International Studies Review (2022), viac008

ANALYTICAL ESSAY

QCA in International Relations: A Review of Strengths, Pitfalls, and Empirical Applications

Tobias Ide
Murdoch University, Australia

AND

Patrick A. Mello
Technical University of Munich, Germany

Qualitative comparative analysis (QCA) is a rapidly emerging method in the field of International Relations (IR). This raises questions about the strengths and pitfalls of QCA in IR research, established good practices, how IR performs against those standards, and which areas require further attention. After a general introduction to the method, we address these questions based on a review of all empirical QCA studies published in IR journals between 1987 and 2020. Results show that QCA has been employed on a wide range of issue areas and is most common in the study of peace and conflict, global environmental politics, foreign policy, and compliance with international regulations. The utilization of QCA offers IR scholars four distinct advantages: the identification of complex causal patterns, the distinction between necessary and sufficient conditions, a middle ground between quantitative and qualitative approaches, and the reinforcement of the strengths of other methods. We find that albeit a few exceptions, IR researchers conduct high-quality QCA research when compared against established standards. However, the field should urgently pay more attention to three issues: the potential of using QCA in combination with other methods, increasing the robustness of QCA results, and strengthening research transparency in QCA applications. Throughout the article, we formulate strategies for improved QCA research in IR.

El análisis cualitativo comparativo (QCA, por sus siglas en inglés) es un método que está emergiendo rápidamente en el campo de las Relaciones Internacionales. Esto plantea preguntas sobre las fortalezas y dificultades del QCA en la investigación de las relaciones internacionales, las buenas prácticas establecidas, cómo actúan las relaciones internacionales con respecto a esas normas y qué áreas requieren más atención. Tras una introducción general a este método, abordaremos preguntas en base a una revisión de todos los estudios empíricos sobre el QCA publicados en revistas de Relaciones Internacionales entre 1987 y 2020. Los resultados muestran que el QCA se ha empleado en una amplia gama de áreas temáticas y es más común en los estudios de paz y los conflictos, políticas medioambientales, políticas exterior y cumplimiento de regulaciones internacionales. La utilización del QCA ofrece a los académicos de Relaciones Internacionales cuatro ventajas distintas: la identificación de patrones causales complejos, la distinción entre condiciones necesarias y suficientes, un

Ide, Tobias, and Patrick A. Mello. (2022) QCA in International Relations: A Review of Strengths, Pitfalls, and Empirical Applications. *International Studies Review*, https://doi.org/10.1093/isr/viac008
© The Author(s) (2022). Published by Oxford University Press on behalf of the International Studies Association. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

punto intermedio entre enfoques cuantitativos y cualitativos, y el fortalecimiento de puntos fuertes de otros métodos. Hemos comprobado que, salvo algunas excepciones, los investigadores de relaciones internacionales llevan a cabo una investigación sobre el QCA de alta calidad en comparación con las normas establecidas. Sin embargo, este campo debería prestar con urgencia más atención a tres cuestiones: el potencial del uso del QCA en combinación con otros métodos, el aumento de la solidez de los resultados del QCA, el fortalecimiento de la trasparencia de la investigación en las aplicaciones del QCA. A lo largo de este artículo, formulamos estrategias para mejorar la investigación del QCA en las relaciones internacionales.

L'analyse comparative qualitative (ACQ) est une méthode connaissant un développement rapide dans le domaine des relations internationales (RI). Cela soulève des questions sur les forces et les écueils de l'ACQ dans les recherches en RI, sur les bonnes pratiques établies, sur les performances des RI par rapport à ces normes et sur les domaines qui nécessitent une attention supplémentaire. Après une présentation générale de la méthode, nous abordons ces questions en nous basant sur un examen de toutes les études empiriques portant sur l'ACQ qui ont été publiées dans des revues de RI entre 1987 et 2020. Les résultats montrent que l'ACQ a été employée dans un large éventail de domaines d'intérêt et que c'est dans l'étude de la paix et des conflits, des politiques environnementales mondiales, des politiques étrangères et de la conformité aux réglementations internationales qu'elle est la plus courante. L'utilisation de l'ACQ offre quatre avantages distincts aux chercheurs en RI: l'identification de schémas causaux complexes, la distinction entre conditions nécessaires et conditions suffisantes, un juste milieu entre les approches quantitatives et qualitatives, et le renforcement des points forts des autres méthodes. Nous constatons qu'à quelques exceptions près, les chercheurs en RI mènent des recherches avec ACQ de grande qualité lorsque nous les comparons aux normes établies. Toutefois, une plus grande attention devrait urgemment être accordée aux trois points suivants dans ce domaine: le potentiel de l'utilisation de l'ACQ en combinaison avec d'autres méthodes, l'augmentation de la robustesse des résultats des ACQ, et l'accroissement de la transparence des recherches dans les applications d'ACQ. Nous formulons des stratégies pour améliorer les recherches avec ACQ en RI tout au long de l'article.

Keywords: best practice, international politics, multimethod, research methods, set theory
Palabras clave: Mejores prácticas, política internacional, métodos múltiples, métodos de investigación, teoría de conjuntos
Mots clés: meilleure pratique, politiques internationales, multiméthode, méthodes de recherche, théorie des ensembles

Introduction

Originated over three decades ago as a methodological alternative to mainstream statistical approaches (Ragin 1987), qualitative comparative analysis (QCA) has enriched the toolbox of empirical research methods. What started as a niche approach has become an established method that is widely used across social science disciplines (Marx, Rihoux, and Ragin 2014). While there have been hundreds of applications each in the fields of political science, sociology, and business studies, the number of empirical applications in International Relations (IR) has grown at a slower pace. However, as figure 1 illustrates, the use of QCA has increased strongly

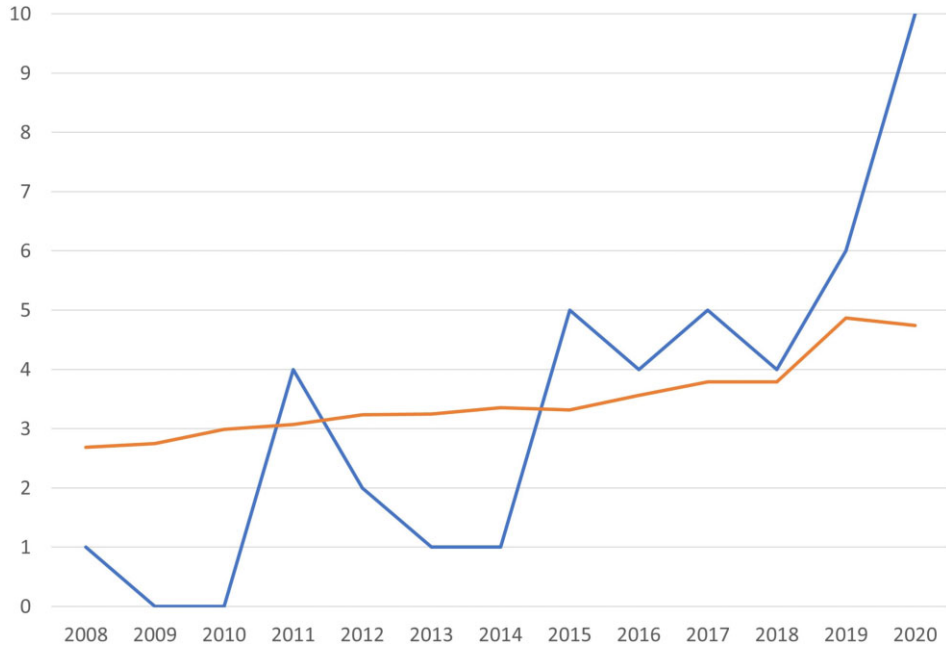

Figure 1. QCA studies in IR journals per year (blue line) and the total number of articles in IR journals (in 1.000s) (orange line).

since 2015 and particularly since 2018. QCA applications have been published in several of the fields' most prestigious journals and contributed to a wide range of debates in IR (see *QCA in International Relations* for further details).

This growing relevance of QCA in IR might not be very surprising. QCA has a reputation of being able to detect complex patterns characterized by conjunctural causation and equifinality. Proponents portray the method as a bridge between quantitative and qualitative approaches due to QCA's ability to integrate both data types as well as its balance between contextual knowledge and generalization. Others conceive QCA as the missing middle that allows analyzing a medium number of cases, an attribute that makes it also well suited to complement more established methods (Goertz 2016; Mello 2021).

