the empirical study *Societal benefits and costs of International Investment Agreements* published by the Organisation for Economic Cooperation and Development (“OECD”)<sup>6</sup>. The study suggests that little evidence has been generated to support many claims about the positive or negative impact of international investment agreements and identifies methodological challenges and suggests areas where further study is needed to draw firmer conclusions. Then she referred to the *Questionnaire on announced U.S. tariff increase on imports of certain steel and aluminium products and possible EU commercial policy measures in response*<sup>7</sup> circulated by the European Commission. The information received through the questionnaire aims to assist the European Commission in assessing the necessity and the parameters of possible commercial policy measures.

---

<sup>6</sup> Available at: [https://www.oecd-ilibrary.org/finance-and-investment/societal-benefits-and-costs-of-international-investment-agreements\\_e5f85c3d-en](https://www.oecd-ilibrary.org/finance-and-investment/societal-benefits-and-costs-of-international-investment-agreements_e5f85c3d-en).

<sup>7</sup> Available at: [http://trade.ec.europa.eu/consultations/index.cfm?consul\\_id=253](http://trade.ec.europa.eu/consultations/index.cfm?consul_id=253).

**Prof. Strong** noted that these pieces give a context to the discussion about the proper and improper research goals and methodologies. For example, the US Senator Rand Paul introduced legislation<sup>8</sup> that would require every federal peer review panel to include two people who are not experts in the subject matter under consideration – one who is an expert in a field different from the field of research of the submitted proposal and is not professionally affiliated with any academic or research institution at present and for at least ten years and a second one focusing on the value the proposed research delivers to the taxpayers. The legislation was criticised for misunderstanding the nature and purpose of the peer review.

**Dr Ballan**, the fifth speaker of this symposium, joined the discussion and asked **Prof. Brekoulakis** how he defines bias and how he distinguishes bias, which is undesirable from diversity. She noted that being sympathetic towards a certain approach or a group of people is not always considered undesirable bias and in some cases can be desirable or at least tolerable. For example, the tendency female judges to give longer sentences to sexual offenders or an arbitrator who is known to have a favourable approach to states.

Following on from **Mr Campbell**'s question, **Prof. Strong** suggested discussing the reasons bias cannot be evaluated directly.

In response, **Prof. Brekoulakis** stated that **Mr Campbell**'s idea to establish a connection between the rule of law/corruption indices and bias is excellent, and he would be interested to see the results. **Prof. Brekoulakis** then explained that as **Prof. Strong** pointed out, bias is difficult to evaluate directly because it is a state of mind, and this is the reason that proxies are typically used to measure it.

Turning to the question about the use of proxies, **Prof. Brekoulakis** stated that the question is not whether proxies should be used to measure bias but whether the *appropriate* proxies are used in international arbitration and what are the limitations of these proxies.

Related, and also in response to the question of diversity and desirable bias raised by **Dr Ballan**, **Prof. Brekoulakis** stated that many arbitration studies have sought to examine bias in ISDS by employing demographics of arbitrators including gender, nationality and education as a proxy of bias. The underlying assumption for these studies is that imbalances about the demographic make-up of ISDS tribunals may result in biased arbitral decision-making.

However, **Prof. Brekoulakis** underlined that the information about demographics is essential, but the relationship between bias in arbitral decision-making and demographics is not straightforward. For example, a recent study on the differences between male and female judges has emphasised that a commitment to feminist values is more significant in the promotion of gender equality in judicial decision-making than the gender of the judge. The same applies to the ISDS. Women and non-western arbitrators who are selected are often educated in the same universities and work in large international law firms such as the male arbitrators from the West. Therefore, it is likely that they share the same views, perspectives and personal values and have similar decision-making attitudes. Thus, the focus on overt characteristics, such as demographics has only limited value in identifying bias in decision-making.

---

<sup>8</sup> Available at: <https://www.insidehighered.com/quicktakes/2017/10/23/rand-pauls-new-target-peer-review>.

**Prof. Brekoulakis** then asked the listserv members to share their ideas and examples from personal research concerning whether there are more appropriate proxies to measure bias in arbitral decision-making.

Participant **Mr Mark Kantor**, an independent arbitrator, joined the discussion and echoed **Dr Ballan**'s inquiry concerning the meaning of bias. **Mr Kantor** pointed out the following potentially overlapping meanings of the word "bias" in the context of ISDS:

- 1) "bias" in favour of an individual party;
- 2) "bias" in favour of a class of parties (e.g., investors v. states, multinationals v. entrepreneurs, or rich states v. poor states) and/or
- 3) "bias" in favour of or against a particular legal position (e.g., the scope of fair and equitable treatment, the process of evolution of Community Infrastructure Levy or written evidence v. witness testimony or broad v. narrow document production).

**Mr Kantor** noted that the value of the proxies depends on the meaning of the term which is in use, however, the empirical studies which he has seen do not seek to draw a difference between the various meanings of the term. Proxies usually do not address the impact of different views of the facts in the dispute, the impact of one or another approach towards a legal doctrine or the difference among treaty texts.

The usual means of applying proxies is to focus on the outcome, determined as a win/loss or damages/no damages. But often the important decisions in an arbitration related to disputed facts. **Mr Kantor** pointed out that the proxies also should deal with differences in the legal instruments (e.g., the scope of the term "fair and equitable treatment in CAFTA<sup>9</sup> v. 1990's U.S. BITs). Often, a difference in approach towards legal doctrines will not have a material impact on the outcome of a particular dispute, precisely because the facts or legal text are decisive regardless. And concerning legal doctrines, if the "bias" measurement is intended to identify "bias" in favour of "my guy", then such a biased adjudicator will surely have a fairly flexible view of legal doctrine. **Mr Kantor** then raised the question how an outcome-oriented proxy would capture the bias of an adjudicator in favour of a particular legal position such as general support to broad document production.

**Prof. Brekoulakis** agreed with the distinction between the three types of bias made by **Mr Kantor**. He further clarified that the first kind of bias ("my guy") is addressed by national laws, institutional rules and court decisions. This kind of bias is unacceptable and legally forbidden, and awards issued by arbitrators inflicted with that kind of bias are significantly sanctioned.

**Prof. Brekoulakis** then referred to the different standards of bias adopted by various courts. The courts in England have adopted slightly different standards of bias, such as "reasonable appearance of bias"<sup>10</sup>, "reasonable suspicion" of bias<sup>11</sup>, "real danger" or "real possibility" of bias<sup>12</sup>. Similarly, in the US different Circuits tend to apply different standards of bias, because the Supreme Court in *Commonwealth Coatings Corp. v Continental Casualty Co* failed to develop a single standard of impartiality. The Second Circuit, for example, requires

---

<sup>9</sup> Central America Free Trade Agreement.

<sup>10</sup> *ASM Shipping Ltd of India v TTMI Ltd of England*, [2005] EWHC 2238.

<sup>11</sup> *R v Mulvihill* [1990] 1 All E.R. 436, 441.

<sup>12</sup> *AT&T Corp. v Saudi Cable Co.* [2000] 2 Lloyd's Rep 127.

