VOLUME 55, NUMBER 2, SUMMER 2014

Behavioral International Law and Economics

Anne van Aaken*

Whereas the rational choice approach to international law has been widely accepted in legal scholarship and international relations theory, challenges to the rational choice paradigm in economic analysis of international law have hitherto not been systematically explored. Nevertheless, behavioral law, economics and psychology have been successfully applied to national law constellations. Behavioral economic insights have furthermore been used in international relations scholarship under the heading of "political psychology," but international norms are neglected. Building on all those insights, this Article explores the potential and challenges of extending the behavioral law and economics approach to public international law and thus to further refine our understanding of international law. It looks specifically at treaty design problems and compliance questions. This ties in with increased use of empirical research in international law, a clear desideratum for evidence-based international law.

INTRODUCTION

Whereas the rational choice approach to international law has been widely accepted in legal scholarship and international relations (IR) theory, challenges to the rational choice paradigm in economic analysis of international law have hitherto not been systematically explored.[1] Nevertheless, behavioral law and economics has been successfully applied to national law and is now widely accepted. Behavioral economic and psychological insights have furthermore been used in international relations scholarship under the heading of "political psychology." However, the latter research does not take into account international norms and is mostly confined to individual actors. This Article explores the benefits and challenges of extending the behavioral

* Prof. Dr. Anne van Aaken, Professor for Law and Economics, Legal Theory, Public International Law, and European Law, University of St. Gallen, Switzerland. Email: anne.vanaaken@unisg.ch. The paper was first presented at Duke Law School in the CICL Lecture in early September 2013, extending a thought that I formulated, after co-organizing a conference on Public International Law and Economics, in Anne van Aaken, *Towards Behavioral International Law and Economics? Comment on Kenneth W. Abbott*, 2008 U. ILL. L. REV. 47 (2008); the papers of the conference are published in: *International Law and Economics, Symposium Issue*, 2008 U. ILL. L. REV. (2008). The author would like to thank the participants of the CICL Lecture for very helpful comments, especially Larry Helfer, Jonathan Wiener, and Rachel Brewster. The author would also like to thank Anthea Roberts, Jean Galbraith, Guillermo Otálora Lozano, and Tomer Broude for very helpful suggestions; Ariel Steffen, Tobias Lehmann, Luis Montilla, and Nora Dushica for editorial help at my chair; and the editors of HILJ for excellent editorial work and very helpful suggestions.

1. Just after my talk, I became aware of the SSRN Working Paper by Tomer Broude, *Behavioral International Law* (Hebrew U. of Jerusalem Int'l L. F., Working Paper No. 12-13, 2013), *available at* http://ssrn.com/abstract=2320375, forthcoming in the University of Pennsylvania Law Review. I read his paper after finishing mine, and include references to his paper here. We essentially share the main methodological thoughts and caveats but also have differences to which I allude in the paper. We are currently working together on a book, BEHAVIORAL ECONOMICS AND INTERNATIONAL LAW, forthcoming from Oxford University Press.

law and economics approach to public international law. With a view to the "unity of knowledge,"[2] it seems time to draw on those different insights to further refine our understanding of international law.

The rational choice paradigm as used in economics has been thoroughly challenged since the 1970s by psychological experimental research and has considerably changed huge parts of economics.[3] Meanwhile, economic analysis of law took off in the United States, continuing to use the rational choice paradigm.[4] It was not until the late 1990s that behavioral economics reached economic analysis of *national* law, mainly in the United States[5]; it took longer in Europe.[6] At around the same time, economic analysis started to be applied to *international* law in the United States,[7] but this has, with very few exceptions,[8] been hitherto confined to rational choice. This is even

2. Edward O. Wilson, Consilience: The Unity of Knowledge (1998). *See also* Herbert Gintis, The Bounds of Reasons: Game Theory and the Unification of the Behavioral Science (2009).

3. Colin F. Camerer, *Behavioral Economics: Reunifying Psychology and Economics*, 96 Proc. Nat'l Acad. Sci. USA 10575 (1999); Matthew Rabin, *Psychology and Economics*, 36 J. Econ. Lit. 11 (1998).

4. This is true no matter whether the Chicago School or the Yale School is meant. *See* Richard A. Posner, Economic Analysis of Law (2d ed. 1977); Guido Calabresi, The Costs of Accidents: A Legal and Economic Analysis (1970); Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 Harv. L. Rev. 1089 (1972). This paradigm comes mainly in two forms: either for individual decision-makers in decision theory or in game theory as a theory of strategic interaction. Both are relevant for international law and both have been transformed by behavioral economics.

5. Christine Jolls et al., *A Behavioral Approach to Law and Economics*, 50 Stan. L. Rev. 1471 (1998); *see also* Christine Jolls & Cass R. Sunstein, *Debiasing Through Law*, 35 J. Legal Stud. 199 (2006); Cass R. Sunstein, Behavioral Law and Economics (2000); Richard H. Thaler & Cass R. Sunstein, Nudge: Improving Decisions about Health, Wealth, and Happiness (2008).

6. Anne van Aaken, Rational-Choice in der Rechtswissenschaft: Zum Stellenwert der Ökonomischen Theorie im Recht (2003); Recht und Verhalten: Beiträge zu Behavioral Law and Economics (Christoph Engel et al. eds., 2007).

7. Kenneth Abbott, *Modern International Relations Theory: A Prospectus for International Lawyers*, 14 Yale J. Int'l L. 335, 348–54 (1989) (discussing methodology in international relations theory that is relevant for this paper); Jeffrey L. Dunoff & Joel P. Trachtman, *Economic Analysis of International Law*, 24 Yale J. Int'l L. 1 (1999); Jack L. Goldsmith & Eric A. Posner, The Limits of International Law (2005); Anne van Aaken, *To Do Away with International Law? Some Limits to 'The Limits of International Law'* 17 Eur. J. Int'l L. 289 (2006) [hereinafter van Aaken, *To Do Away with International Law?*]; Robert E. Scott & Paul Stephan, The Limits of Leviathan: Contract Theory and the Enforcement of International Law (2006); Andrew Guzman, How International Law Works: A Rational Choice Theory (2008); Joel P. Trachtman, The Economic Structure of International Law (2008); *and* Eric A. Posner & Alan O. Sykes, Economic Foundations of International Law (2013). For an overview, see Anne van Aaken, *International Law: Rational Choice*, Oxford Bibliographies Online (2012), http://www.oxfordbibliographies.com/view/document/obo-9780199796953/obo-9780199796953-0051.xml?rskey=X6w0Io&result=21&q=.

8. Van Aaken, *Towards Behavioral International Law and Economics?*, *supra* note *; Jean Galbraith, *Treaty Options: Towards a Behavioral Understanding of Treaty Design*, 53 Va. J. Int'l L. 309 (2013); Lauge N. Poulsen & Emma Aisbett, *When the Claim Hits: Bilateral Investment Treaties and Bounded Rational Learning*, 65 World Pol. 273 (2013); Lauge N. Poulsen, *Bounded Rationality and the Diffusion of Modern Investment Treaties*, Int. Stud. Q. 1 (2013) [hereinafter Poulsen, *Bounded Rationality*]; Lauge N. Poulsen, Letting Down the Guard: Developing Countries, Investment Treaties, and Bounded Rationality [hereinafter Poulsen, Letting Down the Guard] (forthcoming); Emilie M. Hafner-Burton et al., *Decision Makers Preferences for International Legal Cooperation* 1 (Laboratory on Int'l L. & Reg., Working Paper No. 13, 2013), *available at* http://ilar.ucsd.edu/assets/001/504387.pdf (using experiments with students and confirming the results with a smaller group of real decision-makers and holding that

more puzzling since international law and IR scholars have increasingly cooperated[9] and, in IR theory, the so-called political psychology approach has a long tradition.[10]

In consequence, there is no systematic analysis of international law using behavioral economics, that is to say, there is no behavioral international law and economics (BIntLE). The aim of this article is to take a first step in sketching its methodological foundations, by noting the challenges and giving some examples where BIntLE could be fruitfully applied. In spite of the methodological challenges, the proof of the pudding is in the eating: the success of this approach will be measured by its applications and usefulness for the design and interpretation of international law—and last but not least, in the empirical work validating (or not) the hypotheses generated by either international law and economics or BIntLE.[11]

I proceed as follows. First, I will provide a short overview of economic analysis of international law, behavioral economics, and political psychology in international relations. Second, I will highlight some potential challenges to the approach with a view on the relevant unit of analysis and propose

"under some conditions certain types of university student convenience samples can be useful for revealing elite-dominated policy preferences"). *See also* Broude, *supra* note 1. For the human rights sphere, the approach has been used for longer. *See* Andrew K. Woods, *A Behavioral Approach to Human Rights*, 51 Harv. Int'l L.J. 51 (2010); Ryan Goodman et al., *Social Science and Human Rights*, *in* Understanding Social Action, Promoting Human Rights 1 (Ryan Goodman et al. eds., 2012).

9. *See* Anne-Marie Slaughter, *International Law and International Relations*, 285 Collected Courses of the Hague Acad. of Int'l L. 9 (2000/2001); Jeffrey L. Dunhoff & Mark A. Pollack, *Reviewing Two Decades of IL/IR Scholarship: What We've Learned, What's Next*, *in* Interdisciplinary Perspectives on International Law and International Relations: The State of the Art 626 (Jeffrey L. Dunhoff & Mark A. Pollack eds., 2013).

10. For an early example, see Robert Jervis, Perception and Misperception in International Politics (1976). For overviews, see James Goldgeier & Philip Tetlock, *Psychological Approaches*, *in* Oxford Handbook of International Relations 462 (Christian Reus-Smit & Duncan Snidal eds., 2008); Janice Gross Stein, *Psychological Explanantions of International Conflict*, *in* Handbook of International Relations 292 (Walter Carlsnaes et al. eds., 2002); Barbara Farnham, Avoiding Losses, Taking Risks: Prospect Theory and International Conflict (1994); Rose McDermott, Political Psychology in International Relations (2004).

11. For an overview of the increased empirical work in international law, see Greg Shaffer & Tom Ginsburg, *The Empirical Turn in International Legal Scholarship*, 106 Am. J. Int'l. L. 1, 1 (2012) ("What matters now is the study of the conditions under which international law is formed and has effects."). Now, there are even further calls for experiments. *See* Adam S. Chilton & Dustin H. Tingley, *Why the Study of International Law Needs Experiments*, Colum. J. Transnat'l L. (forthcoming), *available at* http://ssrn.com/abstract=2343471 (2013). For a survey on IR and experiments, see Alex Mintz et al., *Experimental Approaches to International Relations*, 55 Int. Stud. Q. 493 (2011); Rose McDermott, *New Directions for Experimental Work in International Relations*, 55 Int. Stud. Q. 503 (2011). McDermott also formulates the traditional objections to the use of experiments in IR at 504: "the major concern raised about the use of experiments in political science regards perceived problems related to achieving acceptable levels of external validity and generalizability . . . . While many international relations scholars find laboratory experiments intriguing or instructive, they may not readily see how controlled laboratory studies translate into a more comprehensive understanding of the complex political and social world they seek to explain." Thus, "Realists do not use experiments, nor do constructivists, and yet both have theoretical roots in disciplines, economics and sociology respectively, which rely heavily on experimentation to test central hypotheses. Liberalism, which arguably should find experiments most sympathetic because of their utility in manipulating fine grained aspects of trade and financial policy, has rejected the method as well."

methodological solutions. Third, the potential application of BIntLE in law-making (international law design), including choice of sources and other examples, as well as in compliance will be illustrated. The last part concludes.

I. Three Pillars of Behavioral International Law and Economics

BIntLE draws on three main pillars, all of which hitherto have stood unconnected to each other. BIntLE aims to draw on them and bring them together in order to generate insights for international law. In the following, (1) economic analysis of international law, (2) behavioral economics, and (3) political psychology in IR will briefly be described.

A. Economic Analysis of International Law

When applied to law, including international law, rational choice (RC) analysis mainly comes under the headings of "law and economics," "institutional economics," "constitutional economics," "political economy," or "economic analysis of law." Those fields are unified by their common methodology; they differ slightly in their focus. International law has been a rather late area of analysis, but research is currently accelerating. Analyzing international law through the RC perspective has become a joint enterprise by economists, international lawyers, and rational-choice political scientists, focusing on more precise questions of international law scholarship intended to inform doctrinal scholarship as well.[12] Although RC analysis has a long tradition in IR scholarship—usually with a focus on international cooperation—detailed analysis of international law (including soft law) has only more recently become more prominent, taking up questions of relevance for treaty design as well as doctrinal development.

The rationality assumption has long been the basis of positive as well as normative economics. Positive (neo)classical economics assumed rational actors, although never denying that there might—occasionally—be lapses from rationality. However, these were deemed not to matter: traditional economics is concerned with average behavior. Since positive economics is also about prediction of future behavior, it was argued that what matters most is the accuracy of the prediction of the models, not the reality of its assumptions.[13] Furthermore, normative economics (that is, welfare economics) did retain the rationality assumption, since everything else could endanger the

12. *See supra* note 7.

13. *See* Milton Friedman, *The Methodology of Positive Economics*, *in* Essays in Positive Economics 3 (M. Friedman ed., 1953). He argues for this because of simplicity (omitting all irrelevant details) in being able to predict at least as much as an alternative theory, as well as requiring less information, *id.* at 13, 14, and fruitfulness in the precision and scope of its predictions and in its ability to generate additional research lines, *id.* at 10. He is surely right that complexity needs to be reduced for any research—if

autonomy of individuals, permitting paternalistic state action. Those reasons are not easily dismissed, and a lot of (international) lawyers implicitly assume rationality of actors, as well: for example when they talk about casting national interests into treaty making[14] or about the effects of state responsibility.

A classical definition of rationality as used by economists is given by Gary Becker: "[A]ll human behavior can be viewed as involving participants who [1] maximize their utility [2] from a stable set of preferences and [3] accumulate an optimal amount of information and other inputs in a variety of markets."[15] The central tenets are utility maximization, stable preferences, rational expectations, and optimal processing of information. Those assumptions were transferred without much ado to collective actors such as firms, states, or international organizations (IOs). It is thus either assumed that potential biases cancel each other out on the aggregate level or do not even occur within corporate actors, although there has been early critique of this handling of rationality: "the principle of rationality, unless accompanied by extensive empirical research to identify the correct auxiliary assumptions, has little power to make valid predictions about political phenomena."[16]

Following the traditional international law assumptions that have prevailed ever since the Westphalian Peace, the nation-state has mostly been analyzed as a unitary actor, or what has been described as a "black box" state.[17] RC analysis has been used to diagnose substantive problems and frame better legal solutions, explain the structure or function of particular international legal rules or institutions, and re-conceptualize or reframe particular institutions or international law generally, such as customary international law.[18] Many tools of economic analysis are used in the endeavor: game theory, contract theory, price theory, principal-agent theory, and collective action theory (such as public choice or political economy analysis), as well as the notions of externalities and common goods.[19] It would exceed the scope of this article to describe all previous research in detail; instead, some of the research will be incorporated in the part on BIntLE when suggesting how it can enrich (or contradict) the RC approach in international law.

---

one wants to see everything, one sees nothing. But the argument here is that one sees more when using behavioral economics and that the higher complexity is therefore justified—up to a point.

14. *See* Galbraith, *supra* note 8, at 310.

15. Gary S. Becker, The Economic Approach to Human Behavior 14 (1976).

16. Herbert A. Simon, *Human Nature in Politics: The Dialogue of Psychology with Political Science*, 79 Am. Pol. Sci. Rev. 293, 293 (1985).

17. John J. Mearsheimer, *Structural Realism*, *in* International Relations Theories: Discipline and Diversity 71, 72 (Tim Dunne et al. eds., 2006).

18. *See* George Norman & Joel P. Trachtman, *The Customary International Law Game*, 99 Am. J. Int'l. L. 541 (2005); Edward T. Swaine, *Rational Custom*, 52 Duke L.J. 559 (2002).

19. *See* Trachtman, *supra* note 7.

B. Behavioral Economics (Psychology)

The rationality assumption has been called into question by cognitive psychologists and behavioral economists. Herbert Simon early criticized the "economic man" who is omniscient and maximizing, deploring the distance between the insights of psychology and the assumptions used by economists, developing instead what he called a behavioral model of rational choice.[20] Psychologists (and economists) have taken up his criticism, most prominently among them Daniel Kahneman, Amos Tversky[21] and Gerd Gigerenzer,[22] exploring systematic heuristics and biases running counter to the rationality assumption, searching for a more realistic model of human behavior. Whereas Kahneman and Tversky tend to interpret deviations from the rationality assumption as a normative problem (deviation from rationality as elaborated on in normative decision theory is seen as a costly human error that must be corrected), Gigerenzer proposes that a model of inferences (which uses simple psychological mechanisms, such as clues) might make human beings "smarter."[23] This debate must inform legal design: a "rational design" should strive to provide norms that lead to "better" decision-making in regard to the assumed preferences of the relevant actors. For states, this should—as a principle—arguably be the normative idea of rationality since only this assumption generates stable expectations in the international realm and avoids potentially costly mistakes (for example, in security constellations). De-biasing, at least in the cognitive sphere, should thus be an aim of international law in principle. But there are also deviations from the RC assumptions, which international legal design may want to use (for example, fairness preferences).

Some of the heuristics and biases analyzed in behavioral economics that are potentially relevant to international law will be described below. A common categorization of biases following Jolls et al. is to separate bounded

20. Herbert A. Simon, *A Behavioral Model of Rational Choice*, 69 Q.J. Econ. 99 (1955); Simon, *supra* note 16 (comparing two theories of human rationality that have found application in political science: procedural, bounded rationality from contemporary cognitive psychology, and global, substantive rationality from economics. The word "satisficing" is a conjunction of the words satisfy and suffice and means a strategy or cognitive heuristics that entails searching through the available alternatives until an acceptability threshold is met (instead of using the best available alternative by optimizing)). For international law, see Abram Chayes & Antonia Handler Chayes, *On Compliance*, 47 Int'l Org. 175, 179 (1993) (who equally assume that states are satisficing rather than optimizing (or maximizing)).

21. Starting with prospect theory and further extending in many other "biases." *See* Daniel Kahneman & Amos Tversky, *Prospect Theory: An Analysis of Decisions Under Risk*, 47 Econometrica 263 (1979) [hereinafter Kahneman & Tversky, *Prospect Theory*]; Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 Science 1124 (1974) [hereinafter Tversky & Kahneman, *Judgment Under Uncertainty*].

22. Gerd Gigerenzer & Peter M. Todd, Simple Heuristics that Make Us Smart (1999); Gerd Gigerenzer & Daniel G. Goldstein, *Reasoning the Fast and Frugal Way: Models of Bounded Rationality*, 103 Psychol. Rev. 650 (1996); Gerd Gigerenzer & Reinhard Selten, Bounded Rationality: The Adaptive Toolbox (2002).

23. For an excellent discussion on those schools from a normative angle, see Mark Kelman, The Heuristics Debate 19–118 (2011).

rationality, bounded willpower, and bounded self-interest.[24] We adopt these categories with a view to actors in the international sphere (problems this approach poses will be discussed in detail in Part II).

1. *Bounded Rationality (Cognition)*

The standard models of RC theory, seen as a model of instrumental rationality, have minimal assumptions about human cognition such as completeness, transitivity, and independence of irrelevant alternatives.[25] These assumptions do not reveal anything about the content of preferences; they rather form the basis for a formal concept of (thin) rationality. We will deal with the content of preferences under Part I.B.3 below and concentrate on cognitive biases in this part.

In standard RC models of choices under uncertainty (the "expected utility" concept),[26] it is assumed that preferences are independent from the circumstances in which they are revealed, which implies that the description and format of alternative decisions have no influence over the decision itself: actors should always choose the same alternative, no matter how (identical) data is presented. However, based on many experiments, prospect theory revealed that this is incorrect. It describes the way people choose between probabilistic alternatives involving risk. Individuals are loss averse, that is, they have an asymmetrical attitude towards gains and losses (loss aversion). Their utility is less increased by gains than by averted losses. In other words, the utility from a 100 USD gain is less than the averted loss of 100 USD. This differs from risk aversion, which can be entirely rational, as long as it stays consistent on every level.

What a gain or a loss is depends on the so-called "framing effect" or "wording effect" that contradicts the axiom of independence (that is, that preferences remain stable and are independent of presentation or choice of words). The "framing effect" or "wording effect" states that logically equivalent presentations of a circumstance lead individuals to different choices. Typically, decisions vary depending on how circumstances are presented, either as positive or negative; this is valid for decisions made under uncertainty as well as certainty.[27] Decisions about medical intervention are a typical example. The standard model would predict that patients

24. Jolls et al., *supra* note 5, at 1476.

25. For a comprehensive treatment, see SHAUN H. HEAP ET AL., THE THEORY OF CHOICE: A CRITICAL GUIDE 4 (3d ed. 1997); GINTIS, THE BOUNDS OF REASONS, *supra* note 2, at 4. Independence of irrelevant alternatives means that the relative attractiveness of two choices does not depend upon the other choices available to the individual.

