Set-theoretic and Modal Notions in Social Research

G. Rohwer, May 2010

Abstract. The article deals with the suggestion, made by Charles Ragin, that theoretical statements in social research most often can be formulated as statements about sets and set relations. In contrast to this view, it is argued that theoretical statements in social research often require modal notions referring to possibilities and probabilities which cannot be formulated in terms of sets and set relations. In order to show this, the article reformulates Ragin's set-theoretic approach in the conceptual framework of statistical variables, and then goes on to argue that social research is often interested in modal generalizations (probabilistic and deterministic rules) which require a fundamentally different conceptual framework.

In a series of publications, Charles Ragin has proposed a "set-theoretic approach" to social research that is based on the idea that observations can often be represented by sets of cases demarcated by common properties (Ragin 1987; 2000; 2008). This representation allows one to use set-theoretic notions, like set inclusion, for investigating relationships between properties as they are manifested in given sets of cases. Corresponding techniques have been called "Qualitative Comparative Analysis" (Ragin 1987; 2000) and, intended as a more general heading for a variety of similar techniques, "Configurational Comparative Methods" (Rihoux and Ragin 2009). As a method of data analysis, this approach can certainly be useful, in particular when one is interested in explicitly comparing a (small or medium) number of actually observed and identifiable cases. In addition to this application of set-theoretic concepts, Ragin has suggested that many, or even most, theoretical statements made in social research can be formulated in terms of sets and set relations. He says

> that almost all social science theory is verbal and, as such, is formulated in terms of sets and set relations. When a theory states, for example, that "small farmers are risk averse," the claim is set-theoretic: small farmers constitute a rough subset of risk-averse individuals. (Ragin 2008: 13)

In contrast to this view, I argue that theoretical statements made in social research[1] often require modal notions referring to possibilities and probabilities and cannot, then, be formulated in terms of sets and set relations.[2]

In sections 1 and 2, I briefly bring to mind Ragin's set-theoretic approach and show that this approach can be reformulated in a conceptual framework for descriptive statistics that is based on the notion of statistical variables. In section 3, I distinguish two kinds of generalizations: descriptive generalizations which can be formulated as statements about sets (populations), and modal generalizations having the form of proba-

[1] There is no clear demarcation of "theoretical statements." I assume, as a minimal condition, that theoretical statements are different from statements describing observed facts.

[2] In this article, I only consider the modal notions 'possible', 'probable' and 'necessary/sufficient'; for a broader view see White (1975).

bilistic or deterministic rules which cannot be formulated as statements about sets; and I argue that social research is often interested in modal generalizations. In section 4, I criticize Ragin's suggestion that there is a basic contrast between his set-theoretic approach and statistical methods. I consider regression functions as a common framework, and argue that the distinction between descriptive and modal generalizations also applies to regression functions. I go on, in section 5, to show that the methods of Qualitative Comparative Analysis (QCA) are basically techniques for the construction of regression functions, broadly understood. In section 6, I argue that also statements about sufficient and/or necessary conditions involve modal notions and cannot be formulated in terms of sets and set relations, but require the supposition of deterministic rules. Finally, in section 7, I briefly consider the question of how to formulate hypotheses about causal conditions of observed facts, and contrast Ragin's notion of coverage with an approach in terms of probabilistic rules. The article ends with a brief conclusion.

1 Statements about Sets

How to understand statements about sets in social research? Here is one of Ragin's examples.

> When researchers argue, for example, that "religious fundamentalists are politically conservative," they are stating, in effect, that they believe that religious fundamentalists form a rough subset of the set of political conservatives, and may even go so far as to argue that their fundamentalism is the cause of their conservatism. (Ragin 2008: 14)

Ragin suggests to think of two sets, a set of religious fundamentalists, say F, and a set of political conservatives, say C, and then to understand the theoretical claim (leaving aside, for the moment, any causal connotations) as the statement that most elements of F are also elements of C.

In order to understand Ragin's suggestion, one needs to understand the sets. It is obviously easy to think of a set F, consisting of religious fundamentalists, and a set C, consisting of political conservatives, such

that no element of F is an element of C. Simply assume that F and C consist of different individuals.

In empirical research, this problem can be avoided by starting from a common reference set, say Ω, consisting of people actually observed. Then one can define F as the set of all members of Ω who are religious fundamentalists, and C as the set of all members of Ω who are political conservatives. It may then turn out that F is a subset of C, and stating this would be a statement about Ω, namely:

$$\text{For all } \omega \in \Omega\text{: if } \omega \in F\text{, then } \omega \in C \tag{1}$$

Of course, there could be exceptions such that F is no longer a subset of C. As Ragin (2008: 45) acknowledges, in empirical social research one most often finds such exceptions. One might then say that most members of F are also members of C. This would be again a statement about Ω. However, in contrast to (1), it cannot be formulated as a statement about the individual members of Ω. Instead, it must be formulated as a statistical statement, that is, a statement about (absolute or relative) frequencies; for example:[3]

$$\mathrm{P}(C|F) = \frac{\#\{\omega \in \Omega \,|\, \omega \in C \text{ and } \omega \in F\}}{\#\{\omega \in \Omega \,|\, \omega \in F\}} \geq 0.9 \tag{2}$$

As suggested by Ragin, $\mathrm{P}(C|F)$ can be interpreted as a measure showing the degree to which F is a subset of C; and having this interpretation in mind, he calls it a measure of consistency (Ragin 2008: 45). In statistical parlance, it is simply a conditional frequency.

