How Annotation for Transparent Inquiry Can Enhance Research Transparency in Qualitative Comparative Analysis

Markus B. Siewert, *Bavarian School of Public Policy at the Technical University of Munich*

According to the American Political Science Association (APSA 2012, 9) ethics guidelines, "researchers have an ethical obligation to facilitate the evaluation of their evidence-based knowledge claims through data access, production transparency, and analytic transparency so that their work can be tested or replicated." Yet, how transparency is best promoted remains controversial in political science (Elman, Kapiszewski, and Lupia 2018; Jacobs et al. 2021; Kapiszewski and Karcher 2021).[1] This article explores how Annotation for Transparent Inquiry (ATI) can be used to enhance the transparency of scholarship that uses Qualitative Comparative Analysis (QCA). My experiences draw on the application of ATI in a paper that I presented at the ATI Challenge Workshop in 2018 organized by the Qualitative Data Repository. This article demonstrates how ATI represents a multifunctional approach to augmenting research transparency.

ABOUT THE RESEARCH PROJECT: CONDITIONS FOR LEGISLATIVE SUCCESS OF US PRESIDENTS

The research project for which I used ATI examined under which conditions US presidents are more or less successful in getting their policy preferences enacted into law. The analysis examined more than 100 pieces of important domestic legislation during the presidencies of William J. Clinton, George W. Bush, and Barack H. Obama. Building on previous research by Rudalevige (2002), Barrett and Eshbaugh-Soha (2007), and Beckmann (2010), I constructed a graded measure of success ranging from outright successes to complete failures. To assess gradations of success and failure, I conducted a qualitative content analysis of various textual sources including newspaper articles, summaries, and reports from *Congressional Quarterly* (*CQ*) *Weekly* and *CQ Almanac*, as well as other evaluations of each piece of legislation. My analytical framework included four explanatory factors that usually are understood to affect whether a president's position prevails in the legislative arena: (1) partisan composition of Congress, (2) public approval of a president, (3) going-public activities, and (4) use of bargaining strategies by the White House. Restated, the analysis sought to identify the conditions—or combinations thereof—that are necessary and/or sufficient for a president's legislative success and failure.

To answer this question, I applied QCA, which is a highly formalized approach to case-oriented comparative research. QCA models relationships of necessity and sufficiency in terms of set relations and is particularly apt in identifying patterns that are configurational, equifinal, and asymmetrical (Mahoney and Goertz 2006; Ragin 2008; Schneider and Wagemann 2012). QCA blends qualitative elements such as a strong focus on concept formation, the conscious construction of cases, and a focus on different types of case studies, and quantitative aspects such as the translation of information into dataset observations, the application of an algorithm to identify factors that are redundant across cases, and the use of model parameters to gauge the quality of set relations.

HOW RESEARCH TRANSPARENCY IN QCA BENEFITS FROM ATI

Table 1 is an overview of how ATI can be used throughout the QCA stages. It illustrates the main added benefits of ATI with regard to the transparency triad—that is, data access, production transparency, and analytical transparency (APSA 2012, 9–10; Büthe and Jacobs 2015, 51–56; Elman, Kapiszewski, and Lupia 2018, 32–35). Each stage is described in more detail in the following discussion.

Concerning *research design*, in addition to footnotes and appendices, ATI provides a space to add background information that is relevant but not essential—that is, to contextualize a QCA in the wider context of the field of study. Concretely, ATI might be used to further describe the construction of the case sample by pointing out whether cases were added or eliminated throughout the analysis or to discuss ambivalent theoretical expectations. These uses of ATI, of course, are not unique to QCA and can be used in scholarship based on other methods. In this study, I found ATI to be particularly useful to explore in greater detail fundamental assumptions underlying the analysis. Examining the success of a president throughout the legislative process, for instance, depends on the assumption that it is not possible to uncover the "true" legislative

doi:10.1017/S1049096521000597

Table 1

How Using ATI Can Enhance Research Transparency in QCA

Stage in the QCA Protocol	Potential Use of ATI	Dimension of Transparency Addressed: *Data Access*	*Production Transparency*	*Analytical Transparency*
Research Design	- Providing additional information on case selection - Discussing alternative theoretical expectations or ambiguous assumptions		++	+
Calibration of Sets	- Providing access to original data sources - Giving examples to illustrate coding decisions and to justify the calibration rationale	++	++	+
Analysis of Set Relations	- Elaborating on technical details - Discussing alternative analytical decisions and how they impact the analysis and results	+		++
Interpretation and Presentation of Findings	- Displaying alternative strategies and robustness tests vis-à-vis the original analysis - Linking results back to cases by providing short analytical case narratives - Providing additional visualizations or pointing readers to further sources	+		++

