Computer Engineers Look at Qualitative Comparative Analysis

Ali Muhammad Ali Rushdi*, Raid Mohammad Salih Badawi
Department of Electrical and Computer Engineering
King Abdulaziz University
P. O. Box 80204, Jeddah 21589, Saudi Arabia
*Corresponding author: arushdi@kau.edu.sa

(Received February 11, 2019; Accepted April 22, 2019)

Abstract
Qualitative Comparative Analysis is a variant of Boolean Analysis that complements quantitative or statistical methods in many scientific disciplines. Therefore, its technicalities resemble those of other variants of Boolean Analysis, such as the one employed by computer engineers for digital design. This paper offers a brief look at Qualitative Comparative Analysis from a Computer-Engineering perspective. Critical observations on some technicalities of Qualitative Comparative Analysis are presented, with an aim to initiate constructive and fruitful intellectual debate that might subsequently lead to desirable enhancements and improvements.

Keywords- Qualitative comparative analysis, Boolean analysis, Karnaugh map, Contradictions, Logical remainders, Prime implicants, Minimal sum, Complete sum.

1. Introduction
Qualitative Comparative Analysis (QCA) was first introduced in the seminal paper of Ragin et al. (1984) and became well established via the celebrated text of Ragin (1987). It is a comparative case-oriented research technique based on Boolean algebra and set theory, essentially intended for the analysis of a medium number of cases (Marx et al., 2013), and deliberately replicating the logic of case-study analysis in a mathematical framework different from that of statistical theory (Achen, 2005). Qualitative Comparative Analysis has now branched into three interrelated variants, namely crisp-set Qualitative Comparative Analysis (csQCA), multi-value Qualitative Comparative Analysis (mvQCA), and fuzzy-set Qualitative Comparative Analysis (fsQCA) (Marx et al., 2014; Roig-Tierno et al., 2017). Though the three variants have many common features, each variant has its own distinctive characteristics. Qualitative Comparative Analysis has the ambitious aim of bridging the gap between (and embodying some key strengths of) the qualitative and quantitative approaches typically used in humanities and social and political sciences (Rihoux, 2003). The utility of QCA has now extended dramatically, and QCA has recently permeated new areas (other than its conventional domains) such as engineering (Jordan, et al., 2011), management (Kan et al., 2016) and medicine (Rushdi and Rushdi, 2018b). Achen (2005) sees both methodologies of QCA and conventional quantitative methods, like both genders, as genuinely distinct but legitimate and complementary.

Marx et al. (2013) observe that the use of csQCA has been the focus of much criticism, and they identify and analyze two such criticisms. The first criticism concerns a key assumption on which csQCA is built, namely that contradictions “naturally” occur and that no omitted variables are excluded from an analysis. The second criticism concerns the sensitivity of the analysis to individual cases. They also note that these two criticisms are interrelated and have mutual dependence since they both touch on the selection of conditions and cases. Several other criticisms also exist, including issues of measurement and calibration, and the inadvertent use of undisciplined
simplifying assumptions. Lieberson (2004) suggests that QCA might be unable to distinguish randomly assigned values that have no meaning from real data. Lucas and Szatrowski (2014) admit that ‘existing research has demonstrated the use of QCA on real data,’ but then argue that ‘such data do not allow one to establish the method’s efficacy, because the true causes of real social phenomena are always contestable.’

Our observations in this paper do not pertain to the essence or nature of QCA, but are related to some technicalities of its implementation, with a stress on aspects of the parent variant of csQCA. These technicalities happen to be shared (and presumably mastered) by disciplines other than that of QCA. Prominent among such disciplines is our field of Computer Engineering, whose subfield of digital design is definitely the most conspicuous contemporary application of Boolean algebra (See Appendix A). Hopefully, the QCA community would not mind our intervention as outsiders, would welcome our comments, observations, and questions, and would tolerate our objections and criticism as a sort of constructive and fruitful intellectual debate.