The example suggests a general distinction between two kinds of statements about sets:

a) Statements, like (1), which can be formulated as statements about the individual members of a set. In this article, such statements will be called *quantified statements*.

b) Statements, like (2), which characterize a set by describing (aspects of) a frequency distribution defined for the set. Such statements will be

[3] I use $\#M$ to denote the number of elements of a finite set M.

called *statistical statements*. (In this article, the term 'statistical statement' will be restricted to statements about frequency distributions or quantities derived from such distributions.)

It is well possible, however, to set up a conceptual framework that includes both kinds of statements by using statistical variables which represent properties of elements of a set. I define a *statistical variable* as a function $X : \Omega \longrightarrow \mathcal{X}$. Ω is a set of objects (called the reference set of the statistical variable); $\mathcal{X}$ is a property space, i.e. a collection of attributes that can be used to characterize the elements of Ω;[4] and the statistical variable X assigns, to each object $\omega \in \Omega$, a value $X(\omega) \in \mathcal{X}$ that characterizes the object. Given a statistical variable $X : \Omega \longrightarrow \mathcal{X}$, its frequency distribution is a function, say P, that associates, with each property set $A \subseteq \mathcal{X}$, the proportion of members of Ω having a property value in A; formally defined: $\mathrm{P}(A) := \#\{\omega \in \Omega \,|\, X(\omega) \in A\} \,/\, \#\Omega$. Due to this definition, statistical variables provide a well-suited starting point for descriptive statistics concerned with the description of frequency distributions in empirically identifiable sets of cases.[5]

This framework also allows one to consider relationships between sets of cases in a general way. One can start, for example, from a two-dimensional statistical variable, say $(X, Y) : \Omega \longrightarrow \mathcal{X} \times \mathcal{Y}$. Then, for any property sets $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$, one can consider relationships between the corresponding object sets. Perhaps the relationship can be described by a quantified statement:

$$\text{For all } \omega \in \Omega\text{: if } X(\omega) \in A\text{, then } Y(\omega) \in B \tag{3}$$

Most often, however, such a statement will not be true and one has to consider the conditional distribution

$$\mathrm{P}(B|A) = \frac{\#\{\omega \in \Omega \,|\, X(\omega) \in A \text{ and } Y(\omega) \in B\}}{\#\{\omega \in \Omega \,|\, X(\omega) \in A\}} \tag{4}$$

[4] It will be assumed that the attributes have a numerical representation such that $\mathcal{X}$ can be considered as a set of real numbers or vectors. In particular, for binary variables, the property space will be assumed to be $\{0, 1\}$.

[5] For additional discussion of statistical variables as a conceptual framework for descriptive statistics see Rohwer (2010: chap. 1).

In fact, using statistical variables as a conceptual framework, there is no need to separately consider quantified statements. They are simply special cases of statistical statements. For example, the statement (3) is equivalent with $\mathrm{P}(B|A) = 1$.

I conclude that Ragin's set-theoretic approach, insofar it is based on ordinary ("crisp") sets, is formally identical with a statistical approach that starts from binary statistical variables.

2 Why Using Fuzzy Sets?

Ragin is aware of the limitations of a set-theoretic approach that is based on binary variables. He has therefore proposed to extend the approach by using fuzzy sets (Ragin 2000; 2008; 2009). However, since a generalization based on statistical variables is already available and in widespread use, one may well ask whether this strategy is worthwhile.

In order to discuss this question, I first stress that a "fuzzy set" is not another kind of set, or an extension or generalization of the ordinary notion of set. In fact, it is not a set at all, but a function having an ordinary set as its domain. Formally, a fuzzy set can be defined as a pair (Ω, g) where Ω is an ordinary set, and g is a function from Ω into the interval $[0, 1]$ of real numbers. The function g is then interpreted as providing, for each element $\omega \in \Omega$, a "grade of membership" value $g(\omega)$. However, since there is no analogy with "being a member of a set," the expression is metaphorical. All objects that get a membership value by g are regular elements of Ω in a set-theoretic understanding; and a further "set" that must be imagined in order to speak of "gradual membership," does not occur in the formal definition of a fuzzy set.[6]

I do not deny that metaphorical talk of "grades of membership" could be useful in some fields of application. However, it is not needed when fuzzy sets only serve as a tool for transforming statistical variables into sets. Ragin proposes to introduce fuzzy sets by starting from statistical variables and "calibrating" their property spaces (Ragin 2008: part II). Formally, starting from a statistical variable, say $X : \Omega \longrightarrow \mathcal{X}$, its calibration

[6] See, e.g., Smithson and Verkuilen (2006: 7, 19).

consists in constructing a function $\phi : \mathcal{X} \longrightarrow [0,1]$, that maps the variable's property space into the interval $[0,1]$ of real numbers. As a result one gets the fuzzy set (Ω, X') having a membership function defined by $X'(\omega) := \phi(X(\omega))$.

