European Journal of International Law
© The Author(s), 2022. Published by Oxford University Press on behalf of EJIL Ltd.
All rights reserved. For Permissions, please email: journals.permissions@oup.com

Behavioural Economics and ISDS Reform: A Response to Maria Laura Marceddu and Pietro Ortolani

Thomas D. Grant[*] and F. Scott Kieff[**]

Abstract

Academic investigators have used behavioural economics, a method developed originally to study consumers and their sentiments towards products, to study matters of public policy. A recent article in the European Journal of International Law *– 'What Is Wrong with Investment Arbitration? Evidence from a Set of Behavioural Experiments' – gives a detailed summary of a series of experiments performed in order to study public sentiment towards investment arbitration. The investigators, Maria Laura Marceddu and Pietro Ortolani observe that public sentiment improves towards the outcome of a dispute settlement procedure when survey respondents are told that the procedure was a 'court' with tenured judges, and it worsens when they are told that it was 'arbitration' with temporary appointees. From their observations, Marceddu and Ortolani conclude that an international investment court, such as that which the European Union promotes, is a good idea. We suggest, however, that a further inquiry should investigate in greater detail public understanding of what qualities the individuals who serve as judges or arbitrators ought to display, as distinct from the institutional format in which dispute settlement takes place.*

1 Introduction

Who should decide investment disputes? The question gains practical urgency as the European Union seeks to replace investment arbitration with international investment courts.[1] However, in the one question, two discrete questions are contained.

[*] Fellow, Lauterpacht Centre for International Law, University of Cambridge, United Kingdom; former designee, Permanent Court of Arbitration (US National Group). Email: tdg20@cam.ac.uk.

[**] Fred C. Stevenson Research Professor, George Washington Law School, Washington, DC, USA; former Commissioner of the US International Trade Commission. Email: skieff@law.gwu.edu.

[1] See European Parliament (Esther de Lange, Rapporteur), *Legislative Train Schedule: A Balanced and Progressive Trade Policy to Harness Globalisation* (2021), available at https://www.europarl.europa.eu/legislative-train/theme-a-balanced-and-progressive-trade-policy-to-harness-globalisation.

What institutions should conduct investment dispute settlement? And of what individuals should those institutions be composed? Maria Laura Marceddu and Pietro Ortolani, authors of an empirical study on public sentiment towards investor–state dispute settlement (ISDS), recently suggested in an earlier issue of the *European Journal of International Law* that the institutions are what matters – and that replacing investment arbitration with investment courts is the right policy response to a crisis of public confidence in ISDS.[2] We are not so sure.

Reading Marceddu and Ortolani's study with the careful attention it deserves, we think that policy-makers should consider the possibility that what really influences public sentiment are the attributes of the individuals who decide investment cases. A continuation and enlargement of their empirical work could test both Marceddu and Ortolani's hypothesis that it is the institutional architecture that matters and an alternative hypothesis that it is the people who make the decisions that matter as much or more.

We start here by recalling the main findings and recommendations from Marceddu and Ortolani's study (section 2). We then propose an alternative model to explain the empirical evidence that they gathered, and we suggest an experimental design that looks more closely at precisely what features of courts and arbitration shape public sentiment (section 3). We recall that behavioural economics, in its original form, called for detailed investigation of the features of a thing being studied and that the more features addressed in an experiment, the more reliable the experimental results (section 4). In concluding, we suggest that the people who comprise a dispute settlement body are a crucial feature, and we express hope that Marceddu and Ortolani's novel approach to public sentiment and investment arbitration invites a closer look at what qualities in such people – independently of whether they hold the title 'judge' or 'arbitrator' – most instil trust in the public they serve (section 5).

2 A Behavioural Economics Case for Decision-Maker Tenure: From Data, to Model, to ISDS Reform

Describing their study as 'the first-ever set of behavioural experiments concerning ISDS and public opinion',[3] Marceddu and Ortolani tested a survey cohort consisting of 684 respondents selected on criteria that they belong to '[t]he ISDS front' – that is, groups associated with public opposition to ISDS – and that they not be ISDS experts.[4] The respondents read dispute settlement scenarios and were asked which they preferred. From the respondents' answers, Marceddu and Ortolani concluded that the factor that most strongly affects preference is institutional design – in particular, whether the decision-making body is an ad hoc arbitral tribunal or a standing court.