As with any other emerging method in IR (and other fields), three sets of questions regarding QCA emerge: (1) What is QCA and how does it work? (2) What advantages and potential does QCA offer in the field of IR, and what shortcomings exist? (3) How do we recognize good QCA work, does QCA research in IR fulfil these standards, and which areas require further attention? This article addresses these three sets of questions. It is intended to serve as a guideline and orientation for those reading and reviewing QCA papers, those supervising research students using QCA, those thinking about or starting to use QCA in their own work, and those reflecting upon research methods, specifically in IR and adjoining fields of study.

In order to serve this purpose, our article proceeds as follows. We begin with a brief explanation of QCA as a method, thereby addressing the first question posed above. Afterward, we conduct a comprehensive review of empirical QCA applications in IR research. We start with brief descriptions of our sampling strategy as well as trends and topics of QCA usage in IR. Next, we discuss in detail the strengths and pitfalls of QCA in IR, hence addressing the second main question posed above while providing key information for (potential) users and method scholars. Finally,

we turn to the third question by tracing how existing QCA research in IR benchmarks against established good practices, thereby providing important guidelines for readers, reviewers, supervisors, and users. Throughout this section, we offer reflection and guidance on how IR research can benefit from this innovative method and improve its respective methodological standards. By means of conclusion, we summarize our findings and reflect upon future pathways of QCA research in IR.

Qualitative Comparative Analysis: A Primer

While a comprehensive introduction to the method is beyond the scope of our article, this section aims to provide a summary of QCA's defining features and to briefly compare it to other methods.[1] At its core, QCA is a set-theoretic, comparative method for identifying conditions (or combinations of conditions) that are necessary and/or sufficient for an outcome. The set-theoretic analysis is based on Boolean algebra, which enables the construction of truth tables and their minimization to derive QCA solutions, or "recipes" for the outcome, usually supported by specialized software.

A core asset of QCA is the method's ability to account for *causal complexity*, as when there are combinations of conditions that are jointly sufficient for an outcome (*conjunctural causation*) or when multiple pathways of conditions lead toward the same outcome (*equifinality*). Arguably, causal complexity is widespread among social phenomena and especially in IR, where different levels of analysis influence and interact with each other. Yet, methods that seek to identify "average effects" of individual variables will tend to overlook patterns of conjunctural causation (Mahoney 2021). Likewise, searching for a "common cause" will be ineffective in situations characterized by equifinality (George and Bennett 2005, 65).

As mentioned, QCA's analytical procedure is geared toward the identification of necessary and sufficient conditions. The identification of a *necessary condition* implies that the outcome of interest does not occur in the absence of the respective condition. By the same token, the existence of a *sufficient condition* means that the outcome occurs whenever the respective condition is present. Yet, it is important to highlight that necessity and sufficiency do not imply determinism. QCA entails measures of fit that allow for the calculation of imperfect set relations (Ragin 2008).

The complement to necessary and sufficient conditions are so-called INUS conditions, which constitute "an *insufficient* but *necessary* part of a condition, which is itself *unnecessary* but *sufficient* for the result" (Mackie 1965, 245). In essence, an INUS condition means that the respective condition is only sufficient when combined with another condition. INUS conditions feature prominently in many if not most QCA results, as individual conditions are rarely sufficient to bring about the outcome on their own, but combinations of conditions often are.[2]

Conceptions of necessity and sufficiency are prevalent in the social sciences, even when they are used implicitly without invoking these terms (Ragin 1987). This is the case for IR as well. According to one stream of IR theory, for instance, the presence of a hegemonic state is a *necessary condition* for meaningful international cooperation. Without a hegemon, no state is willing to incur the costs of cooperation, while concerns about relative gains prevail. However, the presence of a hegemon is not sufficient for such cooperation to occur, for instance if the hegemonic state is malevolent or not interested in the respective issue area. Hence, certain characteristics of the hegemonic state and its interests might well be INUS conditions for international cooperation.

[1] For comprehensive introduction to QCA, see Schneider and Wagemann (2012), Kahwati and Kane (2020), Mello (2021), and Oana et al. (2021).

[2] There are also SUIN conditions that are "a *sufficient* but *unnecessary* part of a factor that is *insufficient* but *necessary* for an outcome" (Mahoney, Kimball, and Koivu 2009, 126).

Conversely, supporting Kurdish separatists is usually a *sufficient condition* for a state to have tense relations with Turkey, but this is by no means a necessary condition: states can have tense relations with Turkey due to competing interests in Libya or critique of Turkey's authoritarian turn since 2016. Both examples also highlight the relevance of equifinality: there are several routes or pathways toward meaningful international cooperation or tense relations with certain states.

QCA's analytical protocol entails separate steps for the identification of necessary and sufficient conditions. The former are tested individually, typically at the beginning of the analysis. The analytical core of QCA is the ensuing construction of the *truth table*, which contains rows that display the logically possible combinations of conditions and the distribution of the empirical cases across those rows. As a truth table covers all possible combinations of conditions, the number of rows corresponds to 2^k, where k is the number of conditions that were included in a given study. This means that a study with three conditions will have a truth table with eight rows of combinations, four conditions will result in sixteen rows, and five conditions lead to thirty-two truth table rows, and so forth. The key point is that with each added condition the truth table will grow exponentially, and fewer rows will be populated with empirical cases (assuming that the number of cases remains stable while conditions are being added). The underlying issue here is *limited diversity*. This becomes relevant when dealing with *logical remainders* (truth table rows—or combinations of conditions—with no corresponding empirical evidence) to derive QCA solution terms. Hence, it is recommended to limit the analysis to a moderate number of conditions, also to enhance the interpretation of QCA results (Marx and Dusa 2011; Mello 2021).

While the truth table is valuable in and of itself, each row in the truth table is also a statement of sufficiency. Rows that are consistently associated with the outcome can be used for *Boolean minimization*: a procedure where configurations of conditions are compared and minimized to less-complex expressions, following the rules of Boolean algebra. On this basis, three QCA solution terms can be derived: the *conservative solution* that draws solely on the rows backed by empirical evidence, the *parsimonious solution* that allows the software to incorporate logical remainder rows in order to attain a simpler solution, and the *intermediate solution* where directional expectations and the plausibility of logical remainders can be assessed by the researcher (Schneider and Wagemann 2012, 151–77).[3]

QCA solutions can be assessed with two primary measures of fit. *Consistency* indicates the degree to which a solution is in line with the underlying empirical evidence (similar but not identical to significance values in statistical research). *Coverage* indicates the proportion of the outcome that can be explained by a solution (measured in terms of the sum of set-membership values) and can hence be regarded as the QCA equivalent of the coefficient magnitude. Both measures can take values between zero (lowest) and one (highest) (Ragin 2008).

Before the truth table can be configured and analyzed, QCA requires the *calibration* of raw data into a set-theoretic format, using either *crisp sets* (binary scores) or *fuzzy sets* (graded membership values between zero and one) (Ragin 2008).[4] For this, researchers need to define the target set (e.g., "strong economy" or "multilateral cooperation") and then decide for every case whether it holds (full or partial) set membership in the given condition or outcome (e.g., a country is considered to have a strong economy or to engage in multilateral cooperation). Researchers can either manually assign scores to their cases or define three "empirical anchors" to let the software conduct the transformation of the raw data into calibrated scores. While the calibration procedures themselves are relatively straightforward, clear

[3] For further discussions of the different solution types, see Duşa (2019), Baumgartner and Thiem (2020), and Mello (2021, 129–36).

[4] There is also multivalue QCA, but here we focus on the two most common QCA variants.

conceptualization of the targets sets and the empirical anchors is vital (de Block and Vis 2018; Kahwati and Kane 2020; Oana, Schneider, and Thomann 2021).

How does QCA compare to other methods? With its emphasis on causal complexity, QCA is typically applied to examine the "causes of effects" rather than the "effects of causes" (Goertz and Mahoney 2012). While many statistical applications explicitly or implicitly follow the latter approach, seeking to identify the net effects of certain independent variables on a dependent variable, QCA aims to detect the complex interplay of conditions under which an outcome occurs. To be sure, quantitative approaches can also model *interaction effects* between variables (Imai 2017). Yet, this is not the same as conjunctural causation in QCA because statistical analyses typically focus on pairs of variables (if interaction terms are included at all), whereas solution terms in QCA often involve three or more conditions. Moreover, the interpretation of complex statistical interaction effects can pose substantial challenges, whereas QCA solutions reveal combinations of conditions that are consistently associated with the outcome. Recently, efforts have been made to introduce set-theoretic tools into quantitative research (Mahoney 2021), which further underlines the added value of a set-theoretic perspective.