The organization of the rest of this paper is as follows. Section 2 discusses the issue of manual versus automated tools. Section 3 comments on the techniques used to reconcile contradictions. Section 4 debates whether logical remainders could be assigned arbitrary values or not. Section 5 explores the possibility of binary encoding of multi-value variables. Section 6 argues for deriving the complete sum (as well as the minimal sum) of the pertinent Boolean function. Section 7 concludes the paper. To make the paper self-contained, its main text is augmented by two appendices. Appendix A attempts to reconcile the disparate jargons for (essentially the same) concepts used in digital design and QCA. Appendix B is a brief exposition of faithful or true representation of data via incompletely-specified Boolean functions. We deliberately delegate all the mathematical content of this paper to these two appendices. This allows readers to concentrate on the essence of our message without being bothered by rigorous or intricate details.

2. Manual versus Automated Tools
This section debates whether the Boolean minimization needed in QCA should be attained via automated algorithms or by manual means such as the Karnaugh map. We reiterate our argument (Rushdi and Badawi; 2017a) that implementation of QCA via computer programs is not really warranted for the dichotomous or crisp-set QCA (but could not probably be dispensed with for the multi-value QCA or fuzzy-set QCA).

Admittedly, the Karnaugh map is useful only for a small number of input variables. Its variable-handling capability might be extended to medium-sized problems by making it variable-entered (Rushdi, 2018a; 2018b; Rushdi and Ahmad, 2018). The map was initially introduced to handle problems of digital design, where typical real-life problems are large (and even extremely large). Therefore, its role in digital design is essentially limited to providing pictorial insight and achieving other pedagogical purposes such as demonstrating concepts, proving theorems, illustrating procedures, and exploring modular, repetitive or symmetric structures. The task of Boolean minimization of functions with large numbers of inputs is indisputably delegated to automated algorithms, which do not necessarily mimic Karnaugh-map procedures, but have more profound concepts and paradigms.

The case of Boolean minimization for Qualitative Comparative Analysis is actually a different story. In QCA, the number of input variables of ‘real-life’ problems seems to be never large. For example, Marx et al. (2014) surveyed almost five hundred papers of QCA applications in political
science during the period 2003-2011. They found that the number of conditions $n$ (independent variables) used in each of these papers ranged from 3 to 10. Rushdi and Badawi (2017a) found that the average value for $n$ in the aforementioned survey is slightly more than 5 and less than 6. There is no reason to expect that the number of independent variables is significantly different for QCA applications in fields other than political science, or during periods beyond the 2003-2011 period. The aforementioned values for $n$ suit the use of a Karnaugh map, whose conventional form is conveniently used up to 6 variables, while its variable-entered form is conveniently used up to 12 variables.

A further reason for favoring map Boolean minimization to automated one is that QCA is a labor-intensive, interactive, iterative and creative process (Rihoux, 2006). Boolean minimization is not a bottleneck for this process, and its automation does not contribute a significant time saving to the whole process. A not-so-important time saving is achieved via automation at the expense of the loss of the insight, flexibility and control gained via map use, and causes repeated jumping between manual and automated tasks (Rushdi and Badawi. 2017a).

It is really intriguing why the csQCA problems are being handled by automated means. Is this because learning Karnaugh-map procedures is more demanding than using readily-available software? Less intriguing is why the QCA researchers developed (from scratch) their own software to handle Boolean minimization, and were not content to adapt packages borrowed from the digital design field. An obvious possible explanation is that these packages were too large for QCA purposes, used terminology different from QCA jargon, and possibly lacked means to perform certain OCA tasks such as reconciling contradictions.