Source: Author's compilation.
Notes: ++ refers to the main aspect of transparency addressed through ATI; + indicates that ATI also supports this dimension but only secondarily.

preferences of a president. However, presidents might engage in strategic behavior: they can publicly undersell their policy preferences and tailor them toward the feasible. Or, they might deliberately ask for more than what they want or think they can achieve and then sign a compromised version of a piece of legislation with which they are still satisfied. In these scenarios, the substantial success of a president would either be overrated or underrated, respectively. Annotations provided the space to discuss this problem in greater detail, to illustrate the reasoning of the actors with quotes from former White House officials, and to discuss how existing studies addressed this issue.

The *calibration of sets* is a crucial task in QCA. Calibration concerns the translation of raw data into set data and is best understood as an interpretation of qualitative and/or quantitative information against the backdrop of a specific analytical concept (Ragin 2008, 71–84). For instance, because I was interested in the legislative success of a president, I had to determine what counts as a full success (i.e., a bill that includes all of the relevant aspects that are important to a president's agenda) and as a clear legislative failure (i.e., a bill that passed over a president's veto or strongly contradicts a president's policy preferences). Because I applied a graded concept of success, I also had to define how a compromise bill or other gradations of success and failure (e.g., bills that were less than compromise but not an outright failure) appear.

I found ATI to be an effective way to strengthen the concept–measurement linkage at the set calibration stage. Specifically, annotations allowed me to illustrate my coding rationale by detailing the empirical evidence found in the sources. I provided quotes from the original sources to illustrate how the relevant issue initially was framed and then used brief analytical notes to discuss the overall context of the bill and to provide a more general assessment of my coding decision. For example, the passage of the Homeland Security Act of 2002 was a major win for President Bush. The *Washington Post* on July 21, 2002, called the bill a "near total victory," which the Senate "may challenge [...] mainly at its edges, not its heart." Additional quotes from other newspapers and legislative reports echoed this appraisal and stressed that the Bush administration prevailed on most contested issues. In the analytical note, I further discussed how the Bush administration had shifted its position on key aspects of the legislative draft and how this affected my coding decision.

I found ATI to be an effective way to strengthen the concept–measurement linkage at the set calibration stage.

With a dataset encompassing more than 100 bills, it was impractical to include annotations discussing every piece of legislation. However, I referred to the appendix, which included legislative histories for each bill for which I summarized all quotes from the qualitative content analysis. I used ATI mainly to present selected examples that I considered good illustrations of pieces of legislation with a given success score. The basic idea was to address Ragin's question: "What is your case a case of?" What does a complete success (1.0), complete failure (0.0), or an outcome between those two poles (e.g., a set score of 0.75 or 0.25) look like? Another best practice is to discuss in annotations cases that are especially

https://doi.org/10.1017/S1049096521000597 Published online by Cambridge University Press

ambivalent or borderline, where it is unclear whether or not a piece of legislation tilts in favor of a president's position.

In summary, I used ATI mainly in connection with the calibration of sets to corroborate my coding decisions by connecting the original sources to their final assessment and to make my interpretation of evidence more transparent. Doing so benefited my analysis in two main ways. First, by providing annotations, I empowered the scientific community to (re)examine the empirical evidence on which analytical claims are built and to judge my interpretation of it (a benefit also noted by Mayka in this symposium). For instance, by providing direct quotes from different sources that all point in a similar direction, one of my main objectives was to strengthen readers' confidence in my evaluation of evidence. ATI also allowed me to critically engage with ambiguities and contradictory findings and to justify my coding decisions by explaining the weighting of inconsistent empirical cues.

Second, ATI encouraged or even compelled me to reveal the thought processes underlying my inferential claims and to double-check them. In line with experiences highlighted by both Myrick and Milonopoulos in this symposium, annotations indeed can work as safety mechanisms that force researchers to engage (even more) thoroughly with underlying source material, to confront their own potential biases (e.g., cherry-picking empirical cues favoring a certain interpretation), and to (more clearly) specify their inferences.