As a case-oriented approach, QCA clearly resonates with methods such as process tracing or intensive case studies. As such, the combination of QCA as a *cross-case* method with process tracing as a *within-case* method holds potential for multimethod research (e.g., Beach and Rohlfing 2018).

Sampling QCA Studies in International Relations

The sample for our review of QCA research in IR contains all empirical QCA applications that were published in IR journals listed on the Social Sciences Citation Index (SSCI) and accessible through the Web of Science. This constitutes the entirety of QCA articles published in IR from the method's development (Ragin 1987) to the end of 2020. To create our sample, we searched the Web of Science Core Collection for the term "qualitative comparative analysis" on February 2, 2021. A subsequent step limited the results to the period 1987–2020 and the category "International Relations," yielding fifty-nine results. After excluding sixteen publications because they were no journal articles, no empirical studies, did not use QCA, or did not address IR topics, we ended up with a sample of forty-three studies for our analysis (see supplementary material for a full list of studies included and an overview about how we coded them).

We should note that our sampling approach resulted in the inclusion of several studies that could be considered on the fringe of IR as a discipline. Likewise, other studies on relevant IR topics were not considered because the Web of Science does not categorize the respective journals in which these studies appeared as IR outlets. We recognize these limitations. However, alternative sampling strategies would have had to face difficult choices about which journals and studies to include. Rather than introducing our subjective biases on this matter, we decided for a transparent and reproduceable approach.

QCA in International Relations

This section contains three parts. We start with a brief review of recent trends and topics in IR research with QCA to show how widely and in which research fields the method is used. Afterward, we assess the potential of using QCA in IR research, drawing on studies from our sample to demonstrate how scholars have used QCA to make contributions to IR research while also reflecting on the method's limitations. Finally, we assess how QCA research in IR benchmarks against established good practices, how to avoid common pitfalls, and which areas require further scholarly attention.

Trends and Topics

QCA as a method has gained considerable momentum in IR. As illustrated by figure 1, the first IR article using QCA was published in 2008 (Rubenzer 2008), with subsequent studies following in 2011. From 2015 onward, four to five QCA studies appeared in IR journals each year, and 2020 saw a steep growth to ten studies per year. This adds to an also increasing number of QCA studies on IR topics published in outlets not classified as IR journals and hence excluded from our sample (e.g., Ansorg 2014; Ide et al. 2020; Maerz 2020). The number of QCA studies in IR is also growing at a more rapid pace than the overall number of IR journal articles (indicated by the orange line in figure 1).

QCA studies further contributed to some of the most influential journals in the field of IR, including International Organization (ranked second in IR according to the 2019 InCites Journal Citation Reports), Marine Policy (seventh), Global Environmental Politics (tenth), Journal of Peace Research (eleventh), Journal of Conflict Resolution (twelfth), JCMS-Journal of Common Market Studies (thirteenth), International Studies Review (twentieth), and International Studies Quarterly (twenty-second).

That said, compared to regression analysis or case studies, QCA is still an emerging method in IR, indicated among others by "only" forty-three studies in our sample. Still, the growing importance of QCA in IR makes a timely review of applications important in order to reflect on the existing trends and improve future research.

IR researchers use QCA to address a wide variety of questions. We will discuss these in greater detail below when drawing on studies on a range of issues to illustrate our arguments. This demonstrates that QCA can be fruitfully applied in a large number of IR subfields. However, QCA is more widely used in some research areas than in others. To detect such patterns, we count all keywords that were used by the articles in our sample. The two most frequently used keywords are "conflict" and "civil war" (each used five times).[5] Together with "violence" (three), "armed conflict" (two), and "escalation" (two), this indicates that QCA gained considerable prominence in peace and conflict studies. Analyses focusing on environmental issues also draw frequently on QCA, suggested by keywords such as "institutions" (four, often referring to environmental institutions), "climate change" (three), or "political ecology" (two). Another cluster emerges from keywords such as "policy" (four), "foreign policy" (three), "authority" (two), and "alliance" (two), indicating the prominence of foreign policy analysis in our sample. Relatedly, the compliance of states with international regulations (e.g., international criminal tribunals, European Union (EU) law) is another IR research area in which QCA is frequently used, although this is not directly visible from the keywords.

QCA is well suited to contribute to IR debates on these topics. For example, environmental policies and their implementation are shaped by local civil society movements, national governments, regional regulations, and global agreements, among others. Likewise, both grievance and opportunity factors influence the dynamics of armed conflict. Due to its conjunctural logic, QCA is able to identify how factors derived from various levels or analyses or theoretical paradigms interact to produce a certain outcome (Mahoney 2021).

Why (Not) Use QCA?

The forty-three studies in our sample illustrate four distinctive benefits of applying QCA in the field of IR. First, as outlined above, QCA is able to address complex causal relations by studying interactions between conditions rather than the net effects of individual variables. This allows researchers to identify causal relations

[5] With the exception of "politics" (used nine times).

that are conjunctural, that is, dependent on the presence or absence of certain context factors. When assessing the drivers of international environmental regime effectiveness, for instance, Breitmeier, Underdal, and Young (2011) find only weak statistical evidence for the causal relevance of majority voting rules. However, in conjunction with a solid knowledge base on the underlying environmental problem and powerful advocate states, majority voting turns out to be an important predictor of effective environmental regimes. In a similar way, Gromes (2019) argues—and then demonstrates empirically—that in order to explain peacekeeping success, interactions between the characteristics of the peacekeeping mission and the peacekeeping environment need to be considered.

The focus on conjunctural causation also offers possibilities to build bridges between various theoretical approaches in IR. Traditionally, research on civil wars has been divided among approaches focusing on grievances (i.e., the reasons that motivate people to engage in armed violence) and those highlighting opportunity structures (i.e., factors allowing the organization of armed violence) (Taydas, Enia, and James 2011). Rather than demonstrating the higher explanatory value of factors associated with one of these approaches, several studies in our sample utilize QCA to show that it is the interaction of grievance- and opportunity-related conditions that explains conflict dynamics best (e.g., Bara 2014; Lindemann and Wimmer 2018). While different research paradigms typically emphasize different levels of analysis, QCA allows for theoretical integration. For instance, Haesebrouck (2017b) combines systemic conditions with factors related to domestic politics in order to provide an account of burden sharing among NATO member states.

Second, and closely related to the aspect of acknowledging causal complexity, QCA enables IR researchers to distinguish between necessary and sufficient conditions. While doing so might not make sense for every research question, it can resonate very well with a range of theoretical expectations. As Grynaviski and Hsieh (2015, 709) put it: "The benefit of using QCA is that it enables leverage on necessary and sufficient conditions in ways that most statistical methods do not: correlations between oxygen and breathing may be weak, but as a necessary condition it is a strong claim." Both researchers use QCA to test their proposition that a hierarchical international order is a necessary condition for international arbitration.

Third, QCA is a method that occupies the "middle ground" between large-*N* studies with hundreds or thousands of cases and small-*N* studies that engage qualitatively with a handful of cases at most (Ragin 2000, 23). Recent surveys indicate that the median number of cases for QCA applications across most social science fields ranges between twenty and thirty cases (Mello 2021, 41). This means that QCA presents a reasonable alternative for comparative researchers working on topics with intermediate numbers of observations (to be sure, this should never be the sole reason why QCA is used). Therefore, QCA also allows researchers to generalize beyond very few intensively studied cases while at the same time accounting for the specifics of each case. For example, Adhikari and Samford (2013) draw on quantitative data to assess the drivers of conflict intensity in seventy-five districts of Nepal during the civil war while using their qualitative knowledge of the local context to choose and calibrate conditions

A further advantage in this context is that QCA is open to both qualitative and quantitative types of data. Both are transformed into values between zero and one during the calibration and can hence even be combined in the same study. Research on climate change and conflict is characterized by a divide between quantitative and qualitative approaches and as such faces challenges to combine very different data types (Daoudy 2021). In a recent QCA study, Ide et al. (2021) integrate data on conditions as different as relative precipitation, nightlight emissions, and local political cleavages to bridge the gap between quantitative and qualitative approaches.