26. *See* BECKER, *supra* note 15; GARY S. BECKER, ACCOUNTING FOR TASTES 5–7 (1996). In the expected utility theory, expected utility of an option X is comprised of the sum of all outcomes x weighted by their probability p. The expected utility hypothesis is first and foremost a normative model. However, economists often use it as an assumption in descriptive models.

27. *See* Kahneman & Tversky, *Prospect Theory*, *supra* note 21; Amos Tversky & Daniel Kahneman, *The Framing of Decisions and the Psychology of Choice*, 211 SCIENCE 453, 455 (1981); Daniel Kahneman, *A Perspective on Judgment and Choice: Mapping Bounded Rationality*, 58 AM. PSYCHOL. 697, 703 (2003). *See also*

would choose the most secure therapeutic method independent of how the choice is presented to them (as death rates or survival rates). Death rates can theoretically function as lost profits or damages. If a relatively safe therapeutic method is presented to a patient in terms of death rate (that is, potential loss) and an unsafe method is presented in terms of survival rate (that is, potential profit), patients will prefer the unsafe over the safe method.[28] To sum up, the description of a problem can lead to a demonstrably false decision. Interestingly, the "framing effect" affects not only patients and their relatives, but also medical staff (that is, experts).[29] In short, informative statements are never neutral. The "framing effect" leads the individual to focus on certain characteristics of a problem (whilst neglecting others) and points toward certain decisions (and not to others). Thus, a decision can be considerably influenced by external elements, including law, depending on how a problem is presented.[30] This has consequences for the rational design of a legal system if it aims at debiasing.

The "framing effect" is also connected with the so-called "reference point." The choice and determination of a "reference point" will influence individual decisions. The difference to the classical expected utility theory is that profits and losses are not perceived in absolute terms, but in relation to a reference point.[31] This is relevant in regard to the circumstances or events that constitute a reference point as well as in regard to the shifting of a reference point. Individuals tend to react more strongly to changes in circumstances than to changes in absolute values (that is, shifting of a reference point). A temperature change will be felt in a different way depending on whether a person is coming from a cold or a warm room. This is not only a physical phenomenon, but also a psychological one. Primary utility drivers are therefore events, not conditions, in a dynamic process.[32] From a legal perspective, it is interesting to understand how the reference point is determined, because individuals significantly base their decisions on reference points. Usually, the reference point is the status quo ("status quo bias") or the current level of wealth. However, reference points can also be set artifi-

---

James N. Druckmann, *Political Preference Formation: Competition, Deliberation and the (Ir)relevance of Framing Effects*, 98 Am. Pol. Sci. Rev. 671 (2004) for a description of experiments and further references.

28. This outcome is compatible with the "prospect theory". *See* Rabin, *supra* note 3, at 36 (including descriptions of experiments and further references); Amos Tversky & Daniel Kahnemann, *Rational Choice and the Framing of Decisions*, *in* Rational Choice (Robin M. Hogarth & Melvin W. Reder eds., 1987) (experiments with medical students).

29. *See* Druckmann, *supra* note 27. This has been proven also for other experts as in accounting or civil engineering. Chip Heath et al., *Cognitive Repairs: How Organizational Practices Can Compensate For Individual Shortcomings*, 20 Research Org. Behav. 1, 4 (1998) ("experts are often wrong but rarely in doubt").

30. This is related to the phenomenon of the perception and recognition of patterns. When individuals make decisions, they follow patterns that the situation reveals. For the "right" decision, the choice of the "right" pattern is important. When the correct pattern is chosen, individuals are less or not at all subject to biases. *See* Gebhard Kirchgässner, Homo Oeconomicus (Siebeck/Mohr. 1991).

31. Kahneman & Tversky, *Prospect Theory*, *supra* note 21.

32. This is a central hypothesis of the "prospect theory."

cially by contracts,[33] judicial decisions or laws that allocate entitlements and legal claims or the expectations thereof ("aspirational level"). They are thus relevant for treaty negotiations as well as for contracting and performance of a treaty—especially interesting topics for international law, with which we will deal in Part III.

The "endowment effect" also plays a role in the determination of a reference point.[34] This effect leads individuals to value goods differently, depending on whether they possess these goods or not: goods already in their possession will, *ceteris paribus*, have more value than goods that are not yet in their possession (that is, the price to be paid for a good already in an individual's possession is considerably higher than the price the individual is ready to pay to acquire the good).[35] In other words, people who have a good at their disposal express a higher willingness to pay for the good (this is also true for situations in which the individual does not have full right to property in respect of the good in question). The longer a person owns a good, the stronger the effect. Thus, the "endowment effect" directly affects preferences. It may inhibit efficient contracting and consensus of states, which is again an issue of relevance to international law.[36]

Individuals procrastinate and adhere to the status quo. Thus, default options are crucial since people tend not to deviate from them. Depending on how the default is set ("choice architecture"[37]), different choices are made. Organ donation, for example, can be set up as an opt-in (no donation as default) or an opt-out rule (donation as default). As has been empirically observed in cross-country studies (for example, for organ donations), a no-donation default leads to many fewer individuals opting in, in comparison with an opt-out design.[38] This same phenomenon holds for many other instances as well.[39]

Furthermore, individuals show an insufficient capacity to make objective assessments of reality. A systemically distorted understanding of information contributes to a distorted perception of reality. These phenomena include the "anchoring effect": when evaluating information relevant for decision-making, individuals employ information obtained previously even

33. Ernst Fehr et al., *Contracts as Reference Points—Experimental Evidence*, 101 Am. Econ. Rev. 493, 493 (2011).

34. For more details and a summary of the experiments, see Daniel Kahneman et al., *Experimental Tests of the Endowment Effect and the Coase Theorem*, 98 J. Pol. Econ. 1325 (1990).

35. *Id.* at 1329–42.

36. For an RC perspective on the problem of consensus finding in international law, see Andrew T. Guzman, *Against Consent*, 52 Va. J. Int'l L. 747 (2012) (arguing that existing commitment to consent is excessive and that better outcomes would result from greater use of nonconsensual forms of international law).

37. Thaler & Sunstein, *supra* note 5, at 3.

38. Eric J. Johnson & Daniel Goldstein, *Do Defaults Save Lives?*, 302 Science 1338, 1338 (2003).

39. For a comprehensive discussion, see Cass Sunstein, *Impersonal Default Rules vs. Active Choices vs. Personalized Default Rules: A Triptych* (2013), http://ssrn.com/abstract=2171343.

if irrelevant for the decision at hand.[40] The reference point (that is, the information obtained previously) serves as the anchor and can be influenced by the initial representation of the information: different starting points generate different evaluations, all of which are in one way or another biased or distorted.[41] This gives rise to situations in which individuals will neither notice nor pay attention to any new information contradicting the first information they received. Moreover, individuals are prone to evaluating new information incorrectly or interpreting it in such a way that it supports their mental representation of the first information.[42] Thus, the time of registering the information influences the overall effect of the information.

Rational individuals are supposed to maximize their *expected* utility,[43] that is, they should calculate their subjective expected utility using the rules of probability calculus.[44] This assumes also that individuals can adequately evaluate risks relevant for decision-making. However, mistakes are frequent. These are relevant to the legal system, where individuals have to deal with uncertain events. The evaluation of very low probabilities is frequently irrational:[45] they are systematically neglected, especially when individuals deal with risks related to losses.[46] Moreover, a stronger weight is attached to the utility of certain events than to the utility of uncertain events with very high probabilities (that is, events with the same expected utility ["security effect"]). Furthermore, there is "ambiguity aversion": individuals are especially risk-averse when probabilities are not clearly defined.

Also, individuals put too much emphasis on obvious features or events when estimating probabilities ("availability bias"). In doing so, they neglect the base rate probability—simply because a specific event is still fresh in their memory or because it is still perceptible. This explains, for instance, why people will systemically overestimate the probability of terrorist attacks after being significantly exposed to such an event by the media. A similar explanation can be given to the systemic overestimation of the risks of accidents or environmental catastrophes. At the same time, risks that have a high probability but are less salient are underestimated. Plane accidents and car accidents are a case in point, where the risk of the former is overesti-

40. Edward J. Joyce & Gary C. Biddle, *Anchoring and Adjustment in Probabilistic Inference in Auditing*, 19 J. Acct. Research 120, 125 (1981) (observing this "bias" in auditors). In the experiment one group was asked whether they thought fraud by management was committed in more than ten companies out of a thousand. Another group was asked whether they thought fraud by management was committed in more than two hundred companies out of a thousand. When the groups were asked to give an estimate, the "two hundred" group gave significantly higher numbers than the "ten" group.

41. Tversky & Kahneman, *Judgment under Uncertainty*, *supra* note 21, at 1128.

42. Rabin, *supra* note 3, at 26, 29, with further references.

43. *See* Heap et al., *supra* note 25. *See also* John v. Neumann & Oscar Morgenstern, Theory of Games and Economic Behavior (2d ed. 1947).

44. *See* Heap et al., *supra* note 25, at 349.

45. *See* Rabin, *supra* note 3, at 24 (on the law of small numbers).

46. *Id.*

mated (being salient through media reports) while the risk of the latter is underestimated.

Furthermore, many empirical studies show that people are on average overconfident about their future and about their predictions for their future ("excess of optimism").[47] This relates not only to their own situation and their capacities, but also to their evaluation of the control they have over a given situation ("illusion of control"). This phenomenon is evident in the context of car driving, but over-optimism is also relevant in the context of drug addiction.[48] Also, when it comes to judicial disputes, it is rather common that parties overestimate their chances to win in court or for commanders to win a war.

2. *Bounded Willpower*

The problem of self-control (that is, inconsistent behavior over time) is not a bias in the strict sense according to the formal RC model, which evaluates the rationality of a decision at a given moment in time. Inconsistent behavior over time is well-documented by experiments. The standard model of RC theory assumes that human behavior is constituted by *one* utility function. Costs and utility are discounted in the future by a reasonable and constant rate. This means that the difference between the utility of a reward (or the costs of a penalty) when comparing today and tomorrow is *pro rata temporis* the same as between today and tomorrow in one year from now. Similarly, it is assumed that the rate of discount is constant and independent from the way in which the problem is posed. Experiments have contradicted this idea,[49] showing that people's discount rates sharply fall over time. This so-called "hyperbolic discounting" leads to inconsistent behavior over time—an expression of weakness of will. Classical examples are drug consumption or overeating (short-term gain equals long-term pain).

Bounded willpower cannot be assumed to be characteristic of corporate actors[50] per se. However, functional equivalents may be present, since these

47. *See generally* Neil D. Weinstein, *Unrealistic Optimism about Future Life Events*, 39 J. PERS. SOC. PSYCHOL. 806 (1980). For more detail on studies of excess optimism, see Marie Helweg-Larsen & James A. Shepperd, *Do Moderators of the Optimistic Bias Affect Personal or Target Risk Estimates? A Review of the Literature*, 5 PERS. SOC. PSYCHOL. REV. 74 (2001).

48. *See* B. Douglas Bernheim & Antonio Rangel, *Behavioral Public Economics: Welfare and Policy Analysis with Non-Standard Decision Makers*, 52 (Nat'l Bureau of Econ. Research, Working Paper No. W11518, 2005). The authors mention a survey of high-school seniors who smoked cigarettes. Fifty-six percent of students asked claimed that they would not be smoking in five years. Only thirty-one percent did in fact quit smoking in five years.

49. *See* Ted O'Donoghue & Matthew Rabin, *Doing It Now or Later*, 89 AM. ECON. REV. 103 (1999), with further references and on the different terminology for this phenomenon. *See also* George Ainslie, *Précis of Breakdown of Will*, 28 BEHAV. BRAIN SCI. 635 (2005).

50. Corporate actors can be defined as organizations endowed with substantial autonomy from the ultimate beneficiaries of their actions that are able to form a common will within a hierarchy (that is, not only firms, but also IOs, states or NGOs), whereas collective actors generate behavior by a mechanism that directly aggregates the will of these beneficiaries (meaning that there is not much agency slack). *Cf.* FRITZ W. SCHARPF, GAMES REAL ACTORS PLAY: ACTOR-CENTERED INSTITUTIONALISM IN POLICY RE-

may also succumb to time-inconsistency problems. Instead of focusing on classical nudges to circumvent temptations on an individual level,[51] we have to focus on political economy mechanisms to self-bind governments not to abuse their power for their own gains, such as a re-election, thereby following short-term policies that are damaging in the long run (short-termism of politics). The most well-known mechanism on the municipal level is delegation: for example, delegation to independent central banks so that politicians are unable to use inflationary monetary policy for their own (political) benefit.[52] But delegation does not only take place toward national independent institutions, such as courts and central banks, but also to international institutions and organizations, such as international human rights courts.[53] Delegation to the international sphere is then a means to make commitments more credible, since it is less changeable and might have more reputational effects. Although this is of utmost importance to international law, it will not be pursued further here, since an analysis is possible by drawing on RC theory only. Nevertheless, it should be kept in mind that those mechanisms may also play a role in IR.[54]

3. *Bounded Self-Interest*

Many economists often use models of rationality that include assumptions about the substance of preferences (thick rationality). Despite the concept of revealed preferences, it is assumed that all individuals share certain preferences: for example, a preference for survival and a preference for money. For states, military or economic power is usually assumed. In any case, the standard economic models assume routinely that self-regarding interest (often material self-interest) is the sole motivation of all actors. But plenty of experimental research has shown that individuals are strongly motivated by other-regarding/altruistic and social preferences.[55] Concerns for the well-being of others (for fairness and reciprocity) need to be taken into account if

---

SEARCH 54 (1997); Christoph Engel, *The Behaviour of Corporate Actors: How Much Can We Learn from the Experimental Literature?*, 6 J. INST. ECON. 445, 447 (2010).

51. THALER & SUNSTEIN, *supra* note 5.

52. Marc Quintyn, *Independent Agencies—More than a Cheap Copy of Independent Central Banks?*, 20 CONST. POLIT. ECON. 267 (2009); Marc Quintyn & Michael W. Taylor, *Regulatory and Supervisory Independence and Financial Stability*, 49 CESIFO ECON. STUD. 259 (2003).

53. Stefan Voigt et al., *Improving Credibility by Delegating Judicial Competence—The Case of the Judicial Committee of the Privy Council*, 82 J. DEV. ECON. 348 (2007); Stefan Voigt & Eli Salzberger, *Choosing Not to Choose: When Politicians Choose to Delegate Powers*, 55 KYKLOS 247 (2002); *see, e.g.*, Anne van Aaken & Richard Chambers, *Accountability and Independence of International Election Observers*, 6 INT'L ORG. L. REV. 541 (2009) (for international election observers which are used in the most delicate moment of temptation for politicians).

54. *See infra*, Part II (on the potential transmission mechanisms between electorate and politicians).

55. This started with the so-called Ultimatum Game. *See* Werner Güth et al., *An Experimental Analysis of Ultimatium Bargaining*, 3 J. ECON. BEHAV. ORG. 367 (1982). Other games include: the Dictator Game, the Power to Take Game, the Third Party Punishment Game, the Gift Exchange Game and the Trust Game. For details on the experiments, see Ernst Fehr & Klaus M. Schmidt, *The Economics of Fairness, Reciprocity and Altruism—Experimental Evidence and New Theories*, *in* I HANDBOOK OF THE ECONOMICS OF GIVING, ALTRUISM AND RECIPROCITY 621 (Serge Kolm & Jean Mercier Ythier eds., 2006).

behavior in social interactions is to be understood.[56] "[T]he real question is no longer whether many people have other-regarding preferences, but under which conditions these preferences have important economic and social effects."[57] Those preferences are especially important for explaining collective action and multilateral cooperation that are central to international law.[58] Furthermore, individuals may punish free riders even if costly for themselves[59]; this is again highly important for international law, which has only a decentralized enforcement system and thus needs "punishers" (states or nonstate actors). Having other regarding preferences or not maximizing earnings does not mean that a person is not capable of strategic thinking. Rather, behavioral insights modify the preference function.[60]

Altruism is a form of unconditional kindness, that is, a favor given does not emerge as a response to a favor received.[61] Whereas an altruist is willing to sacrifice his or her own resources in order to improve the well-being of others, a spiteful or envious person is also willing to do this in order to punish others. Most individuals are conditionally altruistic or spiteful. A known conditional form of altruism and/or envy is the so-called inequity aversion,[62] a situation in which, in addition to the actor's material self-interest, her utility increases if the allocation of material payoffs becomes more equitable. "Equitable" is, of course, an indeterminate notion and usually depends on the reference point; as we have seen above, this is mostly the status quo.

Furthermore, reciprocity has been recognized by social scientists[63] as one of the main basic social relations that constitute societies and IR.[64] Also, international lawyers have an intuitive affinity for the concept.[65] In its most extensive definition, which is usually used in this literature, it consists of being favorable to others because others are favorable to you (it is thus RC

56. Fehr & Schmidt, *supra* note 55.

57. *Id.* at 617.

58. For an application in human rights law, see Herbert Gintis, *Human Rights: An Evolutionary and Behavioral Perspective*, *in* Understanding Social Action, Promoting Human Rights, *supra* note 8, at 135.

59. For a discussion of the motive to retaliate to punish, see Armin Falk et al., *Driving Forces behind Informal Sanctions*, 73 Econometrica 2017 (2005).

60. *See* Colin F. Camerer, Behavioral Game Theory 11 (2003).

61. Jon Elster, *Altruistic Behavior and Altruistic Motivations*, *in* Handbook of the Economics of Giving, Altruism and Reciprocity, *supra* note 55, at 184; Gary Charness & Matthew Rabin, *Understanding Social Preferences with Simple Tests*, 117 Q.J. Econ. 817 (2002).

62. Gary E. Bolton & Axel Ockenfels, *A Theory of Equity, Reciprocity and Competition*, 100 Am. Econ. Rev. 166 (2000); Fehr & Schmidt, *supra* note 55; Charness & Rabin, *supra* note 60.

63. Serge Kolm, *Reciprocity: Its Scope, Rationales, and Consequences*, *in* Handbook of the Economics of Giving, Altruism and Reciprocity, *supra* note 55, at 371.

64. Beth V. Yarbrough & Robert M. Yarbrough, *Reciprocity, Bilateralism, and Economic "Hostages": Self-Enforcing Agreements in International Trade*, 30 Int. Stud. Q. 7 (1986); Francesco Parisi & Nita Ghei, *The Role of Reciprocity in International Law*, 36 Cornell Int'l L.J. 93 (2003); Guzman, *supra* note 7.

65. Bruno Simma, Das Reziprozitätselement im Zustandekommen völkerrechtlicher Verträge. Gedanken zu einem Bauprinzip der internationalen Rechtsbeziehungen (1972); Bruno Simma, Das Reziprozitätselement in der Entstehung des Völkergewohnheitsrechts (1970).

based). Behavioral economics takes a more differentiated look at the reciprocity principle. It identifies three possible rationales[66]: (1) self-sustaining sequences of mutual favors that can be solely self-interested (as, for example, in a synallagmatic contract that can also be explained by RC); (2) balance (which assumes acts of comparison and matching), often related to equality and fairness concerns (reciprocity *strictu sensu*); (3) liking, because being favored induces liking, which induces favoring (a sort of Facebook effect), or because liking can directly result from being liked. The first rationale is the one used in law and economics where "cooperative" or "retaliatory" behavior in repeated interactions is motivated by future benefits. Behavioral economics concentrates on the second and third rationale, and thus reciprocal behavior in behavioral economics differs fundamentally from the RC approach. Reciprocal behavior in one-shot interactions is often called "strong reciprocity" (since RC would assume defection) in contrast to "weak reciprocity" that is motivated by long-term self-interest in repeated interactions (and in conformity with RC).[67]

Although all three forms of reciprocity will be used in Part III, we will concentrate on the latter two here. People are reciprocal if they reward kind actions and punish unkind ones, that is, a reciprocal individual responds to actions he or she perceives to be kind in a kind manner, and to actions he or she perceives to be hostile in a hostile manner.[68] Preferences do not only depend on material payoffs but also crucially on intentions (that is, on beliefs about why an agent has chosen a certain action). People evaluate the kindness of an action not only by its consequences but also by its underlying intention. Here, trust (or reputation) about the intention can play a role.[69] Traditional game theory does not capture the relevance of beliefs about intentions for behavior, since it assumes that outcomes (and not beliefs) determine payoffs, but ever more experiments show how important the perceived intentions of other actors are for interaction.[70] Behavioral game theory attempts to capture those insights without giving up the strategic interaction basis.[71] If actors are perceived as moral and legitimate, cooperation is fostered; if actors are deemed unfair, cooperation is undermined. This has been well-documented in experiments.[72] Sanctioning behavior—an area that is of utmost importance to international law due to the missing central enforcement structure—can thus be better explained. If sanctions are perceived to

66. Kolm, *supra* note 63.

67. *See, e.g.*, Herbert Gintis, *Strong Reciprocity and Human Sociality*, 206 J. THEOR. BIOL. 169 (2000).