The *analysis of set relations* involves the search for necessary and sufficient conditions (Ragin 2008, 29–43, 124–45). In analyzing set relations, a researcher must make a series of decisions including, *inter alia*, the appropriate benchmark criteria with regard to the main parameters used in QCA, how inconsistent cases should be appraised, which solution strategy should be applied, and which settings of the computer algorithm should be used to produce the final results. Justifying all analytical decisions and discussing the underlying technicalities are considered best practices in the methodological literature on QCA (instead of many; see Wagemann and Schneider 2015). However, doing so draws attention away from the substantive arguments in a study—and even may confuse readers who are more interested in the content of the research than the methodological and technical details.

ATI can serve as an effective remedy for this type of confusion. Authors can use annotations to report the technical details that are essential for judging the rigor of the analysis and findings while also keeping the manuscript focused on the theoretical and substantial arguments (see Myrick's contribution and Milonopoulos's article, which uses the vivid metaphor of a "digital exoskeleton" to capture this aspect of annotations). For instance, I provided an annotation discussing why I used a certain consistency threshold rather than an alternative cutoff point. In another analytical note, I reviewed in more detail the problem of simultaneous subset relations, which arises when a configuration is sufficient for the outcome and the non-outcome at the same time. Finally, analytical notes allowed me to put the R codes right next to each analytical step (i.e., the link to the software script) so that other researchers can understand immediately how I performed my analysis.

In summary, I see ATI as a fruitful way to support the set analysis in QCA because it provides additional space to substantiate crucial technical decisions and discuss the effects that different choices might have on the results. Annotations also ease potential tradeoffs between following methodological best practices, on the one hand, and providing a not-too-technical report of the substantive research on the other.

Annotations also ease potential tradeoffs between following methodological best practices, on the one hand, and providing a not-too-technical report of the substantive research on the other hand.

ATI also can assist with the *interpretation and presentation of findings* in QCA. Every QCA accommodates the use of multiple analytical strategies—for example, the choice of parameter thresholds, the handling of configurations of conditions for which no empirical cases exist in the data (i.e., so-called logical remainders), or certain settings concerning the algorithm. Using different strategies, however, can produce different solutions. Annotations allowed me to present multiple solution terms alongside one another to highlight how the results differed depending on alternative analytical choices as well as to discuss what this meant for the relevance of certain conditions and the robustness of findings. On a similar note, authors also can include visualizations and figures in ATI annotations to assist with communicating findings. For instance, I used analytical notes to provide X-Y plots to illustrate underlying data patterns in combination with a short description.

Finally, annotations can be used to provide background information on selected cases that are theoretically interesting or that have an important analytical role. In QCA, it always is recommended to link the formalistic solution terms back to cases and to provide at least one illustrative narrative demonstrating how the identified conditions played out in a given case. However, this may not always be feasible in a journal article and, indeed, was not in my paper discussed in this article. I therefore focused on typical cases and briefly discussed in separate analytical notes three bills for which the QCA model suggested that a president's position would prevail but that instead turned out to be a compromise bill (i.e., 0.25 in the outcome set of legislative successes) or even less (i.e., 0.125 or 0.0). In doing so, I could point out important scope conditions and counteracting mechanisms that were not part of my explanatory framework.

https://doi.org/10.1017/S1049096521000597 Published online by Cambridge University Press

ATI'S COSTS AND CHALLENGES

Using ATI does not come without costs. The most obvious price is the sheer amount of additional work that using ATI requires—a fact that clearly matters in times of "fast science" and the "publish-or-perish" culture still prevalent in academia (this also is noted by Mayka and Milonopoulos in this symposium). The paper I presented at the workshop included approximately 30 annotations, which added almost 10,000 words to a paper of 11,000 words. I also provided standard supplementary materials including an appendix, datasets, and a software script.

The additional effort that goes into ATI, however, is only partly reflected in the word count. The most time-consuming part of ATI is preparing the empirical information to be provided in annotations in a format that can be understood by and useful for other scholars. How to do this depends on the type of annotation an author hopes to create. Providing details for why 0.8 and not 0.75 was chosen as a consistency cutoff and illustrating the consequences of this decision are straightforward. Conveying my own qualitative assessment of why the Sarbanes–Oxley bill was a legislative defeat for President Bush requires walking readers through the process of weighing evidence, which takes time. Yet, I must do this anyway, so why not do so explicitly?

A different challenge is the temptation to provide too much information (see also the contributions of both Milonopoulos and Myrick). In annotating my paper, I often felt pressure to report more evidence to make my analysis as transparent as possible. However, more details do not always make for better analyses or arguments. There is a fine line between being transparent and discussing ambiguous claims and alternative interpretations, on the one hand, and flooding readers and reviewers with empirical material on the other.

