

*Configurational Explanations*

Alessia Damonte

Accepted for publication with ISBN 978-88-255-1888-7  
Please Note. The printed version may be different

*intentionally blank page*

## Acknowledgments

The book has taken its shape over years and experiences, during which I have built up a remarkable debt of gratitude. I owe much to my students: thank you for all your challenges. Thanks to the guests of the doctoral Summer School ReSPoS and the MeMa seminars in Milan for the many fruitful discussions on causation. Thanks to the participants to the Chess seminars in Durham, the APSA section on Mixed Methods, the PSA Method Conference, and the ecpr joint session of workshops for their comments on previous attempts to get to the point. Thanks to the Compasss community and, among them, to Airo Hino and Adrian Duşa for their kind support in the inevitable moments of bewilderment. Special thanks to Carsten Schneider, Benoit Rihoux, and Charles Ragin for their generous teaching that, long ago, initiated this all.

I gratefully acknowledge the financial support of the European Grant ERC-ADG-2015 694632.

The usual disclaimer applies.

AD  
Milan, October 2019

*intentionally blank page*

# Contents

<b>Introduction</b> .....	<b>15</b>
<b>Part I. THE BROADER PICTURE</b> .....	<b>23</b>
<b>1. Explanations as deductions</b> .....	<b>25</b>
1.1 The shape of a statement .....	26
1.2 Relationships and rules.....	27
1.3 The ideal shape of a valid deduction .....	32
1.4 The “because” that make syllogisms causal.....	34
1.4.1 The four causes.....	35
1.5 Limits of categorical deductions.....	38
<b>2. Inference as propositional calculus</b> .....	<b>40</b>
2.1 Stoic sentences .....	41
2.2 The truth of simple sentences .....	42
2.3 The truth of compound sentences.....	44
2.3.1 Conjunctions.....	44
2.3.2 Disjunctions .....	45
2.3.3 Conditional .....	46
2.3.4 Formal equivalences .....	49
2.4 Schemata .....	53
2.4.1 The first indemonstrable .....	54
2.4.2 The second indemonstrable.....	55
2.4.3 Further schemata.....	56
2.5 Themata .....	58
2.6 Strengths and limits of schemata.....	59
2.6.1 Disconnected arguments .....	60
2.6.2 Redundant arguments.....	60
2.6.3 Fallacies.....	61
2.6.4 Incomplete arguments .....	63
<b>3. Modern explanation</b> .....	<b>64</b>
3.1 Sitting the major premisses on the empirical ground.....	65
3.2 Unraveling complex causation .....	69
3.2.1 Mill’s problem of induction, and his solutions .....	70
a) <i>Space and time of causation</i> .....	71

b)	<i>Causes versus conditions</i> .....	73
c)	<i>Mechanical and chemical causation</i> .....	74
3.2.2	Canons and Methods of induction .....	75
a)	<i>The first canon</i> .....	75
b)	<i>The second canon</i> .....	77
c)	<i>The third canon</i> .....	79
d)	<i>The fourth and fifth canon</i> .....	80
3.2.3	The many issues of complex causation .....	81
a)	<i>Issues of Plurality</i> .....	82
b)	<i>Issues of Composition and Constitution</i> .....	83
c)	<i>the Deductive Method</i> .....	84
3.2.4	Modes of explanation.....	87
a)	<i>First mode: combination</i> .....	87
b)	<i>Second mode: intermediation</i> .....	88
c)	<i>Third mode: subsumption</i> .....	89
3.3	Explanation between induction and deduction .....	90
3.3.1	Deductive models of explanation .....	90
3.3.2	Capturing potentials .....	94
3.3.3	The problem of confirmation in explanations .....	97
3.3.4	Operational solutions .....	100
<b>4.</b>	<b>Back to square one? .....</b>	<b>104</b>
4.1	Relevance, causality, and causation.....	105
4.2	Mechanistic models of causation in philosophy .....	109
4.3	Explanatory mechanisms and their requirements .....	112
a)	<i>Desiderata</i> .....	114
b)	<i>Model specification and how-answers</i> .....	115
c)	<i>Salmon-complete constitutive mechanisms</i> .....	116
<b>Part II.</b>	<b>THE STRATEGY AND THE TECHNIQUE.....</b>	<b>119</b>
<b>5.</b>	<b>Configurations and <i>inus</i> causation.....</b>	<b>121</b>
5.1	<i>inus</i> causation .....	121
5.1.1	Necessity, sufficiency, and vending machines.....	122
5.1.2	Full explanations.....	124
a)	<i>Checking for plurality</i> .....	124
b)	<i>Checking for Constitution and Composition</i> .....	125
5.1.3	Principles of <i>inus</i> composition .....	126
<b>6.</b>	<b>Imagining machines .....</b>	<b>128</b>
6.1	The relevance of stage setting .....	128
6.1.1	Salmon-completeness of nomological machines .....	129
6.2	Cultivating mechanistic imagination .....	130
6.2.1	Learning from correlational literature .....	131
6.2.2	Learning from case-oriented literature .....	133

6.3	Summing up.....	135
<b>7.</b>	<b>A language to render <i>inus</i> causation .....</b>	<b>137</b>
7.1	Presence and absence of a property .....	137
	a) <i>The truth table</i> .....	140
7.2	Compounds .....	142
7.3	Substitutes .....	144
7.4	Necessity and sufficiency.....	145
	7.4.1 Necessity .....	145
	7.4.2 Sufficiency.....	146
7.5	Putting it all together for the analysis .....	147
	7.5.1 Checking directional expectations.....	149
	7.5.2 Building the truth table .....	150
	7.5.3 Assessing the sufficiency of the hypothesis.....	151
	7.5.4 Minimizing the truth table.....	152
	a) <i>Hard and easy counterfactuals</i> .....	153
7.6	Wrapping up.....	154
<b>8.</b>	<b>Modeling membership scores .....</b>	<b>156</b>
8.1	Giving variation a meaning .....	157
8.2	Ragin's improvements to fuzzy scores, and some concerns .....	159
	a) Measurement error .....	161
	8.2.1 The relationship between fuzzy and crisp scores .....	163
	8.2.2 Necessity and sufficiency with fuzzy scores .....	164
8.3	Designing membership scores .....	165
	8.3.1 Establish the universe of reference.....	166
	8.3.2 Operationalize attributes' intension.....	167
	8.3.3 Identify the thresholds for calibration .....	168
<b>9.</b>	<b>The protocol.....</b>	<b>171</b>
<b>Part III . THE EXAMPLE .....</b>		<b>177</b>
<b>10.</b>	<b>What causes corruption, and what can hinder it .....</b>	<b>179</b>
10.1	Define: how do we recognize corruption?.....	180
10.2	Learn: how does corruption come about?.....	182
	10.2.1 Cogs and wheels .....	183
10.3	Formalize: hindering corruption .....	186
	10.3.1 Locating levers and pulleys.....	187
	10.3.2 Setting theoretical expectations .....	190
	10.3.3 Gauging levers and pulleys .....	190
<b>11.</b>	<b>Set-theoretic configurational analysis.....</b>	<b>195</b>
11.1	Calibrate the membership scores .....	195
	a) <i>From raw to fuzzy: calibrating the outcome</i> .....	196

11.2	Assess the claim of individual necessity.....	200
11.3	Analysis of sufficiency, first: assess the hypothesis .....	204
11.4	Analysis of sufficiency, second: Minimizing solutions .....	209
11.5	Plot solutions.....	211
11.6	Return to the configurational hypothesis .....	215
<b>12.</b>	<b>Bottom lines .....</b>	<b>218</b>
	References .....	223



# Figures and Tables

<i>Figure 1.</i>	The square of opposition.....	29
<i>Figure 2.</i>	Venn's rendering of the basic Aristotelian relations.....	30
<i>Figure 3.</i>	The <i>barbara</i> syllogism as transitivity .....	33
<i>Figure 4.</i>	Venn's representation of complex sentences.....	52
<i>Figure 5.</i>	Types of objects by the antecedent and the consequent of a conditional statement.....	98
<i>Figure 6.</i>	Ranking of cases by their raw scores on <CLEAN> .....	197
<i>Figure 7.</i>	Cases' distances on the raw measure <CLEAN> .....	198
<i>Figure 8.</i>	Cases' clusters on the raw measure <CLEAN>.....	198
<i>Figure 9.</i>	CLEAN: raw by fuzzy scores .....	199
<i>Figure 10.</i>	XYplot of RTA and the outcome under different fuzzy transformations	202
<i>Figure 11.</i>	XY-plots from .Z calibration .....	212
<i>Figure 12.</i>	XY-plots from .I calibration .....	213
<i>Figure 13.</i>	XY-plots from .M calibration .....	214
Table 1.	Valid moods of the first figure and related set structures .....	33
Table 2.	Quantity of a hypothesis and conclusive observations .....	102
Table 3.	The truth table for the connective $\bar{\phantom{A}}$ .....	141
Table 4.	The truth table for the conjunction $A \wedge B \wedge \bar{C}$ .....	142
Table 5.	The truth table for the disjunction $A \vee B \vee \bar{C}$ .....	144
Table 6.	The truth table for the implication $A \leftarrow Y$ .....	146
Table 7.	The truth table for the conditional $W \rightarrow Y$ .....	147
Table 8.	Positions that any <i>i</i> th entity $u_i$ can take to any set $A$ and the corresponding membership values $\mu_A$ .....	159
Table 9.	Directional expectations.....	193
Table 10.	Raw data .....	194
Table 11.	Raw variable statistics.....	199
Table 12.	Thresholds applied for calibration.....	199
Table 13.	Fuzzy membership scores .....	200
Table 14.	Analysis of individual necessity.....	201
Table 15.	Fit of transformed conditions .....	203
Table 16.	Observational reports from crisp configurations .....	205
Table 17.	Observations in .Z, .C, .D, .I, and .M reports.....	206
Table 18.	Truth tables from .Z, .C, .D, .I, and .M conditions .....	207
Table 19.	Solution terms from .Z truth table .....	210

*intentionally blank page*

# Introduction

Among the practitioners and the students that meet it for the first time, Qualitative Comparative Analysis<sup>1</sup> easily arises some puzzlement. Neither its technical side nor its lexicon fits well with some conventional wisdom about methods. For instance: although QCA is portrayed as a qualitative strategy, it operates on formal models; however, its models rely on membership scores instead of probabilities and on sets instead of variables. Moreover, its algebra seems to speak more to electronic engineering than to politics, laws, or economics and may defy the statistical mindset. Instead of equivalences, it considers inequations. Its models assume compound factors instead of single ones. The model fit captures the “sufficiency” of the claim that the compound leads to the effect, and is decided in terms of “consistency”. Its solutions depend on counterfactual considerations, but the counterfactuals are listed in a “truth-table”, ask whether an unobserved configuration could have obtained, and are routinely decided following “directional expectations.” QCA can claim explanatory import for inferences drawn without any *ceteris paribus* clause, and its findings do not necessarily improve with the number of the observations.

The ground on which QCA builds its solutions, however, is less outlandish than its technical lingo may suggest. QCA relies on logical operations based on the Boolean algebra that anybody applies when querying a search engine. Its algebraic structures can be rendered as set relationships<sup>2</sup> of partial or complete overlapping. That the latter can be given a causal interpretation, however, depends on a further, and sometimes neglected side of Boolean algebra – its logical nature. Logical

---

<sup>1</sup> Ragin, 1987, 2000, 2008; Schneider and Wagemann, 2012; Duşa, 2018.

<sup>2</sup> Stone, 1936.

algebras are literals linked by connectives that render the structural features of our statements about the world. These statements can be as complex as wished and take many shapes – comprised that of a causal account, or an explanation.

From Aristotle to the recent scholarship of mechanistic modeling, the philosophy of science has long since established that explanations are statements and “because-answers” to questions on why something is as it is. Formal logic enters the picture to ensure the “compellingness” of the causal arguments that support because-answers. Under the assumption that compellingness also depends on structural features of the argument, formal logic narrows on the quality, quantity, and relationships of the explanatory statements. Its syntax provides both the principles of valid accounts and the yardsticks to assess their empirical import.

For ensuring validity, logic takes a toll in terms of information loss. Its formalization invites to abstract reality away unless we are left with few relevant facets. Like maps, these reductions serve a purpose, however. They contain the relevant information to recognize the phenomenon across its diverse manifestations and, were it the case, to operate on it effectively<sup>3</sup>. Even when inadequate, moreover, logical models can draw our attention to what is missing from an account and can be improved by filling the gaps. QCA belongs to this tradition as a technique geared to mold, test, and refine an explanatory model according to the rules of logic. Although not unique in its commitment to this rationale<sup>4</sup>, its reliance on sets does impart a twist to its analytic operations.

A full causal argument consists of three pieces of knowledge: the observation that some phenomenon occurs; the tenet that the phenomenon arises from the flow of behavior, activities, or interactions of the individual entities of a particular class; and the bundle of conditions under which the flow obtains<sup>5</sup>. The “oomph” of classical explanations consists of accounting for the occurrence of the phenomenon by exposing the compelling connections that the flow establishes between certain conditions and the occurrence of the phenomenon. The flow, however, has long been taken for granted under the unproven metaphysical assumption that it followed some capacity that “inhered” to the members of the class. Just the opposite, modernity requires that the standing

---

<sup>3</sup> Craver and Kaplan 2018.

<sup>4</sup> e.g., Pearl, 2015; Morgan and Winship, 2015; Humphreys and Jacobs 2015; Bennett and Checkel 2015; van Evera 1997; Ben-Eliyahu and Dechter, 1994.

<sup>5</sup> Craver and Kaplan, 2018; Salmon, 1984, 1998.

of any statement in science is questioned and tested before it is believed, and none of the three components of an explanation escapes the skeptical attitude. The common wisdom about methods then portrays a division of labor among techniques and approaches that address the phenomenon, the flow, and the relevant conditions as proxies of the elusive concept of generative capacities. The criteria that inform such a division of labor, however, are far from neat or unquestionable.

The portrayal that Ragin<sup>6</sup> invoked to justify the proposal of Qualitative Comparative Analysis is quite conventional in pitting “variable-oriented” correlational studies against “case-oriented” narratives. The contrast has been variously reformulated to emphasize the different goals, mindsets, assumptions, and pitfalls of these scholarships<sup>7</sup>. The contrast inevitably emphasizes some features and may result in ungenerous representations, but it clarifies the place that QCA was intended to take among the existing alternatives. From this perspective, correlations narrow on the connection between the phenomenon and selected facets related to the flow to ascertain whether it holds regardless of the background conditions, which will prove its robustness across contexts. Correlations, hence, imply that has explanatory power what, on average, improves the probability that the phenomenon occurs: but robustness and context-insensitivity alone cannot prevent correlations from mistaking the barometer for the cause of the storm, or the shadow of the flagpole for the reason of the rising sun. Narratives, in contrast, are portrayed narrowing on one or few contexts to advance the classical claim that the explanatory power lies in the local bundle of conditions, as it is their interplay that arises or suffocates the tendency of the individuals to behave, act, and interact. Moreover, they often show how the flow in a correlational model is not the only path to the phenomenon to conclude that, eventually, models lie. However, in-depth renderings of the local intersection seldom leave us with valid accounts, as they may explain that salt dissolved in water because someone cast a spell.

QCA was designed to occupy a middle ground and combine the “best” features of the two strategies. It agrees with case-oriented analyses (hence, its “qualitative” nature) that the flow depends on the local conjunction of the right conditions – and that these “scope conditions” do the explaining. At the same time, QCA provides a solution to the relative inability of local narratives to discriminate between relevant

---

<sup>6</sup> Ragin, 1987, 2000, 2008.

<sup>7</sup> e.g., Goertz and Mahoney, 2012; Berg-Schlosser *et al.*, 2009; Brady, 2008; Ragin 1998; Verba, 1967.

and irrelevant conditions<sup>8</sup>. It tackles the problem with the rules of logic: it models all the possible variations from a rich hypothesis, then ascribes explanatory power to those bundles of conditions that survive established criteria<sup>9</sup>.

It will be shown in the following that QCA can yield useful and credible explanatory accounts, although its solutions hold within the boundaries of the observations from which they have been drawn, pending further proofs<sup>10</sup>. The argument develops in the light of the experience that the tools and practices of QCA are deeply engrained in the philosophical discourse on explanation and its evolution, so that, in turn, the familiarity with this discourse improves the understanding of the power and limits of the technique. Consistently, the argument proceeds vertically through the scholarship. The strategy sacrifices many details and alternative applications of QCA<sup>11</sup> to elucidate the rationale of its explanatory usage.

Thus, chapters one and two are intended to set the stage. The first introduces the Aristotelean definition of explanation and its relationship with the idea of *episteme* as reliable knowledge about causation. It clarifies the role that logical syllogistic structures play in connecting the sparse pieces of knowledge into valid inferences and introduces the minimal notions of categorical logic required to understand the composition and functioning of the inferential machinery. The doctrine of the four causes is then presented as the metaphysical warranty that the Aristotelean system requires to ensure that a syllogism is causal, meaningful, and sound. The second chapter widens the inferential toolbox to the alternative logic developed within the Athenian Stoic school. The chapter is intended as a discursive rough guide to the principles of the propositional logic later developed by Frege, Wittgenstein, Russell, Carnap, Quine, and Hempel, among others<sup>12</sup>. Chapter three accounts for the seminal moment when the modern approach to explanation took its

---

<sup>8</sup> e.g., Compton *et al.*, 2019.

<sup>9</sup> Ragin, 1987, 1999, 2008; Quine, 1952; Verba, 1967; Walker and Cohen, 1985; Salmon, 1998; Cartwright, 1999.

<sup>10</sup> However, see Blair *et al.*, 2019.

<sup>11</sup> e.g., Hino, 2009; García-Castro and Arino, 2016; Blatter and Haverland, 2012; Schneider and Rohlfing, 2013; Rohlfing and Schneider, 2013; Baumgartner, 2013; Baumgartner and Thiem, 2015; Goertz 2017.

<sup>12</sup> All the notions of propositional logic in use in QCA will nevertheless be systematically presented in Part II.

shape around the problem of providing a firmer empirical ground to syllogistic inferences through induction.

After a short tribute to Hume, the chapter enters into some details of Mill's contribution, for many reasons. The *System of Logic* brings explanation on the terrain of modern methodology and tackles several of the problems that will later animate the debate in the philosophy and social sciences. Ragin's idea that QCA is better suited to render "chemical causation" instead of the usual physical chain of events explicitly borrows from the distinction established by Mill when reasoning about the plausible shapes of causation. The same holds of the tenet that the "empirical laws of the effect" are a complex interplay of causes, conditions, and obstructions – or the preference for counterfactual eliminative proofs to ascertain the causal import of a factor. The section also calls for attention to Mill's particular understanding of explanation as the self-standing attempts at "resolving" some actual or hypothetical regular association into equivalent relationships.

Although often neglected, Mill's "modes of explanation" have particular importance for the social sciences as they properly apply to those units that are too complex to undergo experimentation, like societies or policy decisions. Their validity does not require any previous "law-like" knowledge, as their standing is established in the inferences by the evidence that supports them. In short, Mill's modes posit explanation between induction and deduction, where Quine, then Ragin, will later operate. Hempel's rendering is addressed in the last parts of the chapter. The portrayal focuses on a mature version of the "covering law model" that popularized the identity of explanation and prediction based on the principle of expectability. The mature version emphasizes that general law-like statements may provide a weak explanatory basis unless the conditions are brought into the picture under which the law-like statement holds. With Carnap, this version considers that the general statement asserts the capacity or disposition of something to contribute to the occurrence of the effect. Conditions are required as the "contrast agent" that reveals the capacity while accounting for its unleashing. Hempel also identifies the strategy of validation that suits sentences under different quantifiers, and adds "observational reports" to each strategy so to avoid the shortcomings of the naïve approach to validation – which casts light on the use of the truth table in QCA.

The concluding section outlines the contemporary mechanistic approaches to explanation as the last influential proposal offered by

Salmon, Woodward, and the stream of mechanistic modelers in response to the covering law model. The proposal integrates knowledge about causal processes and knowledge about causal structures into a single, conceptually consistent picture – dubbed “mechanism” – to account for the occurrence or the non-occurrence of a phenomenon. The conclusion of the chapters and Part I reports Craver and Kaplan’s current recommendations about the suitable level of abstraction in modeling a mechanism, and the reasons that suggest locating it somewhere between the two poles of phenomenological descriptions and bare functional sketches<sup>13</sup>.

Against this backdrop, Part II presents QCA as a technique for modeling the structural part of a mechanism to an outcome around the assumption of a particular underlying process that remains unobserved. Chapter five clarifies that the contribution of QCA consists of specifying the “black box” – Mackie’s “machine”, Cartwright’s “nomological machine”<sup>14</sup>, or Walker and Cohen’s “scope statements” – within which the flow can emerge and unfold until the outcome. The chapter identifies the analytic capability of QCA in its modeling and testing a “machine” as a complete bundle of relevant conditions to the occurrence of the effect. Chapter six addresses the problem of how to learn from the previous literature, then how to select conditions to render a machine. Chapter seven recalls the part of propositional logic in use in QCA, and clarifies the relationship between set-theoretical and logical constructs. Chapter eight addresses the delicate operation of classifying cases as instances of a configuration with the support of set theory and measurement operations. Chapter nine summarizes the protocol, and opens to Part III.

In the last Part, a configurational model is developed around the accounts of the social mechanism of corruption that is fueled by the perception that the policymaking system is illegitimately biased. The configurational model, then, includes the policy levers and pulleys that can be deployed to block the unfolding of the social mechanism. The protocol is therefore applied to test the hypothesis that the differences in the perception of corruption can be explained by the differences in the perceived effectiveness of the constraints that ensure the government is

---

<sup>13</sup> Craver and Kaplan, 2018.

<sup>14</sup> Cartwright, 1989, 1999; Mackie, 1974, 1977; Walker and Cohen, 1985.



accountable to the citizens. The conclusions summarize the features of a configurational explanatory model and discusses its limits.

*intentionally blank page*

Part I

THE BROADER PICTURE

*intentionally blank page*

# 1. Explanations as deductions

The Western philosophy of science owes the demanding equation of deductive reasoning, scientific knowledge, and causal explanations to Aristotle<sup>1</sup>. In his *Physics*<sup>2</sup>, he defines knowledge as the *episteme* that follows from proper response to a why-questions. The episteme, to him, arises from a “syllogism,” that is

a discourse in which, certain things being stated, something other than what is stated follows of necessity from their being so. I mean by the last phrase that it follows because of them, and by this, that no further term is required in order to make the consequence necessary<sup>3</sup>.

The definition highlights two essential features. First, the hybrid nature: the explanation is an argument and, as such, is composed of sentences; at the same time, its sentences are statements about the world. Second, its compellingness. An explanation orders and connects its statements so that the last appears as an inevitable consequence of the previous ones. As a whole, moreover, the argument is self-standing, as it contains every information required to get to the conclusion. It does not aim to yield new knowledge but to organize the existing one and illuminate unnoticed connections.

The Aristotelean definition entails a bold assumption: the sentences about the world do mirror the world itself, and the rules of compelling arguments reflect the rules that govern the relationship among phenomena in the world. From time to time, the mirroring assumption will be debated, exploited to impose specific worldviews, repudiated as void of

---

<sup>1</sup> See Corcoran, 1974; Barnes, 1991; Harari, 2004; Malink, 2013; Bronstein, 2016.

<sup>2</sup> Aristotle, *Physics*, 194b16-195a3.

<sup>3</sup> Aristotle, *Prior Analytics*, 24b18-20.

any cogency, elaborated upon, and problematized. In no case, however, the Aristotelean machinery that embeds this assumption went disregarded. The reason for such an enduring legacy lies in the unquestionable merit of the *Organon*, which first set the ground for logic as a formal language.

## 1.1 The shape of a statement

The strength of an explanation originates from the formal features of the statements that compose it, and on the structure of their connection. The formal part of the syllogistic system is introduced in *Categories* and *On Interpretation*. The system builds on the assumption that we can formulate either true or false sentences about the world. Their truth, in turn, depends on whether the sentence corresponds to the empirical state of the world “in one of the divisions of time”<sup>4</sup>. Thus, to Aristotle, the statement that Socrates is sitting is true unless Socrates gets up, at which point the assertion becomes false. The change in the “truth value” of a sentence, then, depends on the change in the state of the actual thing, and in its misalignment with the corresponding state of the world<sup>5</sup>.

The Aristotelian sentences are the most elementary part of the discourse that cannot be further divided without losing the capacity to bear a truth value. They always include a property or attribute (*katēgoroumenon*) that takes the place of the predicate  $P$  and a subject  $S$  (*hupokeimenon*) of which  $P$  is predicated (*katēgoreitai*). Besides, all these statements can vary along two dimensions: their quality and their number.

The *quality* of a sentence is decided by its predicate, that is, by whether the property is affirmed or negated. An *affirmative* sentence predicates  $P$  of  $S$  and takes the form ‘ $S$  is  $P$ ’. A *negative* sentence predicates not- $P$  of  $S$  and takes the form ‘ $S$  is not- $P$ ’. The relationship between affirmative and negative sentences is assumed to be exclusive: “It is evident that a single affirmation has a single negation. For the negation must deny the same thing as the affirmation affirmed, and of the same thing”<sup>6</sup>.

The quantity of a sentence, instead, depends on the subject and determines the *number* of a categorical sentence. When the subject refers

---

<sup>4</sup> Aristotle, *On Interpretation*, 17a20-17a24.

<sup>5</sup> Aristotle, *Categories*, 4a21-4b19.

<sup>6</sup> Aristotle, *On Interpretation*, 17b38-18a7.

to one or few entities of a class, the sentence is *particular*; when it covers a whole class, the sentence becomes *universal*: “I call universal that which is by its nature predicated of a number of things, and particular that which is not; man, for instance, is a universal, Callias a particular”<sup>7</sup>.

Therefore, affirmation can be predicated of the whole class, as in ‘every  $S$  is  $P$ ’, or of a part of it, as in ‘not every  $S$  is  $P$ ’: the former portrays a *universal affirmative* sentence; the latter, a *particular affirmative* sentence. The negation, too, can be predicated of a part or the whole of a category. A *particular negative* sentence takes the shape of ‘some  $S$  is not- $P$ ’. A *universal negative* sentence states that ‘every  $S$  is not- $P$ ’ or, indifferently, that ‘no  $S$  is  $P$ ’ – the two conveying the same meaning.

## 1.2 Relationships and rules

The compelling nature of an explanation follows from the logical relationships among the sentences that compose it. These relationships arise from symmetries and asymmetries in the quality and numbers of the sentences independent of their substantive content. In better detail<sup>8</sup>:

1. Universal sentences of different quality are *contrary opposites*: they cannot be true together, although, in the Aristotelean system, they can be false together.  
Thus, it cannot be true that ‘every  $S$  is  $P$ ’ and, at the same time and in the same way, that ‘every  $S$  is not  $P$ ’. With an example, it cannot be true that every swan is white and, together, that every swan is not white. The same holds for the singular subject: it cannot be said of the same swan at the same time in the same sense that it is white, and it is not.  
Known as the principles of the “Excluded Middle” and of “Non-Contradiction,” these impossibilities constitute the standard axioms of classical logic, and any logical system characterizes itself for the position it takes on both.
2. Within the same quality, the particular is *subaltern* of the universal. When the universal is true, its particular shall be true, too – and the false particular makes the universal false. Thus, from the

---

<sup>7</sup> Aristotle, *On Interpretation*, 17a37-17b16.