[2] Marceddu and Ortolani, 'What Is Wrong with Investment Arbitration? Evidence from a Set of Behavioural Experiments', 31 *European Journal of International Law* (2020) 405, at 427.

[3] *Ibid.*, at 405.

[4] *Ibid.*, at 416–417.

As with any data set, the observations that Marceddu and Ortolani obtained through their experiments say little in themselves. Necessary for making sense of a data set is a model to explain or predict[5] the observations that comprise it. To arrive at a model, investigators draw inferences from the data. Referring to their empirical observations, Marceddu and Ortolani posited that people do not have confidence in arbitration but they do in courts. To express this thesis in terms of a predictive model, confidence decreases when it is arbitrators who make the decisions. The policy conclusion that Marceddu and Ortolani advance from that model is that a world investment court is a good idea. Few, if any, data sets of any richness, however, are susceptible to only one model. Moreover, the richer a data set, the more diverse the models that it is likely to support.

3 An Alternative Model and Interrogating the 'Tenure' Claim

From Marceddu and Ortolani's data, we believe that an equally plausible model is this: respondents trust decision-makers having the attributes of people whom they assume are appointed to courts. Accordingly, respondents express a preference for decision-makers whose attributes are those that they assume people appointed to courts have. Conversely, respondents express negative sentiments towards people either whose attributes do not match this assumption or whose attributes are unknown to them.

Expanding on Marceddu and Ortolani's work, investigators should delve deeper into the question of what, precisely, people think makes a court trustworthy – in particular, what attributes of the decision-makers – that is, of the judges – attract trust. We posit that those attributes are found in at least some of the people who might serve as arbitrators and, thus, that institutional design might not be the main factor shaping public sentiment towards these decision-making procedures. If the term 'court' is a proxy for other features of a dispute settlement system, then policy-makers thinking about the redesign of the system should consider precisely what those other features might be.

A *Constructing Models Tied to Respondents' Experience and the Attributes of 'Courts'*

Predicting behaviour towards an object with definable features is the goal of behavioural economics modelling. Danial McFadden, the econometrician whose Nobel Prize work was the main early influence in the field,[6] described how to construct a behavioural economics model: '[P]sychological scales can be treated simply as attributes of alternatives in probabilistic-choice models, provided they are constructed *so as to contribute to the explanation of behavior on one hand and to be tied to the experience of the*

[5] For how data scientists use the term 'predict', including its distinction from 'forecasting', see T.D. Grant and D.J. Wischik, *On the Path to AI: Law's Prophecies and the Conceptual Foundations of the Machine Learning Age* (2020), at 54–57.

[6] Expert resume filed in *Guitierrez et al v. Johnson & Johnson* (US District Court for the District of New Jersey, 15 May 2006).

consumer and the objective attributes of products on the other.'[7] McFadden's concern was with market research and, thus, with the experience of consumers and the attributes of products. However, his insight applies across the subject matter that investigators examine through behavioural economics. According to that insight, the investigator must consider precisely what the attributes are of the physical artefact or other phenomenon towards which the investigation aims to elucidate people's behaviour. As McFadden has said, the scales or models must be constructed so as to 'be tied to the experience of the consumer *and the objective attributes of products*'.[8]

The respondents in Marceddu and Ortolani's experiments equate to the 'consumers'. Courts and arbitration tribunals are the 'products'. We think that the experiences of the respondents, and the objective attributes of courts and arbitral tribunals, call for closer examination. The terms 'court' and 'arbitration tribunal' are centrally important to Marceddu and Ortolani's study. The dispute settlement scenarios in the study offered information as to what those terms mean, but the information was only basic. It is not clear that the respondents were given enough information about the object towards which their sentiment (that is, behaviour) was being measured for the study to show precisely what features of the object most affected their sentiment.