3. Reconciling Contradictions
In QCA, a ‘contradiction’ or a ‘logical-inconsistency’ value “C” is used for a configuration that has a “0” outcome for some observed cases and a “1” outcome for other observed cases (Marx and Duşa, 2011). Rihoux and de Meur (2009) and Jordan et al. (2011) discuss standard guidelines to resolve contradictions before further processing. We argue that further processing does not have to wait for the completion of contradiction resolution. The size of the set of legitimate outcomes could be increased by one to accommodate an extra outcome “C”, and hence represent the status quo (Rushdi, 2018a). Alternatively, the effect of our ignorance of contradictory values associated with our current inability to resolve contradictions can be mitigated by exhausting all possible assignments to the contradictory values (Rushdi, 2018a, Rushdi and Rushdi, 2018b). Admittedly, this leads to many candidate QCA solutions, instead of just one. Further research is needed to establish scientifically-sound criteria to select one solution out of the many available. Such a selection effectively achieves contradiction resolution after (rather than before) processing the data. An example of such ‘a posteriori’ resolution is available in Rushdi and Rushdi (2018b). In this example, a candidate outcome was preferred to another based on certain desirable features of the outcome and on expected voting powers of the input variables. The need to handle contradictions is a strong reason why QCA processing should be manual all throughout.

4. The Debate on Logical Remainders
There are two types of don’t-cares known in digital design, corresponding to certain configurations (truth-table lines or map cells). For one type, the output can possibly be either 0 or 1, and for the other type, the configuration cannot happen (Brown, 1990). Logical remainders in QCA terminology bear some similarity to the second type, as they refer to configurations, in which no cases have been, so far, observed. Among the critiques targeted at QCA, there is one about utilizing
logical remainders in obtaining parsimonious minimal formulas. Rihoux and de Meur (2009) pose the question “Isn’t it altogether audacious to make assumptions about non-observed cases?” The answer to this question is a qualified ‘yes.’ The reason is that an unobserved configuration might not be guaranteed to never happen, as it might become observed at a later time. Crama et al. (1988) assert that when only partial or incomplete observations are available, no method can provide definite answers. The observed data should be treated as temporary or tentative, and must be updated till somehow it could be deemed final. Rushdi and Badawi (2017a) suggest that the only rigorous course of action in case of incomplete data (albeit potentially inconvenient) is to employ a faithful representation for the outcome function. Such a faithful representation takes the form of a partially-defined function whose asserted part is a disjunction of the definite (certain) causes of the pertinent phenomenon, while its don’t-care part constitutes its potential or uncertain causes (See Appendix B).

Rushdi and Badawi (2017a) note that while some QCA researchers refrain from utilizing logical remainders for minimization (or for any other purpose), they are still setting these unknown values to zero, and hence are still making unwarranted or unjustified assumptions about non-observed data. There is no reason to prefer these assumptions to other ones, such as those leading to minimization. In fact, the choice of nullifying logical remainders can be thought of to be equally likely to (and hence equally acceptable as or equally deniable as) all other possible choices for the logical remainders. A faithful representation amounts to an exhaustive listing of all possible choices.

Rushdi and Badawi (2017a) have a more subtle objection to the act of ignoring logical remainders through nullifying them. They note that the logical remainders are deliberately nullified for an arbitrarily-selected form or literal of the outcome variable rather than for its complementary form or literal. This constitutes an unjustified bias in the treatment of a function $f$ and its complement $\bar{f}$. If the disjointed don’t-care part of one form $f$ of the function is negatively asserted (set to 0), thereby equating this form to its asserted part, then $\bar{f}$ is equated to the complement of the asserted part of its complement $f$. An unbiased treatment of $f$ and $\bar{f}$ would necessitate that if $f$ is equated to its asserted part, then $\bar{f}$ should be equated to its own asserted part, which differs from the complement of the asserted part of its complement (See Appendix B). This is not just a matter of elegance, but it is an essential requirement to preserve complementarity between $f$ and $\bar{f}$.