“evident partiality” by arbitrators<sup>13</sup> whereas the Ninth Circuit applies a lower threshold amounting to “an impression of possible bias”<sup>14</sup>. Finally, courts in civil law jurisdictions focus their analysis on “justifiable doubts” of bias.

**Prof. Brekoulakis** further stated that the “bias” in favour of or against a particular legal position does not appear to be encompassed by the existing laws. With the exception of Rule 2.1.1 of the IBA Guidelines on Conflict of Interest (2014) which includes in the waivable red list the example where “the arbitrator has given legal advice or provide an expert opinion on the dispute to a party or an affiliate of one of the parties” (which is a different type of bias for a certain legal position), previously expressed legal opinions are typically considered acceptable bias. The IBA, for example, includes in the Green List the case where “The arbitrator has previously expressed a legal opinion (such as in a law review article or public lecture) concerning an issue that also arises in the arbitration (but this opinion is not focused on the case)”.

The types of bias mentioned above, are not usually the subject matter of empirical studies in arbitration, which tend to focus on the second type of bias - bias in favour of a class of parties, e.g. investors v. States, Rich States v. Poor States, etc.

National laws and institutional rules do not typically encompass this type of bias. While it is accepted as inherent and practically unavoidable, **Prof. Brekoulakis** confirmed the view that this type of bias is important in terms of the legitimacy of the system. However, due to one’s natural inclination favouring or opposing States, it can have harmful effects on arbitration. For example, if the vast majority of the people appointed as ISDS arbitrators tend to favour individual entrepreneurship (as opposed to the default right of States to regulate individual entrepreneurship), arbitration as a whole would be a biased dispute resolution system.

These are the reasons to try to measure this type of bias, and most studies have been looking into how the demographic profile of arbitrators suggests that they are pre-disposed towards investors or States. As **Prof. Brekoulakis** noted in his previous post, he finds demographics to be a crude and unreliable proxy.

**Prof. Brekoulakis** then stated that he agrees with **Mr Kantor**’s concerns about the outcome-oriented proxies. In addition to the fact that they tend to overlook the importance of the factual and legal background of a dispute, **Prof. Brekoulakis** expressed the concern that an outcome-focused study on ISDS bias will typically say that investors won in X% of the cases, and they managed to recover Y% of the requested claims. If the first figure is below 50% and the second figure sufficiently low (e.g. close to 30%) the inference is that ISDS is not favouring the investors as is generally perceived to be the case. However, as **Prof. Brekoulakis** has argued elsewhere<sup>15</sup> statistics on award outcomes is a crude and unreliable indicator of impartiality. Given the tendency of investors to bring claims in ISDS when their investment proves unsuccessful, it is possible that many of the ISDS claims are weak in the first place, and therefore even a 39% success rate is significantly higher than what would be expected if decided by impartial adjudicators. Equally, given that claims for damages tend to be inflated, the average amount of damages awarded is likely to be too high.

---

<sup>13</sup> *Scandinavian Reinsurance v Saint Paul Fire & Marine Ins.*, 668 F.3d 60 (2d Cir. 2012).

<sup>14</sup> *New Regency Productions v Nippon Herald Films*, 501 F.3d 1101, 1108 (9th Cir. 2007).

<sup>15</sup> Brekoulakis, Stavros “Systemic Bias and the Institution of International Arbitration: A New Approach to Arbitral Decision-Making” 4(3) *Journal of International Dispute Settlement* 553 (2013).

**Prof. Strong** supplemented the discussion referring to a paper on the effect of the status quo bias in international arbitration written by her.<sup>16</sup> She also stated that other writers have written on other types of biases, such as unconscious gender or racial biases that might affect credibility determinations regarding witnesses in court proceedings and it would be beneficial to translate these judicial studies into arbitral settings.

Participant **David Chriki** asked whether, in **Prof. Brekoulakis**'s opinion, a carefully detailed statistical analysis that accounts for the circumstances of each case allows using outcome-oriented proxies to identify biases of arbitrators. **Mr Chriki** referred to research<sup>17</sup> he conducted which is focused on the outcomes of fair and equitable treatment ("FET") claims and examined what kind of measures are more likely to be found as "unfair" than others. This was done by applying variables, such as the measures taken by the respondent states challenged by the claimant, the type of FET provision in the agreement, and whether the state claimed the challenged measures served health or environmental concerns or were the result of a financial crisis or a revolution.

In response, **Prof. Brekoulakis** stated that in his opinion, the outcome-oriented proxies tend to say little about bias in decision-making. In addition to his and **Mr Kantor**'s previous comments, any outcome-oriented proxy will have to address and account for the very complex issue of the actual merits of a dispute, which has to serve as a point of reference for any assessment of bias in an outcome-oriented approach. Unless a study can find a way to assess the merits of the case, it cannot provide a meaningful yardstick for bias. For example, if in 100 ISDS cases, the claimants-investors are "right" to claim damages in all of them, then even if all 100 cases are indeed rendered in favour of the claimants, this 100-0 outcome cannot suggest that the arbitrators were biased. **Prof. Brekoulakis** further explained that for a research project to determine who is "right" *in abstracto* and without the benefit of reviewing the facts and exhibits of arbitration, is extremely difficult, if not practically impossible. In addition, **Prof. Brekoulakis** expressed the view that limiting the scope of such empirical research to a specific legal question by taking account of a number of important variables, e.g. measures taken, type of FET provisions, etc., like the one proposed by **Mr Chriki**, is in fact a rigorous approach and he looks forward to reading the findings of the analysis of **Mr Chriki**.

## **Speaker 2: Differences in the Approaches of Empiricists and Lawyers and Different Empirical Approaches**

Second speaker, **Prof. Robbennolt**, continued the discussion emphasising the differences in the approaches of empiricists and lawyers:

- 1) Empirical research is about the systematic observation of the world. In contrast, legal analysis and argumentation that may look to authorities or precedent or rely on logic.
- 2) Empirical research is aimed at testing hypotheses, while legal research often has as its goal the "proving" of a particular view or position.

---

<sup>16</sup> Strong, S.I., Truth in a Post-Truth Society: How Sticky Defaults, Status Quo Bias and the Sovereign Prerogative Influence the Perceived Legitimacy of International Arbitration (March 10, 2017). 2018 University of Illinois Law Review, 2018, Forthcoming; University of Missouri School of Law Legal Studies Research Paper No. 2017-07. Available at SSRN: <https://ssrn.com/abstract=2931137>.

<sup>17</sup> Chriki, David, Is the Washington Consensus Really Dead? An Empirical Analysis of FET Claims in Investment Arbitration (March 19, 2018). 41 Suffolk Transnational Law Review 291 (2018). Available at SSRN: <https://ssrn.com/abstract=3144386>.

- 3) Empirical research is typically concerned with aggregate and generalizable effects, while legal analysis tends to be focused on individual cases – resolving a particular dispute or negotiating a particular deal.
- 4) Empirical research is an incremental process in which studies build on each other, and each contributes a piece of the picture, ought to be read in the context of an entire body of studies, and often raise new questions. In contrast, legal decision-makers are often required to make final (and often binary – liable or not) decisions in individual cases – even when evidence is inconclusive, or questions remain.