The information that Marceddu and Ortolani gave their respondents about the words 'court' and 'arbitration tribunal' related to mechanisms of appointment. For example, in one experiment, they referred to '*the mechanism whereby adjudicators are appointed*'.[9] In another experiment, they infer that a possible explanation of the observed data is that the participants did not like 'the *temporary nature* of investment tribunals and the *untenured character* of the arbitrators'.[10] Marceddu and Ortolani indicate that, from the observed data from their final experiment, the respondents 'reacted better to the [*permanent* court] version of the story'.[11] These three experiments thus identified one objective attribute of the objects concerned – whether the objects were permanent or temporary institutions.

A logician or a jurist might interpret information such as this as stipulative, meaning that she would understand that, once the information is given, extrinsic understandings of the subject matter are excluded.[12] We are not aware that the survey respondents in Marceddu and Ortolani's study were instructed to treat the information that they were given about ad hoc arbitral tribunals and courts in a stipulative way. Even in a courtroom where members of a jury are instructed about what information or preconceptions they must leave at the door, the exercise is highly fraught; courts struggle with the risk that extrinsic information or bias might misshape jury outcomes.[13]

7 McFadden, 'Econometric Models for Probabilistic Choice among Products', 53 *Journal of Business* (1980) S13, at S23 (emphasis added).

8 *Ibid.* (emphasis added).

9 Marceddu and Ortolani, *supra* note 2, at 423 (emphasis added).

10 *Ibid.*, at 422 (emphasis added).

11 *Ibid.*, at 426 (emphasis added).

12 See B.A. Garner, *Lexical and Stipulative Definitions, Dictionary of Modern Legal Usage* (2nd edn, 2001), at 257–258.

13 See *Moore v. Dempsey*, 261 U.S. 86 (1923), Holmes, J.

McFadden, from the start, emphasized that an experimental model should be tied to the experience of the survey respondents. In the experience of the non-expert, the term 'arbitration tribunal' might have some connotations, but the term 'court' is likely to have more. When a respondent comes to a survey such as Marceddu and Ortolani's, the respondent, we postulate, brings a conception of 'court'. The respondent's conception, we surmise, is that the people who staff a 'court' have attributes recommending them for making decisions on matters of public concern. We therefore think that it would be fruitful to construct a slightly different experimental model, one that ties more closely to the respondents' experience of 'court', of 'arbitration tribunal' and of judges and arbitrators. Getting to a more granular understanding of that experience, a future experiment could test further hypotheses as to what characteristics instil public confidence in dispute settlement.

B *A Hypothesis About the People Who Serve*

We hypothesize that public antipathy towards ISDS owes not to the absence of a standing, permanent court, as such. It owes, instead, to the absence of reliable signals of trustworthiness in the decision-makers when the term 'arbitration' is used. In our hypothesis, the word 'court' sends a signal that attracts the public's trust to the people whom the public assume serve on courts. In contrast, the word 'arbitration' either sends little signal at all or it evokes private individuals whose provenance, credentials and intentions are not to be trusted. We hypothesize that favourable sentiment towards decisions that the body adopts correlates to a signal that the decision-makers deserve the public's trust. The title that attaches to the body, we hypothesize, is not necessarily as strong a signal as information about the people who comprise the body.

A further set of experiments would test this hypothesis about courts and arbitration by determining what characteristics in a decision-maker, precisely, lead respondents to express confidence in the decisions reached. If appointment to a court is the sole characteristic that assuages respondents' concerns, then that would support the argument for a tenure-based system of ISDS review. By contrast, if it were to come to light that 'court' and tenure are not the only features that predict public sentiment or the best, then policy-makers ought to consider alternatives to the current reform agenda.

4 Challenging the Choice Model

Lawyers sometimes use Latin phrases because they sound important and, moreover, might deflect the judge from asking for meaningful definitions of the concepts in issue.[14] An alert judge will tell the lawyer to give a meaningful definition.[15] This familiar

[14] See this critique of Justice Coke for that jurist's resort to Latin phrases 'when he lacked good arguments or precedents'. *Sperbeck v. A.L. Burbank & Co., Inc.*, 190 F.2d 449, 450 (2nd Cir., 1951).