<sup>8</sup> Aristotle, *On Interpretation*, 17a37-18a12.

true sentence that 'every  $S$  is  $P$ ', it follows that 'some  $S$  is  $P$ ' is true. The rule is not symmetric, though, as a true particular does not warrant the truth of its universal. Being it true that 'some swan is white' does not grant the inference that 'every swan is white' – whereas the reverse holds instead. Moreover, from the false sentence that 'every  $S$  is  $P$ ', it does not follow that 'some  $S$  is  $P$ ' is false. Being false that 'every swan is white' can mean either that every swan is not white, or that some are not.

In short, *truth only flows downward* – it extends from a true universal to its particular, but not *vice-versa*. Instead, *falsehood only flows upward*: a false particular justifies our inference that its universal is false, but not *vice-versa*.

3. The particulars of different quality are known as *subcontraries*. Given their relationship with their universal, and given the relationship between universals, they cannot be false together, so that at least one is always true; however, they may be true together. It can be true that 'some swan is white' and together that 'some swan is black' if it is said of different swans at the same time or of the same swans at different times; in both cases, none of the respective universal statements can be true.
4. Sentences form *contradictory* opposites if they variate in both quality and quantity. In the Aristotelean system, contradictory statements are connected by a relationship of mutual exclusion, so the one being true necessitates the other being false. Thus, being true that 'every swan is white' makes false that 'some swan is black'; being false that 'every swan is black' entails that it is true that at least 'some swan is white'. The evidence (truth) of a particular contradicts the truth (falsifies) the universal of the contrary quality. This rule of falsification will find application beyond the Peripatetic tradition.

A later convention denotes the universal affirmative as  $a$ , and the particular affirmative as  $i$ , from the two first vowels of the Latin word *affirmo*. Similarly, the universal negative is denoted as  $e$ , and the particular negative as  $o$  from the two first vowels of the Latin word *nego*. Then, the relationships between the possible types of sentences taken pairwise yield the renowned *square of opposition* in Figure 1.

The relationships between the corners in the square license the following inferences independent on the content of the sentences:



A1. true *PaS* means

- *PiS* is certainly true;
- *PoS* is certainly false;
- *PeS* is certainly false.

A2. true *PeS* means

- *PoS* is certainly true;
- *PiS* is certainly false;
- *PaS* is certainly false.

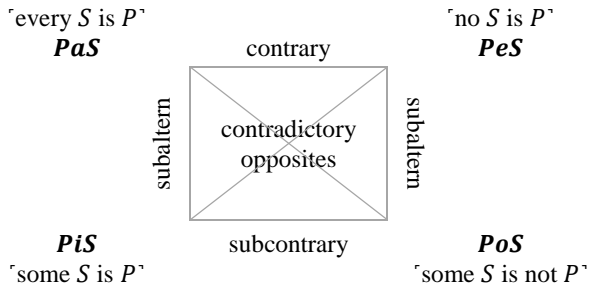
A3. true *PiS* means:

- *PeS* is certainly false,
- *PaS* and *PoS* are possibly true,
- either *PaS* or *PoS* is certainly true.

A4. true *PoS* means

- *PaS* is certainly false,
- *PeS* and *PiS* are possibly true,
- either *PeS* or *PiS* is certainly true.

**Figure 1. The square of opposition**



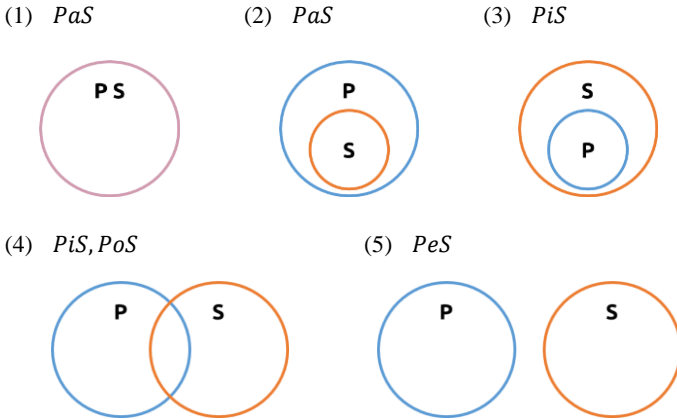
Universals license a higher number of sure inferences than particulars. Particulars, instead, only have the logical power of contradicting – that is, of falsifying – the universal of their subcontrary. Unsurprisingly, Aristotle established universals as the proper ground for *episteme* as sure knowledge. The tenet has proven enduring, at least as a desirable ground to the validity of inferences<sup>9</sup>. The issue appears less problematic in the Aristotelean than in alternative systems. For one thing, categorical logic rested on the assumption of a closed universe, which makes universals easier to establish. Second, categorical logic excludes that we can build a syllogism on accidental phenomena: chance or spontaneity provide no ground to the episteme.

As a further consideration, the relationships between categorical sentences allow for two related interpretations – the *intensional*, and the *extensional*. The intensional reading narrows on those abstract features

<sup>9</sup> Campbell and Stanley, 1963; Cook and Campbell, 1979; *cfr.* Hempel, 1965; Cartwright, 1979.

that qualify a predicate as non-accidental *in se* and opens to considerations about the *possibility* of some occurrence. The extensional interpretation refers to observed actualized phenomena. Along with the extensional reading, Venn<sup>10</sup> will later show that the relationships in the square of opposition can be rendered as a special overlapping of the class of objects *S* and the class of objects with the property *P*. Therefore, *PaS*, *PiS*, *PeS*, *PoS* find correspondence with at least one of the diagrams in Figure 2.

**Figure 2. Venn's rendering of the basic Aristotelian relations**



Source: Venn (1881: 6 ff.)

As Venn underlines, the ambiguity of linguistic statements prevents a clear-cut rendering. The only categorical statement that proves immune to ambiguity is the universal negative *PeS*, which indicates the class of *P* objects and the class of *S* objects do not overlap as in diagram (5) of Figure 2. The picture also intuitively conveys a further property of the negative universal: it ensures that the operation of *conversion* is truth-preserving.

As applied in *Prior Analytics*, conversion refers to the predicate and the subject switching their positions in a sentence. The conversion is truth-preserving when the new sentence maintains its original generality. This is the case of negative universals, as ‘no *P* is *S*’ and ‘no *S* is

<sup>10</sup> Venn, 1881; Stone, 1937.

$P$  hold equally true. That that no human is immortal is as true as that no immortal is human. Conversion does not safely apply to positive universals, instead. The statement 'all  $P$  is  $S$ ' fits diagram (1) as much as diagram (2) of Figure 2, although only diagram (1) supports truth-preserving conversion. That is, one can safely state that all humans are mortal, yet not that all mortals are humans – the class of mortals being broader than that of humans. Direct conversion of a universal positive  $PaS$  is only allowed if we can rule out that its contradictory 'some  $P$  is not  $S$ ' is false, as this proves the overlapping of  $S$  and  $P$  as in diagram (1). Affirmative universals, as diagram (2) displays, only support the conversion into their particular – from  $PaS$  into  $SiP$ . Indeed, if it is true that all humans are mortal yet not that all mortals are human, it nevertheless holds that *some* mortal beings are human. As a general prudential rule, therefore, any affirmative universal is considered as the class of  $S$ -objects that is at least contained in the class of  $P$ -objects. Their identity provides a special case of inclusion that requires empirical probation – for instance, like in geometrical proofs, by showing that absurd conclusions would follow if we assumed the identity false.

The last categorical safe conversion is between positive particulars –  $PiS$  and  $SiP$ . This particular again arises an ambiguity, as it may refer to the class of  $P$ -objects being perfectly included in the class of  $S$ -objects (diagram 3 in Figure 2), or only partially overlapping it (diagram 4 in Figure 2). Only perfect inclusion rules out that  $PoS$  is true. However, the positive particular as such can be logically insensitive to this ambiguity. If it is true that some  $S$  is  $P$ , it also is true that *at least* some  $P$  is  $S$ : so, the partial intersection constitutes the default relationship of reference, of which the full inclusion again provides a special case. Partial intersections, however, also occur by accident, hence cannot be regarded as meaningful as the special case of full inclusion.

The rules of conversion under ambiguity entail a further logical principle<sup>11</sup>. As a chain is no stronger than its *weakest link*, so the valid inference cannot advance claims stronger than the weakest of the relationship among quantities on which it builds.

These few relations and principles provide the ground for establishing truth-preserving explanatory inferences.

---

<sup>11</sup> Malink 2016, Patterson 2002:75 ff.

### 1.3 The ideal shape of a valid deduction

To find which relationships support valid inferences, *Prior Analytics* proceeds deductively from basic stipulations. It establishes every explanatory inference can be reduced to a scheme involving three terms – the subject *S*, the predicate *P*, and a middle term *M*. The three terms are organized into two premisses and one conclusion. The *major* premiss establishes a general rule connecting *P* and *M*; the *minor* premiss instead makes a case relating *S* and *M*; the conclusion reports the claim about the relationship of *S* and *P* that the two premisses should demonstrate. The demonstration depends on *M*, which provides the logical connection between *S* and *P* while falling in the conclusion. Otherwise said, the syllogistic machinery is geared to demonstrate that *S* is *P* because of *M*.

The position of *M* across the premisses determines the *figure* of the syllogism. In the first figure, *M* takes the place of the subject in the major premiss and of the predicate in the minor; in the second figure, it takes the position of the predicate in both premisses; in the third, it always takes that of the subject. A fourth figure, acknowledged yet not discussed by Aristotle<sup>12</sup>, would be later developed in which the middle term is given the position of the predicate in the major premiss and the subject in the minor.

The figures display a different *mood* depending on the combination of quality and number of the categorical sentences that make the premisses and the conclusion. The possible combinations per figure are 64, ranging from the *aaa* of the syllogism with all positive universal statements to the *ooo* of the syllogism of all negative particulars.

Of these possible combinations, Aristotle proved that only a limited number of them were *valid*, that is, truth-preserving. Among the valid syllogisms, those directly associated with the scientific discourse are in the first figure. Known with the mnemonic names of *barbara*, *celarent*, *darii*, and *ferio*, they portray the relations among *P*, *S*, and *M* as displayed in Table 1.

These syllogisms are “perfect” (*anapodeiktos*) as the consequence can be grasped by referring solely to the fundamental relationships between qualities and numbers. Of them, the *barbara* and the *celarent* are of particular import as they portray the fundamental axioms by way of which the *darii* and the *ferio*, too, can be perfected. In the Aristotelean

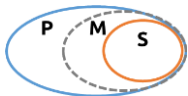

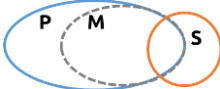

---

<sup>12</sup> Łukasiewicz, 1957.

system, the first two moods of the first figure, the conversion rules, and the principle of non-contradiction applied as argument *ad absurdum* together suffice to prove – or disprove – the validity of any remaining syllogisms, hence of every argument offered as an explanation.

The perfect syllogisms clarify that the middle term explains as it establishes the relevant link. The process is exceptionally transparent in the case of the *barbara* syllogism, which renders the rule of transitivity as chained identities as in Figure 3.

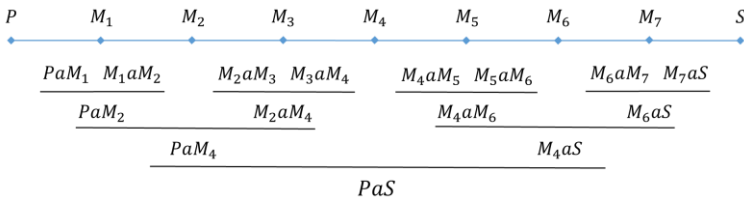
**Table 1. Valid moods of the first figure and related set structures**

Mood	Deduction	Set structure
<i>barbara</i>	$\frac{PaM \quad MaS}{PaS}$	
<i>celarent</i>	$\frac{PeM \quad MaS}{PeS}$	
<i>darii</i>	$\frac{PaM \quad MiS}{PiS}$	
<i>ferio</i>	$\frac{PeM \quad MiS}{PoS}$	

Keys: in each cell of the deduction column, the two statements above the line correspond to the major premiss (on the left), and the minor premiss (on the right); below the line is the conclusion.

Source: Malink (2017); Venn (1881).

**Figure 3. The *barbara* syllogism as transitivity**



Source: adaptation from Malink (2017).

Thus, when the chosen middle terms cannot connect  $S$  and  $P$  in one single step, new premisses can be added as links in a longer chain<sup>13</sup>. Such a mechanism illuminates Aristotelean definition of explanations as self-standing arguments that expose meaningful connections otherwise unnoticed.

The logical chain between  $P$  and  $S$  cannot close if any middle link is missing – which lies the inference open to an infinite regress. In the Aristotelian system, however, such a risk is defused ontologically: the regress ends when the middle term is a cause. Causal middle terms also warrant the syllogism is genuinely explanatory.

#### 1.4 The “because” that make syllogisms causal

From the *Posterior Analytics*, we learn that while all valid explanations are valid syllogisms, the reverse does not hold. To illustrate the point, Aristotle contrasts the following two arguments:

<p>B1    Near things do not twinkle</p> <p style="padding-left: 20px;">Planets do not twinkle</p> <hr style="width: 80%; margin-left: 0;"/> <p style="padding-left: 20px;">Planets are near things</p>	<p>B2    Near things do not twinkle</p> <p style="padding-left: 20px;">Planets are near things</p> <hr style="width: 80%; margin-left: 0;"/> <p style="padding-left: 20px;">Planets do not twinkle</p>
--	--

Both display the same major premiss – the rule connecting “near things” and “not twinkling.” In both, the minor premiss makes the case about “planets.” What changes from B1 to B2 is the term from the major premiss that is used as  $M$  in the minor.

In B1, the minor premiss assigns the position of  $M$  to the property of not twinkling; in B2,  $M$  is the property of being a near thing. Aristotle clarifies syllogism B1 provides the “knowledge-that” (*oti*), establishing in agreement with our senses that planets are near things. Just the opposite, syllogism B2 offers the “knowledge-why” (*dioti*), as its conclusion goes beyond our perception while establishing the reason for an occurrence such as that planets do not twinkle.

The latter explains as it appeals to a *causal* feature of  $S$  – the nature of planets that, in Aristotle’s cosmology, is of near wandering stars. Be-

---

<sup>13</sup> Malink, 2013, 2017.

ing near, to Aristotle, is part of the *essence* of a planet, while not twinkling is a secondary property – an attribute, a mark – that follows from it. Hence, nearness can be invoked as the middle term that turns the valid syllogism into an explanation.

In an explanatory syllogism, “the middle term is an account of the major term, which is why all the sciences come about through definition”<sup>14</sup>. Here, “account” (*logos*) is to be understood as the sentence that defines the *nature* of the subject – the higher property that justifies the compelling linkage between the subject *S* and the predicate *P*<sup>15</sup>. This assumption about *M*, however, seemingly opens a metaphysical breach in the formal construction. It makes proper scientific knowledge depend on the content of the premisses – namely, on the fact that they invoke right, although given, “explanatory factors” or “because” (*aitia*).

#### 1.4.1 The four causes

The Aristotelean ontology is deemed pluralist as it famously considers that we can speak of *aitia* as of four kinds: *material*, *formal*, *efficient*, and *final*<sup>16</sup>.

The material principle narrows on the constituents or the *potential* (*dynamis*) of the subject. Examples are the bronze of a statue or the letters of a syllable. The formal principle refers to its structure instead – understood as its *realized shape* (*entelecheia*), spatial arrangement, or essential relationships binding the constituents together into a whole broader the sum of its parts. Such is the syllable with regard to the letters, or the form to the statue. Together, the formal and material reasons constitute the “nature” of a thing.

The efficient reason captures the capacity that activates in the making of the *explanandum*, such as the skill of the sculptor in molding the statue or the ability of the doctor in ensuring someone’s good health, and more generally, all those factors that initiate or impede a process of change. The final reason recognizes functions and purposes, such as celebrating an athlete of molding a statue or being healthy of someone’s walking. Efficient and final reasons complement the nature of things and account for them becoming *P* or ceasing to be so in the actual

---

<sup>14</sup> Aristotle, *Prior Analytics*, 99a21-3.

<sup>15</sup> Bronstein, 2016; Malink, 2013.

<sup>16</sup> Aristotle, *Physics* 194b24-195a2; 195a15-195a26; Hocutt, 1974.

world<sup>17</sup>. Reasons, too, can be either *potential* or *actual*<sup>18</sup>. Their manifestation is revealed in the process of change and ceases with it: the molding sculptor is the actual reason for the being-molded statue as long as the molding endures. In their potential state, instead, they exist before and after the process: a sculptor's capability survives the molding of the particular statue. A relationship of priority is established as the art – or capacity, or capability – accounts for the skills that qualify the maker.

Together, these four principles are the constitutive *epistemic* dimensions of the Aristotelian substances as the *physis* of everything. Hence, they can be understood as four different criteria to decide on the explanatory relevance of some *M* rather than as a straightforward list of causal entities.

That which “brings about” *P* is anything that qualifies as both an efficient and a final *aitia*. To fully account for the *inherence* of *P* to *S*, however, the efficient and final causes are seldom enough. *S* can become *P*, or manifest it, only when efficient and final factors interact with *S*'s substantial and formal nature, that is, when they affect the features of its constituents and their organization or binding rules. Besides, to Aristotle, too, an explanation is driven by what we want to explain, which also provides the ground on which ‘*S* is *P*’ holds. Therefore, the same factor may enter different explanations with different roles; moreover, different factors can play complementary roles in the same explanation<sup>19</sup>.

The usage of the *aitia* as the criteria for deciding on the explanatory relevance of properties becomes evident in Aristotle's discussion of *accidental* causal factors. In this system, only a general property can close the chain of transitivity and compel a conclusion<sup>20</sup>. The accidental property, instead, becomes explanatory only when conjoined with a proper general factor. Thus, Polykleitos is the valid accidental reason for the statue, because the skills of a sculptor constitute an appropriate, efficient reason for its making, and because being Polykleitos was “accidentally conjoined” with being a sculptor<sup>21</sup>. When the accidental reason cannot be conjoined with a proper general reason, the *explanandum* is ascribed to chance or spontaneity.

---

<sup>17</sup> e.g., Annas, 1982.

<sup>18</sup> Aristotle, *Physics*, 195b17-20.

<sup>19</sup> Moravcsic, 1974.

<sup>20</sup> Aristotle, *Physics*, 195a27.

<sup>21</sup> Aristotle, *Physics*, 195<sup>a</sup>30-35.



Aristotle's *Physics* does not deny that the world has a stochastic element. However, it does not admit valid syllogisms about stochastic occurrences. A set of purely accidental reasons makes the explanation "indeterminable": an accidental maker has "innumerable" properties (such as of being pale, young, a flute player, ...) none of which is connected to the *explanandum* by the sake of its constituents, organization, capacity, or purpose. Because of Socrates' human nature, his death can be expected as inevitable: but its actual occurrence can be explained by his choice of drinking hemlock and by the poisonous effect of hemlock on humans. The color of the cup in which the hemlock was poured or the day it was drunk, instead, count as irrelevant conjoined circumstances on the sole basis of which no explanation could be possible. Syllogisms can only be about "things" that "always come to pass in the same way, or for the most part"<sup>22</sup>.

The explanation of an occurrence, therefore, consists of the conjunctions of the local factors that account for the "necessary" actualization of a general property. The relevance of the particular circumstances, in turn, is decided by their contribution to the actualization of the potential intrinsic to the subject's property. In short, the four *aitia* define the Aristotelean why questions as a matter of accounting for the reasons that make *S* being *P*. The *aitia* also point to the complementaries that the proper explanation should consider: (1) the nature of *S* – its constituents, their bound – that makes it *possible* that *P* inheres in it; (2) the factors – willingness, capacity – that can precipitate *P*'s inherence in *S*. Together, these two facets provide the full definition of the subject<sup>23</sup>.

The explanatory syllogism, moreover, qualifies as an *ex-post* exercise. In the Aristotelean ontology, general properties entail a potential (*dynamis*) that remains unrealized unless some conjunction of particular circumstances is given. The past, therefore, can be given a truth value, whereas the future remains open unless some constellation of circumstances rises that turns the potential into actual. Thus, we can establish why yesterday a sea battle took place, and we can speculate on the circumstances that will make it possible that a sea battle takes place tomorrow. Whether it is true or false that a battle will take place tomorrow, instead, only tomorrow will say<sup>24</sup>.

---

<sup>22</sup> Aristotle, *Physics*, 196b10, 198a5.

<sup>23</sup> e.g., Hocutt, 1974.

<sup>24</sup> Aristotle, *On interpretation* 18b17-18b25; Malink, 2013; Frede, 1992.

## 1.5 Limits of categorical deductions

The Aristotelean syllogism holds when the bundle of causal properties  $M$  groups every instance of  $S$  together as a homogeneous partition of  $P$ <sup>25</sup>. In turn, the holding of the chain of identities depends on the truth-values of the atomic sentences in the premisses, which is ultimately warranted when the bundle of causal properties is a true universal. The bundles of properties, however, suggest that universals are a conditional matter: the reason for  $S$  being  $P$  depends on the efficient and final conditions under which  $S$ 's potential nature can actualize. As the argument of the sea battle emphasizes,  $S$ 's potential remains dormant unless it meets the right conditions, and unactualized potentials leave the truth-values open.

Logical languages can acknowledge the problem of future contingencies by introducing *modal operators*. Modals render the necessity, possibility, impossibility of some occurrence. They signal whether a statement holds in “every possible world,” or accidentally in some of them only<sup>26</sup>. If the statement ‘every  $S$  is  $P$ ’ is a true universal and holds in every place at every time, it can be reformulated as ‘necessarily,  $S$  is  $P$ ’. If the same holds in no place and at no time, we can reformulate the statement as ‘necessarily, it is not the case that  $S$  is  $P$ ’ – which has the same meaning as ‘it is impossible that  $S$  is  $P$ ’.

The Aristotelean categorical logic does support a modal reading. Moreover, the *Organon* acknowledges modal concerns and, implicitly, develops some rules to handle them. For instance, the square of opposition allows maintaining that when the universal  $a$  holds, it entails a necessity shared by any particular  $i$  and the symmetric impossibility of both  $e$  and  $o$ . Hence, if ‘all swans are white’ is necessary, then both ‘every swan is black’, and ‘some swan is black’ become impossible statements.

When we only know the particulars, however, the concepts of possibility and impossibility acquire an epistemic meaning. Thus, the true particular  $i$  entails  $e$  is impossible, yet allows both  $a$  and  $o$  being possible – and so it becomes a “logical necessity” that either one or the other is true. If it is the case that ‘some swan is white’, it is impossible that ‘every swan is black’, and it is either possible that ‘every swan is white’ or that ‘some swan is black’, although not both at once. This

---

<sup>25</sup> Brody, 1972.

<sup>26</sup> Lewis *et al.*, 1951; Rescher, 2007; Hintikka, 1969; Mates, 1968; Kripke, 1959.

necessity is purely logical, as the actual quality of the next swan remains open. Such openness in truth can only vanish if we find the bundle of conditions that, once given, rules out the possibility of a black swan and makes the occurrence of a white swan certain. In themselves, unactualized possibilities only license conclusions of little factual content such as 'therefore, every swan is either white or not' – which is a tautology. As it contains all the possible truth values yet does not specify any, it does certify our ignorance.

Although void of factual content, these statements still allow valid inferences. This point has been of primary interest to a different tradition in logic that will become dominant in modern times, but whose roots lie in another Athenian school of philosophy<sup>27</sup>.

---

<sup>27</sup> Mates, 1949, 1961; Bochenski, 1951; Frede, 1974.

## 2. Inference as propositional calculus

Parallel – and rival – to the Peripatetic, the Stoic school flourished with Chrysippus two generations after Aristotle. Although old catalogs suggest an abundant intellectual production, only a few fragments of the original teaching have survived, mainly through later unsympathetic Greek and Latin sources. Nevertheless, the Stoic fragments exerted an uncontested influence on modern logic<sup>1</sup>.

Given the many positions the Stoic philosophers took on several fundamental issues, theirs cannot be considered a single consistent doctrine. However, the school stands out as distinct philosophical kin by at least two features. First, their lexicon: philological portrayals emphasize the accuracy and continuity in the Stoic usage of technical terms. Second, their ontology: in contrast with the Platonic assumption of metaphysical essences, the Stoics only acknowledged the existence of individual bodies in the natural world, about which our mind develops images (*phantasia*) shared through sentences or “sayables” (*lekta*). Thus, Stoic categories – namely, substratum (*hypokeimenon*), qualities (*poion*), state or condition (*poion pòs echon*), and relative condition (*proion pros ti pos echon*) – are inventories of “what can be said” about individual bodies<sup>2</sup>. Based on fragments, never these categories licensed universal statements as in the Aristotelean logic. Instead, they only applied to some individual subjects – either determined, such as ‘Dion’, or undetermined, such as ‘this’ and ‘that’, ‘the first’ and ‘the second’. Subjects, in turn, could be known: as the unique bundle of substratum, qualities, and conditions that single them out as individual bodies in the world.

Such a position has remarkable consequences on logic. The Stoics considered the Aristotelean type of deductions – based on transitive

---

<sup>1</sup> e.g., Mates, 1961; Kneale and Kneale, 1965; Łukasiewicz, 1952.

<sup>2</sup> Rist 1969: 55-56.

identities and warranted by ontological assumptions about universals – be “unmethodically” valid arguments. In contrast, they developed an understanding of logical validity independent on the number of the subject. Valid inferences follow from the connection among complete statements arranged into distinctive moods, or *schemata*, when the connection proves consistent to few rules basic rules, or *themata*.

Validity in Stoic logic, in short, mainly depends on the relationships within and among sentences and on their fit to given rules of composition. The truth of the inferences about singular occurrences is decided by the rules of formal consistency and independent on the content of the statements. Moreover, the validity of sentences depends on stipulations and is determined independently of their use in a syllogistic chain. Thus, the Stoic logic includes reflexive propositions – such as that *P* is, and that it follows from itself – that the Peripatetic school dismissed as redundant, pedantic, and in any case irrelevant to the syllogistic argument<sup>3</sup>. However, this signals the remarkable analytic effort that the Stoic made to define the object of their logic – the *lekta*.

## 2.1 Stoic sentences

Although insensitive to the substantive content of sentences, valid Stoic inferences do set requirements. *Schemata* properly apply to a particular subclass of *lekta* – the ones that have a *complete* enunciation, i.e., that contain both the subject and the predicate and hence are self-standing. Of the complete *lekta*, moreover, the Stoic logic excludes several complete *lekta* from the proper scope of logic – such as those conveying questions, commands, oaths, salutations, suggestions, wishes, or curses. Valid inferences only rest on complete *assertoric* sentences, as they can bear truth-values.

Complete assertoric propositions can be *simple* or *non-simple*. The two kinds differ in their constitutive elements: a non-simple proposition is made of propositions; just the opposite, a simple proposition is that which does not contain any further proposition. Non-simple propositions of interest to logic come in three basic types depending on the connective (*syndesmos*) that binds them together:

---

<sup>3</sup> Frede, 1974:23.