It is seen that this construction of a fuzzy set is basically the same as a transformation of the variable's property space. The resulting fuzzy set is formally equivalent with a (transformed) statistical variable.[7] Such variables will subsequently be called *f-calibrated variables*.

It follows that, in order to apply set-theoretic notions, fuzzy sets are not needed. Instead, one can directly start from, and work with, statistical variables. Moreover, statements about fuzzy sets derived from statistical variables can directly be formulated as statistical statements. To illustrate, I consider the subset relation for fuzzy sets. Assume two fuzzy sets, say (Ω, X') and (Ω, Y'). The first is a fuzzy subset of the second one, if the following statement holds: For all $\omega \in \Omega$: $X'(\omega) \leq Y'(\omega)$. This statement is equivalent with the statistical statement $\mathrm{P}(X' \leq Y') = 1$.

Like the ordinary subset relation, also this fuzzy subset relation will very often not hold. Starting from somehow calibrated statistical variables X' and Y', $\mathrm{P}(X' \leq Y')$ will most often be less than 1. In parallel to the interpretation of $\mathrm{P}(C|F)$ in section 1, Ragin has therefore proposed to think of $\mathrm{P}(X' \leq Y')$ as a measure of consistency, indicating the degree to which the fuzzy subset relation holds.[8] However, referring to fuzzy sets (and a metaphorical talk of a roughly holding fuzzy subset relation) is obviously not required to make sense of this statistical quantity.

Leaving aside the question whether the rhetoric of fuzzy sets might be useful, there also is a theoretically interesting difference. In the case

[7] This conforms to the following remark made by Ragin and Pennings (2005: 424): "[A] fuzzy set can be seen as a continuous variable that has been purposefully calibrated to indicate degree of membership in a defined set." The "defined set" is implicitly taken as the ordinary set of all cases having a membership score equal to 1.

[8] See Ragin (2008: 48). There is another definition that takes values of the membership functions into account (Ragin 2008: 52):

$$\frac{\Sigma_\omega \min(X'(\omega), Y'(\omega))}{\Sigma_\omega X'(\omega)} = \frac{\text{mean value of } \min(X', Y')}{\text{mean value of } X'}$$

Obviously, also this definition can be given a statistical formulation.

of binary variables, the measure of consistency is defined as a conditional frequency, and this allows one to interprete the measure conditionally on values of one (the independent) variable. This is not possible with measures of consistency for two f-calibrated variables (fuzzy sets), say X' and Y', because these measures concern the common distribution of X' and Y'. Of course, one could formally condition on values of X'. For example, referring to an individual case ω having the value $X'(\omega)$, one could consider the conditional frequency $\mathrm{P}(X' \leq Y'|X' = X'(\omega))$; but this quantity cannot be interpreted in analogy with the unconditionally defined consistency measure $\mathrm{P}(X' \leq Y')$. A theoretically important consequence is that these measures cannot be used to assess the causal relevance of individually attributable properties (values of X'). (I take up the argument in section 5.)

3 Factual Statements and Generalizations

Ragin's set-theoretic approach, like the more general approach that is based on statistical variables, can well be used as a conceptual framework for factual statements about sets of observed cases. For example, having observed 100 individuals of whom 20 are religious fundamentalists, and found that 16 of these religious fundamentalists are political conservatives, one can state that $\mathrm{P}(F) = 0.2$ and $\mathrm{P}(C|F) = 0.8$. These are factual statements about a set Ω consisting of 100 observed individuals.

Further questions concern generalizations. Two kinds must be distinguished. One kind starts from the idea that the set of observed cases, say Ω, can be considered as a subset of a larger set, say Ω^*, often called a population in this context. Generalization then consists in making a statement about Ω^* that has the same linguistic form as the factual statement about Ω. Statistical variables can help to clarify this approach. One starts from a statistical variable $X : \Omega \longrightarrow \mathcal{X}$, defined for the observed set Ω, and assumes a correspondingly defined statistical variable $X^* : \Omega^* \longrightarrow \mathcal{X}$, defined for the population Ω^*. The factual statement about the observed distribution of X is then generalized to a corresponding statement about the unobserved distribution of X^*.

In the example, Ω consists of 100 observed individuals, and the sta-

tistical variable is (X, Y); X records whether an individual is a religious fundamentalist ($X = 1$) or not ($X = 0$), and Y records whether it is a political conservative ($Y = 1$) or not ($Y = 0$). The observations can be used to make a factual statement about the distribution of (X, Y), e.g. $\mathrm{P}(Y = 1 \,|\, X = 1) = 0.8$. Assuming then that Ω is a sample from a population Ω^*, being the reference set of a correspondingly defined variable (X^*, Y^*), the generalization consists in making an analogous statement about the distribution of (X^*, Y^*), e.g. $\mathrm{P}(Y^* = 1 \,|\, X^* = 1) \approx 0.8$.

This approach will be called *descriptive generalization* because the generalization can be formulated as a descriptive (quantified or statistical) statement about a population of cases. The approach requires that the population, Ω^*, for which the generalization is intended, is a set of empirically identifiable cases, implying that they actually exist or have existed in the past. Moreover, in order to think of Ω as a random sample from the population Ω^*, this set can only consist of cases that actually exist while the sample is drawn.