68. Armin Falk & Urs Fischbacher, *A Theory of Reciprocity*, 54 GAMES & ECON. BEHAV. 293 (2006).

69. Furthermore, preferences may also depend on the type of opponent, so-called type-based reciprocity. *Cf.* David Levine, *Modeling Altruism and Spitefulness in Experiments*, 1 REV. ECON. DYN. 593 (1998); Fehr & Schmidt, *supra* note 55.

70. Armin Falk et al., *Testing Theories of Fairness—Intentions Matter*, 62 GAMES & ECON. BEHAV. 287 (2008).

71. It rather expands classical game theory by adding emotions, mistakes, limited foresight and leaning. *Cf.* Camerer, *supra* note 60, at 3.

72. Falk et al., *supra* note 70 (providing further references concerning the experiments).

reveal selfish or greedy intentions, they destroy altruistic cooperation almost completely, whereas sanctions perceived as fair leave altruism intact. Fair sanctions are especially those that sanction defectors in public-good games (that is, climate change constellations or commons constellations like overfishing). In those common pool resource (CPR) experiments,[73] it was also shown that the possibility to communicate and to sanction alleviates the inefficient excess appropriation of the resource (which would be predicted by RC theory). The experiments thus challenge theories of human cooperation that neglect: (1) reciprocity *strictu sensu*, (2) the distinction between (perceived) fair and unfair sanctions, (3) the intentions of the other players, and (4) the "type" of the actor, although they are "probably relevant in all domains in which voluntary compliance matters."[74] Needless to say that this is especially the case in international law.

C. *Psychological Approaches in International Relations*

IR theory has not been untouched by behavioral economics or psychology. On the contrary, in the handbooks of IR, there has always been an overlap of psychology onto IR;[75] the beginning of this line of research is usually ascribed to the seminal scholarship of Robert Jervis.[76] Political psychologists in IR challenge realist theories that assume that states maximize power or security, as well as institutionalist theories that assume that states maximize wealth or utility. They do not stop there but extend criticism to constructivists[77] (who usually assume that everything depends on how human beings construct their world and their normative understanding).[78] However, they assume that at a foundational level constructivism is more compatible with

73. For the classical rational choice prediction concerning CPR, see Reinhard Selten, *The Chain Store Paradox*, 9 THEORY AND DECISION 127 (1978). *See also* Armin Falk et al., *Appropriating the Commons: A Theoretical Explanation*, *in* THE DRAMA OF THE COMMONS (Elinor Ostrom et al. eds., 2002) (finding that there is less appropriation in CPR and more contibution to public goods if the institutional set-up allows for informal sanctions and communication). For example, see Elinor Ostrom, *A Behavioral Approach to the Rational Choice Theory of Collective Action—Presidential Address of the American Political Science Association 1997*, 92 AM. POL. SCI. REV. 1 (1998).

74. Ernst Fehr & Bettina Rockenbach, *Detrimental Effects of Sanctions on Human Altruism*, 422 NATURE 137, 137 (2003); *cf.* Falk et al., *supra* note 70.

75. Goldgeier & Tetlock, *supra* note 10; Gross Stein, *supra* note 10.

76. JERVIS, *supra* note 10.

77. Constructivists stress ideas and norms, the management of the norms, and the perception of fairness and legitimacy. *See generally* Emanuel Adler, *Constructivism and International Relations*, *in* HANDBOOK OF INTERNATIONAL RELATIONS, *supra* note 10, at 95. *See also* Alexander Wendt, *Constructing International Politics*, 20 INT. SEC. 71, 71–72 (1995):

> Critical IR "theory," however, is not a single theory. It is a family of theories that includes postmodernists (Ashley, Walker), constructivists (Adler, Kratochwil, Ruggie, and now Katzenstein), neo-Marxists (Cox, Gill), feminists (Peterson, Sylvester), and others. What unites them is a concern with how world politics is "socially constructed," which involves two basic claims: that the fundamental structures of international politics are social rather than strictly material (a claim that opposes materialism), and that these structures shape actors' identities and interests, rather than just their behavior (a claim that opposes rationalism).

78. *See* Wendt, *supra* note 77, at 73:

political psychology.[79] The call for "behavioral IR" is well articulated.[80] Political psychologists in IR take the cognitive biases described above into account, arguing that bounded rationality and emotions, as well as procedural and distributive justice, should be included and tested in IR theories.[81] They test whether they have better explanations than realists or institutionalists, both of whom use the RC paradigm. The defining characteristic of the political psychology approach is the focus on the individual level,[82] that is, on central or elite decision-makers. They stress the contextual effects: history, development, and learning are assumed to be important to dynamic political processes.[83] Uniqueness of situations and the condition and contextual aspects of a particular decision-maker with a unique history in a specific situation is the subject of analysis. Explanation, rather than prediction, is the goal of the exercise.

There are some shortcomings of the political psychology approach in IR: first, the analysis is mostly confined to international conflict situations focusing on security issues using mainly prospect theory and framing.[84] This is just a small part of international law. Furthermore, IR theory applies those theories almost exclusively to individual decision-makers,[85] not to states as corporate actors. Implicit in this argumentation is that there are certain individuals, such as presidents or dictators, who are crucial for for-

---

> Social structures have three elements: shared knowledge, material resources, and practices. First, social structures are defined, in part, by shared understandings, expectations, or knowledge. These constitute the actors in a situation and the nature of their relationship, whether cooperative or conflictual. A security dilemma, for example, is a social structure composed of intersubjective understandings in which states are so distrustful that they make worst-case assumptions about each other's intentions, and as a result define their interests in self-help terms. A security community is a different social structure, one composed of shared knowledge in which states trust one another to resolve disputes without war. This dependence of social structure on ideas is the sense in which constructivism has an idealist (or 'idea-ist') view of structure. What makes these ideas (and thus structure) "social," however, is their intersubjective quality. In other words, sociality (in contrast to "materiality," in the sense of brute physical capabilities), is about shared knowledge.

Cf. MARTHA FINNEMORE, NATIONAL INTERESTS IN INTERNATIONAL SOCIETY 128 (1996):

> The fact that we live in an international society means that what we want and, in some ways, who we are are shaped by the social norms, rules, understandings, and relationships we have with others. These social realities are as influential as material realities in determining behavior. Indeed, they are what endow material realities with meaning and purpose. In political terms, it is these social realities that provide us with ends to which power and wealth can be used.

79. J. M. Goldgeier & P.E. Tetlock, *Psychology and International Relations Theory*, 4 ANN. REV. POL. SCI. 67, 83 (2001).

80. Alex Mintz, *Behavioral IR as a Subfield of International Relations*, 9 INT. STUD. REV. 157 (2007).

81. Goldgeier & Tetlock, *supra* note 79, at 68.

82. JERVIS, *supra* note 10.

83. MCDERMOTT, *supra* note 10, at 6 et seq.

84. FARNHAM, *supra* note 10. The research is concentrated on case studies like Roosevelt and the Munich crisis, the Iranian hostage rescue mission, etc. *See also* ROSE MCDERMOTT, RISK-TAKING IN INTERNATIONAL POLITICS (1998).

85. *See, e.g.*, MCDERMOTT, *supra* note 10, where only Chapter 9 (20 pages) is dedicated to group decisions.

eign policy decisions to the exclusion of others. Most important, norms have no room in the analysis; international law plays no role.

Political psychologists have challenged some RC approaches, namely game theoretical approaches to conflict as well as the notion of reputation, both crucial to international law and economics and compliance theories.[86] Let us turn first to issues of conflict and security (1.) and thereafter to the role of ideas (2.).

1. *Prospect Theory*

Psychological approaches use prospect theory to understand how states or their leaders react to crises. Prospect theory posits that under certain conditions, decision-makers are either more or less willing to take risky decisions: small probabilities of failure are treated as equivalent to zero, high probability of success is treated as certain, and losses are felt more heavily than gains. From a psychological perspective, individuals (and states) may be in either a loss or a gain frame—and thus react differently.[87] When states are in the domain of losses (for example, Serbia "losing" Kosovo), they are more likely to take an irredentist approach (that is, they are taking more risk). When they are in the domain of gain (for example, when Serbia is offered membership of the EU), they are more likely to accept the status quo.[88] Also, the endowment effect comes into play: it is far easier to prevent a state from taking something it does not already possess than to induce it to give something up. This does not come as a surprise for international lawyers dealing with territorial disputes.[89] States that perceive themselves to be in deteriorating situations, therefore, might be willing to take excessive risks[90]—something to be taken into account in the disputes about the islands in the South China Sea, for example. Loss aversion also explains why states may exhibit the equivalent behavior to sunk costs, for example, pro-

86. Guzman, *supra* note 7; Andrew T. Guzman, *A Compliance-Based Theory of International Law*, 90 Cal. L. Rev. 1823 (2002) [hereinafter Guzman, *A Compliance-Based Theory*].

87. Jeffrey Berejekian, *The Gains Debate: Framing State Choice*, 91 Am. Pol. Sci. Rev. 789 (1997) (holding that states in a gains frame pursue absolute gains and are risk averse, while states in a losses frame pursue relative gains and are risk acceptant. He illustrates this with the behavior of the European Community in the formation of the Montreal Protocol).

88. Goldgeier & Tetlock, *supra* note 79, at 70.

89. For an RC empirical analysis, see Paul Huth & Alyssa Prorok, *International Law and the Consolidation of Peace Following Territorial Changes*, APSA 2011 Annual Meeting Paper 2 (2011), *available at* http://ssrn.com/abstract=1900709 ("Despite the potential positive externalities of peaceful territorial change, the domestic political consequences for leaders who *lose* national territory can be severe. Accepting an unfavorable change to the territorial status quo may jeopardize a leader or regime's political survival, thereby generating incentives for the incumbent to seek the return of lost territory."). A BIntLE explanation for this could be the endowment effect. International law is somehow taking account of that by the criterion of "effective control" and the time factor (display of state administrative authority and the exercise of de facto jurisdiction over territory on a consistent basis over an extended period of time); *see also* Ian Brownlie, The Rule of Law in International Affairs: International Law at the Fiftieth Anniversary of the United Nations 136–42 (1998).

90. Jack S. Levy, *Prospect Theory and International Relations: Theoretical Applications and Analytical Problems*, 13 Pol. Psychol. 283, 285 (1992).

longing failed policies (such as wars). This can be crucial not only in dispute resolution but also in legal design and how legal claims and entitlements are framed.

2. *The Role of Ideas and Beliefs*

Political psychologists would also suggest that the claim that states will secure gains of cooperation by creating institutional frameworks whenever it is beneficial to them (institutionalist thinking in IR) has to be taken with a grain of salt: they deem such a claim to be at the same time too restrictive and too expansive. Indeed, behavioral economics would assume that it is too restrictive since regimes can also arise because of shared norms of fairness or the desire to punish free-riders. The claim is too expansive because actors might either be spiteful or, through shared mindsets, blind to potential benefits.[91] The role of ideas and beliefs in international politics, as posited by constructivists and institutionalists, can thus be judged in a more fine-grained way by drawing on cognitive psychology. If there is systematic slippage between policy-guiding mental representations of reality and reality itself, cognitive psychology might help to serve as a helpful guide in debiasing the slippage. Goldgeier and Tetlock suggest—especially between epistemic communities—transparency: "[t]he potential explanatory role of psychological constructs expands rapidly as we move from domains where the design of institutions is guided by 'well-known facts' and 'solid science' to those where the expert community is deeply divided and there is ample opportunity for cognitive and emotional biases to taint evaluations of evidence and options."[92] The thriving of cognitive psychology in "grey zones" that are open to interpretation is well known. This is interesting in several constellations for international law: first, when designing international treaties; second, when interpreting international law; and third, in order to understand compliance with international law. We will analyze some of the insights of political psychology in more detail when utilizing it for BIntLE.

D. A Synthesis of Approaches

Bringing those strands of research that use behavioral economics or psychology together promises added insights. Political scientists as well as international lawyers have long struggled between different paradigms of either functional analysis using mainly game theory based on rational choice on the one hand or an analysis of how ideas and beliefs "construct" realities and policies on the other.[93]

---

91. Goldgeier & Tetlock, *supra* note 79, at 78. *See also supra* note 36.

92. *Id.* at 79.

93. Robert Keohane, After Hegemony 111–16 (1984) has made use of bounded rationality but only in passing. Duncan Snidal, *The Limits of Hegemonic Stability Theory*, 39 Int'l Org. 579, 584 (1985) has argued that modern international relations theory can only purport to explain state behavior if its assumptions, however simplified, accord with reality, and that if not, they must be tested and refined.

BIntLE can build bridges between those paradigms by having a closer look at "how states (or actors relevant for international law) really behave."[94] It may concentrate on the individual decision-makers and thus extend (normative) decision theory that rests on RC. This approach has mainly been used by national behavioral law and economics as well as by the political psychology approaches in international relations—the focus lies on the individual and its biases, for example, elite decision-makers.

BIntLE may also focus on strategic interactions, as has been mostly done hitherto in IR and international law and economics, using the RC assumption when applying game theory to model interaction between states. Incorporation of behavioral insights into game theory goes under the heading of behavioral game theory. Behavioral research does not call into question strategic behavior as such. It looks at interaction of individuals as well as corporate actors such as firms, but takes into account emotions, (cognitive) mistakes, limited foresight, learning and etc. However, international law was not an object of analysis until now. Many of the insights of behavioral economics were generated by setting up games in the lab; a lot of the experiments, for example, concerning social preferences (Ultimatum Game), are behavioral game theory. As Camerer holds, "[B]ehavioral game theory is about what players actually do."[95] Thus, it lends itself to inquire how behavioral assumptions may change the strategies of states, negotiators and outcomes of games. BIntLE can thus draw on international law and economics and its game theoretic tools by enriching those with behavioral insights. It is cautious to explore exactly where "deviations" from a thick or thin notion of rationality lie: in the preference functions (for example, social preferences, altruism or spite), in the calculations of how to achieve certain elements of the preference function (for example, cognitive errors), or informational failures, satisfying, learning, etc.

II. The Challenges to BIntLE

Before laying out some potential applications of BIntLE, some caveats should be discussed, including some clarifications. Any application of *individual* decision theory to international relations or international law faces problems. Behavioral decision theories are most directly relevant to situations in which an individual makes choices among several alternatives presented by an external environment. Experiments in the lab, where the psychological insights into decision-making were first developed, are highly structured, and few of those conditions present in the lab are present in the

94. For a description of the political science approaches relevant for international law, see Emilie M. Hafner-Burton et al., *International Relations for International Law*, 106 Am. J. Int'l L. 47, 60 (2012).

95. Camerer, *supra* note 60, at 3 (italics and internal quotation marks omitted).

real world of IR.[96] The external validity of those insights for corporate actors is the most important problem for BIntLE.[97]

RC theory is both a normative theory on how actors should behave and also an assumption about the behavior of actors. In its first role, it poses no special problem for descriptive and analytical research. The problem arises only in the analytical part of the research needed to make prognoses on the behavior of actors in IR, and thus the effects of international law (which in turn is needed for design). Here, RC draws on axioms, whereas behavioral economics investigates "how actors really behave." On the individual actor's level, the criticism of assuming unrealistic behavior thus can be raised against RC but less against behavioral economics. What about a state's behavior? RC theory would maintain the Friedman argument that it is not the realism of assumptions that matters but the quality of the model and its predictions.[98] This argument is to be taken seriously—but it could still be, as argued here, that behavioral economics helps to give more adequate explanations and thus predictions, even on the aggregate level of corporate actors. Enhanced complexity is the price to pay, including the disadvantage that models cannot be solved analytically anymore but demand simulations, that is, computers.

The application of behavioral economics insights depends first on the actor who is to be analyzed (individual or group), and second on the transmission mechanism from the individual behavior to the aggregate behavior of a group (for example, a state). Even if it were accepted that individuals may deviate from the RC model in reality, it has been argued that this does not matter for international law and economics scholars since "individual cognitive errors might have few if any macro effects on international relations."[99] This is grossly ambiguous: is it considered that potential individual biases would cancel each other out on a macro level, that is, the state level (just as in the efficient market hypothesis[100])? Or does it mean that governmental decision-makers do not succumb to deviations of rationality? If so, why not? Due to group decision-making? Does rationality also depend on the governmental regime? Can politicians rationally exploit and use biases in the popu-

96. *See also* Levy, *supra* note 90, at 292.

97. In the same vein, Broude, *supra* note 1, discusses this problem extensively. While I am not as pessimistic as he is concerning the application, but I caution, as he does, about an unqualified extension. Having said that, it is by no means new that individual behavior is used for the explanation of state behavior in international relations theory. KENNETH N. WALTZ, MAN, THE STATE, AND WAR: A THEORETICAL ANALYSIS 16 (3d ed. 2001).

98. *See supra* note 13.

99. Goldsmith & Posner, *supra* note 7, at 8. In the same vein, see José Alvarez, *Do States Socialize?*, 54 DUKE L.J. 961, 969 (2005) (warning that using behavioral economics in IR might lead to "pop psychology without people" and would obscure the mechanisms by which states act, just as those are obscured when realists explain state behavior without resorting to internal mechanisms). The discussion in Levy, *supra* note 90, is still the most thoughtful one. Following this discussion is also Galbraith, *supra* note 8, at 354.

100. *See* Ronald J. Gilson & Reinier H. Kraakman, *The Mechanisms of Market Efficiency*, 70 VA. L. REV. 549 (1984).

lation, and what does that mean for behavior of the state on the international plane?

There are several constellations that need to be distinguished; each will be analyzed in turn below.

A. Whose Rationality? The Problem of the Unit of Analysis

IR gets ever more complex. The realm is populated by ever more actors, being not anymore only about the interaction of states; there are also ever more nonstate actors on the scene, such as NGOs, multinational enterprises, and last but not least, IOs.[101] Also, ever more mechanisms are used: hard law, soft law,[102] market mechanisms,[103] and transnational private regulation.[104] Much of the relevant behavior in IR does not originate in isolated individuals but it is embedded in institutional arrangements. Nevertheless, RC approaches to international law and IR theory usually assume that the state is (1) a rational actor, and (2) a unitary actor. Both of these assumptions can be and have been challenged.

Either for the sake of reducing complexity or because it is assumed that this embeddedness changes bounded rational behavior of individuals to rational behavior on the aggregate level, rational behavior is ascribed to an institution that is viewed as a corporate actor (for example, the state).[105] Thus, usually we look at the "behavior of the aggregate, rather than (exclusively) the behavior of those interacting at its interior."[106] Even political psychology scholars sometimes assume that states succumb to the same biases as individuals and thus ascribe behavior to a state.[107] However, there are many research questions in IR and international law that might require different levels of observation: sometimes the corporate actor state per se, sometimes governments, sometimes the population, sometimes judges, sometimes elite decision-makers, and others. The question of who really acts in international law (and whose behavioral assumptions therefore have to be scrutinized) is usually not answered in detail.[108] For the sake of clarity, it

101. *See generally* Non-State Actors as New Subjects of International Law (Rainer Hofmann ed., 1999); Non-State Actors as Standard Setters (Anne Peters et al. eds., 2009).

102. *See generally* Joost Pauwelyn et al., Informal International Lawmaking (2012).

103. *See generally* Anne van Aaken, *Effectuating Public International Law Through Market Mechanisms*, 165 J. Inst. Theor. Econ. 33 (2009).

104. *See generally* Colin Scott et al., *The Conceptual and Constitutional Challenge of Transnational Private Regulation*, 38 J. L. & Soc. 1 (2011).

105. For details, see Engel, *supra* note 50.

106. *See id.* at 446.

107. *See* Levy, *supra* note 90, at 284.

108. For an excellent example of how this can be done, see Goodman et al., *supra* note 8 at 3 ("We were accordingly interested in research insights on topics such as cognitive errors in decision-making, psychological and evolutionary pressures for good and evil, the influence of social structure on collective beliefs, effects of social marketing campaigns in creating individual preferences, and modes of governance in effectuating resistance and compliance."). The book generates exactly those insights which allow for a much more fine-grained analysis of actors and processes. The contributions investigate the human actor (cognitive psychology and social psychology), means of communication (for example, communication

makes sense to treat different actors separately, since this enables us to formulate concretely where behavioral economics might be able to contribute to the analysis of international law. For clarification, the transmission mechanism between individuals and groups (for example, the state) is depicted in the following.[109] We use the terms micro level (individual) and macro level (states, judges sitting in panels, IOs or other corporate actors).

		Rationality	
Unit of Analysis	Macro level	I Rational corporate actor	IV Bounded rational corporate actor
	Micro level	II Rational individual	III Bounded rational individual

If the relevant unit of analysis is the individual actor (micro)—for example, an elite decision-maker—we face no special problems that would deviate from the well-established field of behavioral economics and may use the behavioral economics lab experiments, although the (institutional) context must of course be specified to simulate elite decision-makers.[110] If that is the case, we are mostly in the situation of needing to attribute the decision-making to the state (and are thus regarding decision-makers as agents).[111]

Let us turn to the macro level. The relevant unit of analysis is a corporate actor: for example, a state. The actor on the macro level (Box I) can be rational either because all of its individuals are rational (Box II) or because the bounded rationality of the individual actors (Box III) cancels out each on the macro level. For example, it could be that individual international judges are boundedly rational but the group deliberations in court lead to a rational decision outcome. In the same manner, it could be that a head of state is trapped in a loss frame, but—due to his or her advisors—a rational decision would be taken.