- a) a *conjunction* (*sympeplegmenon*) is the non-simple proposition compounded by the connective 'and' (*kai*), as in 'it is day, and it is light'.
- b) a *disjunction* (*dieteygmenon*) is the one compounded by the connective 'or' (*e*), as in 'it is day, or it is night'. In the Stoic usage, a disjunction compounds *exclusive* alternatives. The inclusive disjunction, allowing for one, the other, or both alternatives together, is dubbed a pseudo-disjunction (*paradieteygmenon*).
- c) the connective 'if' (*ei*) or 'if indeed' (*eiper*) yields the *conditional* (*synemmenon*) non-simple proposition, and signals that the part not immediately introduced by the connective, dubbed "the second" regardless of its appearance in the discourse, is the consequence of the proposition directly introduced by the connective, which is generally indicated as "the first." Thus, a conditional can take the shape 'if (*ei*) the first, the first' or 'the second, if indeed (*eiper*) the first'.

The Stoic syntax also recognizes further non-simple propositions – among which, the ones compounded by the connective 'because', or by the connective 'since'. Fragments, however, leave the reader clueless about their standing, making room to the hypothesis that they may have been denied the same import of the three basic connectives, as they can be formally reduced to the others<sup>4</sup>.

Furthermore, the Stoic logic discriminates among different types of negation, depending on whether it applies within a proposition, or to the whole of the proposition. Within a proposition, the denial of the subject (*oudeis*) voids the predicate of any scope without exception, as in 'no-one walks'. Privative prefixes (*steretikon*) only deny the predicate instead, as in *aphylanthropos*, "un-kind." The negation of a whole proposition follows from an introductory 'not' (*ouchi*), which governs its quality and yields its opposite (*apophaticon assiomata*), as in 'not: it is day, and it is night'.

## 2.2 The truth of simple sentences

Negation is intertwined with the concept of truth and falsity of a proposition. Fragments suggest that the Stoics define as true "what is real

---

<sup>4</sup> Mates 1961:55.

and contradicts something” and as false “what is not real and contradicts something” – where “things that contradict are those of which one goes beyond the other by a negative”<sup>5</sup>. Truth, then, can be “in” or “about” propositions. A proposition is true in itself when the portrayals it conveys about the world (*phantasia*) is accurate and, at the same time, corresponds to some actual state of affair. Thus, the proposition ‘this oar is bent’ is correct yet false if the oar is straight but partially underwater. The truth or falsity of a proposition may also be impossible to establish – for instance, when the predicate correctly applies to a part only of the subject. For example, some men are Greek, and some others are barbarian: so, the generic man can be neither truthfully said Greek nor barbarian. The truth “in” a proposition, therefore, is decided by the alignment of the predicates that the subject can truthfully bear with the truth of the particular predicate that actual sentence attaches to the subject<sup>6</sup>. Otherwise said, the Stoic truth in a proposition is a modal matter. However, it is not rendered through modal operators. Instead, the issue is addressed by stipulating the meaning of the connectives – hence, by originating different logical systems or languages.

These developments seemingly stemmed from Aristotle’s sea battle argument, which the first head of the Stoic school, Diodorus Chronus, developed as a trilemma, known as the “Master Argument”<sup>7</sup>. The Argument maintains that the following three tenets are jointly inconsistent – namely:

- (1) what is past and true is necessary;
- (2) the impossible does not follow the possible;
- (3) what neither is nor will be is possible.

The Stoics never challenged the claim that the three statements are inconsistent, although the motivation is uncertain. The point was, to regain consistency, one tenet must be rejected as false. Fragments indicate a debate spurred about which was the illicit one. Diodorus himself refused the third, and classified propositions as *possible* when true, *either* now or in the future; *nonnecessary* when false, either now or in the future; *necessary* when true now, and not false the future; *impossible* when false now, and not true in the future.

---

<sup>5</sup> Bett, 2005:106.

<sup>6</sup> Reesor, 1965.

<sup>7</sup> e.g., Gaskin, 1995; Michaels, 1976; Reesor, 1965; Mates, 1961; Bochenski, 1951.

Against Diodorus, and more in line with the Peripatetic positions, Chrysippus rejected the second tenet of the trilemma instead. Possible events could never occur, and a possible statement could never prove true because of its circumstances despite the subject has the potential to be. Thus: 'the log at the bottom of the ocean is combustible' qualifies as an impossibility, hence a false statement, under Diodorean rules, and as a possibility, hence a true statement, under Chrysippean rules. The Chrysippean truth entails the assumption of the homogeneity of nature – so that what happens once can happen again – and counterfactual considerations about circumstances. According to his criteria, hence, *possible* are those propositions in which the subject has a true predicate unless external concurrent circumstances do not prevent it from being true. *Nonnecessary* are those propositions in which the subject receives a false predicate unless external concomitant conditions do not prevent it from being false. *Necessary* are those propositions that are true or do not admit of being false, and *impossible* are those propositions that are false or do not admit of being true.

When applied to simple propositions, Chrysippean modalities can be understood as the intensional counterpart of the extensional Peripatetic categories. Impossibility yields the negative universal, necessity, the positive one. Possibility and nonnecessity, instead, recall positive and negative particulars. This correspondence, however, weakens when the rules are applied to decide on the truth-value of singular non-simple propositions.

## 2.3 The truth of compound sentences

The rules of composition apply to each of the *lekta* compounded by *conjunction*, *disjunction*, and *conditional*.

### 2.3.1 Conjunctions

Fragments report a conjunction is true when *all* of its conjuncts are true. To those who argued a conjunction could still be true when the majority of its conjuncts are true, the Stoic responded that

just as in life we do not say that the piece of clothing that is sound in most parts, but torn in a small part, is sound (on the basis of its sound parts, which is most of them), but torn (on the basis of its small torn part), so too the conjunction, even if it has only one false component



and a majority of true ones, will as a whole be called false on the basis of that one<sup>8</sup>.

The response suggests, the Stoic rule of composition maintains that a compound is qualitatively different from the sum of its part. The truth-value about a non-simple proposition renders the truth of the special relationship that the connective establishes among its components. The conjunction, from this perspective, provides a straightforward logical strategy to identify a body as the bundle of properties and conditions that qualify it. The statement ‘Scipio was the son of Aemilius Paullus and twice consul and censor, and the defeater of Hannibal’ would be false because the first three statements are true of Scipio Aemilianus, while the latter is true of his adoptive grandfather instead<sup>9</sup>. On a similar vein, ‘it is day, and I am conversing’ is a false statement unless all the corresponding states of affairs are jointly given in the world – and, according to the Diodorean understanding, only if they jointly hold at any time and cannot change.

Interestingly, the assignment of a truth value to a conjunction of propositions follows the rationale of the weakest link that we found embedded in the chains of identities that shape the Aristotelean syllogisms. The Stoic logic considers that the principle of the weakest link establishes the formal equivalence of conjunctions and transitive identities. Formal equivalence will prove a convenient property of propositions, as it turns inference into a straightforward calculus – at least, a more straightforward one than the Aristotelean oppositions and conversions.

### 2.3.2 *Disjunctions*

A disjunction is true if any of its disjuncts is true; an exclusive disjunction is stricter as it is true when only one of the disjuncts is true. Its rationale recalls a decision among rival alternatives. The fragments agree, Chrysippus maintained this was proper of any rational animal – for instance, of a chasing dog that comes

to a crossroad and, having tracked down the two roads along which the wild animal did not go, starts off at once along the third without tracking down it [... The dog] is implicitly reasoning as follows: ‘The

---

<sup>8</sup> Bettl, 2005:126.

<sup>9</sup> Mates, 1961:54.

animal went either this way or this way or this or this; but neither this way nor this: therefore this way'<sup>10</sup>.

To be valid, therefore, the alternatives in an exclusive disjunction have to be mutually *incompatible* and jointly *exhaustive*. Of the two, incompatibility was recognized higher import, as it had the convenient property of ensuring that, when the disjuncts are two, they are contradictory to one another. Thus, 'it is day, or it is night' qualifies a true disjunction as either one or the other disjunct is true, yet not both. From a truth-functional perspective, the incompatibility of two alternatives establishes them as contraries – same objects related by the negation connective<sup>11</sup>. Hence, alternatives entail identities – although through negation.

### 2.3.3 *Conditional*

According to fragments, a Stoic conditional is a non-simple proposition in which the connective conveys the second component is a consequence of the first. Here, the debate about the standards for truth has long proved as thorny as consequential. Fragments<sup>12</sup> report four criteria were advocated within the Stoic school.

The Philonean criterion establishes a conditional relationship that is always true except when its first is true and its second is false. So, the link "is true in three ways and false in one way" – namely,

- (a) it is *true* when both the first and the second are true, as in 'if it is day, it is light';
- (b) it is *false* when the first is true, and the second is false, as in 'if the soil exists, it flies'.
- (c) it is *true* – although "vacuously" so – when the first is false and the second is true, as in 'if the soil flies, it exists';
- (d) it is *true* when both the first and the second are false, as in 'if the soil flies, it has wings'.

---

<sup>10</sup> Annas and Barnes, 2000: 20.

<sup>11</sup> Bett, 2005:110.

<sup>12</sup> Mates, 1961:47 ff; Kneale and Kneale 1962: 128 ff.; Frede, 1974:4ff; Bobzien, 1996:185ff; Annas and Barnes, 2000: 95ff; Priest and Hyde, 2003; Bett, 2005:112ff.

We will find the Philonean conditional in modern formal logic as the “material implication” codified by Whitehead and Russell<sup>13</sup>, and the attribution of truth values to a connective as the truth-function ‘*true-false-true-true*’ formalized by Wittgenstein<sup>14</sup>.

Over time, the Philonean truth-value assigned to (*c*) has spurred lively philosophical debates. On the one hand, it mirrors the principle, enduring in classical logic, that *ex falso quodlibet* – anything can follow from a false premiss. On the other hand, it seems too lax to render a real-world implication – particularly when the implication is endowed with the causal meaning that follows from the “if indeed” connective.

The Diodorean criterion maintains the conditional is true that neither was nor is able to begin with a true proposition and finish with a false one. This is a stricter standard than the Philonean in two respects.

‘If it is day, Socrates is conversing’ is Philonean-true if, at the time of its utterance, it was day and Socrates was having a conversation. It is Diodorean-false, instead, as it is possible that, during the day, Socrates turns silent: thus, the conditional can have a true first and a false second.

‘If it is night, it is day’ again qualifies as true during the day according to Philo, as the conditional would have a false first and a true second. In Diodorean terms, it qualifies as false even during the day because the day can turn into night, and when uttered at night, the conditional has a true first and a false second.

In short, the Diodorean criterion requires the conditional is true that holds “eternally”<sup>15</sup>. It finds its modern match in the *strict conditional* – the implication that holds in every possible world<sup>16</sup>.

The allegedly Chrysippean criterion appeals to “correctness” instead. It maintains that a conditional is sound when the opposite of the second “conflicts” with the first. In short, a conditional is Chrysippean-valid if it supports counterfactual reasoning. The definition resonates with the Peripatetic doctrine in proving a syllogistic connection *ad absurdum* as it was a geometrical theorem. The proof suggests that the syllogistic chain of identities and the Chrysippean conditional render

---

<sup>13</sup> Whitehead and Russell, 1927.

<sup>14</sup> Wittgenstein, 1922.

<sup>15</sup> Quine, 1960.

<sup>16</sup> e.g., Lewis, 1918.

the same kind of dependency. Aristotle's 'is' and Chrysippus' 'if-then' are homologous as they apply to terms with the same extension.

A typical illustration of the Chrysippean conditional considers the statement: 'if not: atoms constitute things, then atoms constitute things'. Both Philo's and Diodorus' criteria would license it as true: the antecedent does not and can never hold true, the consequent does and can always hold true. By the Chrysippean criterion, however, the conditional is unsound. The contrary of the second, 'atoms constitute things', is 'not: atoms constitute things', which is identical to the first and hence fails the requirement of contradiction. In contrast, the conditional 'if it is day, it is day' is Chrysippean-true, as the opposite of the second reads 'not: it is day', which does contradict the first.

The criterion entails that a conditional relationship mirrors a real world connection (*synartesis*) between the first state of affairs and the second state of affairs. The rule rejects connections between inconsistent terms, such as 'if Socrates is a triangle, then Socrates is a philosopher', that the Philonian would license instead, and from which the Diodorean is not immune, as is could happen with the assertion 'if atoms constitute things, then Socrates is a philosopher'.

We find the rationale of the Chrysippean rule in the modern counterfactual or subjunctive conditional<sup>17</sup>.

Last, fragments report a position generally ascribed to those Stoics who decide on truth based on the content of the statement – in this, closer to the essentialism in the Peripatetic positions. To them, the “duplicated” conditional 'if it is day, it is day' is in itself unsound because conditionality requires that the first contains the second, and nothing can be contained in itself. In the shape of the axiom of reflexivity, this implication is instead accepted and proven in many, although not all, of the modern “possible world semantics”<sup>18</sup>.

These different definitions of conditionality prove especially consequential as they define the boundaries within which complex statements can be treated as formally equivalent, and inferences can be drawn.

---

<sup>17</sup> e.g., Goodman 1947, 1955.

<sup>18</sup> Cresswell 2017; Stakelum, 1940.

### 2.3.4 Formal equivalences

Conjunction, disjunction, and conditionality suggest that simple statements can entertain three kinds of relationships with each other, corresponding to as many types of relationships in the world<sup>19</sup>. They can be mutually incompatible, partially compatible, or fully compatible.

*Incompatibility* entails simultaneous coexistence is as impossible as simultaneous absence. Captured by the exclusive disjunction, it recalls the relationship between Aristotelean contradictory opposites – a universal and a particular of different quality. *Consequence* rises among those states that always coexist. It is captured by the conjunction and conveys that simultaneous presence is as necessary as simultaneous absence – relative to the time constraints to which the definition is committed. Then, incompatibility and consequence are one the mirror image of the other. *Partial compatibility*, instead, is captured by the inclusive disjunction; it conveys a weaker if not an accidental conditionality and is similar to the relationship between Aristotelean subcontraries.

An enduring feature of Stoic propositional logic is that different complexes of sentences can convey the same relationship among states of affairs. Statements compounded by conjunction, disjunction, and conditional, then, become *inter-definable* following rules of formal equivalence.

Many sources<sup>20</sup> report that Chrysippus applied the property of inter-definability to conditionals and oracles. These fragments allow inferring that he recommended oracles such as

‘if anyone is born under the Dog Star,  
then he will not drown in the sea’ [0.0s]

be rephrased as pseudo-disjunctions, as under any stricter interpretation, even a single drowning sailor born under the Dog Star would have proven the oracle false. Seemingly, the argument went as follows. He claimed the conditional could be restated as

‘not both: born under the Dog Star,  
and drowning in the sea’ [0.1s]

which, in turn, can be rendered as

---

<sup>19</sup> Stakelum, 1940:48-53.

<sup>20</sup> Long and Sedley, 2012:232.

‘born under the Dog Star,  
or not drowning in the sea’

[0.2s]

If the latter is understood as a pseudo-disjunction, anybody will make it true that meets at least one of the two conditions. The weak disjunction does not establish compelling connections of consequentality between two sentences. Hence, it can preserve the assumption that the oracles are true without being necessarily-true – which leaves room for free will, human responsibility, and open futures.

Inter-definability was later formalized as *laws of complementarity*<sup>21</sup>. The laws developed from the principles that

- [C1] the negation of a negation affirms,
- [C2] the negation of a conjunction is a disjunction and *vice versa*.

The two principles reduce the essential axioms required to develop a logical language to three – namely, identity, negation, and either conjunction or disjunction<sup>22</sup>. Whitehead and Russell, in laying the foundation of modern propositional logic, accept inter-definability as a consequence of their *axiom of reducibility*. The axiom commits to the assumption that propositions can be reduced to truth functions, although on pain of a loss in modality, intension, and meaning. About such a costly trade-off, they noted that

This axiom has a purely pragmatic justification: it leads to the desired results, and to no others. But clearly it is not the sort of axiom with which we can rest content. On this subject, however, it cannot be said that a satisfactory solution is as yet obtainable<sup>23</sup>.

As an illustration of the working of the reduction, let us adopt the next convention in notation, as follows.

- ( $\alpha$ ) Let us indicate the connective of the negation with the logical ‘not’ indicated either as an overbar or a curl so that  $\bar{P}$  means the same as  $\sim P$  and reads ‘it is not the case that  $P$ ’, or ‘not  $P$ ’ for short. Extensionally, the negation of the sentence  $P$  can be depicted as the shadowed area of the Venn diagram ( $\alpha$ ) in Figure 4.

---

<sup>21</sup> De Morgan, 1847.

<sup>22</sup> Tarski, 1956; Quine, 1960, 1982.

<sup>23</sup> Whitehead and Russell, 1927:xiv.

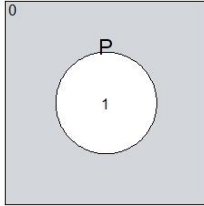
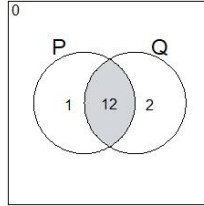
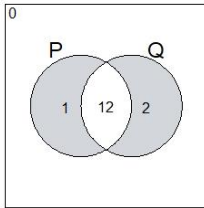
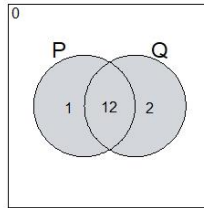
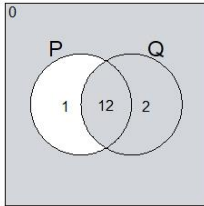
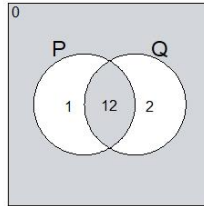
- (β) Let us indicate the connective of the conjunction with the logical 'and', indicated by a wedge. So,  $P \wedge Q$  reads 'it is the case that  $P$  and it is the case that  $Q$ ' or, for short, ' $P$  and  $Q$ '.  
Extensionally, the compound of  $P$  and  $Q$  can be depicted as the shadowed area of the Venn diagram (β) in Figure 4.
- (χ) Let us indicate the connective of the disjunction with the logical 'or', indicated by a vee, so that  $P \vee Q$  reads, 'it is the case that  $P$  or it is the case that  $Q$ ' or, for short, ' $P$  or  $Q$ '.  
It is worth noting that, unless otherwise stipulated, the vee is for the inclusive disjunction: hence, it covers anything that is  $P$  yet not  $Q$ , plus anything that is  $Q$  yet not  $P$ , plus anything that is both  $P$  and  $Q$ . The inclusive disjunction entails the more demanding exclusive disjunction, as the latter covers the same kinds of objects except for those which are both  $P$  and  $Q$ .  
Extensionally, the holding of the exclusive disjunction corresponds to the shadowed area of the Venn diagram (χ.1) in Figure 4; that of the inclusive disjunction, to the shaded area of the Venn diagram (χ.2) in the same Figure.
- (δ) Let us indicate the connective of the Philonean or material conditional with the horseshoe open toward the antecedent, so that  $P \supset Q$  reads 'if it is the case that  $P$ , then it is the case that  $Q$ ' or, for short, ' $Q$ , if  $P$ '.  
Extensionally, the Philonean conditional holds in the shadowed area of the Venn diagram (δ) in Figure 4.

Now, we can read  $P$  as the statement ' $x$  is born under the Dog Star',  $Q$  as the statement ' $x$  will not drown in the sea', and, in each,  $x$  as the post-holder for the same subject – understood either as a subject of the same type, or the same individual. As both the statements in the oracle structure refer to the same subject, however, we can drop the reference to  $x$  in the following. Given these stipulations, we can render the oracle statement [0.0s] as the formula [0.0] below:

$$P \supset Q \quad [0.0]$$

The suggestion to rephrase it as [0.1s] yields [0.1]:

$$\sim (P \wedge \sim Q) \quad [0.1]$$

**Figure 4. Venn's representation of complex sentences***(α) not P**(β) P and Q**(χ.1) either P or Q**(χ.2) P or Q**(δ) if P, Q**(ε) iff P, Q*

*Keys:* The sentence  $P$  holds in the areas marked as 1; else, its negation,  $\sim P$ , holds instead. The sentence  $Q$  holds in the areas marked as 2; else, its negation,  $\sim Q$ , holds instead. In the area marked as 0, only negations hold. In each diagram, the compound stated in the subheading is true in any shadowed area.

The formula reads that it is not the case – it cannot be – that the sentence  $P$  be true and  $Q$  be false. It requires that  $Q$  certainly occurs under  $P$  and that, in the absence of  $P$ ,  $Q$  does not happen. In short, the two propositions are true together or false together. This means the same as the disjunction in [0.2]



$$\sim P \vee Q$$

[0.2]

[0.2] reads that either  $P$  is not, or  $Q$  happens. The slippage in meaning occurs as [0.2] draws attention to anything that is  $\sim P$ , regardless of whether it is  $Q$  or  $\sim Q$ , and to anything that is  $Q$ , regardless of it being  $P$ . The equivalence relaxes the requirement of being  $P$  for having  $Q$ . This relaxation will later be deplored as a source of paradoxical assumptions about the probative import of red herrings, or white shoes, to provide evidence that all ravens are black<sup>24</sup>.

In short, the transformation of [0.0] into [0.1] is generally accepted across logical languages and entails the counterfactual reading of the conditional. Instead, the equivalence of [0.0] as [0.2] is rejected by those understandings that convey a difference between a true relationship of causation, such as ‘if it is heated, then it will burn’, and conversational compositions such as ‘if you want cookies, then you’ll find some in that box’ – both of which fit the material horseshoe, and bear its truth function.

The departure of these logics from the truth-functional definition of the conditional is often signaled by the introduction of different operators for conditionality, such as Lewis’ strict analytic implication and the corresponding “real connection” ( $\rightarrow$ ) reading ‘if, then in result’<sup>25</sup>.

## 2.4 Schemata

The issue of the inter-definability of connectives is especially consequential, as it opens to inference by propositional calculus. The operation consists in assigning a truth value to a component of a compound once we know the truth value of the others and the basic rules for calculating the truth values. The Stoic *schemata* provide the prototypical structures that display the application of these rules for valid inferences. They were deemed “*indemonstrable*” as axioms that “have no need of demonstration”<sup>26</sup>.

Here again, in each schema, the argument is structured into at least three sentences: the first stating the rule that governs the relationship among simple terms, the second making the case about one of the terms, and a third drawing the conclusion. Fragments report that at least five

<sup>24</sup> Good, 1967; Maher, 1999; *cfr* Hempel, 1965.

<sup>25</sup> Lewis 1946:184, 248; Lewis and Langford, 1959; Quine, 1960.

<sup>26</sup> Bett, 2005:132.

of these indemonstrable schemata were in use in the Stoic canon of logic.

### 2.4.1 The first indemonstrable

Later known as the *modus ponendo ponens* or *method of affirming*, the schema reads as in [1.a] below:

[1.a]

If the first, the second.  
 But: the first.  
 Then: the second.

The usual illustration runs as follows: “If it is day, it is light. But: it is day. Then: it is light”. Given the notation ( $\alpha$ ) to ( $\delta$ ), the first indemonstrable can be formalized as in [1] below:

$$\begin{array}{r} P \supset Q \\ P \\ \therefore Q \end{array} \quad [1]$$

the symbol  $\therefore$  reading ‘therefore’ and introducing a logical conclusion.

The formalization clarifies the structure of the argument. The conditional  $P \supset Q$  plays as the major premiss, establishing a general rule. The argument, then, makes the case that  $P$ . The conclusion follows as the inference of the truth value of the unknown component, given the truth function of the conditional and the knowledge from the case.

Notwithstanding the Stoic labels for the schemes, truth-functional logic can prove the first indemonstrable is a formally valid inference by appealing to inter-definability. The proof runs as follows. The conditional in [0.0]  $P \supset Q$  can be restated as the disjunction [0.1]  $\sim(P \wedge \sim Q)$ . The first indemonstrable allows us assuming the substitutability of  $Q$  with  $P$ : hence, we can rewrite [0.1] as [0.1a]  $\sim(P \wedge \sim P)$ . The new conjunction, then, denies the non-simple sentence  $P \wedge \sim P$ . However, the conjunction of a proposition and the negation of itself is a contradiction – that is, a compound that contains every quality and, hence, is a latent construct without corresponding observed instances<sup>27</sup>. Its negation, thus, portrays the principle of noncontradiction, to which classical logic is committed as it preserves the axiom of the identity of the subject. When developed, it yields [0.2a]  $\sim P \vee P$ , which renders the same

---

<sup>27</sup> Wittgenstein, 1922: 4.463.

as the principle of the excluded middle. [0.2a] contains all the possible truth-values of  $P$  at once; hence, it is trivially true ahead of any actualization, as it can never be false. Such a transformation of a true conditional into a tautology proves the first schema is formally valid<sup>28</sup>. Nevertheless, as Wittgenstein himself had to conclude, this proof says nothing about the actual grasp that the conditional has of the real world.

Saying something about the world is the particular task of the practical application of logic as the methodological blueprint of modern research<sup>29</sup>. From this perspective, the *modus ponens* clarifies the *empirical expectation* about the relationship between  $P$  and  $Q$ . The schema states that, whenever we observe  $P$ , we can expect  $Q$ . Hence, we can prove the conditional true by *confirmation* if we select instances displaying  $P$  and find them non-accidentally associated with  $Q$ <sup>30</sup>.

#### 2.4.2 The second indemonstrable

Later known as *modus tollendo tollens* or *method of denying*, the second *schema* stipulates that, as in [2.a]:

If the first, the second. [2.a]  
 But: not the second.  
 Then: not the first.

The typical Stoic example that illustrates its usage reads: 'if it is day, it is light. But: it is not light. Then: it is not day'. The second indemonstrable, hence, can be formalized as in [2] below:

$$\begin{array}{l} P \supset Q \\ \sim Q \\ \therefore \sim P \end{array} \quad [2]$$

The argument again considers that a conditional relationship exists between  $P$  and  $Q$ , but then it makes the case that the consequent is false:  $\sim Q$ . The conclusion infers that, therefore, the antecedent in the conditional is false, too: it is not the case that  $P$ .

---

<sup>28</sup> Wittgenstein, 1922: 4.46, 6.1.

<sup>29</sup> e.g., Mill, 1843.

<sup>30</sup> e.g., King *et al.* 1994:137; Carnap, 1949.

From the truth-functional perspective, the logical validity of this argument, too, can be demonstrated. The schema establishes the substitutability of  $\sim Q$  with  $\sim P$ . Applied to [0.1]  $\sim(P \wedge \sim Q)$ , the substitution again yields the relationship [0.1 a]  $\sim(P \wedge \sim P)$ , whose development is [0.2a]  $\sim P \vee P$ , which again proves the argument logically valid although factually void.

From the pragmatic viewpoint of methods, the *modus tollendo tollens* is the contrapositive of the *modus ponendo ponens* – that is, its negation. As such, it provides the terms under which the empirical import of a conditional hypothesis can be tested by *falsification*. It entails a conditional hypothesis holds unless instances are observed that contradict it. The strategy proves especially useful when the universe of reference is infinite, and the confirmation from the *modus ponendo ponens* cannot prove conclusive<sup>31</sup>.

### 2.4.3 Further schemata

The remaining indemonstrables all develop from exclusive alternatives established as conjunctions or disjunctions. They again will preserve their value over time as the blueprint of inference, as especially Mill' canons show.

The *third indemonstrable*, known as *modus ponendo tollens*, concludes the opposite of a conjunct from the negation of a conjunction of two. Fragments illustrate it as in [3.a]:

Not: the first and the second. [3.a]  
 But: the first.  
 Then: not the second.