**Prof. Robbennolt** then noted that there is a variety of different empirical approaches – interviews, surveys, experiments, content analysis. Each of these methods has its advantages and disadvantages since they prioritise different aspects and have a different focus. The trade-offs vary – but they are always there. All of this means that no individual study will be dispositive of a given research question.

**Prof. Robbennolt** further explained that the confidence in the robustness of the findings of empirical studies could be increased by coming at a research question using a variety of different empirical approaches. For example, if an archival study of actual cases can find an association between two variables in real-world cases and an experimental study can find that same association in a controlled setting, one can have more confidence in the result that it had used only one or the other method. This joint reading of studies across methodologies to get a nuanced picture of the phenomena at issue is known variously as convergent validity, converging operations, or triangulation. One triangulate in on an effect by approaching it from different angles. However, sometimes one comes at a research question from different angles by using different methods and the studies were done using various methods produce contradictory findings. In some instances, by exploring the reasons for those differences, it is possible to generate new hypotheses about the boundaries of particular phenomena, or variables that moderate an effect. **Prof. Robbennolt** advised that in designing studies researchers need to think about the advantages and disadvantages of different methodological approaches and how those strengths and weaknesses relate to the studies that have already been done, the availability of data, the nuances of the legal and research questions, etc.

#### *Q&A Session*

In the Q&A session that followed, moderator **Prof. Strong** asked whether certain types of empirical studies are better at generating particular types of conclusions/data than others. She expressed her opinion that interviews are likely to lead to anecdotal data rather than broadly applicable data.

**Prof. Strong** also asked how to tell whether the data is reliable or not and how to structure the questions to ensure that the data is useful and/or the surveyed population is appropriate. She then asked the members of the listserv about their thoughts when they receive a survey request, see survey results or statistical analyses with mathematical formulae and references to “n” values.

In response, **Prof. Robbennolt** agreed that particular methods could be better for answering specific questions. Referring to **Prof. Strong**'s question regarding interviews, **Prof. Robbennolt** explained that interviews could be particularly good at obtaining rich and detailed information from respondents because the latter may have more opportunity to elaborate on their answers, to be able to answer in their own words and provide information

on their own perceptions and experiences. Moreover, the interviewer can probe for more information or clarifications. She pointed out that interviews might be better than surveys for research questions for which detailed information is essential. Furthermore, interviews might be the right way to figure out whether actual practices match up with a picture of policies or practices that might be obtained by looking at official protocols or policy manuals.

**Prof. Robbennolt** then admitted that interviews also have their limitations. For example, respondents can only report their perceptions, their conscious rationales, and other things of which they are consciously aware, but because they might not be aware of the influences on their behaviour, they will not always be able to report on those fully. It is crucial to have protocols for probing answers and clarifying questions so that the interviewer does not influence the responses. Moreover, respondents may be less comfortable giving answers about sensitive topics to a live person, surveying with privacy protections more likely to get honest answers. **Prof. Robbennolt** agreed with **Prof. Strong**'s statement that because interviews take longer to conduct, sample sizes tend to be smaller. The trade-off here tends to be obtaining richness from a lower number of respondents. She then clarified that paying attention to the size of the sample (often indicated by the letter "n") and the composition of the sample can help one assess how useful and how generalizable the data are.

**Prof. Robbennolt** referred to her previous post, stating that when doing empirical research, it is useful to triangulate with multiple types of methods. It is also worth thinking about the ways that methods might be useful in sequences.

**Prof. Robbennolt** then stated that her previous comments also relate to **Prof. Strong**'s question concerning reliability. She recommended that to be sure that the generated data one relying on is reliable and valid, it is useful to know as much about the underlying methodology as possible. The reliability of a measure is the extent to which it is stable – producing the same or similar result every time it is used to measure the same underlying condition. **Prof. Robbennolt** gave an example with a bathroom scale – if it is reliable, it will show the same weight if one step on it (assuming no change in weight). In the context with the interviews, one wants to use procedures that would give the same information from the interviewee, if used repeatedly. That means that the protocols that are used – the questions that are to be asked, the probes that are to be used to follow up, the responses that are given for clarification – should be consistent across interviewers and interviewees. Similarly, if recording information from case files, one wants to make sure that the instructions are sufficiently clear that multiple people could use them and extract the same information from the files.

**Prof. Robbennolt** then drew attention to the validity of the data obtained – whether the study generates accurate and credible data and conclusions. She noted that data could be reliable without being valid. For this reason, one should know how well the methods used are measuring (referring to the question about proxies discussed previously), how well they can be generalised beyond the particular study, how well other influences are accounted for, etc. That means that one should pay attention to how survey or interview questions are worded and who are the respondents.

**Prof. Robbennolt** then asked the listserv members what types of studies they have questions about and where do they see potential issues with reliability and validity.

Participant, **Mr Kantor**, joined the discussion and stated that he often refers to empirical studies when commenting on international arbitration matters; however, he is sceptical about the reliability of many of them. His concerns are related to the obvious potential problems of those studies:

- 1) skewed selection of respondents - for example, surveying lawyers in areas where non-lawyers are the real decision-makers and/or surveying the wrong lawyers – interviews or surveys of lawyers regarding the role of BITs on FDI decisions are the archetype;
- 2) the impact of self-selection by respondents leading to studies showing the responses primarily of students and/or strongly-opinionated individuals and/or individuals with too much time on their hands and/or individuals lacking true personal knowledge or holding biases;
- 3) leading questions;
- 4) questions with real-world answers not included among the study options provided to respondents;
- 5) questions with multiple correct answers in the real world but only either/or choices in the study;
- 6) inadequate response populations;
- 7) technology impacts - phone v. email v. social networks or internet v. personal interviews;
- 8) absence of any effort to identify or publicise error bars;
- 9) drawing conclusions beyond the study data, etc.

**Mr Kantor** then referred to an empirical study intended to capture arbitrators' bias that reviewed outcomes in ISDS awards resolving issues on a list of "hard questions" developed by the author. The author had listed issues that may have been "hard" as a policy matter for that author but was not necessarily "hard" in terms of the text of one of the relevant legal instrument from which a significant number of the cases had been drawn or in terms of precedents applying that text. In the text of that legal instrument, the drafters of that instrument had resolved the policy issue in a manner different from the preferred result of the author. Therefore, according to **Mr Kantor**, the results of the study's award analyses had little value. Since the outcomes from that one legal instrument all went in one direction and that one instrument played a large role in the award population, the results of the study "demonstrated" that the arbitrators were biased in favour of the party who benefited from that outcome. But this weakness in study design would only be apparent to someone with knowledge of the instruments at issue in the award population. Others would not have the background to see the impact of that one legal instrument in the study methodology.

**Mr Kantor** then asked in addition to being cautious in relying on studies done by persons who have a personal or policy preference in the outcome, what other practical advice **Prof. Robbennolt** have for aspiring empiricists to minimise these design problems before their empirical study "goes live".