[15] For a recent, and extreme, example, albeit one of a *pro se* litigant (that is, a litigant proceeding without a lawyer), see *Benson-Staebler v. City of New York*, slip op. (US District Court for the Eastern District of New York, 16 June 2020), at 1.

courtroom experience is, at least in a general way, analogous to what McFadden was getting at when he called for models that are tied to the objective attributes of the subject matter being studied: a word can describe but can also obscure; a model can invoke a term or concept but fail to reflect what lies beneath. So have Marceddu and Ortolani addressed the objective attributes that lie beneath popular opinion about courts and arbitration, or do those terms inadvertently obscure the terrain that calls for study?

When parties and experts in courtroom settings have challenged particular choice models, they have drawn attention to omissions of 'a critical parameter'.[16] A critical parameter is *per se* missing, if an experiment did not interrogate the respondents closely enough. A critique along that line convinced a US district court in *Laumann v. National Hockey League* to reject the plaintiffs' choice model.[17] The plaintiffs in *Laumann* alleged that professional sports leagues were engaged in unlawful price fixing. The plaintiffs called an expert, and the expert prepared a study concerning sports fans' behaviour. The study concluded that the price that fans pay for broadcasts, if the leagues were to stop their alleged misconduct, would drop considerably.[18] The Court, however, did not buy it. The Court judged that the plaintiffs' expert's choice model displayed a flaw that was 'quite fundamental and fatal' – namely, that it '[did] not rely on sufficient data about consumer tastes and preferences' to support the conclusion that that expert had inferred from it.[19]

According to McFadden, whom the defendants called as expert, the plaintiffs' expert's model was unreliable because it needed to 'find out more about what [the consumers'] tastes are, whether they would consider buying or not at various suggested prices'.[20] In short, the plaintiffs' expert did not ask enough questions of his survey respondents about enough objective features of the subject matter concerned.

5 Conclusion

McFadden's observations in *Laumann* and in his writings on the behavioural economics method have general salience to any choice model. The better model is the one that more closely interrogates the objective attributes of the thing concerned and the experience of the individuals surveyed towards that thing. We think that the features of 'court' and of 'arbitration' and the *a priori* assumptions of the relevant public about those terms require closer examination if empirical claims about popular discontent

[16] *Oracle America, Inc. v. Google Inc.*, unreported (US District Court for the Northern District of California, 2 May 2016), at 7 (expert critique of choice model); see also *Advocates for Transportation Alternatives, Inc. v. U.S. Army Corps of Engineers*, 453 F.Supp.2d 289, 303 (US District Court for the District of Massachusetts, 2006) ('[t]he Federal Transit Administration… noted that the [mode choice]… model used by the MBTA to compare bus and commuter rail alternatives was "insensitive" to important factors generally addressed in [mode choice]… models').

[17] *Laumann, et al. v. National Hockey League et al. and Office of the Commissioner of Baseball, et al.*, 117 F.Supp. 3d 299 (US District Court for the Southern District of New York, 2015).

[18] *Ibid.*, at 302–303.

[19] *Ibid.*, at 315.

[20] *Ibid.*, at 313, quoting McFadden's trial testimony.

towards ISDS are to justify the institutional transformation that advocates of ISDS courts pursue. We hypothesize that public discontent towards ISDS correlates more strongly to information about who serves in dispute settlement organs than to institutional design. A hypothesis that we believe should be tested is that, when respondents receive signals that the community has confidence in the people comprising a decision-making body, their sentiment improves.

The question of what attributes in a decision-maker instil public trust could be the focus of experimental inquiry. Institutions of social significance, investment dispute settlement included, are in a state of ferment. A particular concern is that the people who serve in such institutions do not represent the communities whom those institutions affect. ISDS is, to say the least, not the most diverse or inclusive institution, and, yet, ISDS decisions are significant to society at large. A fresh look at the factors that shape public sentiment towards ISDS might consider whether ISDS decision-makers would earn more trust if they more closely reflected the public whose interests they affect. At the very least, it might be asked whether attributes of public-mindedness, such as a record of service in positions of public trust, would instil greater confidence in ISDS decision-makers.

Marceddu and Ortolani have furnished a useful study. It invites further experiment. The behavioural economics method suggests how investigators might come to a more refined understanding of public sentiment towards decision-makers, a research agenda relevant to ISDS reform and many other areas of public policy as well.