A quite different kind of generalization has the linguistic form of rules. To illustrate, I refer to a random generator, for example, throwing a die. By activating the random generator (i.e. throwing the die in a specified way), one can generate an event that can be described by a number in the set $\mathcal{Z} = \{1, \ldots, 6\}$. Doing this n times, the result can be represented by a statistical variable, say $Z : \Omega \longrightarrow \mathcal{Z}$. The reference set Ω consists of identifiers of the n events, say $\omega_1, \ldots, \omega_n$, and for each event ω_i, $Z(\omega_i)$ describes the outcome.

It follows that describing the distribution of Z makes a factual statement about Ω, the set of realized events. However, this statement does not describe the random generator as a method of generating events. In order to describe this method, another kind of statement is required. In this example, assuming that the die is not biased, it can be formulated as:

> If the random generator is activated, there are six possible outcomes $(1, \ldots, 6)$, all having the same probability. (5)

This statement does not describe a set of events generated with the random generator. Moreover, there is no sense in which it describes a set, however defined. In particular, there is no set of "possible events" that might be

generated with the die. Of course, one can refer to the elements of $\mathcal{Z}$, but these are event types, not "possible events;" and what is more, (5) is not a statement about this set. Instead, (5) is a modal statement, meaning here that it conditionally refers to possibilities, and quantifies the possibilities with probabilities.

Probabilistic rules like (5) will therefore be called *modal generalizations*. In general, such rules say what might be the case, or might happen, given specified conditions, adding some evaluation of the probabilities of the possible outcomes. Continuing with Ragin's example, a probabilistic rule could be:

> If somebody is a religious fundamentalist, it is highly probable that he or she is a political conservative. (6)

This is not a statement about any set of individuals, neither a set of actually observed nor a set of hypothetically imagined individuals. It is a probabilistic rule, and in contrast to a descriptive generalization, it cannot be formulated as a quantified or statistical statement.[9]

I suppose that theoretical interests in social research often aim at modal generalizations. In fact, Ragin himself has stressed the goal of finding "regularities."[10] Given this interest, one needs a conceptual framework for the formulation of modal generalizations that is in important respects different from the conceptual framework of the set-theoretic approach.

4 Regression Functions and Rules

In sections 1 and 2, I showed that Ragin's set-theoretic approach can be reformulated in the framework of statistical variables. Quite differently from this understanding, Ragin has suggested to think of a fundamental contrast between his set-theoretic approach and statistical methods. Presumably, this is due to his belief that "the correlation coefficient" is "the

[9] This will be true even if the formulation refers to a population of cases, e.g. 'It is highly probable that a person, randomly drawn from the set of religious fundamentalists who currently live in Germany, is a political conservative.' This is not a statistical statement about the set of religious fundamentalists who currently live in Germany, but describes a random generator using the specified set in its definition.

[10] See Ragin (2006: 309). See also Rihoux (2006: 682), Berg-Schlosser et al. (2009: 11).

cornerstone of conventional quantitative research" (Ragin 2008: 6; see also Ragin 2000: 45-46). He then says:

> A key contrast is the difference between the correlation (and most other measures of association), which is symmetrical by design, and the set relation, which is fundamentally asymmetrical. This distinction is important because set-theoretic analysis, like qualitative research more generally, focuses on uniformities and near uniformities, not on general patterns of association. (Ragin 2008: 7)

I doubt that the correlation coefficient is "the cornerstone of conventional quantitative research." If there is any statistical method predominating in social research, it is the construction of regression functions based on a distinction between independent and dependent variables. The interest focuses on dependency relations between variables which are asymmetrical by definition. Insofar this interest also motivates Ragin's set-theoretic approach, there is no essential difference. In fact, viewed as a technique, QCA (Qualitative Comparative Analysis) is a version of regression analysis.

Before showing this in the next section, I briefly discuss how to understand regression functions. In parallel to the distinctions discussed in the previous section, one needs to distinguish between descriptive and modal formulations of regression functions.

In order to discuss the difference, I begin with a regression function derived from a statistical variable, say $(X, Y) : \Omega \longrightarrow \mathcal{X} \times \mathcal{Y}$. Defining Y as the dependent and X as the independent variable, a regression function assigns to each value $x \in \mathcal{X}$ the conditional frequency distribution of Y given x; formally: $x \longrightarrow \mathrm{P}[Y|X = x]$. Note that the right-hand side is itself a function ($y \longrightarrow \mathrm{P}(Y = y|X = x)$). However, if Y is a binary variable it suffices to consider the function

$$x \longrightarrow \mathrm{P}(Y = 1 \,|\, X = x) \tag{7}$$

having the domain $\mathcal{X}$, the property space of X. To each value x in $\mathcal{X}$, this function assigns the conditional frequency $\mathrm{P}(Y = 1 \,|\, X = x)$.

For example, if (X, Y) is the variable introduced in the previous section, values of the regression function might be given by $\mathrm{P}(Y = 1 \,|\, X = 1) = 0.8$ and $\mathrm{P}(Y = 1 \,|\, X = 0) = 0.4$, respectively. In general, a regression function shows how the frequency distribution of a dependent variable (or a quantity derived from this distribution, e.g. its mean) depends on values of the independent variable (which, of course, can consist of several components).