**Prof. Robbennolt** agreed that it is always important who are the people surveyed, both when one does a survey and when evaluating a survey. She explained that the population of interest varies from survey to survey depending on the research question. **Prof. Robbennolt** pointed out that one should think about the differences between the people surveyed and the people of interest and the extent to which those differences are likely to have influenced the key results of the study. She noted that in some instances the differences matter more than in other. For example, research on the understandability of jury instructions in the U.S. that used college

students as respondents were subject to criticism because this set of respondents did not match the profile of people sitting on juries – there are differences in age, level of education and other characteristics. However, given that the most relevant difference is the level of education, since the respondents are relatively highly educated, and the results of the study are that the instructions are difficult to understand, this might not be a fatal problem. If relatively highly educated people do not understand the instructions, likely, the jurors who are less well educated will also not understand the instructions. In other instances, if looking at legal decisions that tend to be made by legally educated people – college students may not be useful respondents to survey.

**Prof. Robbennolt** then referred to the second issue raised – how to get from a particular group of interest to a sample of respondents who come from that group. In a perfect world, one would like to have a random sample of people from that group, such that any person in the group has an equal chance of being selected to participate as any other person and the sample is representative of the group. This is not always achievable but is the gold standard. A big issue is that people will choose whether to respond – and there may be differences between those who do respond and those who do not call nonresponse bias. **Prof. Robbennolt** confirmed that as **Mr Kantor** has noted those differences might be relevant to the sorts of responses obtained. She advised that researchers should work hard to get a good response rate. Researchers should also think about the look or length of the survey at the design stage because they can also influence the response rate. It can also be useful to compare the survey respondents to the broader group of interest on characteristics that one know about, such as basic demographics, to see how well the sample matches the group on those characteristics. Those may not be the characteristics that one is most worried about in a particular instance but can give some sense of whether the sample is skewed.

**Prof. Robbennolt** then added a few more problems with survey questions to those listed by **Mr Kantor** - questions that use jargon or important terms are not defined, double-barreled questions that ask two or more things at once, questions that respondents will have difficulty answering due to the limits of memory or access to the relevant information, or questions that respondents might be unwilling to answer.

**Prof. Robbennolt** shared her experience that when drafting survey questions, she finds it useful to think about what the respondent has to do to provide useful information to a survey question. First, the respondent needs to understand the question – both literally and pragmatically. This means that one needs to pay careful attention to the wording and context of the question, what is communicated by the response options or scale points, etc. Second, the respondent has to access information in his memory to respond. This means that one wants to draft questions that the respondent will be able to answer given the bounds of memory. The wording of the questions can influence the connections that are made, and one should consider the earlier point made that people can only report things of which they are conscious. Third, the respondent has to map his response to the response options given. It is important to spend some time thinking carefully about those response options – making sure that the options cover all the bases, that respondent can choose more than one option if that is appropriate, that there is a “do not know” option, that scale labels are carefully chosen, etc. It is also important to have other people to review and comment on the draft survey and then to pretest. It can often be useful to have a respondent take the survey and talk aloud as they do so, so that one can get a sense of the thought processes at play. This is a good way to learn something about how the respondent understands or misunderstand the questions. **Prof. Robbennolt** noted that most people underestimate the amount of time and planning to design

a good study. Often people think the statistics are the hard part, but the design of the study is a really important step.

**Mr Kantor** then asked, from the perspective of a user of an empirical study assuming that that reader of a study know less than the respondents or even the author about the subject matter in the study, what steps can a responsible non-expert reader take to assure that the study is adequately designed and therefore can be relied upon?

In response, **Prof. Robbennolt** stated that as a consumer of empirical research, it is essential to pay attention to the details of the methodology, the respondents surveyed, the response rate, the wording of the questions, the response options. She noted that most peer-reviewed journals would require the author(s) of the study to report the details of the methodology. Those relying on the studies should read these details – not merely rely on the description of the findings and those reviewing studies for journals should be requesting missing details. If these details are missing, one should ask the author of the study to provide them. She pointed out that the question also highlights why it is vital for scholars and practitioners to have a basic understanding of empirical methods – even those who are not going to be conducting these kinds of studies themselves. A basic understanding is necessary in order to know what kinds of questions to ask about the studies that you encounter – and is useful not just for distinguishing flawed from high-quality empirical research, but also for being able to attend to the inevitable trade-offs and subtleties in interpretation that are always going to be present even in well-done work. It is also essential for those who are experts on the underlying substantive legal doctrine or practice to weigh in on how the study does or does not get those details right.

### **Speaker 3: Incorporating Systematic Content Analysis in Legal Research**

Third speaker, **Ms Maryam Salehijam**, continued the discussion and tried to convince the listserv members of the value of incorporating systematic content analysis (“SCA”) in legal research. She pointed out that while the need to integrate empirical research methods in legal scholarship remains a debated topic, empirical legal scholarship has expanded dramatically in the past decade. However, legal researchers rarely conduct their own empirical research but usually read and interpret legal material without a specific approach as to the selection of materials and the method used to conclude.

She noted that SCA is a research method commonly utilised by social scientists. It can be applied in the analysis of a variety of texts, ranging from interview transcripts to legal texts such as case law, legislation and contracts. Since SCA is a research technique, and thus, its application does not necessarily result in an empirically oriented paper, it can be used by scholars to enhance traditional legal research.

**Ms Salehijam** referred to the 2008 study of Hall and Wright, which found that the application of SCA by legal scholars was increasing annually. However, according to **Ms Salehijam**, the lack of resort to SCA is due to its functional difference to traditional legal research. SCA provides objective and scientifically verifiable findings. Therefore, it cannot be used to evaluate the correctness of judicial opinions or legislation. She quoted Hall and Wright stating that SCA “does not fully capture the strength of a particular judge's rhetoric, the level of generality used to describe the issue, and other subtle clues about the precedential value of the opinion. To some extent, the method trades depth for breadth”.

SCA would be suitable in cases where the researcher seeks to analyse legal material in a replicable manner and from an objective stance. It provides researchers with an alternative route when traditional legal research does not provide answers. ***Ms Salehijam*** gave the example that there were no judicial opinions, legislative texts and few scholarly works that address the parties' obligations under an agreement to mediate. She explained that the choice for SCA was fitting, as the parties' agreement to resort to mediation is the basis of their rights and obligations under the contract. Through an application of SCA, she was able to organise and decode 172 ADR agreements.

Moreover, according to Hall and Wright, studies that rely on SCA are more likely to be cited on average than those that do not, and SCA empowers the researcher to justify a claim numerically, which is easy to comprehend, refer to and reproduce.

***Ms Salehijam*** pointed out that SCA's various stages in legal research can be divided into five steps:

- 1) determination of a suitable research question or hypothesis for SCA;
- 2) identification and collection of sufficient data for analysis;
- 3) coding of the data, which has its own stages;
- 4) drawing of conclusions/observations; and
- 5) reporting the findings in a manner comprehensible by the legal community.

She then noted that there are a number of challenges that legal researchers unfamiliar with non-legal research may face in utilising SCA:

- 1) lack of familiarity and training;
- 2) time-consuming;
- 3) demanding data collection; and
- 4) fear of criticism.