5. Binary Encoding of Multi-valued Variables

Multi-value Qualitative Comparative Analysis (mvQCA) has been the focus of many publications recently (Cronqvist, 2006; Rohlffing, 2012; Rushdi, 2018a; Thiem, 2013; Vink and Van Vliet, 2009; 2013). An important issue concerning mvQCA is the debate about expanding the input space when binary variables are used to encode multi-value ones. Rushdi (2018a) notes that the expansion in the input domain is the primary consequence of encoding multi-valued variables as binary variables. Such an encoding can be viewed as an act of switching from mvQCA to csQCA. Once the expansion in input space takes place, one realizes that the expansion is due to the introduction of unused configurations that definitely never happen, and hence can be arbitrarily assigned don’t-care values. The use of don’t-cares in the extra input space is a secondary and beneficial effect, rather than a primary and harmful one. The increase in size of the input domain is inadvertent indeed, but its harm is diluted (rather than aggravated) by having the extra don’t-care configurations. The appearance of don’t-cares is not a disadvantage but is an asset as it facilitates the minimization process and partially remedies the inconvenience caused by the larger domain size.
6. Complete sum versus Minimal Sum

The common QCA practice uses Boolean minimization to obtain the minimal sum (See Appendix A) of the pertinent output or effect function (Thiem and Duşa, 2013a; 2013b; Duşa and Thiem, 2015). This practice is definitely useful since it obtains the most compact or parsimonious characterization of the function. The minimal sum must cover the asserted part of the function (and might cover some or all of its don’t-care part) with a minimum number of prime implicants having a minimum total number of literals. Each of these prime implicants is minimally sufficient for the given output. Non-prime implicants are not included in the minimal sum since they are not minimally-sufficient for the output (albeit being sufficient ones). The compactness of the minimal sum is essentially a result of the fact that it is minimally necessary for the given output.

In general, the minimal sum does not contain all prime implicants (Crama and Hammer, 2011). The disjunction of all prime implicants is called the complete sum (Muroga, 1979), and is known also as the Blake Canonical Form (Brown, 1990). The complete sum lists explicitly all minimally-sufficient entities explicitly, and hence its role supplements that of the minimal sum which characterizes the function, but does not necessarily exhaust all minimally-sufficient entities. In fact, the minimal sum and the complete sum are just two different formulas or forms of the same function, and it is possible to go from one form to the other if the function is completely specified (i.e., if it does not have logical remainders), but this can be achieved at a cost. While the minimal sum is more parsimonious for identifying a Boolean function, the complete sum is more convenient for implementing Boolean reasoning (Brown, 1990; Rushdi and Rushdi, 2018a). Therefore, we recommend that QCA should report both types of sums, and not just the minimal one. Such a recommendation is useful for a comprehensive listing and assessment of all determinants or causes of the pertinent effect.

In line with the observation above, we feel a need for QCA to deduce more useful information from the Boolean function it constructs. For example, each pertinent variable might be assigned a Boolean importance metric, such as the Banzhaf index (Rushdi and Ba-Rukab, 2017a) or the Shapley-Shubik index (Rushdi and Ba-Rukab, 2017b). A pre-knowledge or a pre-supposition of the relative values of one of these metrics can aid in resolving contradictions (Rushdi and Rushdi, 2018b). Other potential deductions that can be made by QCA include deciding whether the pertinent function is independent of some of its (supposed) arguments, investigating whether this function is positive or negative in each of its arguments, checking whether the function is partially symmetric in some of its arguments (Rushdi and Badawi, 2017a).

7. Conclusions

A novel scientific discipline is typically established through the innovative efforts of paradigm pioneers, who have the prudence and courage to venture into unexplored territories under dominant darkness. However, the proponents of a newly-born discipline should not outright reject constructive criticism that might help to streamline and enhance their discipline. The QCA discipline is no exception. In this paper, we try to offer a helping hand by presenting views that seem alien to mainstream QCA. We hope that our views may serve as beneficial decorations to an already magnificent structure.

Appendix A: Concepts Common to Digital Design and QCA

This appendix attempts to reconcile the disparate jargons for minimization concepts used in digital design and QCA. More information is available in classical texts on Boolean functions, digital
design, and Boolean reasoning such as Brown (1990), Crama and Hammer (2011), Fletcher (1980), Hill and Peterson (1993), Lee (1978), Muroga (1979), and Roth and Kinney (2014). Detailed exposition of the required concepts is offered with a stress on QCA context in Rushdi (2018a) and Rushdi and Badawi (2017a; 2017b).