A regression function derived from statistical variables makes a descriptive statement about the variable's reference set. If this reference set, say Ω, is a set of observed cases, it will also be a factual statement, and one can think about descriptive and modal generalizations. A descriptive generalization assumes that Ω is a (representative) sample from a population Ω^* for which analogously defined statistical variables can be assumed, say $(X^*, Y^*) : \Omega \longrightarrow \mathcal{X} \times \mathcal{Y}$. The regression function for the population is

$$x \longrightarrow \mathrm{P}(Y^* = 1 \,|\, X^* = x) \tag{8}$$

Completely parallel to the formulation (7), it makes a descriptive statement about the population Ω^*.

A quite different generalization takes the form of a stochastic regression function. Its domain is, again, the property space of the independent variable, say $\mathcal{X}$. However, the regression function no longer refers to statistical variables defined for some reference set, and therefore cannot be formulated in terms of conditional frequency distributions. Instead, a stochastic regression function assigns to each value $x \in \mathcal{X}$ a conditional probability distribution for the dependent variable. If the dependent variable is binary, the regression function can be written analogously to (7)

$$x \longrightarrow \mathrm{Pr}(\dot{Y} = 1 \,|\, \ddot{X} = x) \tag{9}$$

where now Pr is used instead of P to indicate a probability distribution. A further consideration concerns the variables used in the notation of the stochastic regression function. They cannot be understood as statistical variables representing (observed or hypothetically assumed) facts but must be given a modal interpretation, and I therefore use a different notation. This is obvious for the independent variable, denoted $\ddot{X}$. Since no reference is made to any specific sample or population, there is no frequency distribution for this variable. Instead, its values can be fixed arbitrarily by the person who intends to use the regression function as a rule for making

inferences. Consequently, there is also no probability distribution for this variable.

Also the dependent variable, denoted $\dot{Y}$, does not represent values that are realized in some reference set of actually existing cases, but must be given a modal interpretation. The regression function formulates a statement about possible values of this variable and quantifies the possibilities in terms of probabilities. This is done conditionally on values of the independent variable (which can be assigned arbitrarily). Consequently, there is no (unconditional) probability distribution for the dependent variable. Nevertheless, in contrast to the independent variable, $\dot{Y}$ is a stochastic variable allowing statements about conditional probability distributions.

In contrast to statistical variables, the variables $\ddot{X}$ and $\dot{Y}$ do not relate to a set of actually existing cases, but serve to formulate theoretical relationships using modal notions. I find it helpful to think of these variables as referring to a generic object or situation. For example, one might say: Consider (imagine, think of) a person having the value $\ddot{X} = 1$, i.e. who is a religious fundamentalist. This statement does not refer to a person that can be empirically identified but to a generic individual that is partially defined by the property 'being a religious fundamentalist' (any number of further properties could be added). Then one might further ask whether this person also is a political conservative, or in terms of variables, whether a variable $\dot{Y}$, that can be defined for the generic individual, has the value 1. Obviously, since the person referred to does not exist, the question cannot be answered by a factual statement, but requires modal considerations.

This understanding of the variables is in accordance with interpreting a stochastic regression function as a probabilistic rule that can be used to make conditional inferences. Moreover, the formal framework of regression functions allows one to think of the conditional probabilities as quantities which can be numerically specified. This is not required by the general notion of a probabilistic rule. As illustrated by the example formulated in (6), vague qualifications often suffice. However, if one is interested in more precise estimates, the framework of stochastic regression functions allows one to use observed conditional frequencies as estimates of the conditional probabilities.

5 Regression with Binary and F-Calibrated Variables

Based on the understanding of regression functions discussed in the previous section, one can easily see that the methods proposed by Ragin under the heading of Qualitative Comparative Analysis (QCA) are techniques for the construction of regression functions. Assume that data are given by a statistical variable consisting of $m+1$ binary components:

$$(X_1, \ldots, X_m, Y) : \Omega \longrightarrow \{0,1\}^{m+1} \tag{10}$$

where Y represents an outcome, and one has decided to consider $X_1, \ldots, X_m$ as possibly relevant conditions for $Y = 1$. Considered as a technique for data analysis, QCA consists in constructing a regression function

$$(x_1, \ldots, x_m) \longrightarrow \mathrm{P}(Y{=}1 \,|\, X_1{=}x_1, \ldots, X_m{=}x_m) \tag{11}$$

The function provides, for each combination of values of the independent variables, a conditional frequency distribution of the dependent variable. It may turn out that these conditional frequencies always equal 1; one then gets as a special case a Boolean function. However, most often at least some of the frequencies will be less than 1, and a proper regression function is required.

Now assume that the dependent variable is f-calibrated, say Y' (having values that could be interpreted as scores of membership in a fuzzy set), and that there is a further f-calibrated independent variable, say Z'. Analogously to (11), one could start from the general regression function

$$(x_1, \ldots, x_m, z) \longrightarrow \mathrm{P}[Y' \,|\, X_1{=}x_1, \ldots, X_m{=}x_m, Z' = z] \tag{12}$$

and consider, for example, characterizations of the right-hand side by conditional mean values of Y', given values of the independent variables. However, being mainly interested in degrees to which subset relations hold (measures of consistency), Ragin's QCA with fuzzy sets takes a different route. While the right-hand side of (11) already is a measure of consistency, there is no direct analogy with f-calibrated variables. Definitions of measures of consistency for f-calibrated variables require a reference to their common distribution. For example, with Z' and Y', one could use $\mathrm{P}(Z' \leq Y')$, a quantity derived from the common distribution of both

variables (see section 2). Regression functions for QCA with f-calibrated variables therefore have the general form

$$(x_1, \ldots, x_m) \longrightarrow \mathrm{con}(Z', Y' \mid X_1 = x_1, \ldots, X_m = x_m) \tag{13}$$

where the right-hand side denotes a measure of consistency derived from the common distribution of Z' and Y', conditional on given values of the independent variables.