***Ms Salehijam*** noted that law schools could play a prominent role in the incorporation of SCA in legal research by providing researchers with the confidence to employ new methods. Because the utility of SCA is not emphasised in legal education, many young legal researchers and students remain oblivious about the added benefits of SCA. It is thus of the essence that law schools take on the obligation to introduce their students to enriching research methods that fall outside traditional legal research.

In a subsequent post, ***Ms Salehijam*** further explained that the choice for her to use SCA was motivated by the need to have more in-depth knowledge of the subject of her research. Using SCA was challenging for her because she had to self-learn, there was a lack of resources for legal scholars wishing to employ this method, and she had to familiarise herself with SCA specific software – Nvivo. The next step was to collect the data for analysis. She had to determine whether her study aims at covering an entire population or a (representative) sample. She noted that legal scholars employing SCA tend to limit the scope of their research based on the subject matter, geography, language and/or time. For example, Diamond and Muller limited their study to law journals written in English published from 1998 to 2008. ***Ms Salehijam***'s parameters were somewhat open: "all Alternative Dispute Resolution (ADR) agreements provided by surveyed respondents and gathered via desk research". Her questionnaire received more than 600 respondents. However, many could not provide an agreement due to confidentiality or practical impossibility. Parameters differing from the

traditional components are needed when it is impossible to estimate the population. To illustrate, it is impossible to know how many commercial contracts require ADR when a dispute arises have been concluded in any given time range and geographical scope as these agreements are not publicly recorded. Lastly, even with a clear scope, the total population may be too large to analyse. She further explained that she has completed the coding and is in the final stage of SCA, which requires her to write her findings.

### *Q&A Session*

In the Q&A session that followed, participant **Mr Kantor** confessed that he did not have a clear notion of what is involved in SCA and how the software package relates to SCA method and asked for clarifications.

In response, **Ms Salehijam** explained that some consider SCA as “a sophisticated form of accounting”. The researcher can detect words and concepts in the data to draw inferences and find patterns. For example, if **Ms Salehijam** wants to see how often judges refer to the right of access to justice in refusing to enforce agreements to mediate in all cases since 1990 in England, she would use SCA to code the material for this instance. The distinguishing feature of SCA from traditional legal research is that the collection of the data and the subsequent analysis must follow a specific path. To enable counting, SCA involves coding. The researcher assigns the codes to words or concepts in the material. They can be used to track subject-matter, reasoning, arguments or superficial aspects of data such as length, author, language, etc. There are multiple approaches to coding that reflect the aim of the researcher. Coding also must be distinguished into stages. She then referred to the four stages of coding observed by Hall and Wright:

“(1) [...] create a tentative set of coding categories a priori. Refine these categories after thorough evaluation, including feedback from colleagues, study team members, or expert consultants. (2) Write a coding sheet and set of coding instructions (called a "codebook"), and train coders to apply these to a sample of the material to be coded. Pilot test the reliability (consistency) among coders by having multiple people code some of the material. (3) Add, delete, or revise coding categories based on this pilot experience, and repeat reliability testing and coder training as required. (4) When the codebook is finalised, apply it to all the material”.

**Ms Salehijam** then referred to **Mr Kantor**'s question concerning Nvivo and explained that its purpose is to aid the researcher in organising, analysing and finding insights in unstructured or qualitative data. It enables the researcher to organise and manage his material to gain insights into the data and to delve deeper into the various aspects of the data.

**Mr Kantor** stated that in his opinion identifying the operative phrase and similar phrases is the hardest task. He then asked how SCA takes account of linguistic nuance - phrases that are similar to the chosen phrase but have a slightly different emphasis or meaning.

**Ms Salehijam** answered that the choice of codes (the phrases) is initially a subjective one. To ensure that the coding is reliable, SCA experts made reliability checks through having a second pair of eyes and code a sample of the data already coded. This why it is essential that the researcher records his codes in a codebook detailing what each code refers to. Sample codebooks are freely available online and provide a good indication of how detailed codes can be. Codes can be accompanied by sub-codes, which allows for more nuanced findings.

For example, for the code “payment of the third party neutral” Ms Salehijam has the subcodes “equal share between the parties” and “one party pays all of the costs”.

#### **Speaker 4: Computational Analysis of Law and Big Legal Data**

Fourth speaker, Prof. Wolfgang Alschner, continued the discussion and drew the attention of the listserv members to the increasingly important type of empirical content analysis – the computational analysis of law and specifically, of big legal data.

Prof. Alschner noted that lawyers and legal scholars have to deal with an ever-growing amount of legal texts and it is virtually impossible to read them all and to digest the information contained therein. Computers can help lawyers and legal scholars to better understand vast amounts of unstructured legal texts by treating text as data.

Computational analysis of law is a systematic and replicable review of legal documents based on clearly articulated, objective criteria. It is in many ways similar to traditional content analysis, but there are also significant differences.

First, computers do the actual document review. On the one hand, this makes computational analysis highly scalable – one can analyse entire populations. On the other hand, computers cannot interpret texts. Therefore, when texts are translated into numbers and not read by humans, some potentially valuable contextual information is lost in translation. Depending on the research question, this may or may not be problematic. In any event, this means that computational analysis and traditional hand-coding by humans are complements, not substitutes.

Second, while computational approaches can be fruitfully applied in traditional deductive research for the testing of theory-based hypotheses, it is particularly rewarding to use them for inductive, exploratory analysis. By engaging in data-driven investigations and by letting legal texts speak for themselves, we can often find patterns we did not actively look for nor suspect that they existed.

Prof. Alschner pointed out that computational legal analysis has some challenges. Researcher time that is saved in manual hand-coding often needs to be invested *ex ante*, e.g. when cleaning texts for data processing, or *ex post*, e.g. when validating the computer’s findings from unsupervised machine learning algorithms. Moreover, biases can be aggravated through computational analysis. Missing data, for instance, can severely distort corpus or network characteristics leading to a false conclusion drawn from incomplete data. Finally, getting data and acquiring the skills to analyse it can be daunting for legal scholars. Notwithstanding these challenges, the computational legal analysis is an exciting new frontier in empirical legal research.

For a more extensive introduction to computational analysis, Prof. Alschner recommended to the listserv members to see Wolfgang Alschner, Joost Pauwelyn and Sergio Puig, “The Data-Driven Future of International Economic Law”, *Journal of International Economic Law*, Vol. 20, No. 2, 2017, pp. 217-231. To the listserv members interested in learning how to conduct a computational analysis of law he recommended taking a look at the teaching material of his “Legal Data Science” class available online at [www.legaldatascience.org](http://www.legaldatascience.org).

## Q&A Session

In the Q&A session that followed, **Prof. Strong** stated that in her opinion the computational analysis is best used on materials where one would expect to see a relatively high amount of consistency, for example, investment treaties, where the language is expected to be relatively similar. However, this method would be more problematic in cases involving more diverse language, such as judicial decisions, which can describe the same concept in very different terms. The problem would be exacerbated even further if one was attempting to research jurisdictional boundaries since functionally similar concepts might use different names. She then asked whether her understanding is correct.