The schema generalizes the example: 'not: it is day, and it is night. But it is day. Then, it is not night' and can be formalized as in [3] below:

$$\begin{array}{l} \sim(P \wedge Q) \\ P \\ \therefore \sim Q \end{array} \quad [3]$$

The argument renders an exclusive alternative as the situation in which  $P$  and  $Q$  cannot be jointly given.

---

<sup>31</sup> Mill, 1843; Popper, 1937; Hempel, 1965.

The same is established by the *fourth indemonstrable*. From an exclusive disjunction and one of the disjuncts, the argument concludes the other disjunct – as in [4.a]:

Either the first or the second. [4.a]  
 But: the first.  
 Then: not the second.

The illustrative example states ‘either it is day, or it is night. But it is day. Therefore, it is not night’. The formalization follows the structure in [4]:

$$\begin{array}{l} P \vee Q \\ P \\ \therefore \sim Q \end{array} \quad [4]$$

The third and fourth indemonstrables conclude the same ( $P$ ) after observing the same ( $\sim Q$ ) despite the premiss being different. This licenses the conclusion that the negation of the conjunction and the exclusive disjunction of two terms convey the same underlying relationship between the terms.

The last indemonstrable of sure Stoic origin will be later known as the *modus tollendo ponens*, to indicate that the schema affirms by denying. From an exclusive disjunction, the argument allows inferring the one disjunct from the negation of the other. Fragments portray it as in [5.a] below:

Either the first or the second. [5.a]  
 But: not the second.  
 Then: the first.

often illustrated as ‘either it is day or it is night. But it is not night. Therefore, it is day’. The structure can be formalized as in [5] below:

$$\begin{array}{l} P \vee Q \\ \sim Q \\ \therefore P \end{array} \quad [5]$$

The structure abstracts the schema of the chasing dog argument that Chrisyppus reportedly used to support his definition of the exclusive disjunction.

The *schemata*, then, can be understood as the operational definitions of the connectives employed in the Stoic logic. Thus, they are extensions of the relationships in categorical logic, as in [A.1] to [A.4]. The proper Stoic syllogistic reasoning, however, can be found in the analysis of complex arguments.

## 2.5 Themata

The Stoic “analysis,” much like modern analysis, consists of reducing complex arguments to tautologies. The five indemonstrables were routinely applied to transform the initial argument to the disjunction of a term and its negation. Their application was codified into specific procedures, called *themata*, and structures, called *theorema*.

Only two reliable examples of Stoic analysis have reached us. The first is found in Sextus Empiricus<sup>32</sup>, who reports the following:

If the first and the second, the third. [6.a]  
 Not the third,  
 but the first.  
 Then: not the second.

The unfolding of the analysis can be formalized as in [6] below:

$$\begin{array}{l}
 (P \wedge Q) \supset R \\
 \sim R \\
 \therefore \sim(P \wedge Q) \\
 P \\
 \therefore \sim Q
 \end{array}
 \quad [6]$$

The analysis of the argument clarifies that, whenever  $R$  follows from  $P$  and  $Q$ , but not- $R$  is observed, the conjunction shall be false; this, however, only can be when at least one of the two conjuncts is false –

---

<sup>32</sup> Bett, 2005: 135-136; Mates, 1961:77.

no matter if  $P$  or  $Q$ . In the illustration,  $P$  is observed: therefore,  $Q$  must be false for the argument to hold.

The other recognized example allows maintaining that if a third proposition is deduced from two conjuncts, and one conjunct follows from a fourth proposition, then the fourth proposition with the other conjunct yields the conclusion. The *thema* runs as in [7.a]:

If the first and the second, the third. [7.a]  
 If the fourth, the second.  
 But: the first and the fourth.  
 Then: the third.

The unfolding of the analysis can be formalized as in [7] below:

$$\begin{array}{l} (P \wedge Q) \supset R \\ S \supset Q \\ \therefore (P \wedge S) \supset R \\ P \wedge S \\ \therefore R \end{array} \quad [7]$$

The procedure justifies substituting a proposition with some equivalent statement until it dissolves<sup>33</sup>.

## 2.6 Strengths and limits of schemata

Arguments that unfold along valid *schemata* and *themata* were deemed *valid* and *conclusive* in the formal sense.

Valid patterns could not ensure the arguments were meaningful, however. Meaningful or *demonstrative* arguments are valid structures that, moreover, cast light on an unclear consequence. Thus, 'if it is day, then it is day, but it is day; therefore, it is day' is conclusive according to many criteria, despite its poor content. In contrast, arguments of the kind: 'if one has a healed wound, then he has a scar, but he has no scars; therefore, he has no healed wounds' are conclusive, valid, and demonstrative at once.

---

<sup>33</sup> The procedure is now performed in proof theory, for instance, through semantic tableaux: e.g., Horrocks, 1996; Carnielli, 1987.

Nevertheless, the formal validity of *schemata* and *themata* can at least ensure that arguments escape four major structural faults – namely, disconnectedness, redundancy, bad patterns, and deficiency.

### 2.6.1 *Disconnected arguments*

*Disconnectedness* arises when a conclusion is drawn, but the premisses are unrelated to one another or the conclusion. ‘If it is day, it is light. But wheat is sold in the market. Therefore, Dion is walking’ illustrates a rule, a case, and a conclusion that share no ground, and that the indemonstrables can dismiss as invalid. Indeed, a schema reading  $P \supset Q, S, \therefore R$  provides no element for decision.

### 2.6.2 *Redundant arguments*

*Redundancy* is disconnectedness of a particular kind. It arises when superfluous terms are inserted in the argument. The usual example elaborates on the indemonstrables with the conditional and includes three terms: an antecedent  $P$ , such as ‘it is day’; a consequent  $Q$ , such as ‘it is light’; and the redundant insertion  $R$ , such as ‘*Dion is walking*’.

When the redundant term is added to the case in a schema based on the first indemonstrable, we have

$$\begin{array}{l} P \supset Q \\ P \wedge R \\ \therefore Q \end{array} \quad [1.b]$$

reading ‘If it is day, it is light, but it is day *and Dion is walking*, therefore it is light’. From a formal perspective, the schema is unaffected by the insertion. The conditional in the argument does not provide us with any expectation about the contribution that walking Dion brings to the relationship between day and light. Thus, the validity of the argument does not depend on the truth value of Dion’s state. The pertinent antecedent is observed as true, and true is the conclusion, too. The redundancy of the insertion, moreover, is evident as the conclusion is obtained regardless of the state of the superfluous conjunct.

Validity is preserved even when the redundant term is added to the antecedent of the conditional, which makes the insertion relevant to the argument instead. A modern version of the riddle<sup>34</sup> portrays a salt dissolving ( $Q$ ) if put in water ( $P$ ) *after casting a spell* ( $R$ ). The antecedent

---

<sup>34</sup> Kyburg, 1965; Salmon, 1984; Fetzer, 2000.



requires that the case is made that the whole antecedent be true, as in [1.c]

$$\begin{array}{l} (P \wedge R) \supset Q \\ P \wedge R \\ \therefore Q \end{array} \quad [1.c]$$

The argument reads: ‘if one puts salt in water *and casts a spell*, then the salt dissolves; but it is the case that one puts salt in water *and casts a spell* – therefore, the salt dissolves’. The argument is valid as the conclusion follows. However, the argument is valid despite the redundant term makes the premiss substantively false as a false premiss does not affect the truth of conclusions.

That this schema draws the right conclusions on the wrong ground, however, arises a concern. It shows that the first indemonstrable – hence the *modus ponens*, or any pure strategy of confirmation – cannot discriminate between a real conjunct and a redundant insertion in the antecedent of a conditional. In short, as especially Mill would later underline, when the antecedent is a conjunction, confirmation cannot properly support causal claims – and leaves the explanation on shaky ground.

### 2.6.3 Fallacies

The next structural faults, *bad schemata*, are better known as *fallacies*. They occur when the argument violates the closer valid schema. The known Stoic examples refer to the first and the second indemonstrable.

Recall the pattern of the second indemonstrable maintains that [2]  $P \supset Q, \sim Q, \therefore \sim P$ . The related illicit schema, later known as “denying the antecedent,” unfolds as in [2.f] below:

$$\begin{array}{l} P \supset Q \\ \sim P \\ \therefore \sim Q \end{array} \quad [2.f]$$

The fallacy rules out as invalid those inferences that lead to the negation of the consequent, having observed the negation of the antecedent. In other terms, it considers illicit reasoning that ‘if it is day, it is light; but it is not day; therefore, it is not light’. The rejection of this

pattern of inference as fallacious preserves the truth function of the material conditional. From the substantive perspective of methods, moreover, the rejection of the inference as fallacious is justified by a pragmatic consideration: otherwise, we would be presuming, yet not proving, that the day is the sure and irreplaceable antecedent of light.

The other violation challenges the first indemonstrable, with an illicit schema dubbed “affirming the consequent.” The schema unfolds as in [1.f] below:

$$\begin{array}{l} P \supset Q \\ Q \\ \therefore P \end{array} \quad [1.f]$$

The conclusion is faulty as we cannot conclude the antecedent from the hypothesis of a conditional and the case of the consequent. ‘If it is day it is light; but it is light; therefore, it is day’ does not qualify as valid: the light about which the case is made may come from other sources. The schema is illicit as, again, it stipulates yet does not prove the assumption that  $P$  is the certain and irreplaceable antecedent of  $Q$ .

The inference from the consequent to the antecedent and from the negation of the antecedent to the negation of the consequent becomes licit when the relationship between  $P$  and  $Q$  is biconditional. The relationship is signaled by

- ( $\varepsilon$ ) the *iff* connective. Indicated by a double-headed arrow  $\Leftrightarrow$  or with the triple bar  $\equiv$ , it entails that the antecedent implies the consequent and *vice versa*:  $(P \supset Q) \wedge (Q \supset P)$ . Hence,  $P \equiv Q$  reads ‘if and only if  $P$ ,  $Q$ ’. Extensionally, it holds in the shadowed area of the Venn diagram ( $\varepsilon$ ) in Figure 4 – which also shows it be the complement of the exclusive disjunction ( $\chi.1$ ).

Biconditionals, in short, establish *identities* between the antecedent and the consequent. This symmetry allows for commutations and substitutions that widen the interdefinability of sentences and ease the inference. True biconditionals, however, may prove circular unless they draw upon the identity of an object and its qualifying parts. Thus, both arguments: ‘if and only if it has three straight sides, it is a triangle. But it has not three straight sides. Therefore, it is not a triangle’, and: ‘if,

and only if it has three straight sides, it is a triangle. But it is a triangle. Therefore, it has three straight sides' hold and are demonstrative. 'If, and only if it is day, it is day', instead, is formally valid yet not demonstrative at all.

#### 2.6.4 *Incomplete arguments*

The last source of inconclusive arguments lies in *deficiency* – what we can now identify as a problem of omitted variables or ill-structured problems. In the original fragments, the issue is illustrated with the fifth indemonstrable; however, it follows from the violation of exhaustiveness and, hence, it applies to any argument based on disjunction. Sextus Empiricus, seldom generous with Stoic logic, endorses the principle that the conclusion from an incomplete list of alternatives is flawed. He, for instance, rejects the argument that:

'Either wealth is a bad thing or wealth is a good thing; but wealth is not a bad thing; therefore wealth is a good thing'. For wealth's being an indifferent thing is missing in the disjunction, so that the sound version is, rather, like this: 'Either wealth is a good thing or a bad thing or an indifferent thing; but wealth is neither a good thing nor a bad thing; therefore it is an indifferent thing'<sup>35</sup>.

As in the case of redundancies, deficient arguments show that inferences are sensitive to the structure of the content of their "major premiss" – especially when the deficiency is embedded in a logically complex theoremata.

Themata, schemata, and predicate calculus provide an inventory of the logical problems and the solutions that will shape the modern philosophy of science and methods. In the passage to the modern discourse on explanation, the primary issue has become the tenability of the major premiss and the criteria to establish it. The issue will prove consequential on the whole explanatory machinery.

---

<sup>35</sup> Bett, 2005: 174.

### 3. Modern explanation

Over time, the original works of the Stoic school went almost lost, while the Peripatetic doctrine was employed to support dogmatic positions about the necessity of being. Under the pressure of political and religious censorship, Aristotelianism institutionalized as the “received view”<sup>1</sup>. In 1787 Kant could famously declare that, since Aristotle, logic had “been unable to take a single step forward, and therefore” seemed “to all appearance to be finished and complete” – which also read as a verdict of ossification<sup>2</sup>.

The Aristotelean teaching fell into overt disrepute, indicted for a tendency to cultivate *inutilis subtilitas*. The fault was diagnosed in its reliance on ill-defined syllogistic premisses, from which science had to be freed:

The syllogism consists of propositions, propositions consist of words, and words are counters for notions. Hence if the notions themselves (this is the basis of the matter) are confused and abstracted from things without care, there is nothing sound in what is built on them. The only hope is true induction<sup>3</sup>.

Supported by the progress of mathematics and statistics, the new “philosophy of nature” embraced the idea of the world as a self-moving mechanism and confined inferences on the empirical terrain. The primary motivation became pragmatic, and efforts narrowed down on how to establish notions and propositions to “control and regulate future events by their causes”<sup>4</sup>. The principal test bench of induction became experimental. The original structures of explanation, however, survived

---

<sup>1</sup> Israel, 2001; Grant, 1978.

<sup>2</sup> Guyer and Wood, 1998:106 B viii; Kneale and Kneale, 1962.

<sup>3</sup> Bacon, 1620 [2000]: I, xiv.

<sup>4</sup> Hume, 1748, section vii; Bacon, 1620; Wiener, 1932; Rossi, 1996.

in modern scientific practices as the rationale to establish the tenability of causal claims.

### 3.1 Sitting the major premisses on the empirical ground

The modern detachment of the scientific discourse from metaphysical assumptions is often associated with Hume and his explicit rejection of the ultimate causes of being as a proper scientific concern.

Borrowing from Skeptical arguments, Hume influentially conceded such ultimate causes may exist, yet contended their knowledge was accessible to the human mind. Popper<sup>5</sup> would later dub “the doctrine of the psychological primacy of repetition” Hume’s tenet that the mind can only learn from the direct experience of particular occurrences and incrementally, from the cumulation of evidence. Indeed, the cornerstone of the doctrine is the tenet that a single unprecedented occurrence is not sufficient to generate knowledge. Single unusual events only compel hypotheses about their generation, as

every effect is distinct from its cause. It could not, therefore, be discovered in the cause; and the first invention or conception of it, *a priori*, must be entirely arbitrary. And even after it is suggested, the conjunction of it with the cause must appear equally arbitrary; since there are always many other effects, which, to reason, must seem fully as consistent and natural<sup>6</sup>.

In this doctrine, knowledge of the world arises after multiple observations have induced the idea into the mind that “species” of similar objects are related. Hempel maintained that the human mind recognizes three fundamental relationships as meaningful: *Resemblance*, *Contiguity*, and *Causation*. He dismissed *Contrast* or *Contrariety* from the set of the primary relationships, as it could be derived from the combination of causation and resemblance: “where two objects are contrary, the one destroys the other; [...] and the idea of the annihilation of an object implies the idea of its former existence”<sup>7</sup>.

---

<sup>5</sup> Popper, 1959:420.

<sup>6</sup> Hume, 1748, section IV.

<sup>7</sup> Hume, 1748, section IV, note *e*.

It may fall beyond the mind's capacity to establish whether these relationships are genuinely ontological principles that shape the "matters of fact" in the world. Nevertheless, resemblance, contiguity, and causation provide the epistemic principles that our mind applies to arrange events into consistent narratives. As the work of poets and historians clearly shows,

the most usual species of connexion among the different events, which enter into any narrative composition, is that of cause and effect; while the historian traces the series of actions according to their natural order, remounts to their secret springs and principles, and delineates their most remote consequences. [...] And always, he is sensible that the more unbroken the chain is, which he presents to his readers, the more perfect is his production. He sees that the knowledge of causes is not only the most satisfactory, this relation or connexion being the strongest of all others, but also the most instructive; since it is by this knowledge alone we are enabled to control events and govern futurity<sup>8</sup>.

Despite their epistemic quality, Hume believed the fundamental relationships afford faithful images of the world. He reasoned the reliability of these images is "so essential to the subsistence of all human creatures" that Nature must have implanted them as the "instinct" of carrying forward "the thought in a correspondent course to that which she has established among external objects"<sup>9</sup>. Such an instinct – namely, "custom or habit" – does not open us the knowledge of the "hidden powers and forces on which the regular course and succession of objects totally depend." However, it persuades us *believing* that future occurrences will conform to past patterns. Such a belief is different from mere fiction because it ultimately rests on actual experiences and observations – which gives our inferences "a more vivid, lively, forcible, firm, steady conception [...] than what the imagination alone is ever able to attain"<sup>10</sup>.

Given the correspondence of our mind's images to the actual world, and given the mechanism through which our beliefs take their shape, we identify causes in those objects that seem "entirely uniform and constant in producing a particular effect." We know that fire is the cause of burning, water of suffocation, and impulse of motion by the regularity

---

<sup>8</sup> Hume, 1748, section IV, note 3.

<sup>9</sup> Hume, 1748, section V.

<sup>10</sup> Hume, 1748, section V.

of their joint occurrence. Causes are those which arise “universal laws” of association and admit “of no exception.” These constant conjunctions properly support the supposition that some necessary connection exists. Necessary connections, in turn, arise when one “*object Cause*” exerts its power to produce the other “*object Effect*” so that

all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, if the first object had not been, the second had never existed<sup>11</sup>.

Hume defines causation as a generative connection that we can assume after observing a regular association between an objects-cause and an objects-effect. To characterize its shape, he resorts to two well known logical tenets, as the formalization of his definition of necessary connection shows.

Let us indicate the class of the objects-cause with  $P$  and the class of objects-effect with  $Q$ . The stance ‘all the objects, similar to the first, followed by objects similar to the second’ then takes the shape of the Stoic first indemonstrable [1]  $P \supset Q, P, \therefore Q$ . In Humean terms, it says that the implication qualifies as causal when  $P$ -type occurrences regularly – certainly – anticipate  $Q$ -type occurrences. The rephrasing ‘if the first object had not been, the second never had existed’, instead, embodies the second indemonstrable as the counterproof that the association of  $P$  and  $Q$  is causal. However, Hume invokes it in its fallacious pattern [2.f]  $P \supset Q, \sim P, \therefore \sim Q$ . The pattern licenses the expectation of no  $Q$ -type occurrences, based on the observation of no  $P$ -type occurrences. In itself, this expectation can expose us to consequential surprises<sup>12</sup> – unless  $P$  is proven the unique type of cause of  $Q$ . Nevertheless, the clause establishes that Hume’s idea of a necessary connection is, again, a biconditional relationship.

Evidence of a biconditional relationship, hence, can convince us that objects-cause and objects-effect are *necessarily* connected by some fundamental power. Although we cannot know whether such a relationship embodies any ontological necessity, this evidence satisfies our *idea* of necessity: and this is enough to give our expectations the character of a firm belief. However, our mind also tends to develop expectations from “irregular and uncertain” observations. In this regard, Hume’s philosophy becomes prescriptive.

---

<sup>11</sup> Hume, 1748, section VII.

<sup>12</sup> Nassim, 2008.

He concedes that irregular evidence licenses a probabilistic expectation about the most frequent “effect” given an intermittent “cause.” However, he warns the reader against dismissing alternative effects as irrelevant.

Though we give the preference to that which has been found most usual, and believe that this effect will exist, we must not overlook other effects, but must assign to each of them a particular weight and authority in proportion as we have found it to be more or less frequent<sup>13</sup>.

Expectations of irregular phenomena based on frequencies come with further considerations. Irregularities are not evenly distributed; instead, they change with *circumstances*. The prospect of frost is high in January “in almost every country of Europe,” although it “approaches to a certainty in the more northern kingdoms”<sup>14</sup>. Similarly affected by local circumstances are the uniformities and regularities about human behavior: “Even the characters which are peculiar to each individual have a uniformity in their influence”<sup>15</sup>. Irregular connections, he concludes, are less the actual manifestation of chance than of the limits of our mind.

To Hume, too, the understanding of irregularities as the limits of our knowledge marks the distance between the “philosophical” ascription of causation and the everyday ascription by habit:

philosophers, observing that almost in every part of nature there is contained a vast variety of springs and principles, which are hid by reason of their minuteness or remoteness, find that it is at least possible the contrariety of events may not proceed from contingency in the cause, but from the secret operation of contrary causes. This possibility is converted into certainty by farther observation, when they remark that, upon an exact scrutiny, a contrariety of effects always betrays a contrariety of causes and proceeds from their mutual opposition. [...] Thus, for instance, in the human body, when the usual symptoms of health or sickness disappoint our expectation; when medicines operate not with their wonted powers; when irregular events follow from any particular cause: the philosopher and the physician are not surprised at the matter, not are ever tempted to deny, in

---

<sup>13</sup> Hume, 1748, section VI.

<sup>14</sup> Hume, 1748, section VI.

<sup>15</sup> Hume, 1748, section VIII.



general, the necessity and uniformity of those principles by which the animal economy is conducted. They know that a human body is a mighty complicated machine: That many secret powers lurk in it, which are altogether beyond our comprehension [...] And that therefore the irregular events, which outwardly discover themselves, can be no proof that the laws of Nature are not observed with the greatest regularity in its internal operations and government [...] though not easily discoverable by human sagacity and inquiry<sup>16</sup>.

Despite his pessimism in the human capacity to grasp ultimate causal knowledge, Hume maintains that Nature is uniform: any occurrence has a cause, and the same cause under the same circumstances always yields the same effect. What we may mistake for the failure of a regular pattern, then, signals the local interaction among causes of different signs. In short, the Humean philosophy, too, underwrites the long-honored position that actual causation is true although complex and hard to grasp.

### 3.2 Unraveling complex causation

Similar recognition of causal complexity animates the influential work that John Stuart Mill wrote during the 1830s and first published in 1843. In it, the recognition marries with a more optimistic view on the capacity of the human mind to make sense of it – at least, to make enough sense to serve practical purposes.

The *System of Logic, Ratiocinative and Inductive* again motivates from the concern that classical logic had moved out of tune with scientific practices. The *Logic* still concedes that the deductive and syllogistic reasoning provides proper devices to unravel complex causation. However, it emphasizes that deduction only becomes possible after establishing credible statements. The major problem of both modern philosophy and science, then, is to determine such statements without appealing to metaphysical entities. Hence, the *Logic* posits induction as the leading problem that affects deduction and identifies its solution in methods for organizing experience into credible evidence of causation.

From the successful practices of sciences such as physics and chemistry, the *Logic* borrows the idea that the assumption of the uniformity of nature can be deployed to identify compelling evidence of causation.

---

<sup>16</sup> Hume, 1748, section VIII.

Such persuasive evidence can be *found*, as in *observational* strategies, or it can be *made*, as in *experimental* strategies. In both observational and experimental inquiries, however, evidence proves compelling if it is collected or generated to fit those *comparisons* that, alone, allow adjudicating on the causal nature of the relationship between two phenomena.

### 3.2.1 Mill's problem of induction, and his solutions

The distinctive contribution of the *Logic* to the scientific discourse lies in the codification of rules and methods to infer general statements from experimental and observational instances. These rules, moreover, are designed to overcome the limits of the consolidated mode of generalization *per enumerationem simplicem* that affect the probabilistic approaches, too<sup>17</sup>.

Simple enumeration generalizes the connection between two phenomena based on the many instances that, to our knowledge, confirm the link. Mill reasons that the strategy “affords in general a precarious and unsafe ground of assurance, for such generalizations are incessantly discovered, on further experience, to be false”<sup>18</sup>. Simple enumeration, in short, yields general statements that are vulnerable to *instantia contradictoria*. Hence, in Mill's system, it may provide the first step toward generalization, seldom the conclusive one. A reliable generalization, Mill notes, is far more formidable than the inventory of positive instances. It must allow us assuming that

what we found true in those instances holds in all similar ones, past, present, and future, however numerous they may be. We then, by that valuable contrivance of language which enables us to speak of many as if they were one, record all that we have observed, together with all that we infer from our observations, in one concise expression; and have thus only one proposition [...] to remember or to communicate.<sup>19</sup>

Mill reasons that learning can occur independently of general statements. Inference always proceeds from the past to the next particular experience, and we do not need the “maxim” that fire burns to escape the next flame once we got a scald. Nevertheless, knowing the maxim

---

<sup>17</sup> Mill, 1843, III, viii, §7.

<sup>18</sup> Mill, 1843, III, xxi, §2.

<sup>19</sup> Mill, 1843, II, iii, §3.

could have spared us the injury. General statements, then, are convenient implements that convey abridged knowledge. Distilling them is worth as it allows knowledge circulating beyond individual experience.

Abridged knowledge takes the form of definitions, axioms, or laws of nature, establishing “a specimen or paradigm of the whole class of cases” of interest<sup>20</sup>. They connect observable *attributes* – such as those that denote a being as human – and those *properties* that instead are hidden to the eye, or that manifest themselves only once in each case, but about which we want to know – such as being mortal. Abridging statements convey the experience that the former are *marks* of the latter – as in ‘any human is mortal’. In such a shape, these statements “vouch for the legitimacy” of inferring the unobserved property of mortality in any other instance of humanity, given the attribute.

In short, the general statement – the “law” – conveys *sufficient knowledge* that the relationship holds between different attributes based on the properties that connect them at some more profound level. The sufficiency of such knowledge depends less on the number of collected instances than on the nature of the observed regularity. The relationship is credible that points to a causal connection in the real world beyond the frequency of its manifestations. The problem of induction, then, is that of the criteria to discriminate evidence of causation from a noisy background.

#### a) *Space and time of causation*

To advance his solution, Mill reduces the causal relationships to two basic types, depending on whether they manifest themselves as simultaneous events or as a succession<sup>21</sup>. *Simultaneity* characterizes causal relationships in the space dimension, and becomes visible as the regular coexistence of phenomena or as the pattern in the arrangement of some “bodies.” *Succession*, instead, connects phenomena and bodies along the temporal dimension and becomes visible as recurrent sequences. The parallel with the Aristotelean causes of stability and change is striking, but secondary in Mill’s general framework. The crucial point in it is, both types of relationships are “coextensive” with the “laws of numbers” – especially with algebraic rules<sup>22</sup>. This ensures their reduction in transmissible formulas.

---

<sup>20</sup> Mill, 1843, II, iii, §3.

<sup>21</sup> Mill, 1843, III, v, §1.

<sup>22</sup> Mill, 1843, III, vi, §1.

Moreover, he contends that a relationship qualifies as causal when it displays both, as the tenet that *cessante causa cessa et effectus* had long established<sup>23</sup>. The endurance of a cause may keep the effect alive, as high temperatures that keep metals molten; however, the effect may also outlast the cause, as a house that survives its building. Thus, Mill concludes, succession remains the only solid test-bench for assessing the causal nature of a relationship. However, succession may or may not be visible as a lapse between the manifestations of the *antecedent* and its *consequent*. In short, Mill's succession renders the intuition that the occurrence of the latter depends on the former – as if the antecedent was prior.

The centrality of *dependence* as the criterion of causation is confirmed by the further consideration that “invariable sequence [...] is not synonymous with causation, unless the sequence, besides being invariable, is unconditional”<sup>24</sup>. The requisite of “unconditionality” is introduced to discriminate actual from mistaken causes in the sequence toward the effect. The night is not the cause of the day, despite their regular succession, as both the day and the night depend on the position of the sun above the horizon – i.e., both manifestations are *conditional on* something else. The lighting power of the sun, in turn, is an unconditional cause of the day and, therefore, its actual antecedent. To prove unconditionality, Mill advances a counterfactual argument: the lighting power of the sun is an unconditional cause because, were it removed, there would be no day at all.