Both, (11) and (13), are regression functions. There is, however, a remarkable difference (already mentioned at the end of section 2). The function (11) allows one to consider each combination of its arguments, $(x_1, \ldots, x_m)$, as a specific configuration of conditions for the dependent variable, attributable to individual cases. In the function (13), the same is true for values of $X_1, \ldots, X_m$, but not for values of the f-calibrated variable Z'. Of course, Z' can consist of two or more components, and one can consider several measures of consistency based on the different possibilities to combine the components into a single fuzzy set.[11] Each combination of components could then be viewed as a "configuration of conditions" for the dependent variable.[12] However, when used for measures of consistency, these configurations are no longer attributable to individual cases, and it becomes difficult to see how they could be given a causal interpretation.

Notwithstanding this difference that results from Ragin's focus on set relations, one can see that QCA basically consists of techniques for the construction of regression functions. As I mentioned at the beginning, these techniques could be quite useful for the analysis of data. However, formulated in the set-theoretic framework, the resulting regression functions only make descriptive statements about the cases contained in the reference sets of their variables. Theoretical considerations, aiming at prospective predictions or retrospective explanations, cannot be formulated in this descriptive framework, but require an understanding of regression functions in terms of modal variables which allow one to interpret these functions as probabilistic or deterministic rules.

[11] For example, if Z' has two components, $Z' = (Z'_1, Z'_2)$, one can consider the variables $\min\{Z'_1, Z'_2\}$, $\min\{Z'_1, (1 - Z'_2)\}$, $\min\{(1 - Z'_1), Z'_2\}$ and $\min\{(1 - Z'_1), (1 - Z'_2)\}$ derived from Z' by applying standard fuzzy set operations.

[12] See Ragin (2000: 234-239; 2009: 100).

6 Probabilistic and Deterministic Rules

In section 1 it was shown how quantified statements about finite sets of cases can be considered as special cases of statistical statements. Similarly, one might say that deterministic rules can be viewed as special cases of probabilistic rules. Consider, for example, the deterministic rule:

If the bell-push is pressed, the bell will ring. (14)

This can also be formulated as a probabilistic rule where the conditional probability has the value 1. Formally, using binary modal variables $\ddot{X}$ (for 'the bell-push is pressed') and $\dot{Y}$ (for 'the bell is ringing'),[13]

$$\Pr(\dot{Y} = 1 \mid \ddot{X} = 1) = 1 \tag{15}$$

This is a deterministic rule, and it is formulated as a limiting case of a probabilistic rule.

There is, however, an important difference between deterministic rules on the one hand and probabilistic rules (which are now understood as employing conditional probabilities less than one) on the other. Deterministic rules allow one to speak of sufficient and necessary conditions. For example, (15) allows one to say that $\ddot{X} = 1$ is a sufficient condition for $\dot{Y} = 1$; correspondingly, the deterministic rule $\Pr(\dot{Y} = 0 \mid \ddot{X} = 0) = 1$ would allow one to say that $\ddot{X} = 1$ is a necessary condition for $\dot{Y} = 1$.

These notions cannot be used with probabilistic rules. To continue with the example, imagine that we observed that the bell didn't always ring when the bell-push was pressed, and assume that the observations suggest, instead of (15), a probabilistic rule

$$\Pr(\dot{Y} = 1 \mid \ddot{X} = 1) \approx 0.95 \tag{16}$$

Obviously, this rule does not allow one to speak of sufficient conditions. In fact, it explicitly says that $\ddot{X} = 1$ is not a sufficient condition for $\dot{Y} = 1$.

In the present context, the distinction is important because Ragin (like many other researchers in the field of comparative case studies) thinks of

[13] Since we only consider $\ddot{X} = 1$, there is no need for explicitly writing (15) as a regression function.

causation in terms of sufficient and/or necessary conditions.[14] Moreover, he suggests that the set-theoretic approach (QCA) can be used to formulate and investigate hypotheses about sufficient and/or necessary conditions. He says, for example:

> The fact that democratic dyads constitute a perfect or near-perfect subset of nonwarring dyads signals that this arrangement (international relations between democracies) may be sufficient for peaceful coexistence. (Ragin 2006: 292)

And in a more general formulation:

> [I]f cases sharing several causally relevant conditions uniformly exhibit the same outcome, then these cases constitute a subset of instances of the outcome. Such a subset relation signals that a specific combination of causally relevant conditions may be interpreted as *sufficient* for the outcome. (Ragin 2009: 99)

He then adds the remark:

> The interpretation of *sufficiency*, of course, must be grounded in the researcher's substantive and theoretical knowledge; it does not follow automatically from the demonstration of the subset relation. (Ragin 2009: 99)

This remark is certainly true, but circumvents the question how to make the theoretical idea that is used in the interpretation of the observations explicit. If the theoretical idea is that a complex of conditions is sufficient for an outcome, this requires to presuppose a deterministic rule that allows one to explicate the idea;[15] and this rule cannot be formulated in terms of sets and set relations. This framework would allow one to make statements about subsets of a reference set, for example, that all cases exhibiting some complex of conditions also exhibit a specified outcome.[16] But this would

[14] See, e.g., Ragin (1987: 99; 2000: 104; 2008: 20); Rihoux and Ragin (2009: xix, 10). See also Mahoney (2003), Mahoney and Goertz (2006: 232).