**Prof. Strong** referred to the article *Democracy Distracted*<sup>18</sup> by Prof. James Williams that suggested the contemporary overload of information has made attention - not information - the scarce and desired resource. This corresponded to **Prof. Alschner**'s statements that there is now too much data out there for researchers to competently review using standard methodologies. **Prof. Strong** asked whether this suggests that "big data" techniques should be taught in law schools more.

In response, **Prof. Alschner** explained that computational analysis could be used for any legal document review but the technique used to automate manual content analysis will vary. He noted that contracts, statutes and treaties are suitable for automated textual comparisons because they employ repetitive and formulaic language. For example, Dmitriy Skougarevskiy and **Prof. Alschner** use it to investigate the content of trade and investment treaties. He further noted that judicial decisions are even better suited for computational analysis because they are written in natural language rather than the stylised language of statutes or treaties. One can employ the toolkit of natural language processing ("NLP") developed in computer science with little modification. In an EJIL article<sup>19</sup> Damien Charlotin and **Prof. Alschner** investigate the citation network of the International Court of Justice ("ICJ") and deploy various NLP techniques to figure out whether the ICJ uses precedent ritualistically or argumentatively. He admitted that it is more challenging to use automated content analysis across jurisdictions and languages, but there are solutions to automate at least part of the analysis.

**Prof. Alschner** then referred to **Prof. Strong**'s second question and stated that empirical and "big data" techniques should be learnt at law schools. He noted that he is teaching the first "Legal Data Science" course in Canada. He further explained that although only a few students will ever do empirical legal research, they will all have to learn how to leverage data science and artificial intelligence ("AI") – otherwise they risk being replaced by it.

**Mr Kantor** joined the discussion by asking what other database(s) exist and whether the database(s) containing the texts of the instruments, such as treaty texts, are publicly available, proprietary but available for licensing by third parties, or proprietary and not accessible to third parties at all.

**Prof. Alschner** answered that most of the non-commercial databases, such as UNCTAD's Investment Policy Hub or [www.italaw.com](http://www.italaw.com), are optimised for humans, but not for computers. Many of the documents are not machine-readable, and bulk download is cumbersome.

---

<sup>18</sup> Webcast is available at: [https://soundcloud.com/the\\_rsa/are-digital-technologies-making-politics-impossible](https://soundcloud.com/the_rsa/are-digital-technologies-making-politics-impossible).

<sup>19</sup> Available at: <http://www.ejil.org/pdfs/29/1/2836.pdf>.

Commercial databases, by contrast, have texts in a pristine machine-readable format. But they typically restrict bulk downloads of texts through their terms of service. Thus, getting legal texts in large quantities and a decent format is challenging. However, according to **Prof. Alschner**, this is about to change, and the pool of freely available legal texts in a machine-readable format is growing rapidly. He noted that in a joint project with UNCTAD, the texts of the free trade agreement are made publicly available in machine-readable XML. Another project in cooperation with the World Trade Institute is underway to do the same with investment treaties. Furthermore, PluriCourts in Oslo and iCourts in Copenhagen are also working in this direction. Finally, governments and international organisations are also beginning to realise that legal texts need to be adequately formatted to be mined efficiently by computers.

**Prof. Strong** joined the discussion and pointed out that although judicial decisions are public, this does not mean that they are easily obtainable, particularly in a computer-readable form. She shared her experience in doing research that required finding first-instance decisions from state courts, including small courts. Those decisions were not available on standard legal databases and would require going to each court and get the documents, probably on a pay-per-page basis, which given the cost and time involved would make that analysis too tricky and expensive.

**Prof. Strong** then discussed the issue of researching unpublished decisions. She noted that more decisions in the US are either not written or not fully reasoned than are written/reasoned/published decisions. Getting to those decisions is very hard, even in a system that allows the public to access all court materials. Some legal systems are much more restricted in what can be accessed by the public. For example, when she started practising in England, the underlying papers were not available to non-parties.

She invited the listserv members to share their thoughts on what items are public in their home jurisdictions.

**Prof. Strong** then provided an example for analysis of judicial decisions. If one is researching how the court approaches commercial decision-making might search terms like contract, agreement or transaction. However, on the one hand, the analysis may be far too broad because a business tort will also include these terms. On the other hand, the analysis might be far too narrow, since judicial determinations involving commercial disputes also arise in the regulatory context, for example, anti-trust/competition law. **Prof. Strong** asked **Prof. Alschner** whether the software devices he described capture the range of relevant decisions, and if so, how. She then asked how does the software "analyse" the issue once the decisions have been identified.

**Prof. Alschner** clarified that he and most of the people doing this type of research do not use any proprietary software. Instead, they rely on basic programming environments such as *Python* or *R* that are open source.

In response to **Prof. Strong**'s question, he pointed out that first, it would be helpful to group the documents by content. If one has 10 000 court decisions and only want to investigate those dealing with "commercial decision-making" one way to find those decisions is by running a so-called Topic Model across the data. One would specify the number of topics that would expect to describe content variation in one's data well and an unsupervised machine-learning algorithm would then assign the documents to these topics. Based on the most

frequent words by topic, one would then identify the topic(s) that most closely corresponds to what one considers “commercial decision-making” and focus only on those decisions that prominently deal with that topic(s).

Second, one would conduct an in-depth analysis of the selected documents. The analysis depends on what one wants to investigate. If one is interested in authorities cited, then would extract citations and create a citation network. If looking for specific issues raised, then one could repeat the Topic Model approach described above to further group the decisions made by sub-issue. If looking for specific concepts, then one might write rules that would allow the computer to find these concepts using keywords. Or, if one is interested in more abstract concepts that cannot be captured by rules, then one may code a sizable number of documents by hand and then train a supervised machine-learning algorithm to automate one’s analysis for the rest of the documents.

**Mr Christopher Campbell**, Associate Attorney at Willoughby & Hoefler, P.A., joined the discussion and noted that the final line of one of the previous posts of **Prof. Alschner** stroked him with particular interest: "They will all have to know what questions to ask a legal tech company to know whether it is worth buying their product. In short, they will all have to learn how to leverage data science and AI – otherwise, they risk being replaced by it." **Mr Campbell** then asked whether even lawyers who do understand this technology would still be replaced by it and if so, how far off that point is, and how do the lawyers of today stay relevant for tomorrow.

In response, **Prof. Alschner** pointed out that as the saying goes, “prediction is very difficult, especially if it's about the future.” **Prof. Alschner** noted that he would be very cautious, but there are a couple of things he wanted to highlight.

First, he explained that one should not overhype AI. The progress in data science that is underpinning all developments is based on sophisticated statistics, large amounts of data and a lot of number crunching.

Second, computers can do some things much better than humans, such as analysing thousands of documents. Computers thus beat humans when it comes to e-discovery tasks. But humans beat computers at other tasks, such as making creative legal arguments. Moreover, computers are trained on past data, which is why they have a hard time creating something genuinely new. This has two significant consequences. On the one hand, computers replace lawyers concerning tasks that one did not want to do, such as reading tens of thousands of emails to determine whether they are relevant evidence for this case. On the other hand, because the respective strengths of humans and computers are complementary, the most successful lawyers will be those who combine the skills of humans and machines.