Fourth, and finally, QCA can be applied for other purposes than detecting (complex) causal relations, particularly when combined with other methods

(Berg-Schlosser et al. 2009). One of them is case selection. QCA can help scholars to detect typical, outlier, or unexplained cases that require further in-depth study (Young 2019). Another application of QCA is the identification of ideal types or causal pathways. This is demonstrated by Hao and Gao (2016), who identify three distinct pathways to (or types of) democratization in East Asia during the 1980s and 1990s.

As with any other method, QCA is no silver bullet approach. Both the studies in our sample and the general QCA literature (de Meur, Rihoux, and Yamasaki 2009; Marx, Rihoux, and Ragin 2014; Mello 2021) point toward limitations and shortcomings of QCA. To start with, QCA is best suited for medium-*N* research designs, comprising around approximately ten to fifty cases. With smaller samples, the number of observations is too small for meaningful comparison, while for larger samples, knowledge of individual cases becomes infeasible. This is reflected in our sample of IR studies, where 88 percent of the articles feature between ten and seventy-five cases.[6]

On the one hand, this illustrates that QCA provides a real alternative for analyses where case numbers are too large for in-depth case studies yet too small for statistical approaches. On the other hand, such a focus on the "missing middle" could also have problematic implications. Certain events in IR are quite rare (such as global pandemics or the onset of a world war), while for other phenomena, data on country/administrative unit/cell-years result in thousands of cases. However, IR studies with more than hundred cases (Breitmeier, Underdal, and Young 2011; Dong et al. 2015; Haesebrouck 2017a) or even five hundred cases (Bara 2014) demonstrate that QCA can also be applied to larger samples.

Furthermore, the number of conditions that can be effectively employed in a QCA is limited. As discussed above, with every condition added, the number of truth table rows grows exponentially, resulting in problems related to a high number of logical remainders and limited empirical diversity. Beyond a certain conditions-to-case ratio (see below), results can become unreliable (Marx and Dusa 2011). This limitation confines the number of theoretical relevant conditions that can be tested in any given analysis—certainly a challenge for the field of IR, where several grand theories and a fast-growing number of midrange approaches exist (Dunne, Hansen, and Wight 2013). However, this issue is not uncommon to other methods as illustrated by debates about "garbage-can regressions" in IR (Achen 2005, 327). Furthermore, there are ways to deal with this challenge, for instance, by running separate analyses, each testing a particular theory and then comparing the results (Reynaert 2011; van der Maat 2011).

One area that occasionally draws criticism is the calibration procedure in QCA. Critics sometimes suggest that crisp and fuzzy scores in QCA are essentially "arbitrary" and hence open to manipulation by the researcher. Therefore, calibration decisions should be anchored in prior research or substantive arguments (Ragin 2008) and, if several viable options exist, be subjected to robustness tests (Schneider and Wagemann 2012, 275–95). The calibration of raw data also results in a loss of information (this critique applies particularly to the dichotomous, crisp-set variant of QCA). However, the calibration also allows researchers to specify meaningful variation and less-relevant variation in their raw data. For example, one could decide to treat all countries that pass a certain economic threshold as "strong economies," irrespective of numerical differences between them (which might be analytically less relevant on a global scale).

Finally, there are limits of what QCA can do. It is geared toward assessing set relations of necessity and sufficiency between conditions and an outcome. Depending on the research aims, there may be reasons to complement QCA with statistical tests (see Empirical Applications). This is why several studies in our sample

[6] No study analyzes fewer than ten cases.

combine QCA with regression analysis (Ide 2018a; Caspersen 2019) or descriptive statistics such as chi-square tests (Mello 2020). Furthermore, set relations may indicate but do not prove causality (see also Haesebrouck and Thomann 2021). This is one reason why the combination of QCA with other case-based methods, such as process tracing, has gained currency (Beach and Rohlfing 2018). In multi-method research designs, QCA provides cross-case inferences and process tracing allows the identification of causal mechanisms.

Empirical Applications

While gaining in recognition and usage, QCA has also evolved as a method. Nowadays, there is an agreed-upon core of what constitutes good practice QCA (Schneider and Wagemann 2012; Thomann and Maggetti 2020; Mello 2021). In this section, we analyze how well QCA research in IR does when compared to those standards.

To start with, fuzzy-set (fs) QCA is the dominant *QCA type* in our sample of IR studies (twenty-one applications), followed by crisp set (cs) with fifteen applications and multivalue (mv) QCA with five applications. This indicates an important change compared to a few years ago when csQCA was still the dominant type (Rihoux and Marx 2013). We regard this as an encouraging sign—particularly for the discipline of IR—because fsQCA allows for more nuanced calibration decisions. One should note, however, that several studies (Berlin 2016; Boogaerts 2018) draw on fsQCA despite coding their outcome binary—a practice that is considered problematic by some QCA scholars (Schneider and Wagemann 2012). Furthermore, only a minority of articles (45 percent) discuss why the respective QCA type is used and typically there is no elaborated argument as to why one type was better suited than the other.

Until recently, the assumption was that crisp sets generally inflate consistency scores and as such fuzzy sets should be preferred. However, Rohlfing (2020) finds that the relationship between crisp sets, fuzzy sets, and the consistency scores attained in the truth table analysis is "ambiguous," as it can go in both directions. For users of QCA, the bottom line is that the decision whether crisp or fuzzy sets are used should be justified on substantive grounds and explicitly included in a paper. Whenever there is leeway and both types of sets could feasibly be used, robustness tests should be conducted.

Identifying *necessary conditions* for an outcome is considered a key strength of QCA, and it is recommended to conduct tests for necessity at the beginning of each analysis (Schneider and Wagemann 2012). In our sample, 70 percent of the studies report such tests. This indicates that when it comes to the field of IR, there is no "sufficiency bias" (Schneider and Wagemann 2012, 220) of QCA applications. However, it also implies that almost one-third of the studies are either not transparent about their necessity analysis or do not conduct any tests, hence not utilizing a distinct advantage of the method. We suggest that future IR studies applying QCA conduct and report a necessity analysis. Almost half of all studies in our sample (47 percent) searching for them find at least one necessary condition, which illustrates the potential benefit of this procedure.

Nowadays, the commonly accepted consistency threshold for a necessary condition is 0.9 (Ragin 2008; Schneider and Wagemann 2012), implying that in 90 percent of the set-membership values where outcome is present, the condition is present as well. In the sample under study, several articles refer to outdated thresholds, such as the thresholds of 0.65 for a "usually necessary condition" and 0.8 for an "almost necessary condition" (e.g., Adhikari and Samford 2013; Bhattacharya and Burns 2019). Using these thresholds is not advisable for at least two reasons: First, statements of necessity are strong statements. Few scholars would claim that a condition that is present in 65 percent or even 85 percent of the set-membership values for the outcome should be treated as a necessary condition. Second, it is empirically

rather common that conditions pass the 0.65 threshold in necessity tests, hence undermining QCA's ability to distinguish between relevant and irrelevant conditions.

Consistency scores are a key measure because they indicate the fit between the QCA solution and the empirical evidence. It is an established standard to use a minimum consistency cutoff point of 0.75 for the inclusion of truth table rows (Ragin 2008). Lower values are not prohibited, but they would require thorough discussion and justification. We obtained consistency values for the solutions of thirty-seven studies in our sample and they ranged from 0.70 to 1.00. Only three studies yielded solutions with a consistency below 0.75, and in two of them, they were accompanied by further solutions with consistency scores above this threshold. Overall, QCA research in IR does well in generating solutions with sufficiently high consistency scores. That said, four studies use consistency cutoff values below 0.75 (with one being as low as 0.5), indicating that at least parts of the respective solutions are not well in line with the empirical data.

The *coverage* score tells us how much of the outcome under study is explained by the solution. There is no established threshold for coverage as even solutions that explain a small amount of the outcome can be helpful for researchers, for instance, when it comes to understanding certain cases or specific causal pathways. Nine studies in our sample do not report coverage scores, while for the remaining thirty-four the scores are between 0.32 and 1, with an average of 0.77 (implying that 77 percent of the set-membership scores of the outcome are explained by the solution). This is rather high, particularly compared to other fields, such as business studies, where coverage scores as low as 0.039 are reported in peer-reviewed journals (Wagemann, Buche, and Siewert 2016).