In the alternative, a corporate actor could be boundedly rational (Box IV). Again, there might be two reasons for that: either the individual bounded rational behavior (Box III) leads to aggregated bounded rational behavior of the corporate actor—or even if all individuals are rational on the micro level (II), this still leads in the aggregate to bounded rational behavior, as is often

theory and social marketing) and group dynamics (organizational theory, political, economic and social dilemmas, game theory, and social network theory). Thus, a blend of many disciplines is used in order to explain certain phenomena. This is the way to go in my view.

109. Note that this is meant to apply to corporate actors with decision structures, not collective actors without structured decision making. For the latter from an RC perspective, see JAMES S. COLEMAN, FOUNDATIONS OF SOCIAL THEORY (1990).

110. *See* Hafner-Burton et al., *supra* note 8.

111. All persons mentioned in Art. 7 of the Vienna Convention on the Law of Treaties, (May 23, 1969), 1155 U.N.T.S. 331, 8 I.L.M. 679 (entered into force Jan. 27, 1980) [hereinafter VCLT], would be potential subjects for research.

the case in commons constellations.[112] The exact transmission mechanisms will be dealt with below; they can draw on several social science theories.[113]

B. The State as a Rational Actor?

The level of observation in most of the research in IR and international law focuses on the state as an entity (the billiard ball[114]). This level of observation makes sense for a lot of research questions in international law and IR—since it is indeed necessary, due to complexity problems, to analyze the behavior of the aggregate rather than the behavior of those interacting in its interior. But a state as such cannot act. If the fiction of the state acting is used, we attribute individual decision-making to it. It is directly relevant in IR if either the behavior of an individual is relevant in itself, for example, an international judge or arbitrator, or if the individual's acts on behalf of the state are directly attributable without further decision-modifying circumstances or institutions. Single, independent decision-makers like judges do not pose any methodological problem. But other decision-makers do, either because of decision-modifying circumstances (which can be controlled in a lab) or because of personal traits of elite decision-makers. Attribution of behavior to the state[115] can thus be more difficult.

Let us look at the research concerning the decision-modifying circumstances first. Political psychology in IR looks for case studies where leaders or other elite decision-makers have taken certain decisions, often as if there were no decision-modifying circumstances (for example, legal constraints, judicial oversight, re-election constraints, and accountability mechanisms). Tomer Broude's research on international humanitarian law relating to target selection in armed conflict takes into account the bounded rationality of military commanders.[116] Here, the individual decision-maker is the relevant unit of analysis. There are other instances imaginable where behavioral economics insights can be directly applied, although a note of caution is also adequate here: this research does not take into account the different institutional features of different states. The same biases of leaders or other individual decision-makers may lead to other outcomes depending on whether the respective state is authoritarian, a presidential system, a parliamentary sys-

112. *See infra*, Part III.D.3.

113. For an example in the field of human rights, see Goodman et al., *supra* note 8.

114. Anne-Marie Slaughter, Remarks, *The Big Picture: Beyond Hot Spots & Crises in Our Interconnected World*, 1 PENN ST. J.L & INT'L AFF. 286, 288–89 (2012) ("[S]tates were like billiard balls. We tried to prevent them from crashing into each other. We did not, however, look inside them. We did not think we could change what happened inside them and we did not care what happened inside them. Rather the focus in the billiard ball world was on how the states were configured relative to each other. This was the world of the Cuban Missile Crisis.").

115. This attribution is a legal question only. It is assumed here that the behavior of the individual is directly attributable to the state in the sense delineated in Art. 7 of the VCLT or in the Draft Articles on State Responsibility.

116. Broude, *supra* note 1, at 74.

tem, a federal system, or something other.[117] This makes cross-country comparisons more difficult but also generates very interesting research questions.

Does this solve the methodological problem? Can we *tel quel* assume rationality of a collective actor or—for that matter—deviations from the rationality assumption by attribution? As Engel holds, "Of course, the corporate will is generated in ways radically different from the generation of the individual will. Yet both from the outside and from within, they are perceived in a way that makes the analogy to the individual meaningful."[118] For the sake of the argument and accepting that it is necessary for epistemic reasons to deal with corporate actors, we will briefly refer to the existing research on the corporate actor per se (not aggregating individual behavior inside the ball[119]), which attributes individual decision-making directly to the corporate actor. The research uses a principal-agent approach, which assumes that the behavior of the agent directly transforms into the behavior of the principal (the corporate actor) via attribution. Assuming that decision-modifying circumstances are always present, it is relevant to ask "whether the corporate setting moderates the effect established by experimental economics or psychology."[120] There has been some research on the behavior of companies, which might be used *faute de mieux* as a first approximation. They are only to be considered *mutatis mutandis* for states but are reported below.

The research shows evidence for loss aversion in corporate actors. For example, companies take higher risks if their industry performs poorly compared to other industries (loss frame). It seems reasonable to assume that the bias can be present for all corporate actors,[121] including (democratic) states. This would be in conformity with the research of political psychology. For manager-owned companies, there is no difference compared to the insights of prospect theory on an individual level observable. This is different for actors who do not "own" the corporate actor on behalf of which they are acting.[122] The latter is the most relevant analogy for states, since governments do not "own" their countries anymore (as, for example, King Leopold II did with the Congo). What are the experimental findings for non-manager-owned firms? An effect that has been well-documented in managers is over-optimism: managers tend to believe that their companies are better than average and they underestimate their competitors. They are also prone to self-serving reinterpretations of reality and they take credit for good outcomes and lay blame on the environment for bad outcomes.[123] Managers exhibit this behavior to a larger extent than the normal population. Politi-

117. I would like to thank Ralf Michaels for this remark at the CICL Lecture at Duke.
118. Engel, *supra* note 50, at 445.
119. *See infra*, Part II.C.
120. Engel, *supra* note 50, at 451.
121. *See id.* at 450 for further references.
122. *Id.* at 452.
123. *Id.* at 452.

cians might be prone to over-optimism, as well. There are also biases that do not appear for actors acting on behalf of others. The endowment effect, for example, can disappear when subjects do not have to decide on their own property but are instead assigned the role of agents.[124] Whether politicians exhibit the endowment effect on behalf of countries or whether they exhibit it rationally because their electorate has it, for example, on territorial disputes, is a matter of further research.

A further problem arises in external validity of lab experiments with students, since elite behavior may differ; the population may not be analogous. Hafner-Burton et al. have researched trade negotiators, concentrating on their personal characteristics.[125] Behavior of elite decision-makers seems to be different from the normal population concerning some heuristics and biases, not attributable to the decision-making circumstances but rather to the personal traits of elite decision-makers and their experiences.[126] Although experimental evidence is still scarce, it shows that elites tend to be less averse to losses (that is they are better gamblers), in part because they have a higher level of trust and a generally more prone to cooperate. They also make better use of heuristics and have a higher "k-level awareness"—that is, they are better in projecting the rationality of other agents, which is highly relevant for strategic behavior. Full rationality assumes that players have high k-levels (common knowledge rationality); elite decision-makers are thus closer to an assumed ideal. They are also better in processing complex information. Does this mean that we may continue with the assumed ideal of full rationality? Probably not, since being better does not mean being perfect, and since research thus far only shows that circumstances matter, that personal traits matter, and that experience matters. All of it needs more research, which also leads to the next point: the aggregation mechanisms within the corporate actor.

C. Internal Processes: Group Behavior

As has been well researched by IR scholars, a two-level explanation is often needed in order to understand how states behave on the international scene. This leads to a breaking-up of the billiard ball state and to a look at internal (political) processes—that is, the above-mentioned transmission mechanisms. The unit of analysis is still the state. We look into the billiard ball in order to understand the movements of the ball. But there is another

124. *Id.* at 453.

125. *See supra* note 8.

126. *See* Emilie M. Hafner-Burton et al., *The Cognitive Revolution and the Political Psychology of Elite Decision Making*, 11 PERSPECTIVES ON POL. 369 (2013) (arguing that experienced policy elites differ from inexperienced subjects in how they make decisions, rooted in "sophistication," a learned skill that is derived from experience and tends to be greater in elite than non-elite populations).

world in IR now, which Slaughter terms the "Lego World."[127] Here, it is the bits and pieces within the ball that are the unit of analysis.

We look first inside the billiard ball. This is usually done by a two-level analysis,[128] whereby all actors, the electorate (including the different interest groups), and the politicians acting on the international plane are assumed to be rational actors. Those assumptions are used in trade economics, for example. In order to understand why and how states negotiate trade treaties, an internal political economy approach is used.[129] Although the electorate might be subject to biases, those are assumed to cancel out each other in the political process, and a rational state is the result. If this is the case and if the state is the unit of analysis, we are in Box I and there is no further problem; international law and economics could thus be used.

However, the transmission processes may not always lead to a rational corporate actor. Group dynamics are usually at play, since group behavior is more than a mere aggregation of individual behavior. Group decision-making has been looked at only very tentatively in the political psychology of IR. However, social psychology has long been analyzing group decision-making, and behavioral economics has started to experimentally test group effects, as well. Here, just a short reference to the findings of cognitive biases in groups can be given. Basically, groups succumb to the same cognitive biases as individuals. However, they differ in intensity (that is, some biases can be more or less intense in groups than in individuals). A significant effect was originally posited by Janis, applying it to international crisis situations such as the Bay of Pigs, Watergate, and Vietnam.[130] This group-think model is also being partially tested in experiments now.[131] The main insight is that groups develop in-group pressures toward conformity and cohesiveness. Janis outlined several symptoms, such as illusion of invulnerability (leading to over-optimism about choices the group confronts), inducing excessive risk-taking.[132] Group members would also discount and ignore warnings, especially from outside the groups. Inside challengers of it would be sanctioned by social reprimands or isolation. Belief in one's own group's

127. Slaughter, *supra* note 114, at 294–95 ("The governments can be taken apart, put together with corporations, foundations, NGOs, church groups, universities, or any number of social actors in any number of different coalitions.").

128. Robert D. Putnam, *Diplomacy and Domestic Politics: The Logic of Two-Level Games*, 42 INT'L ORG. 427 (1988).

129. *See* Gene M. Grossman & Elhanan Helpman, *Protection for Sale*, 84 AM. ECON. REV. 833 (1994); Gene M. Grossman & Elhanan Helpman, *Trade Wars and Trade Talks*, 103 J. POL. ECON. 675 (1995). For investment law, see Gene M. Grossman & Elhanan Helpman, Foreign Investment with Endogenous Protection (National Bureau of Economic Research Working Paper No. w4876, October 1994), *available at* http://ssrn.com/abstract=1826557.

130. *See* IRVING L. JANIS, GROUPTHINK: PSYCHOLOGICAL STUDIES OF POLICY DECISIONS AND FIASCOES (2d ed. 1982). For a summary of his research, see MCDERMOTT, *supra* note 10, at 249 et seq.

131. For a survey, see James K. Esser, *Alive and Well After 25 Years: A Review of Groupthink Research*, 73 ORG. BEHAV. HUM. DECISION PROCESSES 116 (1998). See also the whole symposium issue on groupthink in *id.*

132. This has been confirmed experimentally. *See* MCDERMOTT, *supra* note 10, at 252.

superiority would make members unwilling to challenge each other's morality. Parts of the theory have been challenged,[133] and parts have been confirmed by cognitive psychology and research in organizational behavior.

The so-called "group attribution error" is the corollary to the individual fundamental attribution error. People have a tendency to believe either (1) that the characteristics of an individual group member are reflective of the group as a whole (for example, a politician's decision or behavior is attributed to the population, although the leader might be a dictator or reign by a minority government), or (2) that a group's decision outcome must reflect the preferences of individual group members, even when information is available suggesting otherwise (for example, situational factors determined the outcome). Groups also aggravate attribution errors concerning type and situation (relevant for, for example, reputation). For out-groups, people simultaneously underestimate the impact of the situation while over-attributing the impact of another's disposition (type) in explaining behavior.[134] Furthermore, the self-serving bias exists at the level of groups as well (called group-serving biases). The self-serving bias manifests itself in that people tend to attribute their success to their skills rather than the situation—but at the same time, they prefer to attribute failure to situational factors or bad luck than to a lack of skills. In groups this may manifest itself in the form of stereotyping or discriminating against outside groups in negative ways.

Yet it is possible that those biases are moderated by specific rules for taking group decisions, by group discussion, and by norms prevalent in a group.[135] There is some literature in cognitive psychology describing a variety of shortcomings that prevent individuals from learning effectively. The literature also provides suggestions for organizational practices that may effectively repair the cognitive shortcomings of individuals ("cognitive repairs").[136] Those include, *inter alia*, developing procedures to circumvent pathologies of group decision-making; encouraging dissent and criticism within a group; group leaders refraining from stating their own opinions and encouraging discussions; and inviting independent experts. All those issues can be relevant to international law, as well, especially in alternative dispute resolution.

Let us thus assume for a moment that groups—for example, the electorate—exhibit biases. This could, for example, be the case in trade negotiations. There is survey evidence that mass public opinion impedes movements to the pareto-efficient frontier in international trade.[137] There are several explanations for that. It could be an RC explanation: that lobby

133. For an overview, see *id.* at 254.

134. Scott T. Allison & David M. Messick, *The Group Attribution Error*, 21 J. Experimental Soc. Psychol. 563 (1985).

135. Engel, *supra* note 50, at 453–55.

136. Heath et al., *supra* note 29.

137. Richard Herrmann et al., *How Americans Think about Trade: Resolving Conflicts among Money, Power and Principles*, 45 Int. Stud. Q. 191 (2001).

groups with a strong interest in protectionism may influence public opinion or have strong parliamentary groups (as is often the case in the agricultural field). Or it could be that individuals are in a frame as an employee, not as a consumer (as assumed by economists), whereby a loss aversion may play (since there are, of course, losers of free-trade; the gains are in the aggregate). Furthermore, a status quo bias might be prevalent.

Politicians may act rationally following their re-election constraint and not move to the pareto-optimal frontier, even though they know the country as a whole would be better off. This would then result in a bounded rational behavior of the state (Box IV). In addition, frames may play a role. A loss frame might, for example, arise if a politician is concerned about re-election and fears loss. From a state's perspective, however, this might be a gain frame (for example, in trade). Framing cannot be as easily determined as in the lab; there are several different frames for several different actors. A second possibility is that the biases are present in the electorate as well as in the political agents. A third possibility is that the electorate is rational but the political agents are not. All of these possibilities are subject to the aggregation problem which might depend also on internal institutions. Of course, politicians and international negotiators can also behave rationally as assumed by international law and economics (Box I).

Let us now turn to the Lego world, a horizontal world. Surely, the billiard ball is still present, but the component parts become more important players.[138] International law does not only play a role in interstate relationships. It plays a role in internal processes as well, and here the micro level starts to count. There might be internal actors exhibiting altruistic behavior. One prominent example of those actors relevant to international law is consumers.[139] They exhibit altruistic behavior when buying "fair trade" products, when rejecting products made with child labor or under severe working conditions. The same can be said in regard to the tragedy of the commons (for example, fisheries). People care about the extinction of species and use information provided by NGOs like the World Wildlife Foundation on endangered fish. The information required for consumers to make an informed decision may be generated through labeling. Whereas consumers might thus exhibit deviations from the classical RC model (fairness and other-regarding preferences), producers can be thought to be rational, taking into account the fairness preferences of consumers. International law can influence individual actors within a state, for example, by framing their choices. Norms and identities are viewed as crucial by constructivist scholars in IR; transmission mechanisms for international norm internalization can happen

138. Slaughter, *supra* note 114, at 295.

139. For an extensive commentary on the use of the market mechanism (including consumers) to effectuate international law, see van Aaken, *supra* note 103; Anne van Aaken, *Trust, Verify or Incentivize? Effectuating Public International Law Regulating Public Goods Through Market Mechanisms*, 104 AM. SOC'Y INT'L L. 153 (2011).

through transnational processes[140] as well as national processes.[141] Although constructivist theories are meant to explain state behavior, they often disaggregate the state in order to do so; the unit of analysis can be individuals and their ideas and beliefs. Behavioral economics explains how norms can influence those (for example, via endowment effects) and how, in the aggregate that might influence, for example, compliance with international law. Here, political psychology approaches, constructivist approaches and behavioral economics meet. This is relevant for treaty making as well as for compliance questions.

III. New Perspectives Provided by BIntLE

Enriching RC models with insights from behavioral economics might not only further refine our understanding and explanations of public international law—possibly contradicting or confirming insights from international law and economics—but it may also help better design international law as a fundamental de-biasing mechanism (if that is normatively desired[142]). Law, including international law, is never neutral: it sets reference points, produces endowments (and therefore the endowment effect) and sets points for perceived fairness. How, exactly, the law achieves those ends is a promising research field.

The burden of proof lies with BIntLE: is its explanatory power higher than that of classical RC theory? Sometimes, results suggest that it does not contradict but rather reinforces RC arguments.[143] Even if BIntLE does not change the algebraic sign found by RC research, it would still have an additional explanatory effect for the behavior of states by making RC explanations stronger. This potential will be illustrated by a few examples: (1) "choice architecture"[144] of sources of international law, (2) international custom, (3) and (4) treaty design, and (5) compliance questions. Although international dispute settlement is a promising venue for research (and largely unproblematic from a methodological point of view), it is beyond the scope of this article.

140. Harold Hongju Koh, *Why Do Nations Obey International Law?* 106 Yale L.J. 2599 (1997).

141. Thomas Risse & Kathryn Sikkink, *The Socialization of International Human Rights Norms into Domestic Practices: Introduction*, *in* The Power of Human Rights: International Norms and Domestic Change. (Thomas Risse et al. eds., 1999); Martha Finnemore & Kathryn Sikkink, *International Norm Dynamics and Political Change*, 52 Int'l Org. 887 (1998); Thomas Risse, *International Norms and Domestic Change: Arguing and Communicative Behavior in the Human Rights Area*, 27 Pol. Soc. 529 (1999).

142. For this discussion, see Kelman, *supra* note 23.

143. For a discussion of the conceptual and methodological challenges of behavioral economics in IR, competing with RC explanations, see Levy, *supra* note 90, at 296–307.

144. Thaler & Sunstein, *supra* note 5, at 3.

A. Sources of International Law

The most basic choice architecture of international law is found in the legal frameworks in which states chose to participate. Those include the classical three sources of law listed in ICJ Statue Art. 38, but one may also add soft law,[145] since this is used ever more often by states.[146] Treaty law and custom have been analyzed from a RC perspective without taking into account the behavioral effects of default rules of consent to norms, although these may influence the choice of frameworks. An RC approach does not take into account the influence of default options on the behavior of states; it assumes that states choose whatever source is more beneficial. Consent is still taken to be the cornerstone of international law. But consent comes in many forms. By taking the insights of BIntLE of how default options frame decision-makers, or provide a "choice architecture," we can look at sources of international law under the perspective of how they fit into the opt-in versus opt-out scheme.[147] Generally speaking, opt-out systems have higher participation than opt-in systems, as shown by behavioral research. Acquiescence or inaction is often the norm in terms of state action, but different frameworks use acquiescence in very different ways by creating different default rules about the consequences of inaction. That may shed light on why different rules for different types of frameworks are needed in order to achieve different goals. This perspective adds a potential explanatory factor of why states choose particular kinds of cooperation frameworks.

Sources of law can be plotted on a spectrum from (i) specific, written, and individual opt-in (treaties); to (ii) less specific, unwritten and general consensus subject to opt-out (custom); to (iii) unwritten and generally binding not subject to opt-out (jus cogens) or soft law, which may have, although non-binding, no opt-out (if it has extraterritorial effect in enforcement like, say, the Financial Action Task Force Recommendations[148]). Behavioral economics shows that the default rule has considerable impact on the behavior of actors.

When do states use each kind of framework? Rationalist institutional scholars usually follow the "hard and soft law" taxonomy of Kenneth Abbott and Duncan Snidal, who propose thinking of international agreements as relatively hard or soft along three dimensions: obligation, precision, and

145. For the classical definitions, see Daniel Thürer, *Soft Law*, *in* 4 ENCYCLOPEDIA OF PUBLIC INTERNATIONAL LAW (Rudolf Bernhardt ed., 2000) and Dinah Shelton, *Soft Law*, *in* HANDBOOK OF INTERNATIONAL LAW (2008).

146. Stefan Voigt, *The Economics of Informal International Law—An Empirical Assessment*, *in* INFORMAL INTERNATIONAL LAWMAKING (Joost Pauwelyn et al. eds., 2012). Much has been written on soft law from an economic perspective. For an overview, see Andrew Guzman & Timothy L. Meyer, *Soft Law*, *in* RESEARCH HANDBOOK ON THE ECONOMICS OF PUBLIC INTERNATIONAL LAW (Eugene Kantorovich ed.) (forthcoming with Edward Elgar) (on file with the author).