Notably, in Mill's view as in Chrysippus' before, the requisite of unconditionality is satisfied independent on third phenomena impeding the relationship – such as an eclipse. This point allows him preserving the definition of a causal relationship as that which always obtains, as failures are ascribed to meddling third causes. The distinction is justified by the analytical possibility of discerning between those factors that raise a causal relationship and those that *impede* it, as each type contributes to the effect differently. Causes are those occurrences that have to be *present* for the relationship to obtain; just the opposite, impeding occurrences have to be *absent* for effect to happen. Hence, causes enter the abridging formula with a plus sign, while hindering factors can be invoked to account for failures as they subtract from the effect.

---

<sup>23</sup> Mill, 1843, III, v, §7.

<sup>24</sup> Mill, 1843, III, v, §5.

b) *Causes versus conditions*

Even after discounting hindrances, Mill notes, an antecedent can seldom be reduced to a simple single factor. Instead, a relationship of causal succession

is usually between a consequent and the sum of several antecedents; the *concurrence* of all of them being the requisite to *produce*, that is, *to be certain of being followed by, the consequent*. In such a case it is very common to single out one only of the antecedents under the denomination of Cause, calling the other merely Conditions<sup>25</sup>.

Both causes and conditions are required to “produce” the effect. The difference between them, according to Mill, only depends on our perception of their different permanence. We usually indicate as the cause that which we perceive as an undue antecedent change or event – such as the slipping foot when falling from a ladder or the sentinel off his post when the enemy surprised the army. Then, we treat as conditions those antecedents that we understand as more stable *states* – such as, weight and the related gravitational attraction, or military establishments and their operations. The “conditions,” however, are no less indispensable to the effect than the “cause.” Although we may take the conditions for granted to the point of omitting them, in scientific ascription “the statement of the cause is incomplete, unless in some shape or other we introduce them all”<sup>26</sup>.

The need for providing complete statements is especially compelling in the light of further consideration. A condition can be a compound obtained at an earlier stage and arranged so that “in given circumstances certain results will take place”<sup>27</sup>. Although they do not yield any immediate visible effect, these compounds are the product of *spatial* relationships that “predispose the constitution to some mode of action”<sup>28</sup>. Education, training, or the making of gunpowder are all examples of this type of causal factor. Due to the new arrangements of their components, the mind, the body, or the chemical mixture are “endowed with the *property*” of contributing to some effect. The new arrangement, in short, entails a *potential* – that Mill defines “the contingent future fact

---

<sup>25</sup> Mill, 1843, III, v, §3, emphasis added.

<sup>26</sup> Mill, 1843, III, v, §3.

<sup>27</sup> Mill, 1843, III, v, §5.

<sup>28</sup> Mill, 1843, III, v, §5.

brought back under another name”<sup>29</sup>. A complete statement of the cause, hence, has to include any such “constitution of things” capable of generating the effect.

c) *Mechanical and chemical causation*

Mill maintains that the relationship among the components of an antecedent and the ultimate effect can be of two types<sup>30</sup>. The one type embodies a classical *mechanical* rationale: causation consists in the transmission of something by interaction among components, and the joint effect is of the same kind as the effects of each separate element, like in a billiard stroke. The other type embodies a *chemical* rationale: as when oxygen and hydrogen combine in making water, the interaction of the components results in a compound that is qualitatively different from its constituents taken in insulation.

Chemical causation, Mill emphasizes, is harder to grasp than the mechanical one. In mechanical causation, the nature of the components is clear from the effect; we can observe the elements and their interactions at the same time; and the result can be easily reduced to the arithmetic sum of the interactions. In contrast, in chemical causation, the nature of the components can only be discovered by disassembling the compound in its constitutive elements. Either we observe the components by disrupting their interaction, or we observe the effect – and the result is not fully captured by the relationship among the quantities of the elements.

The relationships among quantities, however, are that which laws render to convey experience. Then, given the antecedents, laws of mechanical causation allow calculating the consequent in advance and with good approximation – which Mill equates to deduction as “demonstrative science.” The “heteropathic” relationships of chemical causation, instead, limit predictions to those quantities of the antecedents that also characterize the consequent – for instance, mass, or the ratio of the components. To Mill – who recognized the value of statistical renderings yet was unsympathetic with the formalization of logic that De Morgan was developing at the time – the qualitative facets of the effects of chemical causation defy calculation and prediction.

Either entirely or in some vital facet, nevertheless, both mechanical and chemical causation follow a fundamental principle that Mill calls the *Composition of Causes*. By analogy with the physical principle of

---

<sup>29</sup> Mill, 1843, III, v, §6.

<sup>30</sup> Mill, 1843, III, vi, §1.

the composition of forces, the Composition of Causes stipulates that any consequent is connected to its antecedents by regular relationships that can be known and rendered through algebraic formulas. In short, regularity can be conveniently captured by some empirical rule of composition. Being the rule constant, we can know the causes from their effect. These assumptions, Mill maintains, provide the ground of any induction.

### 3.2.2 *Canons and Methods of induction*

Mill defines induction as an analytical operation, consisting of “the resolution of a complex whole into the component elements” (*ivi*, III, vii, §1). A complex whole first requires that we establish some “mental partition” of the world into “single facts” that can be later systematically observed or reproduced as experiments:

the order of nature, as perceived at first glance, presents at every instant a chaos followed by another chaos. We must decompose each chaos into single facts. We must learn to see in the chaotic antecedent a multitude of distinct antecedents, in the chaotic consequent a multitude of distinct consequents. This, supposing it done, will not of itself tell us on which of the antecedents each consequent is invariably attendant. To determine that point, we must endeavour to effect a separation of the facts from one another, not in our minds only, but in nature. The mental analysis, however, must take place first<sup>31</sup>.

To insulate such relevant “facts” and test their connection in nature, Mill holds we must follow the “rule of *varying the circumstances*”<sup>32</sup>. Both observations and experiments suit the purpose of providing such a variation, although from causal relationships with different structures. To distill knowledge effectively, then, Mill establishes different canons of induction<sup>33</sup>, each corresponding to a method – namely, Agreement, Difference, Indirect Difference, Residues, and Concomitant Variations.

#### *a) The first canon*

The first canon stipulates that

---

<sup>31</sup> Mill, 1843, III, vi, §1.

<sup>32</sup> Mill, 1843, III, vi, §1.

<sup>33</sup> Mill, 1943, III, viii.

if two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or the effect) of the given phenomenon<sup>34</sup>.

To illustrate its point, Mill sketches a formalization that in later notation can be rendered as follows. Let be:

- $A$ , the presence of the candidate antecedents of a phenomenon of interest, and with  $\bar{A}$ , its absence;
- $\Phi$ , other circumstances in a particular state, and with  $\bar{\Phi}$  the same circumstances in a different state;
- $Y$ , the presence of the phenomenon of interest, and with  $\bar{Y}$ , its absence.

The first canon then runs as in expression [8] below:

$$\begin{array}{ll} A \wedge \Phi \wedge Y & \text{Observed in one (set of) instance(s)} \\ A \wedge \bar{\Phi} \wedge Y & \text{Observed in another (set of) instance(s)} \\ \therefore A \circ Y & \text{Conclusion} \end{array} \quad [8]$$

where the symbol  $\circ$  indicates a regular – then, possibly, causal – relationship connecting the term on the left to that on the right.

The related method *of Agreement* applies when we cannot resort to the experimental strategy and have to rely on the instances that nature provides. It stipulates that a meaningful connection exists between that antecedent that does “unvary” with the consequent, *everything else being different*. The conclusion “stands on the ground that whatever *can be eliminated is not connected* with the phenomenon by any law” or, in a word, proves “irrelevant”<sup>35</sup>. The conclusion, moreover, can be reached backward if we compare instances selected on the effect  $Y$ , or forward if we compare cases selected on the antecedent  $A$ .

The *backward* selection narrows the induction on two types of instances:  $A \wedge Y$  and  $\bar{A} \wedge Y$ . Their analysis brings to light whether  $Y$  requires  $A$ . However, the backward selection prevents us from knowing whether  $A$  can also be associated with  $\bar{Y}$  – i.e., whether contradictory instances can arise that weaken the relationship. Hence, the conclusions that we can draw from this set of instances are affected by the biases

---

<sup>34</sup> Mill, 1943, III, viii, §1.

<sup>35</sup> Mill, 1843, III, vii, §3, emphasis added.



allowed by simple enumeration and may convey false certainties about the relationship.

The *forward* selection, instead, narrows on two types of instances:  $A \wedge Y$  and  $A \wedge \bar{Y}$ . Their analysis allows establishing whether  $Y$  follows from  $A$  without contradictory instances. However, based on this evidence, we cannot know whether  $A$  is the unique antecedent of  $Y$ <sup>36</sup>.

Regardless of the selection of evidence, moreover, if the effect  $Y$  depends on the compound antecedent  $A \wedge C$  in one instance and on  $A \wedge D$  in another, the method of Agreement always dismisses  $C$  and  $D$  as varying circumstances of  $\Phi$ , hence irrelevant to its occurrence, and ascribes causation to  $A$  alone. If, instead, the effect depends on simple yet functionally equivalent antecedents – that is, on  $A$  in one instance and on  $B$  in the other, so that the complete factor reads as the exclusive “plural” cause  $A \vee B$  – then the method of Agreement fails to recognize any factor as causal<sup>37</sup>.

b) *The second canon*

This canon again relies on elimination but applies the opposite rationale. The foundation of the conclusion is that whatever *cannot be logically eliminated is not connected* with the phenomenon of interest<sup>38</sup>. Thus, the second canon stipulates the following:

if an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon<sup>39</sup>.

As such, then, the second canon can be summarized as in expression [9]:

$$\begin{array}{ll}
 A \wedge \Phi \wedge Y & \text{Observed in a set of instances} \\
 \bar{A} \wedge \Phi \wedge \bar{Y} & \text{Observed in a set of twin instances} \\
 \therefore A \circ Y & \text{Conclusion}
 \end{array}
 \quad [9]$$

---

<sup>36</sup> Mill, 1843, III, x, §1.

<sup>37</sup> Mill, 1843, III, x, §1.

<sup>38</sup> Mill, 1843, III, viii, §3.

<sup>39</sup> Mill, 1843, III, viii, §2.

The related method, dubbed *of Difference*, maintains that a causal relationship connects two co-varying factors when everything else remains the same. The requisite of similarity in the background circumstances entails that the second canon better applies to identical instances except for the causal attribute under analysis. Although Mill does not require the complete similarity of the instances under comparison – “it needs not extend to such as are already known to be immaterial to the result”<sup>40</sup> – he considers that seldom nature can afford instances suitable to this type of induction. The method better applies to experimental studies where the instances can be made similar in all the relevant attributes except for the one under analysis, whose variation is purposefully induced.

Through the control of the relevant circumstances, experiments also allow dispelling the doubt that the relationship between  $A$  and  $Y$  is conditional on third factors. The claim that  $A$  unconditionally causes  $Y$  stands strong after observing that we can induce or suppress  $Y$  just by compelling or suppressing an insulated  $A$ . The manipulation of the cause through which the method of Difference establishes its conclusions so proves causation as an asymmetric relationship of actual dependence of  $Y$  on  $A$ .

In contrast to the method of Agreement, Mill emphasizes, the issue of the Plurality of Causes does not affect the experimental application of the Method of Difference<sup>41</sup>. Even if the exclusive  $A \vee B$  is the real antecedent of  $Y$ , the manipulation of  $A$  can still be detected by observing the generation and suppression of the effect in the smaller set of instances with the attribute.

However, as the prudential wording of the canon admits, the method of Difference has limits, too. If the true antecedent consists of a conjunction – so that  $(A \wedge C) \circ Y$  – and one of the conjuncts, say  $A$ , is causally required yet empirically associated with both  $Y$  and  $\bar{Y}$  like oxygen in combustion, then the method fails to recognize its contribution – unless it is explicitly hypothesized and tested. Hence, the method of Difference can detect causation from simple equivalent antecedents, yet may be blind to “chemically” relevant factors that we can perceive as background conditions. The problem worsens when chemical causation depends on multiple conjuncts, as their effect only materializes when the components are all joined together<sup>42</sup>.

---

<sup>40</sup> Mill, 1943, III, viii, §3.

<sup>41</sup> Mill, 1843, III, x, §2.

<sup>42</sup> Mill, 1843, III, viii, §4.

c) *The third canon*

Chemical causation arises a methodological challenge: the method of Difference may fail to single out the joint constituents, while the method of Agreement is poorly equipped to yield reliable conclusions from substitutes. Mill then develops the *Indirect method of Difference* to exploit the capacity of both and compensate for their weaknesses<sup>43</sup>.

The method relies on the third canon, maintaining that:

If two or more instances in which the phenomenon occurs have only one circumstance in common, while two or more instances in which it does not occur have nothing in common save the absence of that circumstance; the circumstance in which alone the two sets of instances differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon<sup>44</sup>.

So defined, the canon would read as in [10] below:

$$\begin{array}{ll}
 A \wedge \Phi \wedge Y & \text{in instances of } A \\
 A \wedge \bar{\Phi} \wedge Y & \text{in instances of } A \\
 \bar{A} \wedge \Phi \wedge \bar{Y} & \text{in instances of } \bar{Y} \\
 \bar{A} \wedge \bar{\Phi} \wedge \bar{Y} & \text{in instances of } \bar{Y} \\
 \therefore A \circ Y & \text{Conclusion}
 \end{array}
 \tag{10}$$

The Indirect Method consists of applying the method of Agreement twice. The first analysis is performed *forward* on the instances of the antecedent  $A$ , to check whether they agree on displaying the consequent  $Y$ . These instances only contain information to establish that  $A$  is the invariant antecedent of  $Y$ , saying nothing about the antecedents of  $\bar{Y}$ . The second step addresses this problem and performs the analysis on the instances of the absence of the consequent,  $A\bar{Y}$  and  $\bar{A}\bar{Y}$ , to check if they invariably lack the antecedent so that  $\bar{A} \circ \bar{Y}$ .

In short, without ever mentioning them, the Indirect method of Difference applies the first and the second Stoic indemonstrable as two tests of the hypothesis that  $A$  yields  $Y$  – “each proof being independent of the other, and corroborating it”<sup>45</sup>.

---

<sup>43</sup> Mill, 1843, III, x, §3.

<sup>44</sup> Mill, 1843, III, viii, §4.

<sup>45</sup> Mill, 1843, III, viii, §4.

d) *The fourth and fifth canon*

The fourth canon addresses the issue of ascribing causation when an effect is an assemblage of effects of which some antecedent is known, but not all. It stipulates that

Subduct any phenomenon such part as is known by previous inductions to be the effect of certain antecedents, and the residue of the phenomenon is the effect of the remaining antecedents<sup>46</sup>.

The corresponding method of *Residues* is again reminiscent of the Stoic *schemata* for conjuncts and disjuncts. It ascribes the unexplained part of a phenomenon to that further invariant antecedent that acts together with the known one. Mill presents the method as specially suited to pinpoint the constituents of chemical causation. Thus, we can explain water as the interaction of oxygen and hydrogen when we know rust is the effect of oxidation, and we observe that iron rusts when exposed to water while hydrogen appears and water disappears, the overall weight remaining constant.

Mill holds the method of Residues is the “most fertile,” as it compels the search for the important yet initially disregarded “obscure circumstance” beneath the effect. However, Mill warns, we cannot be sure of the contribution of such an additional antecedent to the occurrence of the effect until we properly test it with a method of Difference.

The fifth and last canon addresses the problem of ascertaining the relationship between an effect and a “Permanent Cause” or “indestructible agent.” These types of causes defy the experimental strategy due to the technical impossibility of either excluding or destroying them<sup>47</sup>. These causes, however, do admit variation. Hence, the relationship with the effect can still be ascertained by the strategy of Difference that narrows on the changes in the antecedent and the consequent when any other relevant circumstance remains the same.

The method of *Concomitant Variations* applies the rationale of the Method of Difference on partial changes in the antecedent and the consequent. The corresponding fifth canon stipulates that

---

<sup>46</sup> Mill, 1843, III, viii, §6.

<sup>47</sup> Mill, 1843, III, viii, §6.

whatever phenomenon varies in any manner whenever another phenomenon varies in some particular manner, is either a cause or an effect of that phenomenon, or is connected with it through some fact of causation<sup>48</sup>.

Mill hinges the causal nature of the induction by Difference on the proof that the antecedent is the only varying circumstance. The method of Concomitant Variations, then, can serve as the extension of the method of Difference to render the relationship between factors into the ratios of their quantities – and to better control for the variation of rival or background circumstances.

The quantification approach to Difference introduces new problems, however. As its conclusions depend on measurement, quantification lies induction open to mistaking measurement errors for meaningful changes and *vice-versa*. Better gauges and broader evidence base can overcome this limit, but arise a new issue. When the consequent depends on a plurality of causes, the relationship between one of the antecedents of interest and the consequent can go dismissed if rare. The limit, Mill emphasizes, does not affect the other versions of the methods of Difference, in which every evidence counts instead.

### 3.2.3 *The many issues of complex causation*

In discussing his canons and methods, Mill establishes four requirements that evidence should meet before we can believe ‘ $A$  causes  $Y$ ’ when we cannot provoke and suppress the consequent by manipulating the antecedent.

The first two requirements – that we can call R1 and R2 – render the cases that fit the first and the second indemonstrable, respectively:

- R1. whenever the antecedent, then the consequent:  
if  $A$  causes  $Y$ , under  $A$ , always observe  $Y$ ;
- R2. whenever not the consequent, then not the antecedent:  
if  $A$  causes  $Y$ , under  $\bar{Y}$ , always observe  $\bar{A}$ .

Thus, R1 and R2 establish that evidence of  $A \wedge Y$  or  $\bar{A} \wedge \bar{Y}$  confirms the conditional  $A \supset Y$ . However, Mill repeatedly underlines that R1 and R2 together do not ensure the conditional relationship is causal “unless we can be quite sure either that the instances affirmative of [the effect]

---

<sup>48</sup> Mill, 1943, III, viii, §6.

agree in *no antecedent whatever but A*, or that the instances negative of [the effect] agree in *nothing but the negation of A*"<sup>49</sup>.

Mill's causal factor then requires two counterproofs that our antecedent is the single factor in the unconditional relationship, and hence the sole reason for the occurrence of the consequent. The counterproofs – that we will call R3 and R4 – establish that real causes impose two impossibilities to the diversity of instances – namely,

- R3. the consequent *only* when the antecedent –  
then *Y cannot* be observed under  $\bar{A}$ ;
- R4. not the consequent *only* when not the antecedent –  
then  $\bar{Y}$  *cannot* be observed under *A*.

Hence, according to R3 and R4, evidence of  $\bar{A} \wedge Y$  or  $A \wedge \bar{Y}$  disconfirms the conditional  $A \supset Y$ . The two requisites and the two counterproofs together ensure the antecedent is the single factor responsible for both the occurrence (R1, R3) and the suppression (R2, R4) of the effect – as in ( $\epsilon$ ). Together they operationalize the ideal biconditional relationship like the one between Aristotelean universals<sup>50</sup>. Biconditionality also warrants Mill's *desideratum* that the knowledge of the antecedent allows anticipating the consequent. However, Mill also recognizes that the ideal situation is seldom true in actual settings – because of the Plurality of Causes, their Composition, and their Constitution.

#### a) *Issues of Plurality*

The *Plurality of Cause* challenges the experimental ideal of causation by recognizing that

It is not true [...] that each phenomenon can be produced only in one way. There are often several independent modes in which the same phenomenon could have originated. [...] A given effect may really be produced by a certain cause, and yet be perfectly capable of being produced without it<sup>51</sup>.

As we have seen, when the antecedent takes the form of a disjunction  $A \vee B$ , the alternatives weakens the cogency of the results from the method of Agreement. The solution rests less on increasing the number

---

<sup>49</sup> Mill, 1843, III, viii, §4, emphasis added.

<sup>50</sup> Mill, 1943, II, ii, §1.

<sup>51</sup> Mill, 1843, III, x, §1.

of same instances under analysis – which only serves the purpose of correcting a measurement error – than on increasing their *diversity* – as even one observation can lead to the elimination of a circumstance and the clarification<sup>52</sup>.

The suitable strategy is the Indirect method of Difference. Its second step satisfies R2 and controls whether the absence of the consequent can be ascribed to the sole absence of the antecedent of interest; the ascription does not hold when more than one antecedent is at work as R4 requires. This failed ascription compels the “conjecture” that the antecedent is plural – at least in its manifestations, as to Mill alternative causes signal a common source of causation to which we can ascend by further inquiry. In the more usual case that “we cannot take this ulterior step, the different antecedents must be set provisionally as distinct causes, each sufficient of itself to produce the effect”<sup>53</sup>.

On a general note, Mill imputes the higher credibility of the various versions of the method of Difference to their reliance on the negative instances. These instances strengthen the reason to trust that a cause is *sufficient* to the effect, as they expose non-causal factors that would pass the test of Agreement<sup>54</sup>.

#### b) *Issues of Composition and Constitution*

The main challenges to R4 arise from issues in the *Composition* or the *Constitution of Causes*.

The Constitution occurs when the effect arises by the “chemical transformation” of the causal elements, but we have not identified all the components: thus, some instances may violate R4 because of the absence of omitted factors. The Composition, instead, concerns the effect of which we have identified the full cause, but that in some instances is impeded by intervening causes with opposite sign, again manifesting a violation of R4. In both cases, the cause is a compound by conjunction: but while Constitution compounds include positive factors, Composition compounds also include negative hindrances. In both circumstances, then, the recognition of the relationship is blurred by instances where the effect is void, although the antecedent is observed.

With Hume, Mill suggests these occurrences should first be addressed semantically: general statements as expressions of the law should not convey that the violation necessarily falsifies the relationship

---

<sup>52</sup> Mill, 1843, III, x, §2.

<sup>53</sup> Mill, 1843, III, x, §3.

<sup>54</sup> Mill, 1843, III, x, §2.

between a single factor and the effect. However, this caution follows from the consideration that a single factor does not capture the whole of the causation:

It is clear that the general proposition, though it would be true under a certain hypothesis, would not express the facts as they actually occur. To accommodate the expression of the law to the real phenomena, we must say, not that the object moves, but that it *tends* to move, in the direction and with the velocity specified. We might, indeed, guard our expression in a different mode, by saying that the body moves in that manner unless prevented, or except in so far as prevented, by some counteracting cause. But the body does not only move in that manner unless counteracted; it *tends* to move in that manner even when counteracted<sup>55</sup>.

Thus, correct general statements about single causal factors refer to their *tendency* to obtain<sup>56</sup>. The difference between mature and rough generalizations, to Mill, lies in the fact that the former deploys an “accurate nomenclature” of tendencies and forces, while the latter speaks of rules and exceptions. Truly general statements have no exception – only “disturbances” that affect their unfolding in actual contexts.

Then, Mill advances practical recommendations for investigating causal conjunctions. The *Logic* contends that any inductive method alone can succeed in unraveling it. Counteracting causes may count as idiosyncratic circumstances, which inductive methods disregard to narrow on constitutive antecedents. The proper method to grasp causation as the complex of causes, conditions, and disturbances proceeds deductively and integrates induction with two additional operations: ratiocination, and verification.

### c) *the Deductive Method*

The processes that Mill envisages as deductive resembles the Peripatetic syllogism in their structure, yet not in their aim. Instead of rearranging previous knowledge to cast new light on their connections, it casts old light – that of the experience abridged into laws – on a new event. The dismissal of metaphysical grounding implies that deductions can be ampliative. Were syllogisms not applicable to new instances, they would establish tautologies – which he understood as a sophistic

---

<sup>55</sup> Mill, 1843, III, x, §5, original emphasis.

<sup>56</sup> See also Mackie, 1965, 1974, 1977; Hempel 1965.



trick to induce begging the assumption instead of proving an argument<sup>57</sup>.

The syllogistic structure, then, brings possible “inconclusiveness to light” in the inference from known particulars to a new one. A correct conclusion “is not an inference drawn *from* the formula, but an inference drawn *according to* the formula”<sup>58</sup>. The Deductive method applies these structures “to find *the law of an effect* from the laws of the different tendencies of which it is the joint result”<sup>59</sup>.

The first step consists of identifying the *tendencies* at work beneath the effect. These can be captured with the method of Concomitant Variation as *empirical laws* – that is, regularities “respecting which we cannot yet decide whether they are cases of causation or mere results if it”<sup>60</sup>. As alternative points of departure, Mill suggests some application of the method of Difference to comparatively more uncomplicated instances or a separate round of deductions. Mill also considers that this first stage can be suppressed when the deduction proceeds from a *hypothesis* that assumes the relationships of causation instead of having them proved<sup>61</sup>. The *Hypothetical deduction* then directly commences from the second step.

*Ratiocination* consists of “determining from the laws of the causes what effect *any given combination* of those causes will produce”<sup>62</sup>. To Mill, this operation is “a process of calculation in the wider sense of the term” that may involve calculation in the narrower sense – mathematical or geometrical – when our knowledge of tendencies allows it. That of ratiocination, in short, is the stage for setting expectations about complexes and compounds and can be drawn at different levels of precision.

All that is essential in it is reasoning from a general law to a particular case, that is, determining by means of the particular circumstances of that case what result is required to fulfil the law.

[...] By such ratiocinations from the separate laws of the causes, we may, to a certain extent, succeed in answering either of the following questions: Given a certain combination of causes, what effect will follow? and, *What combination of causes, if it existed, would produce a given effect?* In the one case, we determine the effect to be expected

---

<sup>57</sup> Mill, 1943, II, iii, §2.

<sup>58</sup> Mill, 1843, II, iii, §4.

<sup>59</sup> Mill, 1943, III, xi, §1, emphasis added.

<sup>60</sup> Mill, 1843, III, xi, §2.

<sup>61</sup> Mill, 1843, III, xiv, §4.

<sup>62</sup> Mill, 1843, III, xi, §2, emphasis added.

in any complex circumstances of which the different elements are known: in the other case *we learn according to what law –under what antecedent conditions– a given complex effect will occur*<sup>63</sup>.

The ratiocination stage, then, generates “conjectures” about the composition of tendencies and circumstances but void of actual import, and about which we may even be unsure of their completeness. To rule out the possibility that we omitted some relevant components, the method of Deduction requires verification as its third and essential stage.

*Verification* consists of retrieving the conjectured balance in actual cases.

To warrant reliance on the general conclusion arrived at by deduction, these conclusions must be found, on careful comparison, to accord with the results of direct observation wherever it can be had. If, when we have experience to compare with them, this experience confirms them, we may safely trust to them in other cases of which our specific experience is yet to come. But if our deductions have led to the conclusion that from a particular combination of causes a given effect would result, then in all known cases where that combination can be shown to have existed and where the effect has not followed, we must be able to show (or at least to make a probable surmise) what frustrated it: if we cannot, the theory is imperfect, and not yet to be relied upon. Nor is the verification complete, unless some of the cases in which the theory is borne out by the observed result, are of at least equal complexity with any other cases in which its application could be called for<sup>64</sup>.

Actual instances that would have been too complicated for serving to induce single general tendencies now become “a new experiment” of the conjecture – “a new trial of the principle in a different set of circumstances” that, moreover, “occasionally serves to eliminate some circumstance not previously excluded”<sup>65</sup>. When the Deductive method applies to hypotheses, moreover, the verification stage can amount to a complete induction if the “very complete” compound meets all the requisites of the Indirect method of Difference<sup>66</sup>.