As discussed above, when the *conditions-to-cases ratio* grows, issues related to limited diversity emerge. This does not have to be a problem, as there will often be a certain degree of limited diversity. However, when there is not enough empirical evidence to test certain theoretical expectations, then the algorithm may produce false-positive findings, leading to unreliable results (Marx and Dusa 2011).

In the IR literature we review here, the median number of conditions is five and the median number of cases twenty-eight, with a median conditions-to-cases ratio of 5.6. This is well in line with established benchmarks. For instance, Marx and Duşa (2011) recommend having at least eighteen to twenty cases for five conditions and Mello (2021, 28) advises a ratio of at least five cases per condition in such a research design. With solid qualitative knowledge of their sample, researchers might also study the same numbers of conditions with a slightly smaller number of cases and vice versa (Schneider and Wagemann 2012). That said, several studies still use a large number of conditions with a low-to-medium number of cases, hence decreasing our confidence in their results. Particularly pronounced examples include ten conditions with eleven cases (resulting in at least 1,013 logical remainders) and six conditions with ten cases (resulting in at least fifty-four logical remainders). IR researchers should be aware of this "many conditions, few cases" issue.

Testing robustness is an emerging yet still insufficiently elaborated issue in QCA research. More than a decade ago, Skaaning (2011, 391) already argued that robustness tests "should be regarded as an important, and maybe even indispensable, analytical step" to ensure that QCA results are not driven by (arbitrary) choices of the researcher. Scholars have proposed to check the robustness of results vis-à-vis different frequency and consistency thresholds, modified samples of cases, different sets of conditions, and alternative calibration decisions (Skaaning 2011; Schneider and Wagemann 2012). Limited guidance on how to interpret robustness checks is available, and the problem is certainly less acute for small samples where researchers are very familiar with their cases. Nevertheless, discussing the robustness of findings is an important part of QCA research (see also Oana, Schneider, and Thomann 2021).

Our findings show that the field of IR falls short of meeting this criterion. Of the studies in our sample, only 44 percent explicitly mention that they conducted any robustness tests. Many of these studies limit themselves to a few alternative analyses, often focusing on one particular aspect like a calibration decision or the replacement of one condition. Others have conducted more comprehensive robustness tests with twenty (Ide et al. 2021) or even thirty (Haesebrouck 2017a) alternative analyses. If summarized by a comprehensive table (in the appendix) and briefly discussed (in the main article), these tests can help communicating to readers how stable the obtained QCA results are (see Mross, Fiedler, and Grävingholt 2021 for a recent example), hence increasing trust in the analysis. Any viable alternative decisions regarding calibration and the analysis should be mentioned and, ideally, be addressed with separate robustness tests.

The *type of solution* to be used remains a contested issue in QCA (Haesebrouck and Thomann 2021; Oana, Schneider, and Thomann 2021).[7] Baumgartner and Thiem (2020) argue that only parsimonious solutions should be used for causal inference because the complex and intermediate solutions may both contain redundant elements and are thus regarded as incorrect. However, the parsimonious solution itself may rest upon "untenable assumptions" about logical remainders, and it should thus be used with care (Schneider and Wagemann 2012, 176). Moreover, Duşa (2019, 1) finds that an intermediate solution that incorporates directional expectations "emerges as the best hybrid that is suitable for causal analysis." This ties in with the emphasis on an intermediate solution that rests upon "plausible counterfactuals" (Mello 2021, 139).

The majority of QCA applications in IR seems to follow arguments in favor of the intermediate solution, which is discussed as the main solution by seventeen studies. The parsimonious solution is used nine times and the complex solution only in seven articles (presumably because the latter solution type is often descriptive and hard to interpret). While many studies are transparent about the solution type they use (thirty-three out of forty-three), there is usually limited discussion around why a certain solution is focused upon. In light of the unresolved debate, we recommend more elaboration on the choice of solution type. At the same, we caution against tendencies to reject research results solely on the basis of which solution type was emphasized. All three solutions are acceptable, but they come with different treatments of logical remainders and that issue needs to be addressed in a straightforward manner.

Above, we identified the possibility to combine quantitative and qualitative *data* as a major advantage of QCA. To be sure, utilizing just one type of data is perfectly fine if appropriate to the analysis. However, there are several studies in our sample that draw on both quantitative and qualitative data sources to produce innovative insights. Berlin (2016), for instance, studies why some states comply with arrest warrants by the International Criminal Tribunal for Rwanda (ICTR), while others do not. He finds that the presence of constituencies lobbying for noncompliance (drawn from the qualitative literature) in combination with low foreign-aid dependency (indicated by quantitative data on aid inflows) is an important causal pathway to noncompliance. This pattern would have remained undetected if only quantitative or only qualitative data had been used.

We encourage IR scholars to take an example from this and similar studies (e.g., Mello 2012; Young 2019; Jensen, Seate, and James 2020; Ide et al. 2021) and explore possibilities to combine quantitative and qualitative data via QCA. In our sample, only eight out of forty-three studies use this possibility, while the remaining thirty-five draw (almost) exclusively on one data type.

[7] This is a larger debate that would go beyond the scope of this article. Interested readers should consult the cited contributions.

Another issue related to data is *limited variation*. Little variation of a condition (or of the outcome) within the sample studied could result in findings that are driven by the underlying data structure rather than observed patterns. This is by no means a specific QCA problem but poses a challenge to all empirical studies. However, when reviewing QCA studies submitted to IR journals, we occasionally encounter issues related to limited variation, and the issue also emerges in our sample.[8]

For instance, the political constraints condition used by Jano (2016) shows calibrated values below 0.5 for all thirty-five cases. This essentially means that no country in this sample is a (partial) member of the set of countries with strong institutional constraints to political changes. Consequentially, the absence of political constraints is a necessary condition (consistency of 0.99) and plays a prominent role in the solution formula as well. However, the same would be true for any condition absent in the entire sample (e.g., the country having a direct flight to Sydney), no matter if it is causally relevant for the outcome or not. Berlin's (2016) high foreign direct investment condition faces a similar challenge, as it is only present in two out of twenty-six cases.

There are three, partially overlapping strategies to address limited variation. First, apart from reporting the truth table and their raw and calibrated data, researchers could revisit their research design to replace conditions that show little variation (and thus do not properly differentiate between the selected cases) with conditions that serve as difference makers. Second, if QCA users detect a condition that shows little variation, they could turn it into a *scope condition* for the analysis while also reflecting what this means for the generalization of findings. Alternatively, and third, calibration decisions could be reconsidered to emphasize variation of a condition. This is no eclectic process, however: calibration decisions still need to be in line with theoretical expectations and/or the data structure.

An additional advantage of QCA outlined above is its high potential for *combination with other methods*—such as regression analysis (Vis 2012; Meuer and Rupietta 2017) or process tracing (Beach and Rohlfing 2018; Mross 2021)—in order to triangulate results and yield nuanced insights. In our sample, QCA has been combined several times with case studies. The latter proved to be highly complementary to QCA as they produced insights that went beyond yet enriched the QCA findings. Researchers used case studies to test the plausibility of the causal pathways indicated by the QCA (Zimmermann 2016), to examine causal mechanisms underlying the QCA solutions (Young 2019), to explore contradictory cases that remained unexplained by the QCA (Yoo 2017), or to investigate the temporal ordering of conditions identified by the QCA (Lindemann and Wimmer 2018).

Statistical approaches can also enrich and complement QCA insights. Researchers have used statistical analyses, for instance, to test whether QCA findings hold when using different data analysis techniques or larger samples (Bara 2014) or to complement set-theoretic tests with descriptive statistics where expectations derive from a mostly quantitative literature (Mello 2020). Given that QCA works best for medium-*N* analyses and is not commonly used for estimating the net effects of individual conditions, testing the QCA results in larger samples and specifying the effect of individual conditions are areas where statistical approaches can complement QCA very well (Vis 2012).