Despite increasing numbers of scholars who consider transparency to be a "meta-standard" (Elman and Kapiszewski 2014, 43–44) or a "fundamental norm" (Moravcsik 2014, 665), highlighting analytical uncertainty and inferential ambiguities in our work and openly discussing alternate or difficult decisions might still lead others to judge the work to be inherently flawed and inconclusive. The utility of approaches such as ATI therefore ultimately rests on a culture that welcomes transparency and views annotations as a laudable attempt to make research more reliable and replicable, not as a sign of weak research (Myrick makes a similar argument in this symposium).[2]

A particular challenge for me was deciding *what* to annotate. Currently, annotations are displayed automatically only when a paper is read online; reading it offline makes it more laborious to locate them. It therefore is possible that readers will not see the information included in an analytical note.[3] Against this backdrop, authors using ATI should think about a hierarchy reflecting the importance of information that they could include in their paper (see Mayka's contribution for stimulating ideas on this issue). I approached this issue by repeatedly asking myself, "Is this piece of information essential or interesting to the reader?" If the answer was "Yes," I usually placed the information in a traditional footnote to ensure that it would be part of the paper. If the answer was a clear "No," I completely omitted it. I included in ATI annotations information that a reader who is interested in the substance of my research might find interesting but not crucial as well as technical information about methodical aspects of the analysis. Initially, the thought to annotate my work did not come naturally. Consequently, as a first step, I marked passages, claims, and different types of information throughout the manuscript that seemed to be potential candidates for annotation. In a second step, I decided whether to include the information in an annotation, in a traditional footnote, or in the appendix.

A related question is *when* in a research project an author should start thinking about applying ATI (see also Mayka and Milonopoulos in this symposium). Given the timing of the ATI Challenge vis-à-vis my work, I had to prepare the annotations for my paper after I completed the data collection and already had presented different versions of it at conferences. Doing so required me to review all of my records to find fitting quotes and passages to include in the annotations. Sifting through my notes and pages of secondary materials gave me the opportunity to again reevaluate and reexamine the evidence, in line with the iterative nature of qualitative research. Nonetheless, it is probably useful for authors to decide early in a research project whether they will use ATI. This way, they can prepare analytical notes during the data-collection process, which is when much inferential reasoning takes place.

The utility of approaches such as ATI therefore ultimately rests on a culture that welcomes transparency and views annotations as a laudable attempt to make research more reliable and replicable, not as a sign of weak research.

THE FUTURE OF ATI

What are the next steps for ATI? In my opinion, three frontiers stand out. The first is developing best practice strategies for integrating ATI into the workflow of research. Future users of ATI would profit from more in-depth discussions about the role of annotations in the research process, how information must be managed so that it can be easily shared, and other essential aspects related to project and data management. The second frontier is advancing the technical implementation of ATI—for instance, with regard to displaying ATI annotations, developing hands-on tools to facilitate annotating manuscripts using ATI, among other

https://doi.org/10.1017/S1049096521000597 Published online by Cambridge University Press

things. The third frontier is alerting the political science community to ATI. This involves raising awareness of the benefits and costs of ATI—for instance, among journal editors and methodology institutes—and more generally providing (more) training opportunities for scholars in all career stages on strategies and tools to improve research transparency.[4] The ideas, insights, and suggestions in this symposium advance the cause of ATI on all of these frontiers. It will be interesting to see how community standards and a culture of transparency progress in the future—a discussion to which ATI clearly has much to offer.

ACKNOWLEDGMENTS

I thank the two anonymous reviewers for their valuable comments, which made this article much less technical. I also am grateful for the feedback from all participants of the 2018 ATI Challenge Workshop in New York and the roundtable held at the APSA Annual Meeting in Washington, DC, in 2019. Special thanks are extended to William Harder, who served as commentator at the ATI Workshop. Finally, I thank Sebastian Karcher and Diana Kapiszewski for setting up this symposium, providing helpful feedback on earlier drafts, and putting so much effort into making pol-sci research more transparent. ■