---

<sup>63</sup> Mill, 1843, III, xi, §3, emphasis added.

<sup>64</sup> Mill, 1843, III, xi, §3.

<sup>65</sup> Mill, 1843, III, xi, §3.

<sup>66</sup> Mill, 1843, III, xiv, §4.

As ratiocination and verification, the method of Deduction is for *proving* the interaction of the laws about the tendencies. To it, Mill adds a “secondary application”: Deduction is for “explaining the laws” about tendencies<sup>67</sup>.

### 3.2.4 Modes of explanation

Mill establishes the operation of *explaining* consists of *resolving* a “law of the effect” into other general statements, and distinguishes three *modes* of resolution<sup>68</sup> that we can understand as combination, intermediation, and subsumption.

#### a) First mode: combination

In the first mode of explanation, a causal relationship is explained by the tendency of the factors that constitute it plus the “fact of their combination” that renders their “collocation” – that is, their simultaneity in space. To illustrate this kind of explanation, Mill recalls that the law of the motion of a planet consists of the law of the acquired force, of the law of centripetal force, and of their coexistence given certain circumstances of place and time.

The first mode, therefore, corresponds to the Deductive method, and can be portrayed as in [11] below:

$$\begin{array}{ll}
 \Phi \Leftrightarrow Y & \text{Relationship}_0 \\
 \Phi = \wedge \varphi_i & \text{Fact of combination}_1 \\
 \varphi_{1}: A \Rightarrow Y & \text{Constituent tendency } T_1 \\
 \varphi_{2.1}: C \Rightarrow Y & \text{Constituent tendency } T_{2.1} \\
 \varphi_{2.2}: D \Rightarrow Y & \text{Constituent tendency } T_{2.2} \\
 C \vee D = B & \text{Fact of combination}_2 \\
 \varphi_k: \bar{P} \Rightarrow Y & \text{Counteracting tendency } T_3 \\
 \Phi = A \wedge B \wedge \bar{P} & \text{Definition of combination}_1 \\
 \therefore A \wedge B \wedge \bar{P} \Leftrightarrow Y & \text{Resolution}
 \end{array} \tag{11}$$

where the starting relationship can be hypothetical, the double-edged arrow  $\Leftrightarrow$  indicates the connection is biconditional, and the single-edged arrow  $\Rightarrow$  indicates the connection is conditional at least.

<sup>67</sup> Mill, 1843, III, xiv, §4, emphasis added.

<sup>68</sup> Mill, 1843, III, xii, §6.

b) *Second mode: intermediation*

In the second mode, the explanation consists of detecting an intermediate link “between what seemed the cause and what was supposed to be the effect ... so that the cause at first assigned is but the remote cause, operating through the intermediate phenomenon”<sup>69</sup>. As a first illustration, the *Logic* recalls the connection between touching an object and perceiving a sensation. The apparently direct experience was later discovered to occur by the mediation of nerves connecting

our outward organs to the brain. Touching the object, therefore, is only the *remote* cause of our sensation; that is, not the cause, properly speaking, but the cause of the cause; – the real cause of the sensation is the change in the state of the nerve<sup>70</sup>.

The *Logic* considers the two separate connections (of the outward organ to the nerve, and the nerve to the brain) are each more general than their conjunction, as the latter can fail under more circumstances than each component<sup>71</sup>. Moreover, the proximate connection has wider relevance in accounting for the occurrence of the effect because different antecedents can activate it, and without it the signal cannot be transmitted. The chain of relationships that resolves the connection, therefore, can be portrayed as in [12.a] below:

$$\begin{array}{ll}
 \varphi_1 \Leftrightarrow Y & \textit{Relationship}_0 \\
 \varphi_1 \Rightarrow \varphi_2 & \textit{Remote relationship} \\
 \varphi_2 \Leftarrow Y & \textit{Proximate relationship} \\
 \therefore (\varphi_1 \wedge \varphi_2) \Leftrightarrow Y & \textit{Resolution}
 \end{array}
 \qquad [12.a]$$

Chemical causation, too, envisages intermediate causation – although as stages in the process that yields the effect:

The more powerful acids corrode or blacken organic compounds. This is a case of causation, but of *remote* causation; and is said to be explained when it is shown that there is an intermediate link, namely, the separation of some of the chemical elements of the organic structure from the rest, and their entering into combination with the acid. The acid causes this separation of the elements, and the separation of

---

<sup>69</sup> Mill, 1843, III, xii, §4.

<sup>70</sup> Mill, 1843, III, xii, §3, emphasis added.

<sup>71</sup> Mill, 1843, III, xii, §5

the elements causes the disorganization, and often the charring of the structure<sup>72</sup>.

To preserve the parallelism, but somehow confusingly, Mill applies the term “remote” to denote the compound in which a factor is embedded that is responsible for the effect of interest when interacting with key components from other compounds. In turn, the factor that enters the composition or constitution of the effect qualifies as “proximate.”

Nevertheless, the second mode resolves chemical causation into remote factors and proximate links by decomposition and reconstruction of compounds, as summarized in [12.b] below.

$$\begin{array}{ll}
 \Phi \Leftrightarrow Y & \text{Relationship} \\
 \Phi: \varphi_1 \wedge \varphi_2 \Leftrightarrow Y & \text{Decomposition}_1 \\
 \varphi_1: A \wedge \varepsilon_1 \Rightarrow Y & \text{Decomposition}_{1a} \\
 \varphi_2: X \wedge \varepsilon_2 \Rightarrow Y & \text{Decomposition}_{1b} \\
 \therefore A \wedge X \Leftrightarrow Y & \text{Reconstruction}
 \end{array} \tag{12.b}$$

c) *Third mode: subsumption*

The third mode of explanation proceeds in the opposite directions of the first two modes, and consists of the *subsumption* of partial or particular relationships under a more general one that recognizes the pattern across “diverse sets of instances.” The third mode resolves particular relationships into a more general one establishing the class equivalence of antecedents and consequents, as in reasoning [13]:

$$\begin{array}{ll}
 \varphi_1 \Leftrightarrow v_1 & \text{Special relationship}_1 \\
 \varphi_2 \Leftrightarrow v_2 & \text{Special relationship}_2 \\
 \varphi_3 \Leftrightarrow v_3 & \text{Special relationship}_3 \\
 \Phi = [\varphi_1] \wedge [\varphi_2] \wedge [\varphi_3] & \text{Class equivalence} \\
 Y = [v_1] \wedge [v_2] \wedge [v_3] & \text{Class equivalence} \\
 \therefore \Phi \Leftrightarrow Y & \text{Subsumption}
 \end{array} \tag{13}$$

Explaining, then, “is but substituting one mystery for another”<sup>73</sup>. However, the substitution extends the reach of our statements as it resolves the original relationship into more certain ones, or into relationships with a broader scope.

<sup>72</sup> Mill, 1843, III, xii, §3, emphasis added.

<sup>73</sup> Mill, 1843, III, xii, §6.

### 3.3 Explanation between induction and deduction

The further modern rendering of classical explanation – and, for a few decades, the “received view” in the philosophy of science<sup>74</sup> – bears Hempel’s signature<sup>75</sup>.

The model developed in the wake of the analytical revolution that, since Frege, engaged the philosophy of science<sup>76</sup>. The agenda originated in the works of Bertrand Russell and Ludwig Wittgenstein. Along with the Stoic logic, it assumed the world has a knowable logical constitution that language can render. In Wittgenstein’s words, language can align a specific “state of affairs,” that is, a constellation of objects that stand in actual relationships with each other in the world, with the picture that we make in our mind of it as a “situation” – which, in the scientific discourse, could be a “theory.” The language performs the alignment through sentences – molecular structures of atomic components whose logical relationship mirrors the relationship among the corresponding formal objects on the one side, and empirical objects on the other. The promise was, again, that the proper logical syntax would have grasped the isomorphism of situations and states of affairs, and yielded a more profound knowledge of the world as verified theoretical statements.

Such an agenda found its limits in the fact that logic opened to possibilities and alternative worlds, while the empirical domain offered portrayals of necessary actualizations, and language involved the attribution of meaning – so that the verifiable overlapping of the three could only be partial at best<sup>77</sup>. Hempel’s proposal, although far from offering the ultimate solution, proved able to attract the attention of as diverse disciplines as history and psychiatry on the relevance of the alignment of language, logic, and evidence to improve our grasp of causation – even in single occurrences.

#### 3.3.1 *Deductive models of explanation*

Of Mill’s three modes of explanation, Hempel considers the third one as his model of reference. In his understanding, an explanation is

---

<sup>74</sup> Salmon, 1994.

<sup>75</sup> Hempel, 1965; Hempel and Oppenheim, 1942.

<sup>76</sup> Mates, 1961; Kneale and Kneale, 1962.

<sup>77</sup> Wolniewicz, 1999.

an operation that subsumes a single causal occurrence under one or more general statements. His explanatory strategy is better known as “of covering law” – a definition that Hempel himself embraced<sup>78</sup>.

In a famous example, Hempel uses a car’s cracked radiator to illustrate how the covering law strategy works<sup>79</sup>. Once and again, the explanation of a singular occurrence follows from two sets of information: “(i) particular facts and (ii) uniformities expressible by means of general laws”<sup>80</sup>.

The first set  $\mathbf{C}_k = \{C_1, \dots, C_k\}$  includes the  $k$  circumstances  $C$  under which the event occurred. To this set belong conditions such as that the car’s radiator was of cast iron; that cast iron belongs to the class of materials with low tensile strength; that the radiator was filled with water, its lid screwed on tightly; that the car was parked in the street where, during the night, the temperature fell below 25°F.

The second set  $\mathbf{L}_r = \{L_1, \dots, L_r\}$  includes the  $r$  proven laws or “universal hypotheses” – statements with an unlimited scope that hold in the world of reference without exception. To this set belong notions such as that water freezes below 32°F at standard atmospheric pressure; that water increases in volume when freezing; that materials with low tensile strength crack when stressed.

The general structure of an explanation, hence, unfolds as in [14] below:

$$\begin{array}{ll}
 \wedge \mathbf{L}_r & \text{the conjunction of law statements} \\
 \wedge \mathbf{C}_k & \text{the conjunction of condition statements} \\
 \hline
 & \text{deductively entail} \\
 E & \text{the statement about the event in case}
 \end{array} \quad [14]$$

Hempel denotes the expression below the line as the *explanandum* and everything above the line as the *explanans*. Like in a classic syllogism, the *explanans* includes a major premiss that sets the general rules and a minor premiss that make the case. Of the two, the conjunction of the  $k$  circumstances,  $\wedge \mathbf{C}_k$  is that which accounts for  $E$ : the radiator cracked because it was of cast iron and filled with water and was exposed to low temperatures. However, Hempel emphasizes, the conditions of the radiator and the weather together explain the occurrence as far as they provide the *boundary conditions* under which  $\wedge \mathbf{L}_r$  unfolds.

<sup>78</sup> Hempel, 1965.

<sup>79</sup> Hempel, 1942:36 ff.

<sup>80</sup> Hempel 1965:336, 1942.

Otherwise said, Hempel emphasizes the general statements  $\Lambda L_r$  are that which makes the boundary conditions relevant to understand and account for the occurrence of  $E$ . The low temperature contributes to explain  $E$  in light of our knowledge of the rules  $L_W$  that capture the general behavior of water. The rules of water behavior, in turns, gain relevance in virtue of the rules  $L_M$  that capture the behavior of solid materials under pressure.

The ultimate reason for  $E$ , in Hempel's view, lies in the  $r$  laws  $\Lambda L_r$  that jointly obtained under  $\Lambda C_k$  and made  $E$  unavoidable given the circumstances. The conjunction of local conditions becomes explanatory as it compels the local instantiation of the general laws. Counterfactually, "if the law were deleted, the argument would no longer be valid"<sup>81</sup>.

General statements also ensure a further hallmark of Hempel's explanation: its grounding in *expectability*. The information that accounts for an occurrence is the same that makes it predictable. Knowing the general behavior of water and iron containers, we can foresee the cracking as much as we can explain it. The difference between the two heuristics reduces to a mere matter of research focus:

While in the case of an explanation the final event is known to have happened and its determining conditions have to be sought, the situation is reversed in the case of a prediction: here, the initial conditions are given, and their "effect" – which, in the typical case, has not yet taken place – is to be determined<sup>82</sup>.

Expectability makes explanation possible even in those "special" sciences – such as history, economics, medicine, or sociology – interested in stochastic phenomena. Hempel maintains select statistical abstractions do support scientific knowledge, too, as far as they allow anticipating the occurrence of some class of events with a reasonable degree of confidence.

All that matters to Hempel, therefore, is whether the general statements cover the occurrence of interest – that is, whether they apply under the circumstances. The default reasoning again unfolds deductively as in [15.a].

Deduction [15.a] reads as follows. Let  $T$  be a treatment for hay fever symptoms,  $C$  some initial health conditions,  $O$  as the relief, and the in-

---

<sup>81</sup> Hempel 1962: 101.

<sup>82</sup> Hempel 1942:37.



dexed  $C$  and  $O$  as the values that health conditions and reliefs, respectively, take in the  $i$ th actual instance. Thus, the major premiss in [15.a] reads that, under general conditions  $C$ ,  $T$  makes  $O$  to be expected with a probability of  $r$ . When the probability is as high as  $r \cong 1$ , the remedy succeeds almost certainly in the cases that qualify as instances of the reference class  $C$ .

$$\begin{array}{l}
 p(O, TC) = r \quad \text{probability of an outcome } O \text{ given } TC \\
 TC_i \quad \text{degree in which } i \text{ is an instance of } TC \\
 \hline \hline
 \text{entails with probability } r \\
 O_i \quad \text{the value of the outcome of case } i
 \end{array} \quad [15.a]$$

Now, the minor premiss offers the observation of the further case  $i$  meeting  $C_i$  – for instance, John Doe, showing clear signs of a hay fever attack after spending a few hours near a bouquet, with a family history of allergies but otherwise young and healthy – and taking  $T$ .

All in all, the structure suggests we can expect  $O_i$  with probability  $r$  – and in both directions. Prospectively, we can predict John Doe’s  $O$  after  $T$  as he qualifies as a  $C$ -type instance. Retrospectively, we can conclude that John Doe’s  $O$  was almost inevitable because the remedy  $T$  was adequate to John Doe’s disease and conditions.

However, general statistical premisses arise a paradox when the possibility always exists that, together with [15.a], [15.b] holds, too:

$$\begin{array}{l}
 p(\bar{O}, TK) = r \quad \text{probability of an outcome } \bar{O} \text{ given } TK \\
 TK_i \quad \text{degree in which } i \text{ is an instance of } TK \\
 \hline \hline
 \text{entails with probability } r \\
 \bar{O}_i \quad \text{the value of the outcome of case } i
 \end{array} \quad [15.b]$$

The structure of [15.b] reads that the same treatment of type  $T$  can be expected to yield the opposite outcome  $\bar{O}$  with the same probability  $r$  as in [15.a], just because of the different conditions  $K$ . For instance, Jimmy Doe can find no relief in  $T$  because the treatment works when the symptoms are caused by allergens, while his disease has a bacterial origin.

On a more general vein, Hempel eventually recognizes that any outcome becomes almost statistically certain given the “right” reference

class, hence under the right set of boundary conditions<sup>83</sup>. This point has long been raised to contend that statistical equations can support explanation as true universals do<sup>84</sup>. The concern compelled Hempel to two clarifications: first, of the function of the general statements employed as explanatory premisses; second, of the ground on which probabilistic statements can license the inference to the next instance.

### 3.3.2 *Capturing potentials*

The scope of the general statement in an explanation does not warrant its explanatory capacity<sup>85</sup>. General statements confer explanatory power to the conditions in the *explanandum* when they capture a potential that the conditions actualize. Therefore, the general statements support the explanation if they capture the possibility of the effect. It is the potential or the possibility of an effect – not its probability – that allows us expecting it and interpreting some of its non-manifestations as a failure.

Now the problem arises as possibilities are unobservable by definition, hence invite to debatable metaphysical tenets. Regardless of their ontological commitment, different strands in the modern philosophy of science<sup>86</sup> conceive of potentials as objects' theoretical properties. The empirical streams utilize these theoretical constructs as placeholders in an explanatory hypothesis whose standing is decided by the fit of evidence. Hempel's explanation sides with this tradition and with its commitment to the empirical testability of the statements about properties. Testability, however, requires the theoretical property is "operationalized" as an appropriate extensional class. Hempel's argument on this point runs as follows.

Explanations are always relative to some *explanandum*  $Q$ . The theoretical mindset commits to consider  $Q$  as the observable state that a particular object, or unit, or body  $x$  takes after causation unfolded. The further assumption stipulates that  $x$  becomes  $Q$  at time  $t$  provided that, at time  $t - 1$ , it already has the potential, the capacity, or the *disposition* to  $Q$ . In short,  $x$  "Q-ed" only because it was "Q-able" first. Assuming that  $x$  is  $Q$ -able allows identifying the attributes of the reference class. The concept holds if its reference class allows predicting  $Q$ , or if

---

<sup>83</sup> Hempel, 1965: 397.

<sup>84</sup> e.g., Scriven, 1958; Salamon, 1984.

<sup>85</sup> Salmon, 1994.

<sup>86</sup> e.g., Russell, 1927; Ramsey, 1931; Carnap, 1936, 1937, 1946; Scriven, 1958; Quine, 1974; Mackie, 1977; van Fraassen, 2008.

it fits the theoretically “right” boundary conditions that can explain the manifestation of  $Q$  in suitable bodies  $x$  or its failure. The open issue, hence, is how to translate disposition-concepts into testable reference classes. Hempel’s explanation develops a solution that takes the seminal contribution of Rudolf Carnap<sup>87</sup> as its point of departure.

Carnap was “the first empiricist who, not content with asserting the reducibility of science to terms of immediate experience, took serious steps toward carrying out the reduction”<sup>88</sup> He provided an operational and experimental understanding of a disposition. “ $Q$ -able” is the predicate that we introduce to indicate the capacity of a body  $x$  to react “in such and such ways and such conditions, e.g., ‘visible,’ ‘smellable,’ ‘fragile,’ ‘tearable,’ ‘soluble,’ ‘indissoluble’ etc.”<sup>89</sup> Hence, if we assign the meaning “soluble in water” to  $Q$ , we can define it as in [D]

$$[D] \quad Q \equiv (C_1 \supset Q_1)$$

where  $Q$  denotes the solubility of any body  $x$ ,  $C_1$  reads that the body  $x$  is placed into water at time  $t$ , and  $Q_1$  reads that the body dissolves at time  $t$ . The definition, hence, stipulates that “any body is soluble in water if and only if, if it is placed into water, then it dissolves.”

The problem with [D] is that it relies on the horseshoe, which makes the implication vacuously true when the antecedent is false. Otherwise said, the definition qualifies a body as soluble even if it is not, based on the fact that it has never been put into water and never will be – such as, with Carnap’s example, the match I burned yesterday.

To address this paradox, Carnap suggests rephrasing [D] as the *reduction sentence* [RS]

$$[RS] \quad C_1 \supset (Q \equiv Q_1)$$

The sentence switches the connectives and yields a conditional definition: if the body  $x$  is placed into water at time  $t$ , then it is soluble if and only if it dissolves at time  $t$  – else, it is not. As such, [RS] implies a loss in scope and a gain in accuracy in contrast to [D], as it only defines things that meet  $C_1$ , hence, solubility in water.

The meaning of [RS] can be generalized as in [RS1] below:

---

<sup>87</sup> Carnap, 1936, 1937.

<sup>88</sup> Quine, 1951:39.

<sup>89</sup> Carnap, 1937:440.

[RS1]  $C_1 \supset (Q_1 \supset Q)$

[RS1] reads that if any space-time point  $x$  meets the empirical condition  $C_1$ , then if we find the effect  $Q_1$ , then the point has the property  $Q$ . So rephrased, it means that observing  $Q_1$  as expected of an instance from the reference class  $C_1$  suffices to maintain that the property applies to the instances of  $C_1$ .

[RS1] also serves to establish the absence of the theoretical property. If  $x$  satisfies the condition  $C_2$ , then if we find the effect  $Q_2$ , then the point has not the property  $Q$ :

[RS2]  $C_2 \supset (Q_2 \supset \sim Q)$

[RS2] means that observing  $Q_2$  from the reference class  $C_2$  as expected suffices to maintain that the property does not apply to the instances of  $C_2$ .

[RS1] assigns  $Q$  to the points of the class  $C_1 \wedge Q_1$  while [RS2] attributes  $\sim Q$  to the points of the class  $C_2 \wedge Q_2$ . Hence, [RS1] offers a valid reduction for  $Q$  except for those points that satisfy  $\sim(C_1 \wedge Q_1)$ . The negation of the conjunction  $\sim(C_1 \wedge Q_1)$ , then, states the impossibility that defines the property  $Q$ .

On the same vein, [RS2] reduces for  $\sim Q$  except for those points that are  $\sim(C_2 \wedge Q_2)$ . The negation of the conjunction  $\sim(C_2 \wedge Q_2)$  states the impossibility that defines the property  $\sim Q$ .

Together, [RS1] and [RS2] constitute a *reduction pair* for  $Q$  and hold for those cases that, logically or empirically, do not satisfy  $\sim[(C_1 \wedge Q_1) \vee (C_2 \wedge Q_2)]$ .

A special case of empirical reduction arises when  $C_2 = C_1$  and  $Q_2 = \sim Q_1$ . It refers to the case of two objects,  $x$  and  $y$ , such that when placed in water  $x$  dissolves, while  $y$  does not. The two effects license the replacement of the reduction pair with a single “bilateral reduction sentence” [BRS]:

[BRS]  $C_1 \supset (Q \equiv Q_1)$

[BRS] reads that if an object meets the condition  $C_1$ , then it has the property  $Q$  if and only if it yields  $Q_1$  – and *vice-versa*, it yields the effect

if it has the property. The bilateral reduction sentence, Carnap notes, cannot apply to the objects for which  $\sim C_1$  holds, that is, outside the conditions that reveal the property through the reaction. When these conditions are a standard constituent of any state of affairs, like oxygen to regular fire, then we can dispense with specifying  $C_1$  and the reduction sentence “degenerates” into the *definition*:  $Q \equiv Q_1$ . The statement says that a body has the property if it manifests the effect and *vice versa*. As such, the statement is circular, void of factual content, and of little use outside the strict domain of logic.

In contrast to bilateral reduction sentences, *reduction pairs* may preserve empirical import. Reduction pairs assign points or entities to the general class  $C_1 \wedge Q_1 \wedge C_2 \wedge Q_2$ . As no object can be at the same time under two different conditions and yield opposite results, then it also must be that no object of a reduction pair satisfies  $\sim (C_1 \wedge Q_1 \wedge C_2 \wedge Q_2)$ , that is, by the laws of complementarity,  $\sim C_1 \vee \sim Q_1 \vee \sim C_2 \vee \sim Q_2$ .

The disjunction of negated terms does not degenerate into a definition if we further specify  $C_i$  – for instance, by stating that if  $x$  is made of a specific material *and* is put into water, then it will be soluble and dissolve. The addition of new predicates results in a *reduction chain* that improves the determination of the instances as it grows richer.

### 3.3.3 The problem of confirmation in explanations

Carnap’s proposal does not capture potentials or dispositions as such – and the question remains open of how to gauge these and other “unobservables” credibly<sup>90</sup>. Nevertheless, his reduction sentences clarify that a disposition becomes visible as the response that special conditions can elicit from certain entities, and that we can interpret causally in the light of a theory. Thus, reduction sentences set hypotheses about conditions that are counterfactually associated with the potential to obtain. Conditions capture how the past, future, or anyhow unobservable entities *would* react to the “experimental” conditions if tested.

In a reversal of the classical logic, late Hempel’s explanation takes Carnap’s reduction sentences as the elementary logical machinery that (i) allows connecting conditions to effects and (ii) licenses the applicability of theoretical terms about potential to the entities in the same class as those for which the connection proved true. At the same time, Hempel considers that the reduction sentences inevitably establish the conditional implication as a *bona fide* general statement that may never

---

<sup>90</sup> van Rooj and Schulz, 2019; Salmon, 1977; Quine, 1951, 1960, 1969.

lose its state of hypothesis. This conclusion follows from the redefinition of the original “covering laws” approach.

In their first versions, Hempel’s “covering laws” are general statements – such as ‘any raven is black’ – and, logically, universal conditionals – such as ‘if it is a raven, then it is black’ – that stand as (dispositional) hypotheses – meaning that, for any  $x$ , if  $x$  is a raven then  $x$  is black. Hempel considers a common intuitive assumption that the conditional is *confirmed* by any object  $a$  that ‘is a raven’ ( $C$ ) and ‘is black’ ( $Q$ ) and *disconfirmed* when we observe a raven and non-black object ( $C \wedge \bar{Q}$ ). Such an assumption, to Hempel, defies essential requirements of consistency that the language of logic imposes. To make his point, he reformulates the issue as follows.

He maintains the universal conditional statement  $S_1$  ‘if raven, then black’ is formally equivalent to the contrapositive  $S_2$  ‘if not black, then not raven’, so that  $S_1 = S_2$ . If the relationship is a material conditional,  $S_1: C \supset Q$  and  $S_2: \bar{Q} \supset \bar{C}$ , and the hypothesis reads  $(C \supset Q) \equiv (\bar{Q} \supset \bar{C})$ . Assuming the antecedent and the consequent in  $S_1$  and  $S_2$  are binary – either true or false – the hypothesis affords four possible types of instances: (a) ravens and black, (b) ravens and non-black, (c) non-ravens and black, (d) non-ravens and non-black – as displayed in Figure 5.

**Figure 5. Types of objects by the antecedent and the consequent of a conditional statement**

$Q$	(c) $\bar{C}Q$	(a) $CQ$
$\bar{Q}$	(d) $\bar{C}\bar{Q}$	(b) $C\bar{Q}$
	$\bar{C}$	$C$

Now, the intuitive understanding of empirical probation entails that objects (a) confirm  $S_1$  yet are neutral to  $S_2$ , and objects (d) confirm  $S_2$  yet are neutral to  $S_1$ . Thus, the intuitive understanding assigns different import to the same instances depending on how the conditional is formulated. Such an asymmetry, in Hempel’s view, violates a fundamental

logical requirement for the adequacy of any criterion of confirmation – that is, the *Equivalence Condition*.

The Equivalence Condition commits to the principle that “whatever confirms (disconfirms) one of two equivalent sentences, also confirms (disconfirms) the other”<sup>91</sup>. Being  $S_1$  and  $S_2$  hypotheses about the same disposition, the Equivalence Condition stipulates they have to be supported by the same type of instances. Hence, objects ( $d$ ) – be them red herrings or white shoes – *do* bear confirmatory import. Moreover, they support  $S_2$  as  $S_1$  and any further equivalent statement – such as  $S_3$ :

$$S_3: C \vee \bar{C} \supset Q \vee \bar{C}$$

The last equivalent statement originates from the insertion of  $\bar{C}$  as a disjunct in both the antecedent and the consequent of  $S_1$ . The addition does not affect the formal relationship between the antecedent and the consequent. However,  $S_3$  states that if anything is either a raven or not a raven, then it is either black or not a raven. As the antecedent is now a tautology and the implication is material by assumption, the consequent  $Q \vee \bar{C}$  always follows. And, as paradoxical as it may seem, logic dictates that  $S_1, S_2, S_3$  all find support in any black object, be it a raven or not, or in any non-raven object, be it black or not. Moreover,  $S_3$  compels us to make room for further eerie assumptions. On pain of logical inconsistency, we should accept that, with red herrings and white shoes, green mermaids, too, stand as confirmatory instances of our hypothesis about black ravens.