147. We would like to thank Anthea Roberts for this suggestion.

148. Those are recommendations originally by the G-8 states concerning the fight against money-laundering and terrorism financing. Although not all states are part of the norm creation, all states may be black-listed and ultimately shut out of financial markets.

delegation.[149] Furthermore, the transaction costs of treaty negotiations are included as an explanatory variable for whether states enter into treaties in the first place (whereas custom has no such costs).[150] Abbott and Snidal argue that states make deliberate choices about the form of their legal obligations, including by altering the bindingness of the obligation (soft versus hard in the traditional sense); by adjusting the degree of vagueness (usually custom is more vague than treaties); and by choosing whether to delegate enforcement and interpretation.[151] Reducing the strength or depth of an obligation can be useful because it may offer benefits in the form of reduced transaction costs, increased flexibility, and enhanced cooperation.[152] However, the RC frame does neglect the default rule mechanisms in its analysis, even though this might be an important design element in the treaty architecture on two levels or layers: (i) on the level of whether a state is bound by international law (the "whether" on the level of sources); and (ii) on the sub-layer, once bound, how default rules are used within a source of international law, especially treaties (for example, reservations).

How, then, do the default rules play into this reasoning? Opt-in and opt-out mechanisms can be understood to align on a continuum. Jus cogens has no opt-out and no withdrawal possibility, no negotiation costs, and only a very minimal possibility to influence the norm for each state; its violation is comparatively more costly than a violation of custom. Treaties, in contrast, are an opt-in default on the first layer, requiring explicit consensus for being bound. But on sub-layers it also allows for the influence of the content of the norms, either through (re-)negotiation or through reservations (sub opt-out), and they usually allow withdrawal (that is, an opt-out later in time). States can thus opt-in bit by bit and in a way that is more tailor-made to their interests. Nevertheless, on the first layer some treaties are an all-or-nothing package—the WTO, for example, which is a single undertaking (a state has to take the whole package). Treaties that do not allow for reservations[153] or that have mandatory dispute settlement provisions fall into this category. Some treaties contain majority rules for change that bind states without explicit opt-in (the European Union is at the forefront of this mechanism). Treaties may exhibit other sub-layers of opt-in versus opt-out mechanisms, such as the positive list approach in the service sector of GATS,[154] where states have to opt in to a certain commitment. The dispute settlement procedure (delegation) of the WTO has changed from positive consensus (opt-

149. Kenneth W. Abbott & Duncan Snidal, *Hard and Soft Law in International Governance*, 54 INT'L ORG. 421, 422 (2000).

150. POSNER & SYKES, *supra* note 7, at 20.

151. Abbott & Snidal, *supra* note 149, at 427 (submitting that their analysis also holds for custom).

152. *Id.* at 423.

153. *E.g.*, Rome Statute of the International Criminal Court, art. 120, July 17, 1998, 2187 U.N.T.S. 90; Marrakesh Agreement, art. XVI:5, April 15, 1994, 1867 U.N.T.S. 154; Basel Convention on the Control of Transboundary Movements of Hazardous Wastes and Their Disposal, art. 26 (1), March 22, 1989, 1673 U.N.T.S. 126.

154. General Agreement on Trade in Services (GATS), 1994, 1869 U.N.T.S. 183.

in) to negative consensus (opt-out) under the new Dispute Settlement Body.[155] This spectrum can be even more refined, especially on treaties, if majority voting in treaties or a certain form of consensus or decision-making powers of IOs are taken into account. Furthermore, prompts may be used, for example, if opt-outs are required to be renewed after a certain period of time (and failure to renew would count as opt-in). I will turn to reservations, optional protocols and objections to reservations in more detail below.

Custom is in the middle between the extremes of treaties and jus cogens. It is an opt-out system (persistent objector), and it is not negotiable; thus, it is a take-it or leave-it approach. No sub opt-out is possible, only full opt-out. Nevertheless, custom does not often present difficulties, since it has a considerable degree of vagueness and thus permits flexibility in this respect. Being bound is thus relative in comparison with treaties, especially since custom (if not backed up by treaties) is less easily enforceable.

Once a state is "in," another variable to look at is how easy it is to get "out" (that is, the costs a state incurs from withdrawal). Whereas treaties allow exits, custom does so only under difficult circumstances[156] and only when custom evolves (not later)—and jus cogens both is mandatory and does not allow withdrawal at all.

When states choose the architecture of their cooperation, considerations of universality and integrity of norms may play a role. The former calls for opt-out default (and thus custom), ensuring near-universality.[157] If universality is desired, opt-out is thus a better option than opt-in. Custom also ensures integrity, since there is no tailor-made opt-out within the norms. It is all or nothing. For the few norms which have status of jus cogens, universality is ensured and it is hard law in the sense of Abbott and Snidal. It is to be expected that this form of cooperation evolves only for very few norms. For norms in the form of treaties, it is harder to ensure universality, since opt-in is the default. This is one reason why those treaties most often allow for opt-out possibilities within the treaty. A systematic analysis of design features along the trade-offs between treaty depth and breadth—and strategies states use to address different forms of uncertainty, including the default option and its consequences for participation of states in cooperative ventures—promises additional insights that are not captured by a pure RC analysis.

155. Dispute Settlement Understanding (DSU), art. 16:4, 1994, 1869 U.N.T.S. 401 ("[T]he report shall be adopted at a DSB meeting unless . . . the DSB decides by consensus not to adopt the report."); *id.* at art. 17:14 for Appellate Body Reports ("An Appellate Body report shall be adopted by the DSB and unconditionally accepted by the parties to the dispute unless the DSB decides by consensus not to adopt the Appellate Body report . . . .").

156. Curtis A. Bradley & Mitu Gulati, *Withdrawing from International Custom*, 120 YALE L.J. 202 (2010).

157. *Id.* (arguing that this "Mandatory View" of custom that disallows withdrawal is a historical development, and that it may have been part of an effort to bind "uncivilized" states to the international law worked out by a small group of Western powers).

B. Customary International Law

Customary international law (CIL) has also been analyzed from an RC perspective. Its two components, namely the consistent and repeated practice of states (*diuturnitas*) they follow from a sense of legal obligations (*opinio juris sive necessitatis*),[158] have been intensively and controversially discussed not only by legal scholars but also by RC scholars. Goldsmith and Posner use simple game theory in order to explain how CIL arises, why nations "comply" with CIL, and how CIL changes, challenging the traditional view that CIL is unitary, universal, and exogenous.[159] They claim that states' compliance and the norms themselves emerge from the states' pursuit of self-interested policies on the international stage, that the behaviors associated with CIL do not reflect a single, unitary logic but reflect instead various and importantly different logical structures played out in discrete, historically-contingent contexts.[160] Self-interested policies come in different constellations: CIL may develop as a behavioral regularity (i) if by coincidence of interest,[161] each state obtains a private advantage from a particular action (for example, by not destroying foreign fishing vessels because it is too costly for their navy anyway), (ii) if one or more states force other states to engage in actions that serve the interests of the other state (coercion), or (iii) if the states' interests converge but each state's best move depends on the move of the other states (for example, if state A chooses action Y, then state B will also choose action Y as this is in its best interest: coordination[162]). Finally, the theory is skeptical of the existence of multilateral behavioral regularities and does not attribute any behavioral force to CIL; it is epiphenomenal in this view.[163] Due to the lack of information and monitoring devices, genuine multistate cooperation is unlikely to emerge.[164] Therefore, universal CIL is unlikely to occur in multilateral prisoners' dilemma situations, and we cannot expect CIL to solve global commons problems.[165] Multistate prisoners' dilemma games tend to be solved, if at all, by treaty or other international agreement, and not by decentralized evolution (that is,

158. *See* RESTATEMENT (THIRD) OF THE FOREIGN RELATIONS LAW OF THE UNITED STATES § 102(2) (1987). For an overview of the positivist discussion, see Anthea E. Roberts, *Traditional and Modern Approaches to Customary International Law*, 95 AM. J. INT'L L. 757 (2001) as well as Tullio Treves, *Customary International Law*, in MAX PLANCK ENCYCLOPEDIA OF PUBLIC INTERNATIONAL LAW (2006).

159. GOLDSMITH & POSNER, *supra* note 7, at 25.

160. *Id.* at 40–43, 45–78.

161. Coincidence of interest denotes a symmetric game, where each of the players has a dominant strategy and the resulting Nash-equilibrium is pareto-optimal.

162. In a coordination game, there are multiple equilibria and the players are indifferent to which one is actually chosen. Once it is chosen due to a strategy of one player, the other players have no incentive to deviate from the equilibrium.

163. GOLDSMITH & POSNER, *supra* note 7 at 25. *See also* Goldsmith & Posner, *A Theory of Customary International Law*, 66 U. CHI. L. REV. 1113 (1999). For an essentially similar but more attenuated argument, compare POSNER & SYKES, *supra* note 7, at 50.

164. GOLDSMITH & POSNER, *supra* note 7, at 87.

165. *Id*. at 35.

the formation of CIL).[166] In short, CIL can reflect genuine cooperation or coordination only between pairs of states or among a small group of states, but even then, what looks like law-abiding behavior can be explained by self-interest. According to Goldsmith and Posner, in all other cases, CIL reflects self-interested behavior either created by coercion or simply coincidence of interest.

Those claims, in turn, have been extensively challenged by other RC scholars.[167] Apart from a critique arising within the RC framework, here we are interested in whether behavioral economics can contribute to the puzzle of CIL evolution, a question that has been extensively discussed within legal scholarship. As Treves holds: "Customary rules are the result of a process—whose character has been qualified by a number of authors as "mysterious"—through which elements of fact, empirically verifiable, acquire a legal character thus creating rights and obligations for the subjects of international law."[168] This procedural character of CIL has been forcefully described by Maurice Mendelsohn: "The customary process is in fact a continuous one, which does not stop when the rule has emerged . . . . Even after the rule has 'emerged,' every act of compliance will strengthen it, and every violation, if acquiesced in, will help to undermine it."[169] CIL thereby rests on the premise that "states should comply with the legitimate expectations of the international community," whereby the question of what is a legitimate expectation depends on "a norm [that] emerges from what is normal . . . . If the generality of states has regularly behaved in certain ways . . . then a legitimate expectation arises that they will continue to do so."[170] Following Mendelsohn's idea that normative expectations and behavior coevolve, Engel argues that "[t]o properly understand what customary law is and why it governs behavior, one has to adopt an evolutionary perspective. New customary law is not generated, it emerges."[171] Fortunately, this can be and has been experimentally tested by playing public goods games,[172] which

166. *Id*. at 36.

167. Mark A. Chinen, *Game Theory and Customary International Law: A Response to Professors Goldsmith and Posner*, 23 MICH. J. INT'L L. 143 (2001); Edward T. Swaine, *Rational Custom*, 52 DUKE L.J. 559 (2002); George Norman & Joel P. Trachtman, *The Customary International Law Game*, 99 AM. J. INT'L L. 541 (2005). For a methodological critique of their way of using game theory, see van Aaken, *To Do Away with International Law?*, *supra* note 7, at 289.

168. Tullio Treves, *supra* note 158, at ¶4.

169. Maurice H. Mendelson, *The Formation of Customary International Law*, 272 RECUEIL DES COURS 155, 175 (1998). *See also Nicaragua v. United States of America*, 1986 I.C.J. Rep. at 98, ¶186 (stating that it is sufficient for the conduct of states to be generally consistent with statements of rules, provided that contrary state practice had generally been "treated as breaches of that rules, not as indications of the recognition of a new rule.").

170. Mendelson, *supra* note 169, at 185.

171. Christoph Engel, *The Emergence of a New Rule of Customary Law: An Experimental Contribution*, 7 REV. L. ECON. 767, 769 (2011).

172. For an overview over those games, see Ananish Chaudhuri, *Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature*, 14 EXPERIMENTAL ECON. 47, 47 (2011) (showing that "many participants in laboratory public goods experiments are "conditional cooperators" whose contributions to the public good are positively correlated with their beliefs about the average group

are often used in IR theory, for example, for combating climate change or terrorism.[173] RC game theory predicts that no actor will contribute to the public good[174]—certainly not in a one-shot game, but also not in repeated games without punishment or communication. As Engel and Kurschilgen argue, two important empirical questions lie at the heart of the debate: "Do normative expectations have any autonomous effect on people's behavior? If so, does this effect rest on the fact that the underlying norm is perceived to be law?"[175] Although it depends on the type of the players and the context, actors contribute much more than expected by RC theory and punish more than defect, even without communication. This holds true for the beginning of the game. But even in the end round, the contributions are much higher than zero.[176] What happens? Engel investigates how participants react in the subsequent period to the experiences they have had in the preceding period. If participants had contributed less than the average and if they react by increasing their contributions, Engel argues that participants' behavior can be interpreted as an adjustment to the perceived expectation of others to make a higher contribution. If many do not play by what a particular player believes to be the rule, that player may react by herself ceasing to abide by that rule[177]—or vice versa. There is thus a convergence of expectations. Legitimacy of the expectations is measured as perceived legitimacy by investigating how behavior develops if participants are given a chance to express disapproval through costly punishment, decentralized punishment being interpreted "as a (very embryonic) institutional intervention,"[178] since the punisher sends a signal of her discontent with the behavior of another participant. Although there might be consequentialist as well as deontic reasons for punishment, if punishment in the last round is sensitive to the deviation from the contribution level in the penultimate round, then pun-

contribution. Conditional cooperators are often able to sustain high contributions to the public good through costly monetary punishment of free-riders but also by other mechanisms such as expressions of disapproval, advice giving and assortative matching.").

173. Todd Sandler, Global Collective Action (2004).

174. The provision of a public good is usually modeled as prisoners' dilemma game that has only one equilibrium: defection—that is, non-contribution. It is a classical collective action problem—that is, climate change. *See* Posner & Sykes, *supra* note 7, at 230; Garrett Hardin, *The Tragedy of the Commons*, 162 Science 1243 (1968); Mancur Olson, The Logic of Collective Action (1971). For the experiments challenging this pessimistic view, see John O. Ledyard, *Public Goods: A Survey of Experimental Research*, *in* The Handbook of Experimental Economics (John H. Kagel & Alvin E. Roth eds., 1995).

175. Christoph Engel & Michael Kurschilgen, *The Coevolution of Behavior and Normative Expectations: An Experiment*, 15 Am. L. Econ. Rev. 578 (2013).

176. *See* Christoph Engel, *supra* note 171, at 773. *See also* Engel & Kurschilgen, *supra* note 175, at 582 ("If the pattern has been repeated for a while, the *behavioral* belief (that is, how others will probably behave) turns into a *normative* expectation (that is, how others should rightly behave). The belief is no longer purely cognitive. A motivational component is added to it. If an actor deviates from established practice, she violates others' normative expectations. Others regard such behavior not only as anti-social but also as illegitimate.").

177. *See* Engel, *supra* note 171, at 776.

178. *Id.* at 777.

ishment demonstrates first, that punishment cannot be exclusively instrumental and second, that punishers regard negative deviations from the norm as unfair, leading to the conclusion that they interpret the contribution level as a *normative* expectation.[179] Engel and Kurschilgen have conducted a more targeted experiment with two treatments. In the first treatment ("Norm"), participants elaborate normative expectations. In the second treatment ("Law"), expectations are embedded into a legal frame about the creation of customary law. Both treatments are shown to have a substantial positive effect on cooperation, but there is no significant difference between "Norm" and "Law." This changes as soon as soon as they make it possible for participants to sanction each other, at a cost. They conclude that in that case, "the legal frame makes a crucial difference for cooperation as law and sanctions clearly seem to complement each other."[180]

Those findings contradict RC approaches in so far as they rely on forces of "acculturation,"[181] fairness components, and expressive law.[182] They contradict the assumptions of RC game theory in which coincidences of interest or coercion are causal factors for the formation of CIL and stress the importance of the evolution of normative expectations. They are less pessimistic concerning multilateral CIL for commons problems although they confirm the context dependency of CIL.[183] With all the caveats mentioned under Part II, those behavioral findings may give an answer to the puzzle of normativity in CIL and the formation of opinio juris.

C. Treaty Design Generally

RC approaches to international law oftentimes use economic contract theory[184] when analyzing international treaties.[185] Traditionally, the *legal* analysis of contracts takes an *ex post* perspective—that is, it focuses, on the one hand, on rights and obligations after there has been an alleged breach, and on the other hand, on the recovery of losses for the injured party. The *economic* analysis, contract theory, shifts the focus to the *ex ante* decision and

179. *Id.* at 778.

180. Christoph Engel & Michael Kurschilgen, *supra* note 175, at 2.

181. Ryan Goodman & Derek Jinks, *How to Influence States: Socialization and International Human Rights Law*, 54 DUKE L.J. 621 (2004).

182. *See* Robert D. Cooter, *Expressive Law and Economics*, 27 J. LEG. STUD. 585 (1998); Richard McAdams, *A Focal Point Theory of Expressive Law*, 86 VA. L. REV. 1649 (2000).

183. *See* Engel, *supra* note 171, at 779.

184. For a short overview of incomplete contracting in a commercial law setting, see Robert E. Scott, *The Law and Economics of Incomplete Contracts*, 2 ANN. REV. L. & SOC. SCI. 279 (2006).

185. SCOTT & STEPHAN, *supra* note 7; TRACHTMAN, *supra* note 7, at 120 et seq.; POSNER & SYKES, *supra* note 7, at 24 ("When nations come together and agree to cooperate using international law, the situation is much analogous to private actors coming together and cooperating by creating a contract . . . . Accordingly, there are great similarities between the problem of designing an optimal contract and the problem of crafting an optimal body of international law." For an application to WTO law, see SIMON SCHROPP, TRADE POLICY FLEXIBILITY AND ENFORCEMENT IN THE WTO: A LAW AND ECONOMICS ANALYSIS (2009); for an application to investment law, see Anne van Aaken, *International Investment Law Between Commitment and Flexibility: A Contract Theory Analysis*, 12 J. INT'L ECON. L. 507 (2009).

thus to treaty design. It is primarily an analytical approach to explain why parties enter into contracts in the first place and why they draft any particular contract, in light of the actions of courts and tribunals. Generally speaking, states will participate in treaties only if the expected costs of constraining (regulatory) sovereignty do not exceed the expected net benefits from cooperation. In contract theory terms: the participation constraint of states must be met.

The best starting point for analysis is the contract that yields optimal outcomes, that is, the first-best contract that can be used as a benchmark. This "[p]areto-efficient complete contingent contract"[186] is the one parties would write if there were no contracting imperfections such as bounded rationality and unforeseeability, no transaction costs, and no enforcement costs. Such contracts would assign risks, rights, and responsibilities in every possible state of the world. The context of the contract would be free of market imperfections, unforeseen developments, and opportunistic behavior. There would be no need for a third-party adjudicator and additional control. In such a counterfactual situation, parties would, under RC assumptions, maximize their *ex ante* commitment, because there are no assurance problems. This is obviously Panglossian. There are thus two essential questions. First, will states enter into treaties in the first place? Second, if yes, how should treaties be drafted? In both instances, the RC answer is modified by BIntLE.

1. *Consent to Treaties*

The first question has a straightforward answer in RC: states enter treaties if their participation constraint is met, that is, whenever they assume that the gains from cooperation are higher than from non-cooperation. The RC approach uses normative decision theory to reach its results: states and their negotiators should compare the subjective expected value of an agreement to the subjective expected value of non-agreement, taking into account factors such as risk, transaction costs, reputational and relational consequences of each possible course of action before selecting the best option.[187] Guzman argues, from an RC perspective, that consent is often difficult to reach and therefore more forms of nonconsensual lawmaking should be used.[188] His argument is that although consent avoids welfare losses for individual states (*volenti non fit iniuria*), welfare enhancing treaties for global welfare might not be concluded due to a cumbersome status quo bias. BIntLE may add an additional argument for his suggestion of using more soft law and other mechanisms. Korobkin and Guthrie find that some common heuristics are likely to influence the negotiator's decision-making processes in bargaining

186. Steven Shavell, *Damage Measures for Breach of Contract*, 11 BELL J. ECON. 466, 467 (1980).

187. Russell B. Korobkin & Chris Guthrie, *Heuristics and Biases at the Bargaining Table*, 87 MARQ. L. REV. 795, 795 (2004).