The two literals of a Boolean variable \( X_m \) are its complemented form \( \bar{X}_m \) and its uncomplemented one \( X_m \). A product (conjunction) of literals is called a term \( T(X) \) if a literal for each variable appears in it at most once, i.e., a term is an irredundant product (conjunction). The constant 1 is the multiplication (ANDing) identity and is the product or term of no literals. The dual of a term is the irredundant disjunction, called an alterm. The constant 0 is the addition (ORing) identity and is the sum or alterm of no literals. The constant 1 is not an alterm and the constant 0 is not a term.

A term \( T(X) \) is an implicant of a function \( f(X) \) if every \( T(X) \) satisfying \( \{T(X) = 1\} \) also satisfies \( \{f(X) = 1\} \), while the converse is not necessarily true. In QCA jargon, \( T(X) \) is said to be sufficient for \( f(X) \), while \( f(X) \) is said to be necessary for \( T(X) \). Alternative ways of stating that \( T(X) \) implies \( f(X) \) are given by the following equivalent propositions (Rushdi and Badawi, 2017b):

\[
T(X) \rightarrow f(X), \quad \text{(A.1)}
\]

\[
T(X) \leq f(X), \quad \text{(A.2)}
\]

\[
\text{If } T(X) \text{ then } f(X), \quad \text{(A.3)}
\]

\[
f(X) \text{ if } T(X), \quad \text{(A.4)}
\]

\[
T(X) \text{ only if } f(X), \quad \text{(A.5)}
\]

\[
T(X) \wedge \bar{f}(X) = 0, \quad \text{(A.6)}
\]

\[
\bar{T}(X) \vee f(X) = 1, \quad \text{(A.7)}
\]

\[
T(X) \wedge f(X) = T(X), \quad \text{(A.8)}
\]

\[
T(X) \vee f(X) = f(X). \quad \text{(A.9)}
\]

A term/alterm \( T_i(X) \) is said to subsume another term/alterm \( T_j(X) \) if the set of constituents of \( T_i(X) \) is a superset of that of \( T_j(X) \) (i.e., all constituents of \( T_i(X) \) are among those of \( T_j(X) \)). The constituents of a term/alterm are the entities (e.g., literals) ANDed/ORed by it.

A prime implicant \( P(X) \) of a Boolean function \( f(X) \) is an implicant of \( f(X) \) such that no other term subsumed by it is an implicant of \( f(X) \). The prime implicant \( P(X) \) is said to be irredundantly or minimally sufficient for the function \( f(X) \).

The Complete Sum \( CS(f(X)) \) of a Boolean function \( f(X) \) (also called its Blake Canonical Form \( BCF(f(X)) \)) is the disjunction (ORing) of all its prime implicants, and nothing else. The complete sum is a closure, unique and canonical formula for \( f(X) \). An irredundant disjunctive form
\( \text{IDF}(f(X)) \) of a Boolean function \( f(X) \) is a disjunction of some of its prime implicants that expresses \( f(X) \) but ceases to do so upon the removal of one of these prime implicants. Such a disjunction is said to be \textit{irredundantly necessary} for \( f(X) \), since \( f(X) \) implies it while failing to imply any other disjunction subsumed by it.

A \textbf{minimal sum} \( MS(f(X)) \) (\textit{minimal irredundant form} \( MIF(f(X)) \)) of a Boolean function \( f(X) \) is an irredundant disjunctive form for the function with the minimum number of prime implicants such that the total number of their literals is minimum. A minimal sum is said to be \textit{minimally necessary} for \( f(X) \), since it is a most compact (most economic) formula for \( f(X) \).