According to **Prof. Alschner** in the short term, AI will not replace all lawyers, but it will create new jobs, such as legal technologists, legal data scientists, etc. In the long term, technology is likely to alter the way legal systems work fundamentally. More data and greater connectivity including between governments, citizens and companies will likely transform the administration of justice and the way law operates, regulates and is enforced in ways, which are now difficult to foresee. This will also have an impact on the role of lawyers and the legal profession as intermediaries. The main takeaway to drawn from all this is that lawyers and especially law students will have to be more aware of how technology transforms the legal profession to anticipate, leverage and shape this change.

## **Speaker 5: Aspects of Empirical Research Specific to International Commercial Arbitration**

Speaking last, **Dr Suha Jubran Ballan**, discussed the aspects of empirical research specific to investment treaty arbitration.

**Dr Ballan** pointed out that mapping a field and drawing a line between different categories, features, or variables is a challenging task in every kind of empirical research and particularly in international commercial arbitration. She noted that this is due to its diversification, referring to the fact that international commercial arbitration is a de-centralised and pluralistic field.

Diversification opens up new topics of research. The field comprises arbitration centres, default rules of procedure, and expertise of lawyers and arbitrators which has been evolving for more than several decades now. In a diversified and decentralised field, doctrines have been shaped, and precedent developed, and a new field of law is being distinguished.

**Dr Ballan** then referred to the challenges set by diversification in data collection and interpretation. In many areas of law, one can know who the judges in a certain court are, and how many decisions have been decided during the last year. Capturing this data in international commercial arbitration is limited, and doing so requires considerable resources. Due to common cognitive biases, interpreting the data is challenging as well. Interpretation of data may involve reflecting one's pre-conceptions, premises, or an assumption of similarity to contexts one is familiar with. This risk is particularly high in international commercial arbitration.

She then referred to the methods to manage the challenges set by diversification. One method is to conduct empirical research in collaboration between experts from different legal cultures and jurisdictions which may help manage challenges of data collection and cultural bias. Another method is to build upon existing empirical research, sharing databases, and comparing and explaining differences. The starting point would be to discuss which data is covered and which is not the limitations of the data, and its implications. Unnecessary generalisations should be avoided. Proper empirical research is one that acknowledges its limitations and distinguishes between conclusions that are directly reflected from the data, and others that require some dose of interpretation.

**Dr Ballan** then asked the listserv members to what extent diversification is an obstacle to empirical research and how it can be managed.

### *Q&A Session*

In the Q&A session that followed, **Prof. Strong** raised a few questions to **Dr Ballan**. First, **Prof. Strong** referred to **Dr Ballan**'s statement that she has found that the use of case law differs from one institutional context to another. She asked **Dr Ballan** to further elaborate on that and explain how she was able to discern that using empirical methods.

Second, **Prof. Strong** referred to **Dr Ballan**'s comment about the role that cognitive biases can play in decision-making and to the previous discussion during the symposium about the difficulties in unpacking rationales in decision-making from empirical data. She provided an example that a judge/arbitrator might decide against a certain party because of a bias against

that party or because that party just did not have a winning legal argument. **Prof. Strong** then asked how empirical research can help identify the presence of cognitive bias and how one knows what type of bias that is. For example, the bias might be against a particular type of parties (states or investors), or it might be against a certain type of advocate (common law v. civil law, male v. female, etc.).

Concerning **Prof. Strong**'s second question about bias, **Dr. Ballan** explained that it is essential to clarify what is meant by bias. There are two types of bias. The first type is the judicial definition, which tackles bias from a procedural manner. In most legal systems, judicial decision-making should be impartial, independent and open-minded. Parties should be allowed to argue their case and bring evidence, and the judge/arbitrator should listen to their arguments and consider them. Deviating from these rules may invalidate the legal decision. These procedural restrictions may minimise biases but not eliminating them, due to cognitive biases of decision-makers. In such cases, empirical research can be either done through experiments and questionnaires or by focusing on outcomes.

The first method focuses on the 'process' while the second method focuses on 'output'. In **Dr. Ballan**'s view, those methods have different aims. While the first examines the types of cognitive biases that may be processed, the second exposes actual biases that had already happened in judicial decisions. Thus, focusing on outcome statistics is more helpful to identify biases against certain groups of people or certain advocates. The starting point will be the impression of interacting with the system, which may be validated or not through empirical research. It is also possible to study different kinds of cognitive biases, try to think about how they may influence decision-makers, and only then attempt to examine the validity of the impression empirically. Both methods require normative assessments about the findings. For example, a different normative assessment is made for decisions in favour of a certain group depending on the relevance of the identity of that group to the case-matter and on whether it is a judge or an arbitrator. In investor-state arbitration, it is well known that certain arbitrators are inclined to decide in favour of states or investors, but many would argue it is a healthy diversity of the arbitration regime. In court legal systems, a judge that favours certain groups is considered biased, undermining the legitimacy of the system. Distinctions between the dry facts presented as empirical outcomes and normative conclusions should be made clear.

**Dr. Ballan** then referred to **Prof. Strong**'s first question and stated that the research she conducted examined how investor-state arbitration tribunals in different institutional backgrounds use different judicial reasoning in investor-state disputes. Exploring the use of case law shows that NAFTA tribunals refer only to NAFTA case law (90%) while other non-NAFTA tribunals refer to NAFTA case law in a normal percentage. This implies that NAFTA tribunals may have more impact than others in developing case-law. She noted that when she presents those findings, the reaction she got is hardly a surprise. However, when the development of case law and precedent is discussed in the literature, it is considered generally, without referring to the different categories of case law.

**Prof. Alschner** joined the discussion commenting on the findings of **Dr. Ballan** and stated that NAFTA tribunals have a "NAFTA hat" on when they are deciding NAFTA disputes. He pointed out that he has done analysis on citation patterns and found that tribunals constituted under the Energy Charter Treaty ("ECT") only refer to jurisprudence rendered under other ECT disputes in 16% of the cases (as opposed to NAFTA tribunals where **Dr. Ballan** found that 90% of citations are intra-NAFTA). While NAFTA tribunals wear a "NAFTA hat" and

primarily draw on NAFTA precedent to interpret NAFTA obligations, ECT tribunals do not wear an “ECT hat” and draw on non-ECT precedent to interpret ECT obligations. One conclusion may be that investment arbitration case law is not a coherent whole but instead consists of jurisprudential islands, which are more or less interconnected. NAFTA cases speak to ECT cases, but ECT cases do not speak to NAFTA. **Prof. Alschner** stated that the empirical research of **Dr Ballan** helps to bring these previously unknown patterns to light, but it also raises new questions. **Prof. Alschner** was wondering what the cause for those patterns is and what is the reason NAFTA tribunals to wear “NAFTA hats” rather than broader “investment arbitration hats”. He was also wondering whether the underlying procedural rules, underlying treaties or the identity of arbitrators are the reason and what renders some of these islands isolated (few citations to non-NAFTA cases), but influential (many citations from non-NAFTA cases to NAFTA cases) while others are interconnected, but less influential (ECT cases).