188. Guzman, *supra* note 36.

(assuming reciprocal treaties, such as trade negotiations). They distinguish between "judgment" of a negotiator, meaning she must evaluate the content of available options, and "choice," meaning that the negotiator must determine which available option she prefers.[189] They discuss anchoring and adjustment, availability, self-serving evaluations, framing, the status quo bias, contrast effects, and reactive devaluation, all present in negotiators.[190] In the judgment phase, anchoring and adjustment may impact valuation of an option. If, for example, numerical estimates are necessary, as in trade or climate change negotiations, the initial anchor might lead the negotiator to twisted evaluations. Negotiators will also often evaluate the likelihood of various outcomes or scenarios based on the ease with which they come to mind (or are brought into the negotiations). This availability bias can lead to flawed predictions. The self-serving bias infects negotiators too. Depending on which side they negotiate on (usually their home country), the bias can cause negotiators to overestimate the likely benefits from concluding an agreement (or not); in other words, neutrality is impacted in the estimation of benefits and costs.[191] Those biases might also be present in international treaty negotiators. If the concession aversion is added to the list, this makes for a forceful argument for Guzman's suggestion: consent, or the lack thereof, may not be based solely on objective judgment and choice of negotiators but might be blurred, leading to bargaining impasse.[192]

Others have now focused on elite decision-makers in international relations and their bounded rationality in order to explore questions of which kind of treaties are entered into. Two behavioral traits (patience and strategic skills) of individuals served as an explanation: patient subjects were more likely to prefer treaties with larger numbers of countries (and larger long-term benefits), as were subjects with the skill to anticipate how others will respond over multiple iterations of strategic games.[193] This insight again raises questions as to the pure rationality and objectivity of negotiations in international law. When treaties with large numbers and long-term benefits are negotiated (for example, in climate change or multilateral trade treaties), the skills and patience of the negotiators seem to play a big role. Using behavioral economics and extending research into why states enter which treaties might also generate useful insights in the human rights sphere.

Let us turn to an example: investment treaties. Using a prisoner's dilemma game involving competition for capital, Andrew Guzman offered an

189. Russell B. Korobkin & Chris Guthrie, *supra* note 187, at 789.

190. *Id.*

191. *See* Georg Loewenstein & Linda Babcock, *Self-Serving Assessments of Fairness and Pretrial Bargaining*, 22 J. LEG. STUD. 135 (1993) (finding that depending on whether experimental subjects are assigned the role of a plaintiff's or a defendant's lawyer, they reach other conclusions on the outcome of the trial. *See also* Linda Babcock & George Loewenstein, *Explaining Bargaining Impasse: The Role of Self-Serving Biases*, 11 J. OF ECON. PERSPECTIVES 109 (1997).

192. Seminal on that topic including the biases mentioned here is KENNETH J. ARROW ET AL., BARRIERS TO CONLFICT RESOLUTION (1995).

193. Hafner-Burton et al., *supra* note 8.

early rationalist explanation of why developing states enter bilateral investment treaties that might hurt them. Although less developed states would have been better off acting together (for example, by negotiating as a block), they signed bilateral treaties that offer investors much greater protection than did the old rule of CIL. Guzman explains this by pointing to the strong incentives an individual country has to negotiate with and offer concessions to potential investors, thereby making itself a more attractive location relative to other potential host countries.[194] Lauge Poulson investigates bilateral investment treaties by drawing on cognitive heuristics and policy diffusion theory. He suggests that a bounded rationality framework has considerable potential to explain why and how developing countries have adopted modern investment treaties.[195] This follows earlier work by Poulson and Aisbett where the signing of bilateral investment treaties (BITs) is explained by the states ignoring low-probability, high-impact risks (like an investment treaty claim) until the claim hits them.[196] The explanation thus turns on a cognitive error: the low probability neglect,[197] a well-known effect from cognitive psychology, namely that individuals systematically neglect events with low probabilities (especially if they are not salient).[198]

2. *Treaty Design*

Let us now turn to how treaties are written and how they ought to be written. Since we do not live in a Panglossian world, complete contracts foreseeing every possible contingency are impossible to draft and even trying to come close creates high negotiating costs. This, in turn, gives rise to the impossibility to foresee and describe appropriately the contractual outcome for all states of the future world: "[c]ontracts will be incomplete in the sense that they will fail to discriminate between states of the world that optimally call for different obligations."[199] A well-known problem in contract theory is that of dealing with uncertainty.[200] Parties cannot easily design contracts that maximize jointly beneficial investments and at the same time respond

194. Andrew T. Guzman, *Why LDCs Sign Treaties That Hurt Them: Explaining the Popularity of Bilateral Investment Treaties*, 38 VA. J. INT'L L. 639 (1998). Another rationalist account of investment treaties and the reaction of states is given by van Aaken, *supra* note 185.

195. Poulsen, *Bounded Rationality*, *supra* note 8.

196. Poulsen & Aisbett, *supra* note 8.

197. *See* Rabin, *supra* note 3, at 24.

198. CASS SUNSTEIN & RICHARD ZECKHAUSER, *Dreadful Possibilities, Neglected Probabilities*, *in* THE IRRATIONAL ECONOMIST: MAKING DECISIONS IN A DANGEROUS WORLD (Erwan Michel-Kerjan & Paul Slovic eds. 2010).

199. SCOTT & STEPHAN, *supra* note 7, at 76; Jean Tirole, *Incomplete Contracts: Where do We Stand?*, 67 ECONOMETRICA 741, 743 (1994) (defining an incomplete contract as one that "does not exhaust the contracting possibilities envisioned in the complete contract").

200. For a thorough treatment of uncertainty in international relations, especially concerning the capacity of states to implement international treaties in internal policies, see GEORGE W. DOWNS & DAVID M. ROCKE, OPTIMAL IMPERFECTION? DOMESTIC UNCERTAINTY AND INSTITUTIONS IN INTERNATIONAL RELATIONS (1995); and for uncertainty in contracts as applied to public international law, see SCOTT & STEPHAN, *supra* note 7.

appropriately to changing conditions *ex post*. Changing conditions are a prevalent characteristic in international law, since most contracts are made with a long-term perspective in mind.[201] Contract theory distinguishes between uncertainty about the future (unforeseeability), uncertainty about the actions of other players (asymmetrical information), and uncertainty about the meaning and scope of the contractual provisions (that is, textual ambiguity and legal indeterminateness). Information asymmetries between the contracting parties pose the biggest problem: each party has information about itself (private information) that the other does not have, giving rise to potential opportunism. Contract theory thus analyzes problems of adverse selection, moral hazard, and verification.[202] International law is incomplete due to its long-term nature and multilateral character and often lacks specified sanctions for non-compliance. It is therefore paradigmatic for incomplete and difficultly enforceable contracts, which makes a consideration of implicit agreements and social norms even more necessary.[203]

Contract theory finds that overly strict and inflexible contracts may impair the joint surplus of the contracting parties. Thus, a trade-off arises between *ex ante* strong commitment devices, on the one hand, and flexibility *ex post* in order to uphold the efficiency of the contract, on the other hand. It is well acknowledged in contract theory that the advantage of writing a contract with "hard" and precise terms is to ensure credible commitments,[204] because "hard" and precise terms are less open to interpretation and the uncertainty of risk-shifting to the injured party is diminished.[205] But unless the parties can

> fully and accurately anticipate the conditions that exist at the time of performance, a contract containing only 'hard' terms will always turn out to be suboptimal once the future arrives . . . . In short: once conditions change, a contract with hard terms will lead to outcomes that are less desirable than those the parties would have agreed to had they known the uncertainties in advance.[206]

---

201. But, for structural flexibility mechanisms of renegotiation and duration of treaties, see Barbara Koremenos, *Loosening the Ties That Bind: A Learning Model of Agreement Flexibility*, 55 INT'L ORG. 289 (2001).

202. Contract theorists distinguish between observable and verifiable information. The former can be observed by the two parties, but it may still be that the information is not verifiable in the sense that the observing party is unable to establish the fact sufficiently to convince a neutral third party at reasonable cost, for example, the investment tribunal. For details, see SCOTT & STEPHAN, *supra* note 7, at 71–72.

203. Different sorts of uncertainty may impact compliance differently. *See infra* Part III.E.

204. For a seminal, rational choice paper on credible commitments that are especially important in systems without central enforcers, see Oliver E. Williamson, *Credible Commitments: Using Hostages to Support Exchange*, 73 AM. ECON. REV. 519 (1983).

205. Rigidity mostly comes in the form of "hard legal terms," but not necessarily. "Soft terms" or indeterminate legal terms can also be interpreted in a "hard way." Both result in a stricter view of the obligation of the injurer (states), that is, they provide less leeway for actions of the injurer (state).

206. SCOTT & STEPHAN, *supra* note 7, at 77.

Anticipating this, the parties would want to incorporate flexibility to adjust the investment whenever future circumstances render the investment no longer profitable (for either side).[207] More flexibility in turn leads to a weakening of the credibility of the parties *ex ante*. Under RC assumptions, flexible contracts often are preferable to rigid ones.[208]

BIntLE suggests a more differentiated view and reinforces the problem of the credible commitment, although the behavioral economics literature on incomplete contracts is still relatively small.[209] Hart and Moore propose an alternative and complementary view to RC, arguing that, "a contract provides a reference point for the parties' trading relationship: more precisely for their feelings of entitlement."[210] This impacts the normal trade-off between flexibility and rigidity of a contract. A flexible contract is good in that parties can adjust to the state of nature (the RC argument for flexibility), but bad in that there can be a lot of aggrievement and shading (the behavioral economics argument for rigidity).[211] A rigid contract is good in that there is little shading, but bad in that parties cannot adjust to the state of nature.[212] Contracts are always incomplete but flexible contracts are even more permissive with regard to licit behavior. Flexible contracts permit so-called shading[213] in *ex post* performance, whereas under rigid contracts much less shading can occur. Shading will occur if parties to the contract interpret "their" reference point with a self-serving bias, or if they deem contracts unfair. Although parties do not feel entitled to outcomes outside the contract, they may feel entitled to different outcomes within the contract.[214] If a

207. Scott E. Masten & Stéphane Saussier, *Econometrics of Contracts: An Assessment of Developments in the Empirical Literature on Contracting*, *in* THE ECONOMICS OF CONTRACTS, THEORY AND APPLICATIONS 283–84 (Eric Brousseau & Jean-Michel Glachant eds., 2002) ("The more uncertain the environment and the harder it is to accommodate changing circumstances within the contract, the more likely it will be that parties will sacrifice the precision and ease of implementation of definite contract terms for more cumbersome but flexible relational contract terms that define performance obligations less precisely or establish procedures for negotiating adjustments in the terms of trade within the contract.") (internal quotation marks omitted).

208. Fehr et al., *supra* note 33.

209. Botond Koszegi, Behavioral Contract Theory (Nov. 22, 2013) (unpublished manuscript), *available at* http://www.personal.ceu.hu/staff/Botond_Koszegi/JEL_Behavioral_Contract_Theory.pdf), at 52.

210. Oliver Hart & John Moore, *Contracts as Reference Points*, 123 Q.J. ECON. 1, 2 (2008).

211. They assume that a party is aggrieved and shades if he gets a worse outcome than the best possible outcome under the initial contract.

212. Fehr et al., *supra* note 33, at 494.

213. Hart & Moore, *supra* note 210, at 3 (defining shading as follows: "[W]e distinguish between perfunctory performance and consummate performance, that is, performance within the letter of the contract and performance within the spirit of the contract. Perfunctory performance can be judicially enforced, whereas consummate performance cannot. Second, we introduce some important behavioral elements. We suppose that a party is happy to provide consummate performance if he feels that he is getting what he is entitled to, but will withhold some part of consummate performance if he is short-changed—we refer to this as shading.") (footnotes and internal quotation marks omitted).

214. This is especially relevant for renegotiations. Psychologically, this means that a party is not aggrieved if the terms of the initial contract are bad, but is aggrieved if the terms of the renegotiated transaction are bad (in her view). Hart & Moore, *supra* note 210, explain this by arguing that there is a change in circumstances: the final agreement often occurs in situations of bilateral monopoly, whereas the initial contract is agreed to in a more competitive, "objective" setting where the reference point is set

party does not get what it feels entitled to, it is aggrieved and shades by providing perfunctory rather than consummate performance, thereby causing deadweight losses. This has been confirmed in experiments: grievance and shading occur mainly within (and not outside of) a given contract.[215] Thus, whereas flexible contracts dominate rigid ones under RC assumptions, this is not necessarily the case with behavioral assumptions. Given that international law lacks the same monitoring mechanisms as municipal law, it is all the more important that shading does not occur.

The problem is aggravated if one assumes, as Oliver Williamson does, opportunistic behavior by the actors, supposing that under conditions of imperfect information all transactions are affected by the problem of "self-interest seeking with guile."[216] Opportunistic behavior can itself be differentiated.[217] It can be characterized by moral hazard, which represents a deviation from joint surplus maximizing behavior within the terms of an extant contract. Yet another form of opportunism includes efforts to evade performance or to force a renegotiation of a contract's terms by imposing costs on one's contractual partner. Because the incentive to engage in such efforts is likely to be related to the *ex post* distribution of contractual surpluses (e.g., natural resource royalties), parties greatly disadvantaged by the terms of a contract are more likely to want to evade or renegotiate a previous deal. Contracting parties will seek to design contracts to divide *ex post* rents "equitably"[218] to keep the relationship within the agreement's "self-enforcing range",[219] or, equivalently, to achieve what Oliver Williamson has called "hazard equilibration."[220] This is crucial for (international) law. Constructivists try to explain fairness and how fairness impacts IR and states' behavior. In this venture, they meet with political psychology and behavioral economics.[221]

Textual ambiguity can thus be problematic from a BIntLE perspective, since such ambiguity creates space for self-serving biases and other forms of bounded rational behavior. BintLE would call for avoiding textual ambigu-

---

externally. *See also* Fabian Herweg & Klaus M. Schmidt, *Loss Aversion and Ex Post Inefficient Renegotiation* (Nov. 15, 2012) (unpublished manuscript), http://www.kellogg.northwestern.edu/research/math/seminars/201213/KlausSchmidt4-24-13.pdf.

215. Fehr et al., *supra* note 33, at 493.

216. Oliver Williamson, Markets and Hierarchies: Analysis and Anti-trust Implications: A Study in the Economics of Internal Organization 47 (1975) [hereinafter Williamson, Markets and Hierarchies]; Oliver E. Williamson, The Economic Institutions of Capitalism 255 (1985) [hereinafter Williamson, The Economic Institutions of Capitalism] (defining opportunism as follows: "the incomplete or distorted disclosure of information, especially to calculated efforts to mislead, distort, disguise, obfuscate, or otherwise confuse.").

217. Masten & Saussier, *supra* note 207, at 283.

218. Scott E. Masten, *Equity, Opportunism, and the Design of Contractual Relations*, 144 J. Inst. Theoretical Econ. 180 (1988).

219. Benjamin Klein, *Contracts and Incentives*, *in* Contract Economics 149 (Lars Werin & Han Wijkander eds., 1992).

220. Williamson, The Economic Institutions of Capitalism, *supra* note 216, at 34.

221. Goldgeier & Tetlock, *supra* note 79, at 85.

ity as much as possible and instead introduce flexibility explicitly, for example through explicit escape clauses.[222] This insight would be in conformity with the argument of legal scholars that precision of treaties is a factor of legitimacy that fosters coherence of the law and leads to a "compliance pull."[223] Furthermore, contractual dynamics that allow for recontracting with a view to fairness considerations should be taken into account. BIntLE suggests that the danger of self-serving biases in flexible contracts may lead to suboptimal performance. Prospect theory leads us to the importance of reference points and thus also to the endowment effect. Choosing different endowment reference points by states may result in competing perceptions of fairness. This impacts international conflicts that include distributional issues. Treaties can account for that by renegotiation clauses, by delegation of distributional issues to third parties, or by sliding distributional scales that can be adjusted dynamically. A further venue of research is to explore in more detail how contracts affect the preferences of elite decision-makers as well as of the population. This would require a joint effort of behavioral contract theory and constructivist approaches to international law.

D. Some Examples of Treaty Design of International Law

It would be worthwhile to go through all areas of general and special public international law and analyze different treaty provisions by means of RC as well as BIntLE. However, this would clearly exceed the scope of this article. Nevertheless, some examples will be given to illustrate the potential of BIntLE, some drawing on excellent research already conducted in BIntLE.

1. *Reservations, Optional Protocols and Objections*

Reservations are one kind of flexibility mechanisms that allow states to exclude some norms of treaties that (for whatever reason) they do not like. Although general rules are contained in the Vienna Convention on the Law of Treaties (VCLT), treaty practice is diverse. Already the legal treatment of reservations is complicated[224]: "rules governing reservations in the VCLT are complex, ambiguous, and often counterintuitive"[225] and worth a functional view. RC approaches have helped explain not only the rationale of different flexibility mechanisms, but also the importance of different norms contained in Art. 19 et seq. of the VCLT. International law and economics has been analyzing reservations from a "game theory" viewpoint, using Art. 21 (1) of

222. For other explicit flexibility mechanisms, see Laurence R. Helfer, *Flexibilty in International Agreements*, *in* Interdisciplinary Perspectives on International Law and International Relations: The State of the Art, *supra* note 9, at 175.

223. Thomas M. Franck, The Power of Legitimacy among Nations (1990).

224. For doctrinal treatments, see Guide to Practice on Reservations to Treaties, International Law Commission, 75, 63d Sess., U.N. Doc. A/66/10 (2011), ¶ 75, *available at* http://legal.un.org/ilc/texts/instruments/english/draft%20articles/1_8_2011.pdf.

225. Laurence R. Helfer, *Not Fully Committed? Reservations, Risk, and Treaty Design*, 31 Yale J. Int'l L. 367, 367 (2006); *see also* Edward T. Swaine, *Reserving*, 31 Yale J. Int'l L. 307 (2006).

the VCLT to explain why states do or do not reserve.[226] The provision establishes that reservations are reciprocal: between a reserving state and a state that objects to the reservation, the provision of the treaty will not be in force. Therefore, if a state wants to exempt itself from a treaty obligation, it must be willing to let other nations invoke that exemption vis-à-vis itself as well.[227] Parisi et al. use this as an explanation for why human rights treaties[228] have many reservations and few objections compared to reciprocal treaties. Since reciprocity does not apply to human rights treaties, the disciplinary force of VCLT, Art. 21(1) is immaterial. However, others have viewed this flexibility as ultimately fostering treaty accession as well as treaty compliance, thus affirming the traditional legal rationale of achieving universality for multilateral treaties.[229] The international law and economics framework for reservations casts doubts upon the conventional doctrinal wisdom that reservations systematically benefit reserving states and disadvantage non-reserving states. Swaine argues that the possibility of reservations enhances the depth and breadth of treaty commitments for all states and that the act of reserving reveals useful information about a state's reputation and its propensity to comply with international law. Helfer enriches this RC approach by taking a dynamic view on reservations. He argues that the theory of state interests and information usually advanced in international law and economics is essentially static and is not addressing the question of dynamic interaction of states at the date of a treaty's conclusion or to earlier ratifications by other treaty parties. Furthermore, the consideration of how reservations relate to treaties the obligations of which evolve over time in response to rulings by international tribunals or to changes in the geostrategic environment, is neglected.[230]

The analysis hitherto has employed an RC framework. This was recently enriched by a remarkable study by Jean Galbraith that utilized behavioral insights in order to analyze treaty reservations.[231] She uses the insights of Sunstein and Thaler on "choice architecture" in order to analyze the observ-

226. Francesco Parisi & Catherine Sevcenko, *Treaty Reservations and the Economics of Article 21 (1) of the Vienna Convention*, 21 BERKELEY J. INT'L L. 1 (2003). A reservation precludes the operation, as between the reserving and other states, of the provision reserved. An objection thereto leads to the reservation being in operation as between the reserving and objecting state to the extent that it has not been objected to.

227. *See also* Certain Norwegian Loans (Fr. V. Nor.), Judgment, 1957 I.C.J. 9, ¶¶ 35–36. For a description of the case, see Jochen Herbst, *Norwegian Loans Case*, *in* MAX PLANCK ENCYCLOPEDIA OF PUBLIC INTERNATIONAL LAW ONLINE (Rüdiger Wolfrum ed. 2009).

228. For a doctrinal treatment of reservations to Human Rights Treaties, see Konstantin Korkelia, *New Challenges to the Regime of Reservations under the International Covenant on Civil and Political Rights*, 13 EUR. J. INT'L L. 437 (2002).

229. *See, e.g.*, Madeleine Morris, *Few Reservations About Reservations*, 1 CHI. J. INT'L L. 341 (2000) (holding that a flexible reservations regime can be considered a human rights strategy). She is followed by Swaine, *supra* note 225 (outlining the international law and economics approach in a more detailed manner).