[15] To say that A is sufficient for B *means* that, given A, one can infer B; and this requires a deterministic rule to be used for the inference.

[16] It should be clear that reference must be made to a set of cases. It wouldn't make sense to think of sets of causes that could be subsets of sets of effects; see Goertz (2003: 59).

be a descriptive (factual) statement about a set of (observed) cases, not a deterministic rule.

This argument concerns the formulation of theoretical claims. A different question is whether deterministic rules, which are required for thinking of sufficient and/or necessary conditions, are useful in social research. This question will not be discussed in the present article.

7 Causal Conditions of Observed Facts

Social researchers are often interested in finding causes of facts (states, events), observed in a single case or in some set of cases. There are two (sometimes complementary) approaches. One approach investigates the processes that generated the observed facts in the individual cases. This is sometimes called "process tracing."[17] Another approach that will be considered in the present section proceeds in terms of variables. There is a set of cases, say Ω, and for each case $\omega \in \Omega$, one has observed the value $Y(\omega)$ of an outcome variable. To simplify the discussion, I assume that Y is a binary variable. The goal then is to find another variable, X, often consisting of several components, say $X = (X_1, \ldots, X_m)$, such that the value $X(\omega)$ can be considered as a cause, meaning here a condition, or complex of conditions, contributing to the occurrence of $Y(\omega)$.

Given this set-up in terms of variables, I argue that the set-theoretic framework will not suffice, even if the theoretical interest only concerns the observed cases. In order to develop the argument, I start from Ragin's notion of "coverage." (Ragin 2008: 54-57) Assume that theoretical ideas suggest to consider values of X as conditions which possibly contribute to the occurrence of $Y = 1$. Then, for any value x of X, its coverage is defined by

$$\mathrm{cov}(X{=}x) := \mathrm{P}(X{=}x \,|\, Y{=}1) \tag{17}$$

In Ragin's interpretation, $\mathrm{cov}(X = x)$ quantifies the proportion of cases in which the condition $X = x$ caused (contributed to the occurrence of) $Y = 1$.

[17] See George and Bennett (2004).

It is remarkable that 'coverage' is a descriptive notion that can be defined in a set-theoretic framework (with statistical variables), but not with deterministic or probabilistic rules. To illustrate, I consider an example where a bell's ringing ($Y = 1$) can be accomplished by one of two bell-pushes, represented by $X = (X_1, X_2)$. I assume that one has observed 100 situations:

X_1	X_2	cases with $Y = 1$
0	0	0
1	0	90
0	1	8
1	1	2

(18)

It follows that $\mathrm{cov}(X_1 = 1) = 0.92$, meaning that in 92 % of the observed cases pressing the first bell-push caused (contributed to) the bell's ringing; and $\mathrm{cov}(X_2 = 1) = 0.10$, meaning that in 10 % of the cases pressing the second bell-push caused (contributed to) the bell's ringing.

Now assume that we want to construct a model representing a generic situation where one of two bell-pushes could be used for ringing a bell. This could be a deterministic or a probabilistic model. I first assume a deterministic rule

$$\Pr(\dot{Y}=1 \,|\, \ddot{X}_1=x_1, \ddot{X}_2=x_1) = \begin{cases} 1 & \text{if } x_1 = 1 \text{ or } x_2 = 1 \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

Obviously, the model does not allow one to define something like coverage. This would require that one could refer to a distribution of values of the variables $\ddot{X}_1$ and $\ddot{X}_2$. But these are exogenous variables of the model which do not have a distribution. If one only knows the rule (19), one cannot even make some informed guess about how often the first or the second bell-push might have been used. This conclusion holds independently of whether the model is deterministic or stochastic.

In order to make predictions about causes of observed events, one would need another rule. If there could be different causes, as it is the case in the example, this must be a probabilistic rule where the variables have changed their role. Instead of $\dot{Y}$, one has to define an exogenous variable $\ddot{Y}$ that can be used to *assume* that the bell did ring ($\ddot{Y} = 1$). And instead

of the formerly exogenous variables one has to define stochastic variables, $\dot{X}_1$ and $\dot{X}_2$, that allow one to consider conditional probabilities. Then one can think of a probabilistic rule

$$(x_1, x_2) \longrightarrow \Pr(\dot{X}_1=x_1, \dot{X}_2=x_2 \,|\, \ddot{Y}=1) \tag{20}$$

that could be used for probabilistic statements about possible causes.