NOTES

1. In addition to cited works, interested readers are referred to the symposia on "Openness in Political Science" in *PS: Political Science & Politics* 47 (1); "Transparency in Qualitative and Multi-Method Research" in *Qualitative & Multi-Method Research* 13 (1); and "Data Access and Research Transparency (DA-RT)" in the *APSA Comparative Politics Newsletter* 26 (1).
2. Similar arguments concerning accepting, incentivizing, and rewarding transparency are part of the debate about challenges linked to preregistration and open science more generally. See, for instance, Haven and Grootel (2019) and Allen and Mehler (2019).
3. The same applies, of course, to footnotes, online appendices, and so forth, with which only a fraction of readers of an article actively engages. In light of this, the question is: "Why we should invest the additional time and effort ATI requires in the first place?," to which Musgrave and Karcher (2018) provide a good answer.
4. An unsystematic review of the course listings of five established method training programs—European Consortium for Political Research methods school, Essex summer school, Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan, the International Political Science Association summer schools, and the Institute for Qualitative and Multi-Method Research at Syracuse University—reveals that only ICPSR offers specific short courses related to issues of research transparency.

REFERENCES

Allen, Christopher, and David M. A. Mehler. 2019. "Open Science Challenges, Benefits and Tips in Early Career and Beyond." *PLOS Biology* 17 (5): e3000246. https://doi.org/10.1371/journal.pbio.3000246.

American Political Science Association. 2012. *Committee on Professional Ethics, Rights and Freedoms*, "A Guide to Professional Ethics in Political Science." Washington, DC: American Political Science Association. www.apsanet.org/portals/54/Files/Publications/APSAEthicsGuide2012.pdf.

Barrett, Andrew W., and Matthew Eshbaugh-Soha. 2007. "Presidential Success on the Substance of Legislation." *Political Research Quarterly* 60 (1): 100–112. https://doi.org/10.1177/1065912906298605.

Beckmann, Matthew N. 2010. *Pushing the Agenda: Presidential Leadership in U.S. Lawmaking, 1953–2004.* Cambridge and New York: Cambridge University Press.

Büthe, Tim, and Alan M. Jacobs. 2015. "Conclusion: Research Transparency for a Diverse Discipline." *Qualitative & Multi-Method Research* 13 (1): 52–64.

Elman, Colin, and Diana Kapiszewski. 2014. "Data Access and Research Transparency in the Qualitative Tradition." *PS: Political Science & Politics* 47 (1): 43–47. https://doi.org/10.1017/S1049096513001777.

Elman, Colin, Diana Kapiszewski, and Arthur Lupia. 2018. "Transparent Social Inquiry: Implications for Political Science." *Annual Review of Political Science* 21 (1): 29–47. https://doi.org/10.1146/annurev-polisci-091515-025429.

Haven, Tamarinde L., and Leonie Van Grootel. 2019. "Preregistering Qualitative Research." *Accountability in Research* 26 (3): 229–44. https://doi.org/10.1080/08989621.2019.1580147.

Jacobs, Alan M., Tim Büthe, Ana Arjona, Leonardo R. Arriola, Eva Bellin, Andrew Bennett, Lisa Björkman, et al. 2021. "The Qualitative Transparency Deliberations: Insights and Implications." *Perspectives on Politics* 19 (1): 171–208. https://doi.org/10.1017/S1537592720001164.

Kapiszewski, Diana, and Sebastian Karcher. 2021. "Transparency in Practice in Qualitative Research." *PS: Political Science & Politics* 54 (2): 285–91. https://doi.org/10.1017/S1049096520000955.

Mahoney, James, and Gary Goertz. 2006. "A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research." *Political Analysis* 14 (3): 227–49. https://doi.org/10.1093/pan/mpj017.

Moravcsik, Andrew. 2014. "Trust, but Verify: The Transparency Revolution and Qualitative International Relations." *Security Studies* 23 (4): 663–88. https://doi.org/10.1080/09636412.2014.970846.

Musgrave, Paul, and Sebastian Karcher. 2018. "Seven Reasons We Use Annotation for Transparent Inquiry (ATI)." Duck of Minerva [blog], May 15. https://duckofminerva.com/2018/05/seven-reasons-we-use-annotation-for-transparent-inquiry-ati.html.

Ragin, Charles C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond.* Chicago: University of Chicago Press.

Rudalevige, Andrew. 2002. *Managing the President's Program: Presidential Leadership and Legislative Policy Formulation.* Princeton, NJ: Princeton University Press.

Schneider, Carsten Q., and Claudius Wagemann. 2012. *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis.* Cambridge: Cambridge University Press.

Wagemann, Claudius, and Carsten Q. Schneider. 2015. "Transparency Standards in Qualitative Comparative Analysis." *Qualitative & Multi-Method Research* 13 (1): 38–43.