230. Helfer, *supra* note 225.

231. Galbraith, *supra* note 8.

able behavior of states concerning treaty reservations. For the explanation, she draws on the framing of treaty options, especially default options as discussed in the behavioral economics literature. Drawing on a treaty database, she reveals an impressive correlation between framing and state consent: where states have the implied authority to reserve out of ICJ jurisdiction, 95% continue to accept it; and where states have the explicit right to opt out of ICJ jurisdiction, 80% continue to accept it. In contrast, where states can explicitly opt into ICJ jurisdiction, only a mere 5% of state parties do so on average.[232] This replicates many studies on the importance of default rules for individuals.[233]

Her findings do not stop here but rather extend to optional protocols as well: depending on how the option for the states is framed, states will react in different manners. Since human rights treaties have a special underlying incentive structure (well-analyzed by RC scholars like Parisi et al.[234]), it is reasonable to expect human rights treaties to have more reservations than other treaties. Furthermore, under an RC perspective, it should not make a difference whether optional compliance mechanisms in human rights treaties are framed as opt-in clauses in the main text or instead offered through optional protocols since they have the same legal consequences. However, Galbraith finds that ratifying states have proved much more willing on average to embrace these commitments when they are presented in optional protocols (that is, documents separate from the main treaty) than when these commitments are presented in opt-in clauses in the main treaty. She attributes this to salience. This also has potential implications for the current negotiations under the UN Framework Convention on Climate Change (UNFCCC).[235] Using opt-out framing in treaties is preferable to opt-in if universality in large participation is desired. If there is a need to use an opt-in form, it should be framed as an optional protocol whenever possible. Galbraith suggests that framing the UNFCCC provision (UNFCCC Article 4.2 (g)) that allows non-Annex I Parties to take on commitments comparable to Annex I Parties as an optional protocol, rather than an opt-in, might have been more effective.[236]

In sum, she argues that these behavioral effects could result either from direct cognitive biases on the part of state agents, or from these agents' expectations that their constituencies are subject to cognitive biases. Galbraith thereby addresses the problems of whose bias is influencing states'

232. Galbraith, *supra* note 8, at 314.

233. Thaler & Sunstein, *supra* note 5.

234. Parisi & Sevcenko, *supra* note 226.

235. United Nations Convention on Climate Change, *entered into force* March 21, 1994, 1771 U.N.T.S. 107.

236. Foundation for International Environmental Law and Development, Behavioural insights, legal form and the future climate change agreement, Discussion 4 November 2013, Summary notes, *available at* http://www.field.org.uk/sites/field.org.uk/files/papers/FIELD%20Summary%20Behavioural%20Insights,%20Legal%20Form%20%26%20Climate%204%20Nov%202013_0.pdf.

behavior, as discussed in Part II (elite decision-makers attributed to the state). The BIntLE analysis of reservations seems convincing and would call for different treaty design. It does not contradict international law and economics but enriches the understanding of the situations in which states reserve. It thus complements the international law and economics explanation.

While a similar explanation for objections to reservations is still missing hitherto, some conjectures drawing on behavioral economics are possible. Not objecting to a reservation can be seen as the default option: an objection requires the active decision by the government to notify the objection to the depositary of the treaty, usually the UN. Doing nothing counts as acceptance of the reservation. Objections can be stated within twelve months for reservations that are made when signing the treaty (Art. 20 (5) VCLT). States can either explicitly accept a reservation or implicitly consent to it by not issuing an objection within twelve months (the latter being an opt-in rule). Objections are not common in general, but they are even less common in human rights treaties. The latter is to be expected under an RC framework as well, since reciprocity is immaterial in human rights treaties. Also, it is well-imaginable that many states are unable to monitor all reservations made by all countries and determine whether they want to object to them or not. This might be a problem of technical capacity and human resources, especially in small and poor countries. But it might also be a problem of procrastination and default rules: objections are equivalent to an opt-in rule. Thus, if the integrity of treaties is normatively desired, the default option can be changed (if not in the VCLT, then tailor-made in specific treaties). This is especially important for treaties without a court deciding on the validity of the objection.

BIntLE thus suggests additional design options for treaty makers. Now, states either allow reservations on all norms, prohibit reservations on all norms, or allow reservations for some but not others in one treaty. But a more fine-grained approach would use objections through consent-oriented decentralization in the context of VCLT, art. 19(c). For example, they could change the default option of interpreting the omission of an objection as an implicit consent to another default rule where the omission by silence is interpreted as an objection (opt-out rule). This would respect the consent of states while taking into account their biases. It would also help to solve the problem where VCLT, art. 19(c) permits *de facto* (self-serving) auto-interpretation if there is no court available to decide whether a reservation is incompatible with the object and purpose of the treaty. The decentralized solution of VCLT, art. 19(c) can thus be used in the reverse: if a certain percentage of states do not explicitly accept the reservation, it might be deemed incompatible.

Especially for human rights treaties, it has been suggested that reservations should not be allowed or that they should be more restricted.[237] Instead of prohibiting reservations *ex ante*, a procedural solution can be used. An example to draw on is Art. 20 (2) of the International Convention on the Elimination of all Forms of Racial Discrimination,[238] pursuant to which a two-third majority of state parties is needed to invalidate a reservation under VCLT Art. 19 (c).[239] This majority rule may be reversed, such that the consent to a rule where the minority of states' objections (for example, 10%) is enough to invalidate the reservation (reversed consent). This latter rule is especially interesting for regimes for which reciprocity is immaterial (for example, human rights or some parts of environmental regimes) since, from an RC perspective, incentives for objections are weak due to the absence of the reciprocity mechanism of VCLT Art. 21 (1).[240] Absence of protest would then no longer be interpreted as consent. Given the significance of opt-in as opposed to opt-out options in treaty reservations, it seems manifest that the same apply to objections to treaties—even more so since there is less rational self-interest involved than in treaty reservations that concern the state itself.

Instead of reservations and objections thereto, non-mandatory norms in treaties may also be used ("should" or "shall consider" in contrast to "shall adopt"), as, for example, in the United Nations Convention Against Corruption[241], which includes many non-mandatory norms[242] and also allows for reservations. Both are a consequence of insufficient consent of the state parties to the substantive norms by which they would be bound. Whether non-mandatory norms or reservations are used should not matter from an RC perspective. Under BIntLE, the conjecture would be that non-mandatory norms are opt-in norms, whereas reservations are opt-out. Thus, if it is

237. Human Rights Committee, General Comment 24 (52), reservations to the ICCPR, U.N. Doc. CCPR/C/21/Rev.1/Add.6, ¶ 8 (Nov. 4, 1994); Ryan Goodman, *Human Rights Treaties, Invalid Reservations, and State Consent*, 96 Am. J. Int'l L. 531 (2002).

238. International Convention on the Elimination of All Forms of Racial Discrimination, G.A. Res. 2106 (XX), Annex, 20 U.N. GAOR Supp. (No. 14), U.N. Doc. A/6014 at 47 (1966) entered into force on Jan. 4, 1969.

239. "A reservation incompatible with the object and purpose of this Convention shall not be permitted, nor shall a reservation the effect of which would inhibit the operation of any of the bodies established by this Convention be allowed. A reservation shall be considered incompatible or inhibitive if at least two thirds of the States Parties to this Convention object to it." *Id.* art. 20, ¶ 2.

240. Human Rights Committee, General Comment 24 (52), reservations to the ICCPR, U.N. Doc. CCPR/C/21/Rev.1/Add.6, ¶ 17 (Nov. 4, 1994) argues that human rights treaties, "and the Covenant specifically, are not a web of inter-State exchanges of mutual obligations. They concern the endowment of individuals with rights. The principle of inter-State reciprocity has no place, save perhaps in the limited context of reservations to declarations on the Committee's competence under article 41. And because the operation of the classic rules on reservations is so inadequate for the Covenant, States have often not seen any legal interest in or need to object to reservations. The absence of protest by States cannot imply that a reservation is either compatible or incompatible with the object and purpose of the Covenant."

241. United Nations Convention Against Corruption (UNCAC), G.A. Res. 58/4, U.N. Doc, A/58/4 (Nov. 21, 2003), entered into force Dec. 14, 2005.

242. *See, e.g.*, *id.* art. 18 (trading in influence); art. 19 (abuse of functions); art. 21 (bribery in the private sector).

deemed desirable that states follow certain provisions, it is advisable to use reservations rather than non-mandatory norms in the treaties.

2. *Linkage Issues*

Linking treaties has been considered an apt mechanism for concluding treaties (allowing for a wider range of bargaining) and also for inducing compliance (such as linking a human rights treaty or an environmental treaty to a trade treaty).[243] The question of how and why issues are linked in the international sphere has been thoroughly discussed.[244] Also, asymmetric security treaties contain linkages that induce states to join treaties which they otherwise would not necessarily have joined; the same is found in conditions of trade access or the Generalized System of Preferences. Koremenos et al. have several conjectures about issue-linkages in treaties and international cooperation. For example, linkage "not only allows states to increase efficiency but may also allow them to overcome distributional obstacles."[245]

This conjecture can be questioned from a BIntLE perspective. Actors in bargaining situations treat their own concessions as losses and the concessions they receive from others as gains.[246] They will tend to overestimate the value of their own concessions and undervalue those of their adversary, rendering bargaining more difficult and thereby explaining impasse. Parties would therefore risk the negative consequences of a possible deadlock in order to minimize their concessions. This suggests a reluctance to accept losses on any dimension of linked agreements.[247] Bargaining over the allocation of losses leads to less agreement than bargaining over gains, and this asymmetry suggests that the framing of the situation is crucial. Linking issues might thus be viewed a bit more pessimistically than RC approaches suggest.

Behavioral economics thus suggests some limits to the theory of linkages in international treaties that holds that such linkages are open bargaining spaces. It might explain part of the current deadlock at the Doha Round in the WTO, since it reduces the size of the bargaining space. Although the last Ministerial Conference in December 2013 has made a bit of progress on the Doha Development Agenda,[248] many crucial issues are still on the table,

243. Paul Poast, *Can Issue Linkage Improve Treaty Credibility? Buffer State Allliances as a 'Hard Case'*, 57 J. Conflict Resol. 739 (2012) (finding higher compliance rates).

244. *See generally* Barbara Koremenos et al., *The Rational Design of International Institutions*, 55 Int'l Org. 761 (2001).

245. *Id.* at 786.

246. Levy, *supra* note 90, at 290; Kahneman et al., *supra* note 34, at 1345 (calling this the "concession aversion").

247. Levy, *supra* note 90, at 290.

248. The WTO Ninth Ministerial Conference, held in Bali, Indonesia, from the 3rd to 7th of December 2013, adopted the "Bali Package," a series of decisions aimed at streamlining trade, allowing developing countries more options for providing food security, boosting least-developed countries' trade and helping development more generally. Nevertheless, a lot of things highly relevant were left on the table, such as in trade in services, in agriculture and in intellectual property. States have only agreed on issues

such as market access or the elimination of export subsidies in agriculture (supposedly seen as a loss by subsidizing countries).[249] A pluri-lateral approach, such as the one taken with the Agreement on Public Procurement,[250] instead of the approach that "nothing is agreed until everything is agreed" might have led to more agreement, such as in the field of investment or competition.[251] For treaty design, this warns against linking too much. It is, for example, now planned in the regional trade agreements such as the Transatlantic Trade and Investment Partnership (TTIP) or the Trans-Pacific Partnership (TPP), which include trade in goods (including non-tariff measures), services, intellectual property, as well as investment.

3. *Treaties dealing with Commons Problems*

A salient issue in international law is the commons problem, prevalent in all constellations where a good is non-excludable and rivalrous. The RC approach by Hardin[252] articulates a pessimistic view of the "tragedy of the commons." Given the incentive structure of social dilemmas, one would predict inefficient excess appropriation of common-pool resources; this is often found.[253] In situations of very incomplete and difficult-to-enforce contracts, there are ample opportunities for taking advantage of one's exchange partners. This axiom of self-interested behavior implies that the only remedy for the compliance problem lies with the provision of sufficient rewards or sanctions. The effectiveness of these treaties is thus often questioned by international law and economics scholars.[254]

Hardin's rational choice view has been challenged by numerous field studies reported in Ostrom's seminal book.[255] Ostrom takes a behavioral economics approach, emphasizing that people may be able to govern the commons, depending on behavioral factors, institutions, and motivations. Thus, under certain circumstances, people are able to cooperate and improve their joint outcomes, especially when in small groups. Since international law is concerned with cross-border or international commons, Ostrom's research can be scaled up to see under what conditions commons may be sustained on the international plane. Behavioral research has yielded evidence indicating that other-regarding preferences are decisive for explaining col-

---

that hurt nobody. *See* World Trade Organization, Bali Ministerial Declaration, WT/MIN(13)/DEC/W/1/Rev.1 (2013).

249. Tim Josling, *A Post-Bali Agenda for Agriculture*, in Building on Bali: A Work Programme for the WTO 105 (Simon Evenett & Alejandro Jara eds., 2013).

250. Agreement on Government Procurement, April 12, 1979, 1869 U.N.T.S. 508.

251. For that problem, see also Anabel Gonzalez, *The Rationale for Bringing Investment into the WTO*, in Building on Bali: A Work Programme for the WTO, *supra* note 249, at 72.

252. Garrett Hardin, *The Tragedy of the Commons*, 162 Science 1243 (1968).

253. The problem of overfishing is prevalent even in highly institutionalized settings such as the European Union.

254. *See, e.g.*, Posner & Sykes, *supra* note 7, at 232.

255. *See generally* Elinor Ostrom, Governing the Commons: The Evolution of Institutions for Collective Action (1990) (detailing field studies challenging Hardin's rational choice view).

lective action and multilateral cooperation.[256] Can this insight be used for states' behavior in public goods and commons situations (including normative public goods such as human rights)? As behavioral economists would argue, it is the ubiquity of cheating opportunities that renders altruistic cooperation important. Altruistic cooperators are willing to cooperate—that is, to abide by the implicit agreement—even though cheating would be beneficial for them. The research indicates that altruistic cooperation is an important behavioral force.[257] Social scientists have even hypothesized that economic incentives might undermine altruism.[258] Surely, states as such cannot not be altruistic (they do not have emotions). But once we break up the black box and look at the Lego pieces, we can attribute behavior of elite decision-makers or analyze internal mechanisms within the state, including market mechanisms where individual preferences count. States may behave altruistically because of their internal mechanisms, and thereby altruistic behavior of individuals might play a role in international law as well. The unit of analysis is thus not the state but individual behavior relevant to international law in the aggregate. In laboratory experiments, it was possible to isolate two special factors that lead to cooperation in commons problems: sanctioning and communication (the underlying motivation being a preference for reciprocity *strictu sensu* and equity).[259] One may assume that individual fairness preferences might also exist in issue areas covered by international law. People make sacrifices in order to punish free-riders. When given credible information, they would, for example, not buy fish that are close to extinction or refrain from purchasing products produced in an unsustainable manner or through child labor. This is relevant whenever the state is disaggregated and individual behavior has an influence on either civil society behavior or market behavior.[260]

E. Compliance with International Law

IR and international law scholars are deeply divided on the reasons why states comply (or not) with international law.[261] The RC approaches in IR,

256. Fehr & Schmidt, *supra* note 55.

257. Ernst Fehr & Simon Gächter, *Altruistic punishment in humans*, 415 Nature 137 (2002).

258. Bruno S. Frey, Not Just For the Money: An Economic Theory of Personal Motivation 13–19 (1997); Uri Gneezy & Aldo Rustichini, *A Fine is a Price*, 29 J. Legal Stud. 1, 15 (2000) (a field study in a group of day-care centers where introducing a monetary fine for late-coming parents led to a significant increase in the number of late-coming parents).

259. Falk et al., *supra* note 73, at 158–63.

260. *See* van Aaken, *Trust, Verify or Incentivize? Effectuating Public International Law Regulating Public Goods Through Market Mechanisms*, *supra* note 139.

261. Already the notion of compliance is disputed. For a discussion of the different notions, see Jana von Stein, *The Engines of Compliance*, *in* Interdisciplinary Perspectives on International Law and International Relations: The State of the Art, *supra* note 9, at 493; Kal Raustiala & Anne-Marie Slaughter, *International Law, International Relations and Compliance*, *in* Handbook of International Relations 538, 539 (Walter Carlsnaes et al. eds., 2002) ("compliance as a state of conformity or identity between an actor's behavior and a specified rule"). There is consensus that compliance does not

international law, and economics look at external restrictions, rather than preferences, as a reason for compliance. Whereas at one end of the spectrum, states only accede and comply with international law if their utilitarian calculus so demands (logic of consequences), on the other end of the spectrum states comply not only because it is in their own best interest but because they believe there are good reasons to do so (logic of appropriateness). Constructivists would hold that international institutions can change the perceptions and conceptions of state interests. Actors can be guided by moral, religious, or ideological ideals that do not follow consequential reasoning.[262] Also, fairness and legitimacy are considered causal factors in international law compliance. Norms are accepted if they came into being through a fair process and are used and applied in a fair process.[263] Legal clarity and equality are important components of this. This closes the gap to the behavioral economics approaches (for example, in contract theory) as well as the importance of procedures in de-biasing as analyzed in behavioral law and economics. The following parts take the three Rs—retaliation, reciprocity, and reputation—of compliance used in international law and economics by Andrew Guzman in turn and analyze each from a BIntLE perspective.[264]

1. *Retaliation*

Retaliation is the direct sanctioning of a violating state. Retaliation is modeled as a prisoner's dilemma game since it is costly for the retaliating state. Especially in multilateral treaty settings (*inter omnes* norms or violation of custom), states face a classical second order enforcement problem. Every state is better off if other states retaliate and prefers to free-ride. Insights from behavioral economics would mitigate this problem and predict that a greater number of retaliation incidents would occur, be it through collective or unilateral sanctions. Abstracting from the group problem, behavioral economics would suggest that norm violators are punished more often than RC would predict, at least in cases where the intentions of the violating state are viewed as malevolent. Behavioral economics has tested three major potential factors for informal sanctioning, namely, (1) strategic sanctioning for selfish reasons, (2) non-strategic sanctions driven by spitefulness,[265] and (3) non-strategic sanctions driven by the violation of fairness principles. The third

necessarily mean that "international law matters," it just captures conformity, not causality of the law for the behavior of states.

262. Goldgeier & Tetlock, *supra* note 79, at 81.

263. Tom R. Tyler, Why People Obey the Law 108–12 (1990); Tom R. Tyler, *Psychological Perspectives on Legitimacy and Legitimation*, 57 Annu. Rev. Psychol. 375, 382–84 (2006).

264. *See generally* Guzman, *supra* note 7. Guzman draws on Robert O. Keohane, International Institutions and State Power: Essays in International Relations Theory 29-30 (1986).

265. Falk et al., *supra* note 59, at 2024 (defining spiteful sanctions as "those that occur because the sanctioning subject values the payoff of the sanctioned subject negatively, irrespective of whether the sanctioned subject behaved fairly or unfairly, and irrespective of the distribution of presanction payoffs").

factor is the most important in experiments with individuals. The motivational force behind punishment is the desire to harm those who committed unfair acts. In international law, the first factor is often asserted: for example, the United States' intervention in Iraq in 2003 alleging that the United States wanted to secure its oil supplies. Retaliation was attributed to strategic interests (that is, rational choice was the behavioral assumption). But states do not only intervene if it is in their strategic self-interest to do so. Although there might be a mixed motive game, egregious behavior by a violating state can lead to fairness considerations, especially in cases of humanitarian disasters (and especially if reactions of the electorate are taken into account). BIntLE would thus be less pessimistic about retaliation in international law than the RC approach, especially in public good games. This does not mean that fairness considerations alone guide states' behavior. Instead, BIntLE would argue that when, for example, humanitarian intervention is being discussed, fairness considerations would apply greater weight on one side of the scale than under RC assumptions.

Behavioral economics has also shown that unaffected third parties punish violations of social norms quite frequently, but the punishment is less strong than the punishment by an affected second party.[266] This leads to the differentiation between Art. 42 and Art. 48 of the Draft Articles of State Responsibility.[267] The Draft Articles distinguish between the invocation of responsibility of injured and non-injured states. Injured states can invoke any remedy permissible in international law, whereas non-injured states can invoke only specified remedies under Art. 48 (2). The Draft Articles allow erga omnes norms to be enforced by non-injured states in Draft Articles art. 48, although they differentiate between obligations *erga omnes partes* (Draft Articles, art. 48 Para 1 (a)) that target a specific group that has a collective interest (for example, a region) on the one hand, and obligations owed "to the international community as a whole" on the other hand.[268] The latter obligations target either club goods[269] or global public goods.[270] In the latter case, more than in the former, it is expected that enforcement by other states will be weak as the free-rider problem comes to the fore (under RC assumptions, confirmed but attenuated by BIntLE). Effectuating the responsibility of states in cases of obligations erga omnes is very difficult. The problem is mitigated for obligations *erga omnes partes* under certain circumstances. From an RC perspective, the Draft Articles would have needed to

266. Ernst Fehr & Urs Fischbacher, *Third-party Punishment and Social Norms*, 25 EVOL. HUM. BEHAV. 63 (2004).

267. UN GAOR, 56th Sess., Supp. No. 10, at 43, UN Doc. A/56/10 (2001) [hereinafter Draft Articles].

268. *See* JAMES CRAWFORD, THE INTERNATIONAL LAW COMMISSION'S ARTICLES ON STATE RESPONSIBILITY 278 (2002).

269. Club goods are goods that are non-rivalrous but excludable (for example, satellite television). *Cf.* James M. Buchanan, *An Economic Theory of Clubs*, 32 ECONOMICA 1 (1965).