\textbf{Appendix B: Incompletely-Specified Boolean Functions}

An incompletely-specified Boolean function (also called a partially-defined Boolean function) \( f \) is defined in terms of two functions \( g \) and \( h \) by any of the following equivalent statements employing the don’t-care notation (Rushdi and Albarakati, 2014).

\[
f = g \lor d(h) = g \lor d(\bar{g}h) = g \lor d(g \lor h).
\] (B.1)

Each of the three definitions in (B.1) is rigorously understood to be equivalent to the system of two inequalities

\[
g \leq f. \tag{B.2a}
\]

\[
\bar{g} \bar{h} = \bar{g}(g \lor \bar{h}) = \bar{g}(\bar{g} \bar{h}) \leq \bar{f}. \tag{B.2b}
\]

The function \( g \) is called the \textit{asserted} part of \( f \), while the functions \( h \), \( \bar{g}h \), and \( g \lor h \) are called the \textit{don’t-care} part, the \textit{disjointed don’t-care} part, and the \textit{augmented don’t-care} part of the function \( f \), respectively. The symbol \( d(\ldots) \) denotes the don’t-care operator. The complement \( \bar{f} \) of \( f \) is defined in terms of the two functions \( g \) and \( h \) \textit{via} the don’t-care notation as,

\[
\bar{f} = \bar{g} \bar{h} \lor d(\bar{g}h) = \bar{g} \bar{h} \lor d(\bar{g}). \tag{B.3}
\]

A notable observation from (B.1) and (B.3) is that the asserted part \( \bar{g} \bar{h} \) of the complement \( \bar{f} \) of \( f \) is not necessarily equal to the complement \( \bar{g} \) of the asserted part \( g \) of \( f \). Equations (B.2) can be combined to express \( f \) and \( \bar{f} \) in the double-inequality notation as

\[
g \leq f \leq g \lor h, \tag{B.4a}
\]

\[
\bar{g} \bar{h} \leq \bar{f} \leq \bar{g}. \tag{B.4b}
\]

The conditions above can be combined to yield

\[
f \land \bar{f} = d(\bar{g}h), \tag{B.5a}
\]

\[
f \lor \bar{f} = g \lor \bar{h} \lor d(1), \tag{B.5b}
\]

\[
0 \leq f \land \bar{f} \leq \bar{g}h, \tag{B.6a}
\]
\[ g \lor \bar{h} \leq f \lor \bar{f} \leq 1. \]  
(B.6b)

Under the condition \( \{g \bar{h} = 0\} \), the double-inequality notation might be simplified to

\[ g \leq f \leq h, \]  
(B.7a)
\[ \bar{h} \leq \bar{f} \leq \bar{g}. \]  
(B.7b)

Each of the conditions (B.1)-(B.4) is equivalent to a single Boolean equation of the form

\[ g \bar{f} \lor \bar{g} \bar{h} f = 0. \]  
(B.8)

The general parametric solution of this Boolean equation for \( f \) and \( \bar{f} \) in terms of \( g \) and \( h \) is (Rushdi and Amashah, 2011)

\[ f = g \lor p h = g \lor p \bar{g} h = g \lor p (g \lor h), \]  
(B.9a)
\[ \bar{f} = \bar{g} \bar{h} \lor \bar{p} \bar{g} h = \bar{g} \bar{h} \lor \bar{p} \bar{g}. \]  
(B.9b)

where \( p \) is a parameter that belongs to the free Boolean algebra \( FB(g, h) \), which has 16 elements that are the two-valued Boolean functions of \( g \) and \( h \). The general solution in (B.9) is subject to a consistency condition that reduces to the valid identity \( \{0 = 0\} \). The particular solutions of (B.8) are obtained from the general solution in (B.9) as \( f = g \) and \( \bar{f} = \bar{g} \) for \( \{0 \leq p \leq \bar{h}\} \), and as \( f = g \lor h = g \lor \bar{g} h \), and \( \bar{f} = \bar{g} \bar{h} \) for \( \{h \leq p \leq 1\} \). Comparison of equations (B.9) to equations (B.1) and (B.3) reveals that the don’t-care operator \( d \) can be interpreted as a multiplying parameter \( p \).

**Conflict of Interest**

The authors declare that no conflict of interest exists.

**Acknowledgements**

The authors are greatly indebted and sincerely grateful to King Abdulaziz City for Science and Technology (KACST) of Saudi Arabia for providing generous financial assistance to support their research work culminating in the publication of this paper under grant number 1-17-01-009-0002.

**References**