Some studies even embed QCA in broad multi-method designs using three or more methods (Dong et al. 2015; Ide 2018a). That said, almost two-thirds of the IR articles we sampled (65 percent) employ no additional method, and only 7 percent use more than one additional method. This is of course no shortfall as QCA is a legitimate method by itself. Furthermore, there are considerable hurdles to using various methods in a single article, including the requirement to command different approaches, epistemological inconsistencies, and word limits. Nevertheless, as

[8] As well as in QCA studies in other disciplines (Ide 2018b).

outlined in the previous paragraph, our review revealed several ways in which IR studies combined QCA with other methods to yield nuanced findings. We deem it promising for further research to undertake similar efforts.

A major issue we encountered during our review is a lack of *transparency*, at least regarding some essential aspects, by many studies. The numbers reported in this part exclude studies that did not state essential information but where we were able to retrieve them by contacting the authors or rerunning the analysis with the data provided. The consistency cutoff for truth table rows and the type of solution used are crucial for the interpretation of findings, particularly in the light of recent debates (see above). However, we were unable to get access to this information for 16 and 23 percent of the articles in our sample, respectively. The assumptions underlying the intermediate solution were rarely stated. Coverage and consistency are the core measures of fit for QCA applications, but despite additional efforts, we could not retrieve them for 16 and 12 percent of the articles reviewed. Three articles provide no raw data, calibrated data, or truth tables at all, so it was impossible to replicate their results.

Providing such key information is important to ensure that QCA research is transparent and reproducible (Schneider and Wagemann 2012; Oana, Schneider, and Thomann 2021). QCA applications in IR need to improve on this aspect. Examples are set by studies that transparently explain their calibration decisions (Bhattacharya and Burns 2019; Polman 2020) or use extensive online appendices to document their procedures (e.g., Haesebrouck 2017a; Gromes 2019; Witte 2020).

Another area where the observed studies varied greatly is *software* usage. While the fs/QCA software remains most common, there are now a wide variety of different packages and programs with which to conduct QCA, including the Tosmana, R packages, and SetMethods.[9] In the IR literature we reviewed, fifteen studies used fs/QCA, ten studies used R, and one study used Tosmana. Seventeen studies did not indicate which software their analysis relied on. This is problematic for the replication of results because the software varies with regard to default settings and the measures of fit that are reported. IR researchers should, therefore, always state which software and version they use for the analysis.

Finally, our review also examined whether the observed studies incorporated *visualizations* in their articles. Recent work has highlighted the benefits of displaying QCA results and/or theoretical expectations in a visual way. These include the more effective communication of results and enhanced transparency (Rubinson 2019). Only nine studies in our sample included some form of illustration. Examples of this include *X*–*Y* plots for the solution and/or individual paths (e.g., Binder 2015; Mello 2020), *X*–*Y* plots to visualize the interplay of different conditions (e.g., van der Maat 2011; Jensen, Seate, and James 2020), or schematic visualizations of results (Yoo 2017). Clearly, this is an area where QCA research in IR has great potential.

In this section, we picked up a number of core indicators and quality standards for QCA analyses and assessed how they are addressed by applications in the field of IR. Before concluding the article, two remarks are due.

First, this list is by no means exhaustive. There are general quality standards for all empirical research methods, such as sample selection and interpretation of results, which we did not address here in greater depth. Likewise, several issues are specific to one or very few studies in our sample and hence not discussed at length. To give just three examples: (1) Calibrating missing values as zero does by no means exclude them from the sample but rather entails a very strong assumption that a case is completely out of the respective set of cases (Bhattacharya and Burns 2019). (2) Calibrating a case as being neither in nor out of a sample of cases (0.5) results in the omission of the case from the truth table analysis and hence the loss

[9] For a comprehensive list of available software, see the COMPASS website: https://compasss.org/software/.

of empirical information (Tobin 2017). Whenever possible, this should be avoided. (3) Solutions can be long and complex, hence making interpretation hard and indicating that the results are descriptive rather than analytical. The study of Yoo (2017), for instance, returns six different solution paths with each containing two or three conditions, hence forcing the author to limit his discussion to a part of the solution only. Strategies to address this issue include checking whether there are omitted variables or whether the cases are too heterogenous to be covered by a single analysis (Radaelli and Wagemann 2019).

Second, the standards and good practices discussed here are by no means absolute and authoritative, hence depriving IR researchers of analytical flexibility when employing QCA. Deviations from these guidelines are possible but should be made transparent and justified well. Haesebrouck's (2017a, 150) study is illustrative here. It opts for a consistency cutoff value of 0.67 in one of the analyses, which is well below the standard threshold of 0.75. However, Haesebrouck uses his qualitative knowledge to argue that the one case that drives the low consistency value of one truth table row is actually a "false negative" and that this configuration should hence be included.

Discussion and Conclusion

Qualitative comparative analysis (QCA) is an emerging method in the field of International Relations (IR). The number of journal articles using QCA is rising faster than the number of overall IR articles, and the growth of QCA applications has been rapid since 2015. This poses three sets of questions to readers, reviewers, supervisors, (potential) users, and method scholars in IR: (1) What is QCA and how does it work? (2) What advantages does QCA offer in the field of IR? (3) Does QCA research in IR follow established good practices, and if/where not, what strategies are available to improve research? We discuss the first question briefly before extensively dealing with the second and third questions. This discussion is based on a comprehensive review of forty-three empirical QCA applications in IR journals published between 1987 and 2020. We find that scholars have utilized QCA to enrich debates around a broad range of issues on IR, with a particular focus on (1) peace and conflict, (2) global environmental politics, (3) foreign policy, and (4) international regulations.

QCA offers several unique benefits to IR scholars: it allows for the detection of causal complexity characterized by conjunctural causation and equifinality as well as to distinguish analytically between necessary and sufficient conditions for an outcome. QCA strives to establish a middle way between qualitative approaches (as it is sensitive to the particularities of individual cases) and quantitative approaches (as it can provide a modest degree of generalization) and provides a missing middle between large-*N* and small-*N* approaches. It also facilitates the integration of quantitative and qualitative information during a single analysis via the calibration procedure. Finally, QCA can be fruitfully combined with other methods because it allows the identification of ideal types, (typical or outlier) cases for in-depth qualitative studies, and relevant combinations of conditions to be tested by statistical analyses. However, QCA also faces some limitations, particularly when it comes to analysis with very large or very small numbers of cases, when relevant information is lost during the calibration process, or when causal links have to be traced and proven.

Overall, our review suggests that IR does quite well when applying QCA. Most of the forty-three articles meet established methodological standards (or discuss intensively why they deviate from them). This is particularly true for "bread-and-butter" aspects of QCA such as consistency scores, coverage scores, and conditions-to-cases ratios. We detected issues like outdated necessity thresholds or an untenable conditions-to-cases ratio only in a few individual studies and provide suggestions for

how to address them. The large majority of the studies in our sample also derive some genuine benefits from using QCA, hence complementing insights from other methods in their respective IR subfields.

This positive impression holds when comparing IR to other research areas such as public policy (Rihoux, Rezsöhazy, and Bol 2011) or public administration (Thomann and Ege 2020). Unfortunately, there is lack of similarly systematic reviews for other fields that would allow for structured comparisons. That said, we can draw some preliminary insights from comparing our results on IR with Wagemann, Buche, and Siewert's (2016) findings on QCA in business research. Both fields suffer from issues related to insufficient transparency and some studies using untenable conditions-to-case ratios, but these seem to be more widespread in business research. Likewise, good practices regarding consistency thresholds are more widely implemented in IR: Only one out of thirty-six studies includes truth table rows with a very low consistency score (business research: seventeen out of forty-two), and while many solution consistency scores in business research are below 0.5, no solution consistency in our sample is below 0.7. QCA studies in IR (70 percent) also test more frequently for necessary conditions than QCA studies in business research (25 percent).