**Prof. Strong** then asked whether the incidence of citation has something to do with how various decisions are reported. For example, the NAFTA decisions, like the Iran-United States Claims Tribunal cases, were widely reported in specialised print reporters even before the age of the internet. Moreover, NAFTA and Iran-United States Claims Tribunal cases developed relatively early, before the explosion of BIT arbitration, which meant that counsel and arbitrators were familiar with those awards, leading to widespread citation. **Prof. Strong** referred to an empirical project she was conducting and stated that familiarity is an essential element of legal reasoning/argument. This is particularly important given the average age of arbitrators since many came of age during the 1990s or before.

She assumed that there might be a connection that has little to do with the conscious decisions of the arbitrators and more to do with the availability of/familiarity with data. **Prof. Strong** referred to a book and an article<sup>20</sup> on research in international arbitration she has written, but she did not see any ECT reporting series similar to the NAFTA and Iran-United States Claims Tribunal sources. She only found analyses of ECT cases in two monographs, printed in 2006 and 2009, whereas NAFTA reports and treatises dated back to the 1990s, as do materials on the Iran-United States Claims Tribunal. In her opinion, the timing/availability issue might be relevant to the analysis of citation methods.

**Prof. Alschner** agreed that the timing and availability of awards are potential determinants. He also added that the different ways the parties argue cases might be another factor. Tribunals may not be consciously wearing that NAFTA hat. He noted that more empirical research is necessary to explain the causal factors behind the patterns properly.

**Dr Ballan** agreed with **Prof. Strong** that familiarity with the awards might play a role, as well as the fact that tribunals may refer to cases brought by the parties, but this is not the case for the pattern she described in NAFTA’s references. First, in certain cases, the parties do refer to non-NAFTA cases in their arguments, but the tribunal prefers to discuss only NAFTA case law. Only NAFTA tribunals refer in a pattern not consistent with the distribution of awards.

In response to **Prof. Alschner**, **Dr Ballan** explained that formal and informal institutional arrangements shape the legitimate concerns of tribunals, and in the case of NAFTA, tribunals mainly address the local public. Significant to NAFTA, is that most cases are examined by national courts in Canada and the US that the states intervene in the process as a signatory

---

<sup>20</sup> Available at: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1650263](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1650263).

and not as respondents. She noted that computerised analysis might be extremely helpful in such cases, as it allows the examination of more caseload and tests alternative hypotheses in a shorter time.

**Prof. Strong** then suggested the speakers discuss the importance of knowing in advance **how** the shape of particular research will affect the outcome and underlined the importance of the desk research when determining the environment in which the study will be set and to design the research question properly.

**Dr Ballan** agreed that finding the research question to start with is never easy, and it is very frustrating to conduct empirical research and to realise later that the data does not indicate insightful conclusions. She confirmed that it is vital to be familiar with existing empirical research in the field of study. It is also helpful to choose a topic that contributes to existing research even if the data does not eventually bring the expected outcomes.

**Mr Kantor** joined the discussion and stated that research results failing to support the researcher's hypotheses are often not published. But the information that a hypothesis was not supported is often just as valuable to decision-making as the converse. He asked whether **Dr Ballan** has any suggestions for how unsuccessful research can be disseminated without harming the researcher's career.

In response, **Dr Ballan** explained that one way is to try and figure out why the data is different than expected. Sometimes finding the answer may lead to interesting insights. Another way is to set a question that is interesting to answer regardless of the outcome.

**Prof. Strong** then added that one should also think about the concept of the null hypothesis, which is a concept drawn from the hard sciences. The idea is that one should not be trying to prove the thing that expects but should instead try to prove the opposite. That will help to avoid the confirmation bias and other types of analytical errors that lead one to see what he expects to see. For example, when researching class arbitration, she set out to prove that it was an abomination and improper as a matter of procedural justice but learned that it was a procedurally just process, given the nature of the disputes at issue.

In response to **Prof. Strong**'s question about ECT, **Dr Ballan** pointed out that a possible answer may be that the majority of past ECT awards are managed by non-ICSID arbitration centres which are more commercially oriented than ICSID, and their awards are reviewed by national courts in several European jurisdictions. For that reason, a possible explanation is that the forces driving to uniformity in the ECT context are weaker than those in NAFTA or ICSID. She also noted that examining case law means being several years behind and that the developments of recent years are not captured yet.

Participant **Mr Damien Charlotin** joined the discussion and agreed that it is a very challenging experience to spend hours collecting and analysing data to confirm a hypothesis and fail, but it is similar when the data confirms one's hypothesis. In the context of citations to NAFTA jurisprudence, he introduced Data Visualisation as an essential facet of empirical research. He then referred to a study he conducted about the citations to precedent in investment arbitration awards using network analysis software and explained how the data was visualised.

In response, **Prof. Strong** stated that the visual chart he has used is consistent with best practices. She asked whether there are values that are measured horizontally and vertically, in addition to the density elements reflected in the size of the circles.

**Mr Charlotin** clarified that there is no X and Y axis, but the network analysis map is based on a network of relations. The algorithm used looked at all those relations and found the best way to represent them in a two-dimensional area. He then explained the way the algorithm works.

**Prof. Strong** followed up asking how a reader is to determine relative distances between core and non-core data points without a scale and referred to a chart<sup>21</sup>. She noted that for **Mr Charlotin**'s type of chart could be significant differences in how one reads/interprets/applies the data depending on how far away different nodes are from each other.

**Mr Charlotin** agreed that with data visualisation precision and exhaustivity are often trade off for clarity and immediacy. He noted that the key feature for this kind of map is the relation of proximity between nodes of a certain kind. Then he explained that his map allows one to see that the NAFTA nodes are generally much closer one to another than to non-NAFTA nodes and that ECT nodes do not exhibit the same kind of close relationship at all. He further explained that network analysis is built on a range of mathematical and statistical tools and methods, one of which is called homophily. It measures how likely two nodes are to be connected one to another given a characteristic(s) they have in common. He noted that it is likely that the characteristic "NAFTA as a jurisdictional basis" would score high on how likely it predicts the level of neighbourliness between two awards.

**Prof. Strong** then referred to a video<sup>22</sup> from West Wing in which CJ and Josh meet with the Cartographers for Social Equality to discuss how visual representations - this case, the map of the world - can unconsciously affect both perceptions and policy.

Moderator, **Prof. Strong**, closed the symposium by thanking the speakers for taking the time to share their knowledge and insights with the audience.

---

<sup>21</sup> Available at: <https://i.imgur.com/MAw0bdT.png>.

<sup>22</sup> Available at: <https://www.youtube.com/watch?v=QMlp8BeBJgg>.

**Transnational Dispute Management** is a peer-review online journal publishing about various aspects of international arbitration with a special focus on investment arbitration.

This article was made available as a free download from [www.transnational-dispute-management.com/](http://www.transnational-dispute-management.com/).

Young professionals and students with an interest in transnational disputes, international investment law developments and related issues, are hereby invited to join the Young-OGEMID discussion group. Membership is free.

Simply visit [www.transnational-dispute-management.com/young-ogemid/](http://www.transnational-dispute-management.com/young-ogemid/) and fill in the registration form. Other summaries of our Young-OGEMID virtual seminars focussing on career development can also be found there.