Hempel’s tenet spurred a lively debate on the paradoxes of implication that, nevertheless, did not cause any radical change in his understanding of the problem of inference and his commitment to formal logic as the framing language of reference<sup>92</sup>. However, he later conceded that categorical and Aristotelean-like systems were typically able to avoid the green mermaids’ paradox by forbidding the negation of the subject – that is, its qualitative variation. Predicate logic lifted the constraint by reducing subjects to predicates that, as such, can be negated. He then reasoned that an Aristotelean restriction could be fruitfully introduced through a rule requiring the “existential import” of the instances under analysis. The rule, Hempel reasoned, was not to apply to logical hypotheses directly, as the validity of these statements only must

---

<sup>91</sup> Hempel, 1965:13.

<sup>92</sup> Hempel, 1965.

depend on their fit to the axioms of propositional logic. Instead, the requirement of “existential import” had to apply to the *operationalization* of logical statements – that is, to the *observational reports* that allow us to adjudicate on their empirical standing. Late Hempel, so, posited operationalization as the ground where validity serves soundness.

### 3.3.4 Operational solutions

The operational perspective assumes that hypotheses and evidence are rendered by sentences formulated in the same “language of science.” The language allows shaping the connection between a property, such as ‘being black’ or ‘being a raven’, and, on the other, the observable attributes and relationships that qualify an entity as an instance of that property. An *observational sentence*, then, is “the sentence affirming or denying that a given object, or sequence of objects, possesses one of those observable attributes”<sup>93</sup>. The observation sentence, in short, describes the output of a measurement, or of any other pragmatic operations devised to decide on the nature of objects based on their observable features. The *observation report* **R** consists of the finite class of observation sentences.

So equipped, Hempel stipulates that *conclusive confirmation* or *disconfirmation* can be framed as a logical relationship of *entailment* between an observation report and a hypothesis, such that any hypothesis that is entailed by an observation report is also confirmed by it. The entailment condition, in Hempel, builds on a Chrysippean-like criterion of correctness – hence, it moves the issue outside the terrain of formal logic. It stipulates that a hypothesis-statement **H** is confirmed as its conjunction with a false observational report **R** is unsatisfiable, as in [16]

$$\sim(\mathbf{H} \wedge \bar{\mathbf{R}}) \quad [16]$$

The entailment condition allows for the *conclusive confirmation* of a hypothesis – its *verification* or *falsification* – which, however, is limited in scope: “from the information contained in **R** we can infer that the hypothesis **H** does hold within the finite class of those objects that are mentioned in **R**”<sup>94</sup>.

To widen the scope, Hempel adds two further conditions:

---

<sup>93</sup> Hempel, 1965:23.

<sup>94</sup> Hempel, 1965: 36.



- the *Consequence Condition* maintains that if an observation-report verifies a specific class of sentences, then it confirms any sentence that is a logical consequence of that class;
- the *Consistency Condition* maintains that every logically consistent observation report is logically compatible with the whole class of the hypotheses that it confirms.

Consequence and consistency allow extending the scope of confirmation beyond observations through the *development* of the hypothesis. The development of a hypothesis **H** for a condition *C* “states what **H** would assert if there existed exclusively those objects which are elements of *C*”<sup>95</sup>. Hence, it consists of the chained assertions of the hypothesis **H** for all the objects in an observational report. The hypothesis **H** is verified when the sentences in the report **R** entail the development of the hypothesis for the finite class of the objects in the report.

To illustrate the point, let us assume that

- **H**:  $C \supset Q$  renders the raven hypothesis,
- we have a finite class of three objects  $\{i, j, k\}$ .

The *development* of the raven hypothesis then reads  $\mathbf{H}_1: (C_i \supset Q_i) \wedge (C_j \supset Q_j) \wedge (C_k \supset Q_k)$ .

Let also assume that, in the three objects,

- the property *C* is rendered by the observable attribute  $C_1$ , and the observations of  $C_1$  lead to the statements  $C_{1i}, \bar{C}_{1j}, \bar{C}_{1k}$ ,
- the property *Q* is rendered by the observable attribute  $Q_1$ , and the observations of  $Q_1$  in the three objects yield the statements  $Q_{1i}, Q_{1j}, \bar{Q}_{1k}$ .

The report then reads  $\mathbf{R}: (C_{1i} \wedge Q_{1i}) \wedge (\bar{C}_{1j} \wedge Q_{1j}) \wedge (\bar{C}_{1k} \wedge \bar{Q}_{1k})$ .

**R**, then, entails  $\mathbf{H}_1$ ; therefore, we can infer that **H** is satisfied within the finite class  $\{i, j, k\}$  under the conditions of Consequence and Consistency. **R** then confirms **H** for  $\{i, j, k\}$ .

---

<sup>95</sup> Hempel, 1965:36.

In short, Hempel assumes the impossibility  $\sim(\mathbf{H}_1 \wedge \bar{\mathbf{H}})$ , in which the truth value of  $\mathbf{H}_1$  is decided by  $\mathbf{R}$ . Also, Hempel equates the *disconfirmation* of a hypothesis  $\mathbf{H}$  with the confirmation of its denial,  $\bar{\mathbf{H}}$ , decided by the observational report  $\mathbf{R}'$ . A report is *neutral* to  $\mathbf{H}$  if it neither confirms nor disconfirms it. In his note 46, he also suggests that when a report contains a tautological observation about an object, then the object shall be dropped from the reference class.

Then, *verification* and *falsification* are those types of confirmation that apply when observation reports are *conclusive* (Hempel, 1965:39). Whether observational sentences have such a heuristic power depends on the quantity of the hypothesis, as summarized in Table 2.

**Table 2. Quantity of a hypothesis and conclusive observations**

Scope	Hypothesis	Reports	Conclusion
<i>Singular</i>	$C(x) \supset W(x)$	$C(x) \wedge W(x)$	Verified
		$C(x) \wedge \bar{W}(x)$	Falsified
		$\bar{C}(x)$	–
<i>Existential</i>	$\exists x \in X . C(x) \supset W(x)$	$C(x) \wedge W(x)$	Verified
		$C(x) \wedge \bar{W}(x)$	–
		$\bar{C}(x)$	–
<i>Universal</i>	$\forall x \in X . C(x) \supset W(x)$	$C(x) \wedge W(x)$	–
		$C(x) \wedge \bar{W}(x)$	Falsified
		$\bar{C}(x)$	–

*Keys.* Lowercase indicates objects; uppercase indicates attributes.

$\exists$  is the existential quantifier and reads “there exists at least one”;  $\forall$  is the universal quantifier and reads “for all”, “for any”.

$\in$  indicates the relationship of belonging;  $X$  indicates the reference class.

Table 2 says that, in a universe  $\mathbb{U}$ , a *singular* hypothesis about a unique object is both verifiable and falsifiable. The statement  $C(x) \supset W(x)$  can read ‘if  $x$  is a cygnet, it turns white’. The observation report of  $x$  changing into white verifies the statement; of it changing into any other color falsifies the statement, while a report of no change leaves the statement undecided.

*Existential* hypotheses, instead, can be verified yet not falsified. The statement  $\exists x \in X . C(x) \supset W(x)$ , reading that ‘there exists at least one entity  $x$  belonging to class  $X$  (of chicks) such that, if  $x$  is  $C$  (a cygnet), it  $W$  (turns white)’, is conclusively proven by the observation sentence about those chicks that make the conditional true. However, observation reports cannot decide the truth of the statement when there is no  $x$ .

Pure *universal* hypotheses can be falsified when an observation report contradicts them; however, they cannot be verified. The statement  $\forall x \in X. C(x) \supset W(x)$ , reading that ‘for any object  $x$  drawn from the class  $X$ , if  $x$  is  $C$ , then  $x$  is  $W$ ’, is conclusively proven false when we find  $\bar{W}(x)$  – in the cygnet example, if the chick turns into a black swan. A series of observational reports about white swans, however, only makes the statement provisionally true – confirmed, yet not verified.

The parallel is striking of Hempel’s considerations about observational reports with the inferences allowed by the square of opposition in Figure 1. Existential hypotheses concern particulars, while the hypothesis about singular causation mirrors the Aristotelean statements about a closed universe – in Hempel’s case, a universe of one item – whose testing consists of measurement. Hempel’s universal hypotheses do preserve the nature of the Aristotelean universals in that they can be contradicted although not conclusively proven – unless their scope is restricted to particular space-time boundaries.

From the advantage point of observational reports, Hempel’s suggestions indicate that the same observation does have a different proving power – depending on the scope of the hypothesis. Thus, given the unquantified hypothesis that  $C \supset W$ , the conjunction  $C(x) \wedge W(x)$  verifies the singular and the existential understandings of it; the report  $C(x) \wedge \bar{W}(x)$  falsifies the singular and the universal understandings; and reports about red herrings, white shoes, and green mermaids  $\bar{C}(x)$  do not affect the provisional standing of any version of the hypothesis. The quantifier of the hypothesis, *vice versa*, determines the type of observation required to establish the truth of the hypothesis.

## 4. Back to square one?

The portrayals sketched in the previous chapters, although partial, suggest a remarkable continuity in the understanding of scientific explanation. Aristotle identified it in the middle term that qualifies some objects as instances of a universal relationship: planets do not twinkle because they are near things, and near things never twinkle. Hempel identified it in the circumstances that license expectations about the behavior of particular instances: the car's radiator broke because it was made of cast iron and filled with water and exposed to low temperatures, and iron is fragile and water increases in volume when freezing.

A robust strand agrees on identifying explanations with the set of conditions under which the effect occurs. More controversial is the ground on which we can accept that those conditions are relevant to the occurrence of the effect. Intuitively, conditions explain as they signal the presence in the field of some causal power, or tendency, or disposition to yield the effect that they activate – which licenses the inference. The difficulty consists of defining the ultimate criteria that allow us assuming such power inheres in some object and is manifested as some concomitant attribute. The problem is exquisitely Aristotelean, but the modern philosophy of science contends that it can be tackled with metaphysical objects. On pain of circularity, we cannot conclude that not twinkling inheres in near things as the manifestation of their “nearthingness.”

The problem has two sides – ontological, related to the potential of interest, and epistemic, related to the knowledge that we can have of it. Logical languages have been introduced to improve our grasp of both sides. As the previous chapters show, these languages have cast some convincing light on the epistemic side of the problem, showing under which empirical conditions we can maintain a hypothesis orders reality according to logical criteria. Confirmation and disconfirmation apply to

an image of reality that may have no stronger standing than an educated guess. As Hempel and Carnap established with the rules of entailment, the match of observation reports and reduction sentences allows us deciding when a hypothesis is proven and on which ground. The open question asks which conditions we shall include in our hypothesis that captures the potential. This, however, leads us back to the ontological side of the picture.

The contemporary answer to the ontological question comes from strands that associate causation with mechanisms. The strands originate in the broader domain of the so-called “special sciences” – an extensive range that encompasses as many different disciplines as biology and sociology – from scholars’ discontent with the Hempelian “received view”. The covering law model stipulated that laws are required to endow explanatory conditions with the capacity to explain occurrences, establishing standards that many scholars saw as unreasonable or beyond the scope of their research practices.

A variety of philosophical contributions has elaborated on this discontent and developed the idea of a “mechanism” as an alternative grounding to covering laws. Among them, the following three deserve a special mention, again for their broad influence beyond the domain of philosophy of science.

#### 4.1 Relevance, causality, and causation

Salmon<sup>1</sup> develops his proposal around Russell’s idea of a “causal line” as the persistence of something. His proposal, hence, understands causation in line with the Humean tradition, but with a twist. Persistence pertains to *processes* that we can observe as the state of the property of an object at different space-time points. Under this definition, a stone resting defines a process as much as a stone throwing. Causation occurs at the intersection of two processes at the space-time point  $t$ . The intersection is causal when the processes are different at  $t + 1$  and at  $t - 1$ .

Originally, the difference was understood as the transmission of a mark beyond the point of interaction. To ensure the mark is not spurious, Salmon was compelled to add the counterfactual condition that the change would not have occurred, had the intersection not happened –

---

<sup>1</sup> Salmon, 1984, 1998.

which he held in some contempt. Counterfactuals, he considered, demoted the mark principle to an epistemic criterion and again left causation impinging on the actual occurrence of interactions and modifications. In short, counterfactuals make processes dispensable in defining causation<sup>2</sup>.

In the effort to preserve the ontic priority of processes, Salmon then embraced the principle of constant or *invariant quantities* as the feature that qualifies a process as causal. Thus:

A process transmits an invariant (or conserved) quantity from A to B ( $A \neq B$ ) if it possesses this quantity at A and at B and at every stage of the process between A and B without any interactions in the half-open interval  $(A, B]$  that involve an exchange of that particular invariant (or conserved) quantity<sup>3</sup>.

The half-open interval, he specifies, allows “for the possibility that there is an interaction at A that determines the amount of the quantity involved in the transmission.” The new definition implies that “a causal process does not enter into any causal interactions”<sup>4</sup>. The causal process only denotes what happens as a consequence of interaction, and independently on the effect that it sorts out. It captures the speeding bullet that transmits the energy-momentum from the gun to the target. It renders the notion of *propagation* as separate from that of *production*. From this viewpoint, therefore, mechanisms are causal interactions that produce a change in the structure of the causal processes, which later propagate across the space-time region until it delivers.

Salmon’s proposal is relatively silent on how to establish the propagation process, except for the requirement of tracing it in its actual shape. More attention is paid to the relationship between productive interactions and propagating processes, which is empirically validated when the interactions provide a *homogeneous partition* of a *reference class*.

Salmon<sup>5</sup> understands a reference class *S* as a collection of objects identified by some attribute of interest. Its *partition* yields *cells* or “sub-classes that are mutually exclusive and exhaustive within *S* – that is,

---

<sup>2</sup> Kitcher, 1989.

<sup>3</sup> Salmon, 1998:257.

<sup>4</sup> Salmon, 1998: 258.

<sup>5</sup> Salmon, 1989:63 ff, 1998.

every member of  $S$  belongs to one and only one subclass of the partition.” A partition is *relevant* to an attribute  $P$  if the heterogeneity of  $P$  in each cell is different from the heterogeneity in any other cell. Then, the reference class  $S$  is homogeneous to  $P$  when no relevant partitions can be made of it.

The language in Salmon’s proposal is probabilistic and Bayesian. The reference class provides the prior probability  $p(P|S)$ , and the cells from its partition by attribute  $C$  provide the posterior probabilities  $p(P|SC)$ ,  $p(P|S\bar{C})$ . The relevance of  $C$  to  $P$  in  $S$  is decided by comparing the posterior probabilities in each cell and to the prior and is proven when a cell displays a different conditional probability than any others. As a consequence, the symmetry that was typical of law-like generalizations is violated.

That which does the explaining, hence, is the attribute  $C$ , or the intersection of attributes  $\Phi$ , that constitute the ultimate homogeneous cell to which no further partition can prove relevant. The proof of relevance consists in the fact that any new condition  $G$  satisfies the equation in the left side of [17]:

$$[p(P|S\Phi G) = p(P|S\Phi)] > p(P|S) \quad [17]$$

The left side of [17] renders the Markov condition and indicates that  $\Phi$  “absorbs” or “screens off” the relevance of any  $G$  to  $P$  in  $S$ . An explanation, hence, is the conjunction of attributes that identify a group as the ultimate homogeneous partition in the reference class. In practice, with Salmon’s example, the answer to the question “why did those youngsters steal a car?” appeals to those relevant conditions  $\Phi$  that sort the group  $P$  of car thieves out of the reference class  $S$  of teenagers. The answer is *complete* when no further conditions can be added to  $\Phi$ , as even the relevant ones would be screened off by the conditions in the explanation.

Completeness and relevance equip us with the criteria to pinpoint credible explanatory conditions. These are the “proximate causes” that absorb the entire relevance of more remote causes to a phenomenon. The criteria can guide our search for relevant conditions, as well as our decision to end our search. With a reasoning that recalls Mill’s second mode of explanation, Salmon maintains that

A complete set of factors statistically relevant to the occurrence of a given event can be found by examining the interior and boundary of

an appropriate neighboring section of its past light cone. Any factor outside of that portion of the cone that is, by itself, statistically relevant to the occurrence of the event in question is screened off by events within that neighboring portion of the light cone<sup>6</sup>.

Completeness also ensures that the explanation is only made of relevant conditions. Thus, it provides a solution to the famous riddle of the barometer, that a frequentist confirmatory approach, like any simple enumeration, would mistake for the cause of the storm. The Markov condition in [17] would prove instead that the conditional probability of the storm given the reading of the barometer and the pressure is the same as the probability of the storm given the pressure alone: therefore, a barometer is not a relevant part of the causal explanation of the storm.

Salmon does not recognize the set  $\Phi$  any metaphysical import. Relevance is relative to our knowledge of the reference class, the process, and the candidate initiating conditions that we consider for testing. An explanation, therefore, is always *bona fide*, given the limits of our knowledge. He also acknowledges that his principle of relevance, understood as an invariant relation of quantities, brings the explanation back in the riverbed of law-like regularities – so, at a short distance from early Hempel’s and, before, Mill’s explanation. However, in Salmon’s perspective, the invariance of conditions is as local and relative as the actual processes that they explain. What truly matters, in Salmon’s account and to the broader debate on explanation, is the underlying ontology of causation. The causal processes of propagation are the ontic spindles that, on the one side, confer relevance to initiating conditions and, on the other side, make the effect happen. This view suggests an essential distinction between causality and causation.

Salmon’s proposal entails that firing a bullet, the speeding bullet, and the shot target are three distinct parts of the problem. Causation can be grasped as the change in the state of the target due to the firing and explained by the intersection. Causality, instead, is the invariance of the speeding bullet that propagates the force unleashed when firing. Salmon draws attention to the fact that the causality generated at the intersection of the starting conditions may or may not end with the effect of interest, hence “forks” alternatives. However, he never frames them as possibilities, only as forks capable of experimental interpretation. These forks end in a different way to different individuals of a specific reference class. That which we understand as the effect only is the recognition

---

<sup>6</sup> Salmon 1998: 121.



that these responses may be of the same kind, which again can be accounted for by some combination of relevant conditions. These conditions can be found by testing their relevance to the process – ideally, for each fork.

In Salmon’s view, explanations require both statistically relevant conditions and a connecting causal process – else, the explanation would be wanting. Salmon recognized priority to causal processes over conditions as, like Hempel before him, he maintains the propagation confers conditions their explanatory standing. Moreover, the process of propagation remains the key for understanding “how the world works,” which, to Salmon, is necessary to scientific knowledge. Causality, then, is an intensional matter; interventions and counterfactuals, which are extensional instead, are far from required to characterize it – although they improve our grasp of causation.

## 4.2 Mechanistic models of causation in philosophy

The focus on causation, interventions, and counterfactuals is the hallmark of the “mechanistic model” perspective. This again takes its move from the discontent with Hempel’s requirement of universal generalization as the grounding of explanation and predictions alike. However, the mechanistic model approach disagrees with Salmon on the ontic priority of some individual eventless “causal rope” pulled between the generative intersection and its end results. The disagreement is especially evident in the influential contribution of Woodward<sup>7</sup>.

Woodward contends that Salmon’s focus on actual single processes can have any currency beyond the domain of classical physics and beyond simple systems. Faced with

large numbers of interactions among many distinct fundamental causal processes [...], it is often hopeless to try to understand the behavior of the whole system by tracing each individual process. Instead one needs to find a way of representing what the system does on the whole or on the average, which abstracts from such specific causal detail. [...] In this treatment, such characteristically “epistemic” or “inferential” concerns as finding techniques for actually solving the relevant equation governing this aggregate behavior and for avoiding computational intractabilities are of quite central importance. Rather

---

<sup>7</sup> Woodward, 1989, 2003.

than merely mirroring facts about causal interactions and processes the relevance of which for inclusion in the explanation is determined on other grounds, such epistemic considerations seem to have an independent role in determining why this sort of explanation takes the form it does<sup>8</sup>.

Moreover, Woodward considers that giving priority to seamless processes posits inescapable questions about the amount and type of evidence required to support something more than a singular or an existential claim that  $\Phi(x)$  makes  $S(x)$  be  $P(x)$ . The focus tapers the analysis and puts some desirable goals out of reach. For instance, based on single processes, we may not be able to answer why that particular process took place instead of others that were equally possible. Addressing such alternative “what-if” questions is nevertheless desirable in Woodward’s eyes, as these responses reduce the sense of contingency while pointing out that similar processes can be expected under a variety of initial conditions. Besides, the focus on singular processes may prevent our attempts to consolidate dispersed knowledge under one consistent framework, as the concept of mechanism instead can and should do.

Woodward conceives of a mechanism as a *model*, that is, an image that captures macroscopic patterns or regularities by abstracting away from the details of single occurrences. Thus, mechanistic explanations first come in the shape of a theory as a formal model. Construed as a reduction that mimics the aggregate behavior of the actual system, models are proven “true” when they reproduce the observed behavior – as is the case of those algorithms that, from biology to rational choice, find equilibria in the aggregate responses to selection and adaptation pressures. The mechanistic model approach, then, agrees with the intensional rendering that explanations cannot be reduced to content-free logical structures. At the same time, it departs from the intensional rendering as it understands explanation to be the solution to an equation, valid in the strict domain of the driving question about the effect, but of general space-time applicability in that domain.

The main issue that these mechanistic models have to address is the specification of a causal structure. Woodward establishes that

a necessary condition for a representation to be an acceptable model of a mechanism is that the representation (i) describes an organized

---

<sup>8</sup> Woodward 1989:362-363.

or structured set of parts or components, where (ii) the behavior of each component is described by a generalization that is invariant under interventions, and where (iii) the generalizations governing each component are also independently changeable, and where (iv) the representation allows us to see how, in virtue of (i), (ii) and (iii), the overall output of the mechanism will vary under manipulation of the input to each component and changes in the components themselves<sup>9</sup>.

At the core of this image lies an arrangement of elements, each of which displays some observable pattern of behavior, and such that all contribute to the output.

These patterns do not correspond to any law; instead, they are empirical relationships. They bear counterfactuals in the classical, experimental understanding, but apply the principle to Markovian structures. Thus, an acceptable mechanistic model is that in which an intervention on the value of the  $k$ th elements  $X_k$  results in the change of the value of the output  $Y$ , and the change only passes through  $X_k$  and the further elements that are “causally between”  $X_k$  and  $Y$ . According to this definition, a change in the value of the barometer effected by a change in atmospheric pressure does not count as an intervention that proves the barometer has a “causal betweenness” to the storm. An intervention on the barometer, instead, is the attempt at making the storm more or less likely by intervening on it in insulation. The fact that the manipulation of the barometer in a vacuum chamber does not affect the weather is observed as the vanishing correlation between the reading of the barometer and the occurrence of the storm as the observations cumulate.

The causal process, in short, manifests as the correlation that we can make vanishing to disprove something is not causal, or relevant, to the effect of interest under repeated manipulations. Woodward acknowledges that even a real causal relationship may be affected by some disruptive interference; however, it is expected to hold on average under a variety of interventions. For instance, a block sliding along an inclined plane would keep sliding according to the empirical regularities of gravity and friction regardless of average changes in the width of the surface, the weight of the block, or greasing.

The disproving strategy of betweenness proves more useful than a confirmatory tracing strategy, Woodward maintains, because it can unravel complex situations. Counterfactually, we can discover that a kind of protein intervenes on the operations of an enzyme by inhibiting its

---

<sup>9</sup> Woodward 2002:S375.

blocking agent when the double prevention proves robust under a variety of changes.

The counterfactual tests of betweenness, hence, are desirable as they yield information that we can employ to intervene on the causal process beneath an undesired effect.

Invariance [under intervention] thus has the virtue of capturing the idea that what really matters to whether a generalization describes a causal relationship is whether it describes a relationship that is potentially exploitable for purposes of manipulation<sup>10</sup>.

Manipulability compels Woodward to introduce the assumption that the mechanism is made of independent elements and modular connections, which allow “one to trace out the consequences of possible changes in any of them for the overall behavior of the system.” The module can be conveniently assigned a residual effect on the output, support intervention, and be controlled. The major problem that a modeler has to solve consists in identifying the right *decomposition* of the process to the effect. Modularity calls for criteria for establishing the boundaries of any independent element of the mechanism inductively.

Woodward’s model provides an essential map of a mechanism as a consistent system, geared toward explaining its behavior by decomposing – hence, again, reducing – it into recognizable functioning parts that together contribute to yield the effect. The explanation by mechanistic modeling, then, recalls Mill’s first and, possibly, second method of explanation, as in [11] and [12.a], [12.b]. In no sense, however, Woodward’s models can be equated to a “true” image of a complex system – and purposefully so. Instead, his reductions are designed to offer sufficient information to act upon phenomena and change them.

### 4.3 Explanatory mechanisms and their requirements

Salmon’s and Woodward’s approaches to mechanisms may be portrayed as rivals. The one is confirmatory, intensional, and oriented to singular causation; the other is disconfirmatory, extensional, and geared to simulate whole systems. The two provide different responses to the same dissatisfaction for the covering-law model of explanation: Salmon’s account aims to portray the propagation process as close as

---

<sup>10</sup> Woodward, 2002: S371.

possible to the real one, while Woodward develops a map that statistical equations can render and make interpretable. However, both depart from the classical syllogistic construction to emphasize the probative value of observational reports as such – in Woodward’s proposal, as detached as possible from contextual conditions. Nevertheless, both attempt to capture the potential in the observable unfolding of activities.

Unsurprisingly, the contemporary philosophy of science has tried to reconcile the two views<sup>11</sup>. Again for its influence on methods and research, a special mention deserves a late mechanistic strand in the philosophy of science – the eclectic synthesis developed as a reflection on biological and neurobiological explanations.

Within this strand, Craver (2006, 2007) portrays explanatory mechanisms as the structure developed to systematize separate pieces of knowledge and develop a picture for both understanding and manipulation. His mechanisms are “constitutive explanations,” in which “a property or behavior of a system is explained by the properties and activities of its parts.” The underlying assumptions maintain that the interactions of the constitutive entities at the lower-level at time  $t_1$  arise higher-level effects that, in turn, constrain the interactions of the same constituent entities at time  $t_{1+n}$ . In this view, mechanisms offer a comprehensive model of causation as a local order, roughly captured as an input-throughput-output system process that organizes the world. The epistemic question, following the debate between Woodward and Salmon, asks which desiderata – which criteria – a model should meet. Craver equates modeling with finding a function or an algorithm that mirrors the worldly order to be ascribed to the mechanism so that the input-output patterns are close enough to what is observed – at least in their relevant aspects, although in all the relevant ones. The problem that modeling poses to the researcher is that of information loss:

Few models are isomorphic with the phenomenon, given that models typically abstract away from the precise details of the system being modeled, that they typically are only approximate, and that they make simplifying assumptions in order to apply a particular formalism<sup>12</sup>.

Then, different models can be developed of the same phenomenon to render it at different levels of detail. A coarser one accommodates a

---

<sup>11</sup> e.g., Friedman, 1974; Kitcher, 1989.