These are encouraging indicators. However, based on our review, we urge QCA users in IR to pay further attention to three aspects:

1) Multi-method research: As stated above, it is perfectly reasonable to use QCA as a stand-alone method and/or with one type of data (quantitative or qualitative). However, QCA has considerable potential when used to cross boundaries. Our review revealed how scholars complemented quantitative environmental data with case study knowledge (Ide et al. 2021) or in-depth insights from interviews with sociodemocratic characteristics (Jensen, Seate, and James 2020) to advance research frontiers. Likewise, QCA can inspire statistical analyses to check the external validity of results (Bara 2014) or support elaborate case-selection strategies (Yoo 2017). Sixty-five percent of the studies in our sample use no other method, and 81 percent draw on one data type only. This indicates that crossing such boundaries is no easy task: it might require IR researchers to gather data in ways unfamiliar to them (e.g., ethnography or processing of large datasets) or to expand their methodological knowledge. However, there is ample evidence that the yields of such endeavors are worth the efforts.
2) Robustness: The issue of robustness gained prominence in QCA in the early 2010s but has received limited attention since then. While there is agreement on which kinds of robustness tests to perform, there is little guidance on how to interpret them. This might explain why only a minority (44 percent) of IR studies conduct any robustness tests (and even fewer adopt a systematic procedure to do so), but it does not justify this omission. Analytical decisions can affect QCA results, making robustness checks an important part of the research process (Skaaning 2011; Oana, Schneider, and Thomann 2021). IR scholars should systematically test for robustness in future QCA research. Given its long history of engaging with debates about the robustness of research findings (Moravcsik 2014; Brigden and Gohdes 2020), IR is also well equipped to contribute to QCA debates on this issue.

3) Transparency: Being transparent about data, decisions, and results is essential for research to be reproducible, credible, and cumulative. Yet, we note a lack of transparency of QCA research in IR. Despite our efforts to contact authors, locate supplementary material for which no correct link was provided, and rerun analyses, we were unable to retrieve important information for a considerable number of studies. These include analytically relevant decisions on the software used (39 percent), consistency cutoffs (16 percent), and solution types (23 percent) as well as key metrics such as coverage (16 percent) and consistency (12 percent). Such a lack of transparency is specific neither to QCA nor to IR research, but with major initiatives under way to address this issue (Nosek et al. 2015), it requires the urgent attention of scholars. With the vast majority of IR journals nowadays offering the possibility to have extensive online appendices and many QCA scholars already using this (e.g., Lindemann and Wimmer 2018; Gromes 2019; Witte 2020), information on raw data, calibrated data, analytical decisions, and key indicators should be transparently shared.

QCA has proven to be a powerful addition to the toolkit of IR researchers seeking to disentangle complex causal patterns. By following these recommendations, the field of IR can further improve the already high quality of its QCA applications, hence paving the way for further theoretically and politically relevant insights.

Acknowledgment

Tobias Ide's research was funded by the Australian Research Council (ARC) under the DECRA scheme (DE190101268).

Supplementary Information

Supplementary information is available in the *International Studies Review* data archive.

References

Achen, Christopher H. 2005. "Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong." *Conflict Management and Peace Science* 22 (4): 327–39.

Adhikari, Prakash, and Steven Samford. 2013. "The Nepali State and the Dynamics of the Maoist Insurgency." *Studies in Comparative International Development* 48 (4): 457–81.

Ansorg, Nadine. 2014. "Wars without Borders: Conditions for the Development of Regional Conflict Systems in Sub-Saharan Africa." *International Area Studies Review* 17 (3): 295–312.

Bara, Corinne. 2014. "Incentives and Opportunities: A Complexity-Oriented Explanation of Violent Ethnic Conflict." *Journal of Peace Research* 51 (6): 696–710.

Baumgartner, Michael, and Alrik Thiem. 2020. "Often Trusted but Never (Properly) Tested: Evaluating Qualitative Comparative Analysis." *Sociological Methods & Research* 49 (2): 279–311.

Beach, Derek, and Ingo Rohlfing. 2018. "Integrating Cross-Case Analyses and Process Tracing in Set-Theoretic Research: Strategies and Parameters of Debate." *Sociological Methods & Research* 47 (1): 3–36.

Berg-Schlosser, Dirk, Gisèle de Meur, Benoît Rihoux, and Charles C. Ragin. 2009. "Qualitative Comparative Analysis (QCA) as an Approach." In *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, edited by Benoît Rihoux and Charles C. Ragin, 1–18. Thousand Oaks, CA: SAGE.

Berlin, Mark S. 2016. "Why (Not) Arrest? Third-Party State Compliance and Noncompliance with International Criminal Tribunals." *Journal of Human Rights* 15 (4): 509–32.

Bhattacharya, Srobana, and Courtney Burns. 2019. "What's War Got to Do with It? Post-Conflict Effects on Gender Equality in South and Southeast Asia, 1975–2006." *Journal of Asian Security and International Affairs* 6 (1): 55–81.

Binder, Martin. 2015. "Paths to Intervention: What Explains the UN's Selective Response to Humanitarian Crises." *Journal of Peace Research* 52 (6): 712–26.

Boogaerts, Andreas. 2018. "Beyond Norms: A Configurational Analysis of the EU's Arab Spring Sanctions." *Foreign Policy Analysis* 14 (3): 408–28.

Breitmeier, Helmut, Arild Underdal, and Oran R. Young. 2011. "The Effectiveness of International Environmental Regimes: Comparing and Contrasting Findings from Quantitative Research." *International Studies Review* 13 (4): 579–605.

Brigden, Noelle K., and Anita R. Gohdes. 2020. "The Politics of Data Access in Studying Violence across Methodological Boundaries: What We Can Learn from Each Other?" *International Studies Review* 22 (2): 250–67.

Caspersen, Nina. 2019. "Human Rights in Territorial Peace Agreements." *Review of International Studies* 45 (4): 527–49.

Daoudy, Marwa. 2021. "Rethinking the Climate–Conflict Nexus: A Human–Environmental–Climate Security Approach." *Global Environmental Politics* 21 (3): 4–25.

de Block, Debora, and Barbara Vis. 2018. "Addressing the Challenges Related to Transforming Qualitative into Quantitative Data in Qualitative Comparative Analysis." *Journal of Mixed Methods Research* 13 (4): 503–35.

de Meur, Gisèle, Benoît Rihoux, and Sakura Yamasaki. 2009. "Addressing the Critiques of QCA." In *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, edited by Benoît Rihoux and Charles C. Ragin, 147–65. Thousand Oaks, CA: SAGE.

Dong, Bingying, Ling Zhu, Kevin Li, and Meifeng Luo. 2015. "Acceptance of the International Compensation Regime for Tanker Oil Pollution - and Its Implications for China." *Marine Policy* 61: 179–86.

Dunne, Tim, Lene Hansen, and Colin Wight. 2013. "The End of International Relations Theory?" *European Journal of International Relations* 19 (3): 405–25.

Dusa, Adrian. 2019. "Critical Tension: Sufficiency and Parsimony in QCA." *Sociological Methods & Research*. https://doi.org/10.1177%2F0049124119882456.

George, Alexander L., and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.

Goertz, G. 2016. "Multimethod Research." *Security Studies* 25 (1): 3–24.

Goertz, Gary, and James Mahoney. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.

Gromes, Thorsten. 2019. "Does Peacekeeping Only Work in Easy Environments? An Analysis of Conflict Characteristics, Mission Profiles, and Civil War Recurrence." *Contemporary Security Policy* 40 (4): 459–80.

Grynaviski, Eric, and Amy Hsieh. 2015. "Hierarchy and Judicial Institutions: Arbitration and Ideology in the Hellenistic World." *International Organization* 69 (3): 697–729.

Haesebrouck, Tim. 2017a. "EU Member State Participation in Military Operations: A Configurational Comparative Analysis." *Cambridge Review of International Affairs* 30 (2–3): 137–59.

———. 2017b. "NATO Burden Sharing in Libya: A Fuzzy Set Qualitative Comparative Analysis." *Journal of Conflict Resolution* 61 (10): 2235–61.

Haesebrouck, Tim, and Eva Thomann. 2021. "Introduction: Causation, Correctness, and Solution Types in Configurational Comparative Methods." *Quality & Quantity*. https://doi.org/10.1007/s11135-021-01209-4.

Hao, Shinan, and Qiqi Gao. 2016. "East Asian Pathways toward Democracy: A Qualitative Comparative Analysis of 'the Third Wave'." *Journal of East Asian Studies* 16 (2): 239–60.

Ide, Tobias. 2018a. "Does Environmental Peacemaking between States Work? Insights on Cooperative Environmental Agreements and Reconciliation in International Rivalries." *Journal of Peace Research* 55 (3): 351–65.

———. 2018b. "On the Appropriate Use of QCA in Environmental Management Research: A Comment on Hossu et al." *Ambio* 47 (7): 831–32.