In a sense, there is a correspondence between the conditional frequencies, called 'coverage' by Ragin, and the conditional probabilities on the right-hand side of (20). The observed conditional frequencies could be used as estimates of the theoretically supposed conditional probabilities. There is, however, not only the conceptual difference between facts (observed frequencies) and modalities (theoretically assumed probabilities). A probabilistic rule, like (20), also presupposes a theoretical model and, in particular, that all possible causes have a representation in the model. In contrast, statements about coverage, as defined in the set-theoretic framework, do not require this assumption; they simply describe what has been observed.

A further point should be mentioned. Ragin has suggested that the coverage of a (causally relevant) condition can be interpreted sensibly only if the condition has a high consistency (Ragin 2008: 55). For example, following this suggestion, the interpretation of $\mathrm{cov}(X=x)$, defined in (17), would require that $\mathrm{P}(Y=1 \,|\, X=x) \geq \alpha$, where α is some number near to 1. However, one can easily find examples where the suggestion is misleading. Think, for example, of throwing a die; $X = 1$ records the event that the die is thrown, Y records the outcome. For each possible outcome $y \in \{1, \ldots, 6\}$, the consistency, $\mathrm{P}(Y = y \,|\, X = 1)$, is low; but the coverage has a high value, $\mathrm{P}(X = 1 \,|\, Y = y) = 1$, and has a clear interpretation: In all observed cases, the die was thrown before the outcome showed up.

8 Conclusion

The article discussed the suggestion, made by Charles Ragin, that many, or even most, theoretical statements in social research can be formulated in a set-theoretic framework. In order to discuss the suggestion, it was shown that Ragin's set-theoretic approach, including his usage of fuzzy-

set notions, can be reformulated in the conceptual framework of statistical variables. The article then developed the argument that this framework can only be used for descriptive statements (including descriptive generalizations). And it was further argued that social research is often interested in modal generalizations, consisting of probabilistic and deterministic rules, which cannot be formulated in the set-theoretic framework. In particular, this framework does not allow one to formulate hypotheses about sufficient and/or necessary conditions, since these notions presuppose deterministic rules.

The article also criticized Ragin's opposition between his set-theoretic approach and conventional statistical methods. Based on the reformulation of this approach in terms of statistical variables, it was shown that the methods of Qualitative Comparative Analysis (QCA) basically are techniques for the construction of regression functions. Finally, the article briefly considered the explanation of actually observed facts in terms of variables, and showed that Ragin's concept of 'coverage' cannot be used for these explanations.

References

Berg-Schlosser, D., De Meur, G., Rihoux, B., Ragin, C. C. (2009). Qualitative Comparative Analysis (QCA) as an Approach. In: B. Rihoux, C. C. Ragin (eds.), *Configurational Comparative Methods,* 1–18. Los Angeles: Sage.

George, A. L., Bennett, A. (2004). *Case Studies and Theory Development in the Social Sciences.* Cambridge: MIT Press.

Goertz, G. (2003). Cause, Correlation, and Necessary Conditions. In: G. Goertz, H. Starr (eds.), *Necessary Conditions,* 47–64. Lanham: Rowman & Littlefield.

Katz, A., vom Hau, M., Mahoney, J. (2005). Explaining the Great Reversal in Spanish America. Fuzzy-Set Analysis versus Regression Analysis. *Sociological Methods & Research,* 33, 539–573.

Mahoney, J. (2003). Strategies of Causal Assessment in Comparative Historical Analysis. In: J. Mahoney, D. Rueschemeyer (eds.), *Comparative Historical Analysis in the Social Sciences,* 337–372. Cambridge: Cambridge University Press.

Mahoney, J. (2008). Toward a Unified Theory of Causality. *Comparative Political Studies,* 41, 412–436.

Mahoney, J., Goertz, G. (2006). A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research. *Political Analysis,* 14, 227–249.

Ragin, C. C. (1987). *The Comparative Method. Moving Beyond Qualitative and Quantitative Strategies.* Berkeley: University of California Press.

Ragin, C. C. (2000). *Fuzzy-Set Social Science.* Chicago: University of Chicago Press.

Ragin, C. C. (2006). Set Relations in Social Research: Evaluating Their Consistency and Coverage. *Political Analysis,* 14, 291–310.

Ragin, C. C. (2008). *Redesigning Social Inquiry.* Chicago: University of Chicago Press.

Ragin, C. C. (2009). Qualitative Comparative Analysis Using Fuzzy Sets (fsQCA). In: B. Rihoux, C. C. Ragin (eds.), *Configurational Comparative Methods,* 87–121. Los Angeles: Sage.

Ragin, C. C., Pennings, P. (2005). Fuzzy Sets and Social Research. *Sociological Methods & Research,* 33, 423–430.

Rihoux, B. (2006). Qualitative Comparative Analysis (QCA) and Related Systematic Comparative Methods. *International Sociology,* 21, 679–706.

Rihoux, B., Ragin, C. C. (eds.) (2009). *Configurational Comparative Methods.*

Los Angeles: Sage.

Rohwer, G. (2010) *Models in Statistical Social Research.* London: Routledge.

Smithson, M., Verkuilen, J. (2006). *Fuzzy Set Theory. Applications in the Social Sciences.* Thousand Oaks: Sage.

White, A. R. (1975). *Modal Thinking.* Oxford: Basil Blackwell.