270. Public goods are non-rivalrous and non-excludable.

differentiate according to circumstances and applicable remedies in order to incentivize and induce non-injured states in public good constellations to be internationally responsible.[271] But Draft Articles art. 48 (2) (a) confines any non-injured state to request cessation of the wrongful act, and if the circumstances so require, assurances and guarantees of non-repetition under Draft Articles art. 30. Furthermore, non-injured states can claim reparations (including restitution) from the responsible state but such a claim must be made in the interest of the injured state, if any, or of the beneficiaries of the breached obligation. From an RC perspective, this amounts even more to the provision of a public good, since the claiming state acts on behalf of the community as a whole. Few claims under article 48 of the Draft Articles are to be expected, unless the state has a strategic interest. The problem is partially mitigated if BIntLE is taken into account since it is foreseeable that states, too, claim for fairness reasons (they also sanction more than would be expected, since economic sanctioning is costly to the sanctioning state, especially if the sanctions are not targeted since exporters are hurt and possibly imports too). This might have been the case in the South West Africa Cases of the ICJ, where Ethiopia and Liberia sought declarations on behalf of (now) Namibia in 1960 and challenged the continued existence of the mandate for South West Africa and the duties and performance of South Africa as a Mandatory. In the second phase, the Court found in a divided vote that the applicant states could not be deemed to have established any legal right or interest in the subject matter of their claims and accordingly decided to reject the claims.[272]

2. *Reciprocity*

Reciprocity is seen as a central mechanism for compliance in international law: it mainly concerns horizontal effectuation of international law.[273] In RC models,[274] reciprocity is viewed as an external mechanism following the logic of consequences. Game theory in the form of a repeated prisoner's di-

271. This kind of incentive mechanism is used, for example, when incentivizing whistleblowers in the United States. Under the Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, § 922, 124 Stat. 1367, whistleblowers are provided with rewards, ranging from 10 percent to 30 percent of the amounts collected by the SEC. *See also* Securities Whistleblower Incentives and Protections, 76 Fed. Reg. 34, 300 (June 13, 2011).

272. There are a limited number of cases concerning club goods. For example, in the South West Africa Cases, the ICJ decided not to deliver a judgment on the merits of the case, holding that "to a plea that the Court should allow the equivalent of an actio popularis, or right resident in any member of a community to take legal action in vindication of a public interest. But although a right of this kind may be known to certain municipal systems of law, it is not known to international law as it stands at present." South West Africa (Eth. v. S. Afr.), Preliminary objections and judgment, 1962 I.C.J. 46 (Dec. 21); South West Africa (Eth. v. S. Afr.), Second phase judgment, 1966 I.C.J. 46 (Jul. 18) (internal quotation marks omitted). The United Nations General Assembly later decided on the continued presence of South Africa in Namibia. G.A. Res. 2145 (XXI), U.N. GAOR 21st Sess., Supp. No. 16, (Vol. II), U.N. Doc. A/RES/2145(XXI) (Oct. 27, 1966).

273. *See* Stein, *supra* note 261.

274. VCLT art. 60 embodies this principle of reciprocity.

lemma is used to model the strategies of whether or not to comply. In bilateral settings, the idea is straightforward: a potential defector will defect if the gains from future cooperation are smaller than the gains from defection. Hence, such reciprocal treaties are inherently self-enforcing.[275] The threat to react to a potential violation could already prevent this very same violation, given that the shadow of the future is large enough. Threats of mutual defection are more likely to sustain cooperation. RC scholars are more skeptical concerning the force of reciprocity in public good settings, such as fisheries: "at one point it becomes unrealistic to expect that states can produce the cooperative outcome. However, it is difficult to say what this point is."[276] RC scholars would assume that cooperation is more likely when defection by one party can easily be detected. Also, RC would acknowledge that the rules of the game need to be clear—that is, it must be clear what exactly constitutes a defection.[277]

Reciprocal behavior in behavioral economics differs fundamentally from the RC approach where "cooperative" or "retaliatory" behavior in repeated interactions is motivated by future (material) benefits. As described above, reciprocal behavior in non-repeated interactions is termed called "strong reciprocity" (since RC would assume defection) in contrast to "weak reciprocity" that is motivated by long-term self-interest in repeated interactions.[278] This distinction already calls for a more differentiated hypothesis regarding reciprocity in international law. As behavioral economics has shown, reciprocity as a reaction to the violation of other actors is often based on whether an action is perceived as kind or unkind. Thus, BIntLE would predict that in situations of "strong reciprocity," if the other player is perceived as kind (again, perceived intentions matter), the treaty will not be terminated, and the defector will not be punished. If non-contribution to a multilateral treaty is not attributed to bad intentions but instead to technical problems, reactions of other states will be different and "kind." In contrast, if the defection is perceived as intentional as well as unfair and "unkind," stronger reactions are to be expected. Although the aforementioned point of cooperative breakdown is still unknown, this would suggest that the point is further away. That is, more cooperation should be observed—even in multilateral public good games—than the RC approach predicts.

Furthermore, BIntLE can also be applied for the effectuation of international law when breaking up the black-box state; the relevant unit of the analysis is individuals in the aggregate. Reciprocity plays not only interstate but also between nonstate actors. One prominent example would be

275. *See* Beth Simmons, Mobilizing for Human Rights: International Law in Domestic Politics 116 (2009) (providing a definition for self-enforcing agreements).

276. Posner & Sykes, *supra* note 7, at 30.

277. *See id.* at 30; Parisi & Ghei, *supra* note 64.

278. Kolm, *supra* note 63.

corruption fighting as developed in UNCAC.[279] Behavioral economics can inform how to destabilize the reciprocity of illicit corruption contracts and networks. For example, there have been experiments showing that accepting and offering bribes through intermediaries carries a lighter moral cost than direct bribery. Trust destabilizing mechanisms have been shown to be very effective in experiments, thereby undermining the reciprocity underlying corrupt transactions.[280] There is a plethora of institutional mechanisms to fight corruption. The international law on corruption should also be behaviorally informed.

3. *Reputation*

Reputation-building has been characterized as "one of the most talked-about, yet least understood strategic phenomenon [sic] in international relations."[281] Whereas realist RC scholars are more skeptical about the role of reputation in effectuating international law, Guzman[282] and others[283] see it as a main driver for compliance, formalizing reputational concerns as a repeated prisoner's dilemma game. It seems to work under some circumstances but not others.[284] In addition, non-RC scholars attribute causal force to reputation for compliance with international law. In a more optimistic model, the main idea is that a state that violates international law develops a bad reputation that in turn leads other states to exclude the violator from future cooperative opportunities. Anticipating a loss of future gains, states will more often comply with international rules that are not in their immediate interests than would be expected under a realist approach.[285] Also, international law scholars usually attribute compliance to reputational forces, but when "we look at how governments actually make decisions, the reputational costs of violating international rules are likely to be significantly lower than international relations theorists and legal scholars commonly think."[286] As Brewster argues, two questions have been neglected by

279. *See* United Nations Convention Against Corruption, *supra* note 241.

280. For a survey on the experiments and institutional recommendations, see JOHANN GRAF LAMBSDORFF, *Behavioral and Institutional Economics as an Inspiration to Anticorruption—Some Counterintuitive Findings*, *in* ROUTLEDGE HANDBOOK OF POLITICAL CORRUPTION (Paul Heywood ed. 2015) (forthcoming, on file with the author).

281. Dustin H. Tingley & Barbara F. Walter, *The Effect of Repeated Play on Reputation Building: An Experimental Approach*, 65 INT'L ORG. 343, 343 (2011).

282. GUZMAN, *supra* note 7; Guzman, *A Compliance-Based Theory*, *supra* note 86.

283. *See* KEOHANE, *supra* note 93 (naming several reasons for compliance with international law: "reasons of reputation, as well as fear of retaliation and concern about the effects of precedents"). For an encompassing discussion, see Rachel Brewster, *Unpacking the State's Reputation*, 50 HARV. INT'L L.J. 231 (2009) [hereinafter Brewster, *State's Reputation*], and Rachel Brewster, *Reputation in International Relations and International Law*, *in* INTERDISCIPLINARY PERSPECTIVES ON INTERNATIONAL LAW AND INTERNATIONAL RELATIONS: THE STATE OF THE ART, *supra* note 9, at 524 [hereinafter Brewster, *Reputation in International Relations*] (with further references that non-rationalists also rely on more intuitive notions of reputation).

284. *See* Tingley & Walter, *supra* note 281.

285. *See, e.g.*, GOLDSMITH & POSNER, *supra* note 7, at 101.

286. Brewster, *State's Reputation*, *supra* note 283, at 232.

the compliance literature, namely, (1) whose reputation; and (2) a reputation for what? Within the RC approaches,[287] there is consensus neither on whether reputation is ascribed to one actor, namely the state,[288] or to a government, nor on reputation for what—unitary reputation or issue specific reputation.[289] Whereas some scholars see reputation as unitary (ascribed to a state, crossing over different issues), others are less confident and attribute it mainly to governments (since a state's reputation can change quickly with a change in government) and argue that reputation is issue-specific: the latter view considerably weakens reputation's causal force. Furthermore, if a reputation is attributed to a government, the time frame may be much shorter (surely so in democracies) and the reputation may easily be repaired if a new government comes to power. Reputation thus loses its compliance-inducing force. But this can be questioned if one takes into account the fundamental attribution error found by political scientists. Behavior is usually attributed not only to a government (even if the leader acts against the majority will of the people) but also to the people, that is, rather, the state.[290]

Under RC models of compliance, reputation is used as a concept to solve problems, that is, as an independent variable. How reputations form has not been researched, that is to say, reputation forming has not been treated as a problem in itself.[291] But this formation might be crucial to understand states' behavior (of the acting state as well as of other states) and might prove indispensable in treaty design as well. The discussion has hitherto not taken into account psychological research on reputation, with the exception of Brewster.[292] The psychological component cuts both ways in changing the analysis of reputation as a compliance mechanism: games can have not only logical, but also psychological equilibria.[293] Political psychologists are much more skeptical about reputation as a factor in international relations since it is a *relational* concept that is influenced by the actor, the audience, and the situation.[294] They complain that "our theoretical understanding of reputation remains shallow."[295] Reputation is less an issue of the objective behavior of a state than an issue of the perception of that behavior by other states. There are thus several factors that can influence the strength of the

287. For an exhaustive discussion of the RC approaches, see *id.*

288. *See* GUZMAN, *supra* note 7.

289. *See* George W. Downs & Michael A. Jones, *Reputation, Compliance, and International Law*, 31 J. LEGAL STUD. 95 (2002) (arguing for issue specific reputation, just as POSNER & SYKES, ECONOMIC FOUNDATIONS OF INTERNATIONAL LAW, *supra* note 7, at 33, following Brewster, *State's Reputation*, *supra* note 283.

290. *See* Allison & Messick, *supra* note 134.

291. *See* JONATHAN MERCER, REPUTATION AND INTERNATIONAL POLITICS 3–4 (2005).

292. *See* Brewster, *Reputation in International Relations*, *supra* note 283.

293. *See* Goldgeier & Tetlock, *supra* note 79, at 77.

294. *See* MERCER, *supra* note 291, at 6; *see also* Brewster, *Reputation in International Relations*, *supra* note 283, at 531 (partially discussing Mercer's arguments).

295. MERCER, *supra* note 291, at 3.

reputational factor in compliance. Let us first consider the acting state before turning to the perception of that state.

A state (or a government[296]) may care more or less about its reputation, depending on its risk appetite. Additionally, loss aversion may explain why states are more concerned with preventing a decline in their reputation or credibility than increasing it. This might explain the relative scarcity of treaty exits.[297] It might also make the reputational effect of international violations stronger as a compliance mechanism. Prospect theory also suggests that those effects would be even stronger if the threat of loss of reputation were perceived to be certain. Precise treaty mechanisms with clear sanctions may help. Automatic referral to an international court (whose negative decision affects reputation), for example, makes sanctions much more likely (which are therefore perceived as almost certain, and will not be ignored).

Turning to the other states—that is, to how behavior is ascribed and evaluated—it is noteworthy that this perception is relevant to changes in the reputation of a state: "[f]or reputation to be an effective enforcement mechanism, other governments must actually draw the informational inferences that the state is a desirable or an undesirable treaty partner and this must then motivate the acting government."[298] However, how this information is perceived and interpreted is subject to psychological analysis. First, there might be informational problems: the international system is noisy and thus reputations may be difficult to form since different states perceive actors and situations differently. This might make the mechanism less strong than has hitherto been assumed. Informational problems could still be captured by an RC model. How this grey zone of indeterminacy is interpreted and perceived, however, is a question of psychology. Two conditions are necessary to form a reputation: first, the behavior of another state must be explained by dispositional (type) or situational terms, and second, those explanations must be used to predict behavior in the future. Reputation can only form if behavior is attributed to disposition, and the reputation is then used to predict behavior in the future. Since reputation is a judgment about dispositional and not situational factors, only the former can generate a reputation. Thus, if behavior is attributed to the situation, the reputation of the actor is not affected, since situational attribution does not have cross-situational validity: situations, and thus the expected behavior, vary by type.[299] Also, states can have a reputation for different things, for example, being a good global citizen or being compliant with international law. Although these may overlap, it need not be the case.[300]

296. *See id.* at 26; Brewster, *State's Reputation*, *supra* note 283, at 249–57.
297. *See* Laurence R. Helfer, *Exiting Treaties*, 91 VA. L. REV. 1579 (2005).
298. Brewster, *State's Reputation*, *supra* note 283, at 244.
299. *See* MERCER, *supra* note 291, at 6, 36.
300. *See* Brewster, *State's Reputation*, *supra* note 283, at 238.

Furthermore, experiments with elite decision-makers (British policymakers) as well as the general population (in the United States) have shown that individuals are far more likely to oppose policies that would violate international law than otherwise identical policies that would not. International law changes preferences and expectations mainly via anticipated reputational ramifications.[301] It also suggests that a state will try to influence the narrative of its behavior since it might not want to be seen as a "bad type" that leads to less cooperation and more punishment. Psychological and behavioral economics research has shown that much depends on whether a given actor's behavior is attributed to the situation or to the "type." Whereas the Obama administration has argued that it would lose its reputation if it does not intervene in Syria (disposition),[302] many spectators may attribute the behavior to the situation, namely that—given the situation in Syria—an intervention would cause an even worse outcome than that if no intervention takes place (in terms of human suffering).[303] A lot depends on how the allies and adversaries of a state interpret its behavior. Social psychology has long researched the perception of cause (whereas RC scholars have not) by developing theories of causal attribution. Attributions are inferences made to explain events. Attribution to disposition is a necessary condition for a state's reputation to be formed. It has been well-documented that actors attribute events differently depending on whether the behavior comes from another actor within the group or from outside the group: "[d]esirable out-group behavior elicits situational explanations, and undesirable out-group behavior elicits dispositional explanations."[304] The reverse holds true for in-group behavior. The tendency to attribute behavior to dispositional factors and to discount situational factors is well-acknowledged in psychology: a so-called fundamental attribution error or "over-attribution."[305] Mercer connects over-attribution to the salience and anchoring bias of Kahneman/Tversky.[306] Once an initial attribution is made, people are quite resistant to adjusting their views in order to accommodate new information.[307] People also tend to attribute causality; attributing events to happenstance is avoided. Salience explains over-attribution. Thus, if a situation is made salient, people over-attribute to the situation and neglect disposition (and vice versa).[308]

301. Michael Tomz, *Reputation and the Effect of International Law on Preferences and Beliefs* (Feb. 2008) (unpublished draft), http://www.stanford.edu/~tomz/working/Tomz-IntlLaw-2008-02-11a.pdf.

302. This seems to be a common argument of American presidents. *See* MERCER, *supra* note 291, at 2 (discussing, and finding to be erroneous, the reputation for resolve—that is, that others will be more likely to believe one's threat and therefore are less likely to challenge one's own interest).

303. *Id.* ("We cannot sensibly assess the grim trade-off between lives lost and reputation gained until we understand how and when a reputation for resolve forms.").

304. *Id.* at 46.

305. *Id.* at 49 *et seq.* (with further references).

306. *Id.* at 50.

307. *See* Tversky & Kahneman, *Judgment Under Uncertainty*, *supra* note 21.

308. *See id.*

In short, there are many biases and heuristics that may impact a state's reputation. Awareness of this might lead states to alter their communication strategies (for example, in front of international courts). Are there any mechanisms that allow states to strengthen the reputational mechanism? In order to reduce the informational deficit in the international system, mechanisms for better fact-finding and monitoring could be introduced. Although many treaties already contain such mechanisms, treaty makers should be more aware of the importance of those for effectuating the reputational mechanism. Furthermore, over-attribution can be attenuated by accountability mechanisms, that is, the need to justify one's causal inferences by giving reasons for those inferences.[309]

Another mechanism to deal with the fundamental attribution error in international law is a clearer assignment of responsibilities, especially if states are acting jointly or together with other actors, such as NGOs or IOs. Attenuating the attribution errors depends on whether there are legal criteria for attribution, an unclear notion in international law.[310] Although the Draft Articles for State Responsibility[311] outline some, they might be insufficient. The criteria include capacity,[312] contribution, control,[313] and causation (amongst others).[314] If the criteria are not well-defined *ex ante*, noise is created concerning the question of who is to blame and why. This leaves leeway to the interpretation of the observers, and thus diminishes reputational effects. Furthermore, even if the attribution criterion is clear, observability and verifiability of the behavior is often limited. The best option here would be to extend monitoring mechanisms. Although present in current treaties, they ought to be emphasized even more. Only then can murky responsibilities—where causality is not easily observable or verifiable—be clarified, thus allowing for individual state responsibility that is proportional to the harm (whatever the criterion for attribution is).

Since the relevance of beliefs about intentions for the behavior of a given state is crucial in order for a reputation to be formed—and thus for reputation being a force for compliance—legal mechanisms must 1) make those

309. *See* Philip E. Tetlock, *Accountability: A Social Check on the Fundamental Attribution Error*, 48 Soc. Psych. Q. 227 (1985).

310. David D. Caron, *The Basis of Responsibility: Attribution and Other Transsubstantive Rules*, *in* The Iran-United States Claims Tribunal: Its Contribution to the Law of State Responsibility 109, 180–81 (Richard B. Lillich & Daniel B. Magraw eds., 1998).

311. *See* Draft Articles, *supra* note 267; Int'l Law Comm'n, Draft Articles for the Responsibility of International Organizations, ¶ 88, U.N. Doc. A/66/10 (2011).

312. This is exemplified in climate change discussions.

313. This is mainly relevant in peace-keeping missions.

314. For a discussion of shared responsibilities in international law, see André Nollkaemper & Dov Jacobs, *Shared Responsibility in International Law: A Conceptual Framework*, 34 Mich. J. Int'l L. 359 (2013). For a political economy perspective, see Anne van Aaken, *Shared Responsibilities in Internation Law: A Political Economy Analysis*, in Distribution of Responsibilities in International Law (André Nollkaemper and Dov Jacobs eds., forthcoming 2014).

beliefs more reliable, and 2) sort intention from outcome. Blame games are played not only in national law but also in international law.[315]

IV. Conclusion and Outlook

This Article is but a start for outlining where research in international law using behavioral economics (which has been so successfully applied to municipal law, for example, in the US, the UK, and by the European Commission) might go, and where it might encounter difficulties. The goal is not to "*dis*prove" rational choice (or game theory) and its application to international law, but to use its *im*provements to inspire new theories and empirical research about international law (including in the lab). In my view, the parsimony of rational choice still makes it the starting point for research. Since behavioral research adds complexity (something which every academic should try to avoid if a simpler explanation is possible for answering a certain research question), it has to show that it generates better insights and is able to explain phenomena that cannot be explained by drawing on the rational choice approach alone. As stated by Einstein: "Everything should be as simple as it can be, but not simpler." And sometimes simplifying too much can lead to overlooking important insights. In trying to understand international law, RC might sometimes be too simple. BIntLE is able to generate more and better insights into the functioning of international law. BIntLE does not necessarily contradict classical RC approaches, but may confirm or reject and refine such arguments. Because of the weight I put on parsimony, I put the burden of proof on BIntLE to show that it can generate better insights than a rational choice approach to international law.

This Article has identified the foundations on which the promising research agenda of BIntLE can be built. It has identified the problems, as well as the potential, of BIntLE. IR and international law scholars can use the systematic arguments made by behavioral economists and psychologists in order to understand the "choice architecture" in international law as well as the norms and their effect in IR. BIntLE can also help understand internal processes and the behavior of nonstate actors. The predictive power of BIntLE depends on the biases, circumstances, and aggregation mechanisms to which states (or elite decision-makers) are subject. The latter can be institution-specific or circumstance-specific.

Every good theory should be able to withstand rigorous tests. Problems will surely arise for methodological reasons, but also due to lack of data.

315. *See* Christopher Hood, The Blame Game: Spin, Bureaucracy, and Self-Preservation in Government (2011). For experimental confirmation, see Björn Bartling & Urs Fischbacher, *Shifting the Blame: On Delegation and Responsibility*, 79 Rev. Econ. Stud. 67 (2012) (finding that, first, responsibility attribution can be effectively shifted and, second, this can constitute a strong motive for the delegation of a decision right).

States cannot be replicated in the lab. Being aware of the problems does not entail not trying to solve them, by experimental research, by qualitative case studies, or by large-n studies. This has to be done step-by-step, analyzing different fields of general and specific international law. Only in this way can the research be convincing. After all, it is the empirics that validate (or not) the research hypotheses advanced by any theory. In the name of unity of knowledge, in my view, it is desirable to pursue this venue of research, without forgetting what it is built on and what neighboring disciplines are doing.