Ide, Tobias, Juan Miguel Rodriguez Lopez, Christiane Fröhlich, and Jürgen Scheffran. 2021. "Pathways to Water Conflict during Drought in the Mena Region." *Journal of Peace Research* 58 (3): 568–82.

Ide, Tobias, Michael Brzoska, Jonathan F. Donges, and Carl-Friedrich Schleussner. 2020. "Multi-Method Evidence for When and How Climate-Related Disasters Contribute to Armed Conflict Risk." *Global Environmental Change* 62 (1): 1–8.

Imai, Kosuke. 2017. *Quantitative Social Science: An Introduction*. Princeton, NJ: Princeton University Press.

Jano, Dorian. 2016. "Compliance with EU Legislation in the Pre-Accession Countries of South East Europe (2005–2011): A Fuzzy-Set Qualitative Comparative Analysis." *Journal of European Integration* 38 (1): 1–22.

Jensen, Michael A., Anita Atwell Seate, and Patrick A. James. 2020. "Radicalization to Violence: A Pathway Approach to Studying Extremism." *Terrorism and Political Violence* 32 (5): 1067–90.

Kahwati, Leila C., and Heather L. Kane. 2020. *Qualitative Comparative Analysis in Mixed Methods Research and Evaluation*. Los Angeles, CA: SAGE.

Lindemann, Stefan, and Andreas Wimmer. 2018. "Repression and Refuge: Why Only Some Politically Excluded Ethnic Groups Rebel." *Journal of Peace Research* 55 (3): 305–19.

Mackie, John L. 1965. "Causes and Conditions." *American Philosophical Quarterly* 2 (4): 245–64.

Maerz, Seraphine F. 2020. "The Many Faces of Authoritarian Persistence: A Set-Theory Perspective on the Survival Strategies of Authoritarian Regimes." *Government and Opposition* 55 (1): 64–87.

Mahoney, James. 2021. *The Logic of Social Science*. Princeton, NJ: Princeton University Press.

Mahoney, James, Erin Kimball, and Kendra L. Koivu. 2009. "The Logic of Historical Explanation in the Social Sciences." *Comparative Political Studies* 42 (1): 114–46.

Marx, Axel, and Adrian Dusa. 2011. "Crisp-Set Qualitative Comparative Analysis (CsQCA), Contradictions and Consistency Benchmarks for Model Specification." *Methodological Innovations Online* 6 (2): 103–48.

Marx, Axel, Benoît Rihoux, and Charles Ragin. 2014. "The Origins, Development, and Application of Qualitative Comparative Analysis: The First 25 Years." *European Political Science Review* 6 (1): 115–42.

Mello, Patrick A. 2012. "Parliamentary Peace or Partisan Politics? Democracies' Participation in the Iraq War." *Journal of International Relations and Development* 15 (3): 420–53.

———. 2020. "Paths towards Coalition Defection: Democracies and Withdrawal from the Iraq War." *European Journal of International Security* 5 (1): 45–76.

———. 2021. *Qualitative Comparative Analysis: An Introduction to Research Design and Application*. Washington, DC: Georgetown University Press.

Meuer, Johannes, and Christian Rupietta. 2017. "Integrating QCA and HLM for Multilevel Research on Organizational Configurations." *Organizational Research Methods* 20 (2): 324–42.

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

Mross, Karina. 2021. "Disaggregating Democracy Aid to Explain Peaceful Democratisation after Civil Wars." *European Journal of International Security*. https://doi.org/10.1017/eis.2021.36.

Mross, Karina, and Charlotte Fiedler, and Grävingholt, Jörn. 2021. "Identifying Pathways to Peace: How International Support Can Help Prevent Conflict Recurrence." *International Studies Quarterly*. https://doi.org/10.1093/isq/sqab091.

Nosek, B.A., G. Alter, G.C. Banks, D. Borsboom, S.D. Bowman, S.J. Breckler, and S. Buck et al. 2015. "Promoting an Open Research Culture." *Science* 348 (6242): 1422–25.

Oana, Ioana-Elena, Carsten Q. Schneider, and Eva Thomann. 2021. *Qualitative Comparative Analysis Using R: A Beginner's Guide*. New York: Cambridge University Press.

Polman, Daniel. 2020. "Participation of Implementing Agencies in European Administrative Networks." *JCMS: Journal of Common Market Studies* 58 (4): 818–35.

Radaelli, Claudio M., and Claudius Wagemann. 2019. "What Did I Leave Out? Omitted Variables in Regression and Qualitative Comparative Analysis." *European Political Science* 18 (2): 275–90.

Ragin, Charles C. 1987. *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies*. Berkeley, CA: University of California Press.

———. 2000. *Fuzzy-Set Social Sciences*. Chicago: University of Chicago Press.

———. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago, IL: University of Chicago Press.

Reynaert, Vicky. 2011. "Explaining EU Aid Allocation in the Mediterranean: A Fuzzy-Set Qualitative Comparative Analysis." *Mediterranean Politics* 16 (3): 405–26.

Rihoux, Benoît, and Axel Marx. 2013. "Qualitative Comparative Analysis at 25: State of Play and Agenda." *Political Research Quarterly* 66 (1): 167–71.

Rihoux, Benoît, Ilona Rezsöhazy, and Damien Bol. 2011. "Qualitative Comparative Analysis (QCA) in Public Policy Analysis: An Extensive Review." *German Policy Studies* 7 (3): 9–82.

Rohlfing, Ingo. 2020. "The Choice between Crisp and Fuzzy Sets in Qualitative Comparative Analysis and the Ambiguous Consequences for Finding Consistent Set Relations." *Field Methods* 32 (1): 75–88.

Rubenzer, Trevor. 2008. "Ethnic Minority Interest Group Attributes and US Foreign Policy Influence: A Qualitative Comparative Analysis." *Foreign Policy Analysis* 4 (2): 169–85.

Rubinson, Claude. 2019. "Presenting Qualitative Comparative Analysis: Notation, Tabular Layout, and Visualization." *Methodological Innovations* 12 (2): 1–22.

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

Skaaning, Svend-Erik. 2011. "Assessing the Robustness of Crisp-Set and Fuzzy-Set QCA Results." *Sociological Methods & Research* 40 (2): 391–408.

Taydas, Zaynap, Jason Enia, and Patrick James. 2011. "Why Do Civil Wars Occur? Another Look at the Theoretical Dichotomy of Opportunity versus Grievance." *Review of International Studies* 37 (5): 2627–50.

Thomann, Eva, and Jörn Ege. 2020. "Qualitative Comparative Analysis (QCA) in Public Administration." In *The Oxford Encyclopedia of Public Administration* edited by Guy Peters, B., and Ian Thynne. Oxford: Oxford University Press.

Thomann, Eva, and Martino Maggetti. 2020. "Designing Research with Qualitative Comparative Analysis (QCA): Approaches, Challenges, and Tools." *Sociological Methods & Research* 49 (2): 356–86.

Tobin, Paul. 2017. "Leaders and Laggards: Climate Policy Ambition in Developed States." *Global Environmental Politics* 17 (4): 28–47.

van der Maat, Eelco. 2011. "Sleeping Hegemons: Third-Party Intervention Following Territorial Integrity Transgressions." *Journal of Peace Research* 48 (2): 201–15.

Vis, Barbara. 2012. "The Comparative Advantages of FsQCA and Regression Analysis for Moderately Large-N Analyses." *Sociological Methods & Research* 41 (1): 168–98.

Wagemann, Claudius, Jonas Buche, and Marcus B. Siewert. 2016. "QCA and Business Research: Work in Progress or a Consolidated Agenda." *Journal of Business Research* 69 (1): 2531–40.

Witte, Daniel. 2020. "Business for Climate: A Qualitative Comparative Analysis of Policy Support from Transnational Companies." *Global Environmental Politics* 20 (4): 167–91.

Yoo, In Tae. 2017. "Power Concentration or Parity? The Conditional Influence of Power Distribution in Regional Economic Integration." *International Relations of the Asia-Pacific* 17 (2): 203–32.

Young, Alasdair R. 2019. "Where's the Demand? Explaining the EU's Surprisingly Constructive Response to Adverse WTO Rulings." *Journal of European Integration* 41 (1): 9–27.

Zimmermann, Katharina. 2016. "Local Responses to the European Social Fund: A Cross-City Comparison of Usage and Change." *JCMS: Journal of Common Market Studies* 54 (6): 1465–84.