<sup>12</sup> Craver 2006: 357.

broader range of diverse manifestations, although it dismisses more information about their actual diversity under the assumption of irrelevancy. More fine-grained characterizations, instead, provide richer information yet only apply to few or single phenomena. However, Craver maintains that phenomenal accuracy and explanatory power are independent of one another. As such, the two criteria set two different desiderata – namely, *robustness* and *plausibility*.

a) *Desiderata*

To Craver, the explanatory power of a model depends on whether it can render the unfolding of a phenomenon under normal as well as extreme conditions – that is, on whether it can map actual outputs even from degraded inputs or damaged units. A model that only maps the functioning of “the system” under normal conditions is not *robust*; hence it is not explanatory – although it can serve the purpose of prediction. To him, “this is the take-home lesson of the several decades of attack on covering-law models of explanation at the hands of advocates of causal-mechanical models of explanation”<sup>13</sup>.

*Explanatory robustness* and *predictive robustness*, therefore, are assessed on different grounds – possibly, on opposite ones. The purpose that explanatory models serve, however, is no less significant than that of predictive generalizations. Both allow establishing how a system behaves under a variety of contingencies so that this information can support mindful decisions about intervention. Robust maps of phenomena include the “causal organization of the system” and its possible changes under different scenarios<sup>14</sup>. The causal organization, Craver maintains, is rendered as soon as the model includes “all the entities, properties, activities, and organizational features that are relevant” to preserve the functioning of the system across circumstances<sup>15</sup>. Relevance, therefore, depends on the driving question and the scope of the model; explanatory robustness, instead, guards against idiosyncratic bias in mapping.

Explanatory robustness is not the sole criterion that a model shall meet. To be mechanistic, the model should refer to components and activities that mirror the purported components and activities of actual phenomena. In short, the model should provide a plausible portrait of the particular region of the world that it maps. *Plausibility*, however, is a goal that can be reached by successive approximations from starting

---

<sup>13</sup> Craver 2006: 357.

<sup>14</sup> Craver 2006, 359.

<sup>15</sup> Craver 2006, 359.

idealized conjectures – as it has been the case of the atom model, the solar system model, or the neural network model among many others. In the case of mechanisms, the approximation strategy yields different types of accounts in response to different versions of a how-question.

b) *Model specification and how-answers*

The strategy takes a definition of the *explanandum* as its starting point. The definition comes as an idealized “how-possibly” model shaped as a hypothetical *sketch* that specifies the primary connections and inserts the “black-boxes” that justify their emergence and direction. The accurate point of arrival, instead, is a “how-actually” model that provides “ideally complete descriptions of a mechanism” and includes “all of the entities, properties, activities, and organizational features that are *relevant* to every aspect of the phenomenon to be explained”<sup>16</sup>. In between, several “how-plausibly” models provide *schemas* that satisfy pragmatic needs. *Schemas* are mechanistic explanations in which known details have been intentionally abstracted away, and that, therefore, are similar to a testable theory in shape and function<sup>17</sup>.

The passage from lower to higher accuracy “often proceeds by functional analysis, reverse engineering, homuncular explanation, and decomposition”<sup>18</sup>. The passage unfolds along intertwined, although analytically separate lines.

On the one side, the focus is on the *conditions* that shape the phenomenon. The knowledge of a mechanism improves with the understanding of the standard circumstances under which the event of interest can be detected. A better picture of the working of the system also includes the *impeding* conditions under which the mechanism fails. Further details are added when the model accounts for the *modulating* conditions – which, in other contexts, are understood as the moderators, corresponding to those background conditions that affect the intensity of the output. The picture becomes almost complete when these conditions are identified to render non-standard circumstances – those that would never occur in an everyday observational setting. Eventually, the full picture emerges when the byproducts and the side-effects are clarified.

On another side, the focus is on the *entities* that are responsible for the activities resulting in the phenomenon of interest. Here, the passage

---

<sup>16</sup> Craver, 2006:360, emphasis added.

<sup>17</sup> Kaplan, 2011: 348; Machamer, Darden and Craver, 2000: 15 ff.

<sup>18</sup> Craver, 2006: 362.

from idealized possibilities to actual representations occurs when the “real components” are identified. Components are “stock-in entities [...] when they exhibit a stable cluster of properties” that indicate their subtypes and the associated causal powers of interest<sup>19</sup>.

The issue is strictly related to the further analytical dimension, which narrows on the space-time organization of the entities in causal structures and dependencies. These objects are construed within the scope and the limit of technical choices, in which respect Craver recommends an eclectic approach, to generate as many possible maps as techniques are available for the sake of triangulation and validation.

In the end, Craver portrays a mechanistic model as an ongoing endeavor that ideally ends when a detailed stable picture is produced of the phenomenon. His understanding of explanation is maximalist, as the ideal model includes all the classic *aitia* as established by a multiplicity of testing techniques, plus their descriptions and connective processes, so to provide answers to any question about a phenomenon – the what, the how, the where and when, as well as the why.

c) *Salmon-complete constitutive mechanisms*

As a last note, in his more recent contributions with David Kaplan, Craver operates a correction of some earlier position while addressing the question of which details really matter to constitutive mechanistic models<sup>20</sup>. These last works specify the requirements of a “successful” explanatory model as follows:

- (a) the variables in the model correspond to components, activities, properties, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifiable) causal relations among the components of the target mechanism<sup>21</sup>.

However, they also emphasize that

the idea of an ideally complete how-actually model, one that includes all of the relevant causes and components in a given mechanism, no

---

<sup>19</sup> Craver, 2006: 370.

<sup>20</sup> Kaplan and Craver, 2011; Craver and Kaplan, 2018.

<sup>21</sup> Kaplan and Craver, 2011: 611.



matter how remote, negligible, or tiny, without abstraction or idealization, is a philosopher's fiction. Science would be strikingly inefficient and useless both for human understanding and for practical application if it dealt in such painstaking *minutiae*<sup>22</sup>.

So they posit abstraction and, implicitly, mechanistic *schemas* as the unavoidable feature of some usable scientific knowledge of phenomena. They insist that these models cannot prioritize epistemic factors such as description, prediction, or confirmation at the cost of the structural constraints, as the dismissal of the latter yields a faulty knowledge about barometers accounting for the storm. They emphasize that constitutive mechanistic models do not provide the "aetiological" explanation of chained causation. Instead, they render the arrangement that links the phenomena and its basic elements<sup>23</sup>. In short, mechanistic models are those that capture the structure beneath chemical causation. They contend that chemical causation can be "scientifically" rendered as a seamless flow of details, as the seamless flow would obfuscate the structural constraint that makes the model suitable for scientific purposes. Neither models of the flow of activities nor model describing entities would answer the question of why the system produces the output that we observe. The model that counts as an explanation is that which selects particular facets and connects them into the structure ensuring the different activities will yield the output. Purposeful omissions are the epistemic feature that confers models their explanatory value. The point solicits criteria to establish which details have to be included in the model.

Craver and Kaplan identify a criterion in the definition of completeness that Salmon applied to elucidate his concept of relevance, decoupled from Bayesian formulations. A model, along this line, is "Salmon-complete" when it includes all and only those relevant proximate conditions that somehow absorb the relevance of any other. The Salmon-complete model is that which partitions the reference base into cells displaying an "unimprovable" degree of homogeneity at best of current knowledge.

Although, again, Craver and Kaplan maintain the criterion of Salmon-completeness can establish an ideal that real schemata may not satisfy, they add a further consideration about the nature of the elements of a mechanistic model that can push our modeling attempts closer to

---

<sup>22</sup> Kaplan and Craver, 2011: 609–10.

<sup>23</sup> Kaiser and Krickel, 2016.

that ideal<sup>24</sup>. They consider that, within a knowledge horizon, a better mechanistic model is not that which provides more answers to counterfactual “what if” questions, as Woodward suggested. Neither it is that which answers to a higher number of actualist “how” questions. The priority of structural arrangements gives center stage to questions of a third kind, that they label “*r*-questions,” asking what role a particular item plays in the mechanism – its fit in the puzzle.

A scientist who knows many facts about action potentials but does not know that action potentials play a role in neural signalling is missing an important filter on which details are relevant. The components of the action potential only hang together as such because we appreciate that the action potential is a significant feature of neuronal behaviour. No description of the molecular details could be organized into a causally productive unit without taking this higher-level perspective on the bewildering causal mayhem of the neuron and its components<sup>25</sup>.

The *r*-answers learn from how-answers and prove their standing by responding to what-if questions. However, they are shaped to retain those relevant details that, alone, make an element fit into the functional *schema* of a mechanism producing an effect – that is, into our hypothetical generative structure.

This is where QCA enters the picture as a technique capable of modeling Salmon-complete generative structures without appealing to (Bayesian) probabilities.

---

<sup>24</sup> Craver and Kaplan, 2018:24 ff.

<sup>25</sup> Craver and Kaplan, 2018:26.

Part II

THE STRATEGY AND THE TECHNIQUE

*intentionally blank page*

## 5. Configurations and *inus* causation

The recent literature has compelled a change in the logical structure of explanations. The traditional place of the major premiss, which the modern philosophy of science occupied with regularities with a universal scope, is now taken by mechanisms. The change, on the one side, revives the original understanding of explanation as reasoning about the actualization of potentials. On the other, as Mill already noted, it blurs the boundaries between induction and deduction.

Successful explanations of outcome regularities literally entail a hypothesis about the structure that arranges entities and activities so that they “bring about” the effect. Observation reports are employed as the minor premiss to confirm, or disconfirm, the hypothesis, following long-established logical rules that, however, are differently invoked depending on the language. The hypothesis, in turn, is a schema that no longer makes the case of a difference between causes and conditions: all are understood as parts of the local intersection that arises the right process of propagation to the effect and all contribute to making the effect sure or impossible. Still, the hypothesis needs to be specified in-line with some criterion of relevance. Salmon-completeness provides a useful principle, which interestingly chimes in with the concept of *inus* causation recognized by the broader configurational scholarship<sup>1</sup>.

### 5.1 *inus* causation

Recall from Part I that Mill proposed a redefinition of causation as a compound of irreplaceable and replaceable constituents under which,

---

<sup>1</sup> Ragin, 2000:11, 2008; Goertz 2002; Baumgartner, 2009; Grofman and Schneider, 2009; Schneider and Wagemann, 2010, 2012; Rohlfing, 2012; Hackett, 2015; Schneider and Rohlfing, 2016; Duša, 2019; Haesebrouck, 2019.

and only under which, the effect inevitably occurs in the absence of hindrances.

The redefinition encompasses the whole fact of causation, without distinctions between causes and conditions. Analytically, the compound captures all the tendencies as empirical laws connecting each component to the effect and turns the causal power into a local occurrence. Be it true and complete, the compound shall yield evidence from the reference class about two types of observation sentences, the one about the conjunction of the full compound and the effect, and the one about the conjunction of a partial or null compound and the absence of effect, without exceptions. The shape of the relationship, then, corresponds to that of a logical *iff* connective, establishing the identity of two terms that licenses *barbara*-type of inferences. The relationship, in short, is philosophically “necessary” – something that has to occur and cannot occur in a different shape.

### 5.1.1 *Necessity, sufficiency, and vending machines*

Mackie considers that the concept of philosophical necessity has been loaded with a variety of meanings over the centuries, but that two are of particular interest for a better grasp of logical causation<sup>2</sup>. He famously illustrates the two concepts with the example of the three imaginary vending machines named K, L, and M. In his fiction, all the machines sell chocolate bars for one shilling each. Machine K always and only drops one bar after the shilling. Machine L, too, requires a shilling to drop a bar, although sometimes it fails. Machine M, again, drops the bar after the shilling, although it may also eject bars with no apparent reason.

Any of the three machines, then, affords the four possible types of observations portrayed in Figure 5, and that in this context read:

- (a) coin and bar,
- (b) coin and no bar;
- (c) no coin and bar,
- (d) no coin and no bar.

and the different functioning of each machine is captured by the special observation reports **R** that it allows.

---

<sup>2</sup> Mackie, 1974.

The observation report from Machine K is  $\mathbf{R}_K: \{(a) \wedge (d)\}$  – whenever the coin, the bar follows; without the coin, the bar certainly does not drop. The functioning of Machine K, hence, supports the inference that  $\mathbf{H}_K: coin \equiv bar_K$  – that is, a relation of *necessity* and *sufficiency*. Following Mill, however, it should be noted that the defining part from the observation report  $\mathbf{R}_K$  consists less of what is observed than of what is missing. Machine K never yields the two possible types  $(c)$ ,  $(d)$ ; to claim that  $\mathbf{H}_K$  holds unconditionally true, we should rule out that the two observation sentences are possible. In short, we should establish Machine K is failsafe and meets Mill’s requirements R3 and R4. This becomes possible in two different ways. First, we can prove the claim for the reference class and the time of observation – which already can be a remarkable result. Else, we can identify some feature of Machine M that makes the bar dropping unavoidable once the coin is dropped.

In Machine L, not putting the coin makes the dropping impossible; putting the coin makes the dropping possible, although not certain, as some coin can fail. The observation report from Machine L, then, is  $\mathbf{R}_L: \{(a) \wedge (b) \wedge (d)\}$ . The report supports the inference that the relationship is of logical *necessity*, and  $\mathbf{H}_L: coin \subset bar_L$  in the slice of the world under analysis. The claim, however, is better supported when  $(c)$  is ruled out as impossible, and Mill’s requirement R3 is satisfied.

In Machine M, the coin will certainly yield the bar, although there may be free droppings. The observation report, therefore, is  $\mathbf{R}_M: \{(a) \wedge (d) \wedge (c)\}$ , supporting the hypothesis that the shelling is the *sufficient* condition for dropping the bar, and  $\mathbf{H}_M: coin \supset bar_M$  holds. This, again, can be proven within the limited space-time region of our observations; however, it would better be credible had we information that Machine M makes  $(b)$  impossible so that Mill’s requirement R4 is satisfied.

Mackie maintains with Mill that Machine K is ideal as it is fully intelligible. Even if it eventually needs maintenance and degenerates into an L-type, Machine K can be fixed as the source of failure can be located in some defective component. However, a K machine as an irreplaceable, surefire, *single* mechanism to the effect is seldom given due to the possible alternative channels to the effect. The common occurrence in the real world, then, is of a K-machine into an M-machine – a system with many channels of which we only know the one with the coin. This knowledge, however, may prove enough when we can be reasonably sure that we have identified a trigger and ways to fix its failures. Without such minimal knowledge, L-effects overlap M-effects,

and we cannot recognize any causation. The issue with mechanisms, once and again, is that they are out there, but our grasp is based on our intuition of their functioning based on patterns in behavior. As such, the knowledge we have of them is tentative and limited – but it can improve by cumulation.

### 5.1.2 Full explanations

Now, Mackie maintains that, despite our limits, we can aim at a full explanatory schema of some phenomena<sup>3</sup>. He illustrates his point with the well-known example of the burned building<sup>4</sup>.

He notices how easily we can buy any standard conversational explanation of the sort ‘the warehouse burned out because of a short-circuit’. However, he emphasizes, the explanation is far from compelling. A bit of formalization can better expose its weaknesses.

Let  $C$  indicate the short-circuit and  $B$  the burning. The underlying hypothesis about the burning,  $\mathbf{H}_B$ , then reads  $\mathbf{H}_B: C \supset B$ . The reasoning that supports the explanation runs as follows: if  $C$  then  $B$ , my observational report includes  $(a): C \wedge B$ , which confirms the hypothesis – therefore  $B$  because of  $C$ . So displayed, we realize why the explanation is wanting: the observational report is too poor to support the claim that the hypothesis about the burning accounts for the case of the warehouse.

#### a) Checking for plurality

The claim is about singular causation, and we can both falsify and verify it. The observational report  $\mathbf{R}_H: \{(a)\}$  can contribute to the validation of  $\mathbf{H}_B$ : but the convincing report for verification also requires that we consider  $(d)$ -type of observations. In short, as our hypothesis is about singular causation and aims to ascription, we cannot hold it true unless we are sure that  $C$  is irreplaceable.

The claim of irreplaceability, however, stands against the backdrop of a theory about the role that the factor can play in bringing about the outcome, plus some reasons to believe that no relevant alternative factor was available at the same time in the same place that could have played the same role. To keep on with the example: we may maintain that short-circuits are relevant to fire as a source of heat. Then, we have to

---

<sup>3</sup> Mackie, 1965.

<sup>4</sup> The reasoning closely recalls the strategy of adjudication in case studies developed by Van Evera, 1997, the contributors to Bennett and Checkel, 2015; Fairfield and Chapman, 2018.



consider that alternative sources do exist, in the form of unattended pans on a stove, careless smokers, bolts, and the like. Hence, a better hypothesis about the fire should read that the short-circuit  $C$  or some alternative source of heat  $\Phi$  caused the fire  $\mathbf{H}_{B1}: C \vee \Phi \supset B$ . We then can look for something fitting the observational sentence ( $a'$ ):  $\Phi \wedge B$  in the context of the warehouse. If we can rule out  $\mathbf{H}_{B1}$ , we are allowed to maintain that, counter-to-the-facts, without the short-circuit, the fire would not have been. The counterfactual consideration adds the unobserved yet plausible type ( $d^*$ ) to the report  $\mathbf{R}_{H1}: \{(a) \wedge (d^*)\}$ . The new and more convincing version of the explanation reads that 'the warehouse burned down because of the short-circuit – and this is true because nothing else happened that could have done it'.

This story, however, is not complete yet. All that the report  $\mathbf{R}_{H1}$  says is that the short-circuit *could* have produced the fire under the circumstances. As such, eventually,  $\mathbf{H}_1$  is a statement about a potential that may or may not actualize. On this ground, we can only maintain that the burning was more likely in our warehouse than in a twin building where the electrical system was safe.

#### b) *Checking for Constitution and Composition*

To connect the potential to its effect, we have to clarify whether ( $b$ )-type observations were impossible under the circumstances. Here, Mill and Mackie suggest, two different orders of considerations enter the picture.

The first concerns constitutive issues, and calls for attention to the complements to  $C$  in making  $B$  unavoidable. Again, the story improves if we can invoke a theory about what is needed to make the effect happen for sure. The theory would suggest that the short-circuit fails unless it meets fuel,  $F$ , and oxygen,  $O$ . Both, then, had to be present in the warehouse, or the fire would not have been. The second consideration draws attention to the hindering factors – the ones that, Mill said, have to be absent for the compound to obtain. Thus, Mackie warns, the explanation is again incomplete unless we bring a broken sprinkler  $\bar{S}$  into the picture.

Eventually, our improved hypothesis reads  $\mathbf{H}_{B2}: (C \vee \Phi) \wedge (F \wedge O \wedge \bar{S}) \supset B$ . In compiling the observation report,  $\mathbf{H}_{B2}$  compels us to verify that  $F \wedge O \wedge \bar{S}$  were all jointly present in the expected state. Having observed the fuel, the oxygen, and the broken sprinkler, our observational report again reads  $\mathbf{R}_{H2}: \{(a) \wedge (d^*)\}$ , but now we are sure that the missing instances ( $b$ ), ( $c$ ) do not appear in the report as they were

not part of the state of the world – hence, their observation was impossible under the circumstances. As a side note, Hempel did not register the difference between a wanting observation report and one that has dismissed the logically impossible types of observations as actually absent in the situation, which conveys ambiguous information about the empirical standing of the hypothesis. As we will see, the technical protocol of QCA avoids such an ambiguity through its “truth table”, which compels the researcher considering every logically possible observation sentence from a reduction sentence.

The complete convincing explanation, then, specifies that ‘the warehouse burned down because it was crammed with waste tires and leaking barrels of exhausted oil, and the sprinklers were broken when the short-circuit happened – but, were it not been for the short circuit, the burning would not have happened’.

### 5.1.3 Principles of *in*us composition

Together with Mill, Mackie<sup>5</sup> maintains that a “cause” is a structural arrangement and consists of the configuration of several elements in the world. Every disjunct in the disjunction  $C \vee \Phi$  constitutes an alternative element that identifies a machine of causation. Each machine is a minimally sufficient compound: as far as we can rule out its failure – which can be done by discounting obstructions to the “propagation” of the causal process – the compound will certainly yield its effect. The cause, therefore, is the conjunction of replaceable and irreplaceable components  $C \wedge O \wedge F$  that, jointly given and unimpeded, makes the occurrence of the effect unavoidable. The certainty of the effect allows the understanding that the components entertain a constant relationship with each other and with their effect. The relationship, however, defies the standard arithmetic rendering that would reduce their interactions to a sum of quantity. Necessity and sufficiency, Mackie advances, provide a better grasp of the compound. This casts new light on the fire example.

First, the single factors in the explanation are “partial causes.” They cannot obtain without the compound, hence they are individually *insufficient* to the effect. However, the compound cannot obtain when even one of their components is in the wrong state. Therefore, each component is *necessary* – that is, required, nonredundant – to at least one compound on pain of failure.

---

<sup>5</sup> Mackie, 1965, 1974, 1977.

Second, that which qualifies as a “full cause” is the compound. When the conjunction is complete – in the sense that every component is given in the right state, in the right place, at the right time – then the compound is a K-mechanism of the effect and sufficient to obtain it. However, the compound includes replaceable components, which means the same effect can also be reached through alternative K-machines. In this respect, the compound is unnecessary to the effect: alternative “equifinal” arrangements may be possible – and we have to rule out the alternatives if we want to be sure that the ascription is unambiguous.

In the light of these considerations, Mackie dubs single factors in complex causation as “*inus* conditions” – the acronym standing for Insufficient yet Non-redundant parts of Unnecessary yet Sufficient compounds. The acronym summarizes the composition rule of a “nomological” complex as the individual necessity and joint sufficiency of the components to the effect.

*Inus* causation calls for two different types of questions. The first asks which mindset supports conjectures that fit the shape of *inus* causation. This will be the topic of the next chapter. The second question asks which language equips us for the analysis and the refinements of these conjectures. This will engage us in the remainder of this part.

## 6. Imagining machines

The lesson from contemporary mechanistic literature is that even an elliptic explanation about a single factor makes sense to us as long as we have a theory about the generation of the phenomena. Even if we ignore whether the single factor was the only “anomaly,” or whether the context was “structurally disposed” to make the effect happen, the information is credible as the factor has some property, or capacity, that candidates it to play a role in the mechanism. These abstract schemes offer the starting point of an explanatory configurational analysis.

### 6.1 The relevance of stage setting

At the ontological level, explanatory configurations agree with the mechanistic assumption of an abstract world where entities are endowed with some capacity to make something, which we can observe through special manifestations and can indicate as an entity’s attribute. The capacity unfolds into a causal power when effects accompany or follow its manifestation. The further and crucial ontological tenet maintains that any property remains in the state of latency until it meets the “right” conditions. This point can mark a departure from the standard approach to mechanistic causation.

Mechanistic accounts in Woodward’s tradition tend to prize an extensive understanding of robustness, and aim to model underlying mechanisms as interactions independent on background conditions<sup>1</sup>. However, as Nancy Cartwright has long argued, abstractions under the assumption of context independence may put explanations on shaky ground. If, with Mackie, Hempel, and Mill before them, we consider

---

<sup>1</sup> Woodward, 1989, 2003; Pearl 2009, 2015; Hedström and Ylikoski, 2010; Morgan and Winship, 2015.

that potential is a theoretical construct with convenient epistemic properties that, by definition, defies direct observation, then the adequate rendering of mechanisms includes the “right” conditions under which causation occurs.

Some causal powers need to be triggered if they are to be exercised; some may need facilitating even once triggered; some can be inhibited so the contribution they produce is diminished or distorted; and even [...] an interference may stop them from being exercised although all the other conditions are right<sup>2</sup>.

Cartwright’s definition renders the concept of *inus* causation from the perspective of the functions and the roles that have to be fulfilled for the unfolding of the power, as Kaplan and Craver advocate. The model that captures a power is the interplay of triggers, facilitators, and shields that, jointly given, arise a “nomological machine” – an arrangement that warrants and compels the local production of the effect as if the effect obeyed some laboratory law of nature. By definition, therefore, a nomological arrangement includes anything it takes to yield the effect with certainty. A complete nomological arrangement, rather than a single capacity or potential, justifies the expectation about the occurrence or the nonoccurrence of the effect. When complete, it explains similar effects in dissimilar settings; when incomplete, it accounts for the failure of the effect in a setting despite the single capacity was there.

### 6.1.1 *Salmon-completeness of nomological machines*

A nomological machine is a compound of *inus* conditions as both provide an antecedent in a Carnapian reduction sentence about the effect. The *inus* machine, however, also implies that we can develop a Salmon-complete image of it. In the case of the fire, for instance, we can understand the process to the fire as “combustion.” A chemist would better specify it as an exothermic chemical reaction of reduction between a substance that plays as a reductant and an oxidant, which occurs at high temperatures. This definition, however, does not capture the kernel of the decomposition. Instead, it renders the starting conditions – the compresence of a reductant, an oxidant, and high temperature – and tells us that, given heat is one of the products of the process, the reaction tends to reproduce itself unless one of its components goes exhausted.

---

<sup>2</sup> Cartwright 2017:12.

Interestingly, the *inus* components of combustion – the “reductant,” the “oxidant” – are rendered with terms that refer to the role they can play into the process. Since long, we can ascribe the meaning of an interaction to the conjunction connective, so that the result of the conjunction is more than the algebraic sum of the conjuncts. Similarly, the conditional can be given the meaning of a nomological M machine. Then, we can capture the general process with the schema *reductant*  $\wedge$  *oxidant*  $\wedge$  *temperature*  $\supset$  *fire*. In short, although we cannot pin down the reaction as a tangible object, we can identify the *inus* causes of its effect – which provide a reduction sentence of the elusive real thing. Of course, the schema can be made more specific. Alternative K machines can be built from catalogs of reductants and of oxidants as possible *inus* conditions that we can meet in the field.

Moreover, the special reductant and oxidant couple will react at a particular temperature so that the related term in the conditional can be replaced with some source of heat capable of reaching the critical temperature. We can then modulate the level of detail to gain a better grasp of the various versions of the nomological machine. However, we can establish that the representation is Salmon-complete independently on the level of detail.

A Salmon-complete nomological schema yields (*a*)-type and (*d*)-type of instances – else, it is not effective – but also prevents the insurgence of (*b*)-type and (*c*)-type of instances. Those of the (*b*)-type indicate that our schema is replaceable with some alternative mechanisms – or that our definitions of the components are not abstract enough to keep the actual alternatives within the same catalog. This, however, does not invalidate the specification of the schema. More challenging are instead (*b*)-observations. They signal our schema is Salmon-incomplete, and our knowledge of the generation of the effect is unreliable. In short, we have omitted something relevant – such as, the obstructions, like the sprinklers, or the shields that inhibit the obstructions, like the failure of the fire system. It may also be the case that we have ill-designed catalogs, following ill-designed roles, or compound. In short, (*b*)-observations are useful indications that our schema is flawed.

## 6.2 Cultivating mechanistic imagination

The formal criteria suggest a nomological machine is that which may stay dormant yet does not fail when activated, and that we can capture

it as the complete set of relevant conditions required to initiate, support, and shield some elusive process to the effect. How to get this schema as the starting point of a QCA is, as usual, a matter of cultivated imagination in drawing a configurational hypothesis<sup>3</sup>. The methodological literature of any flavor agrees that a hypothesis is a claim about a relationship between specific factors and an effect that builds on previous knowledge. The strands that commit to dis/confirmation also add that the hypothesis should be amenable to empirical probation<sup>4</sup>.

The issue concerns how we can learn from the previous literature so that existing models and findings can fit the configurational mindset and contribute to shaping a nomological machine.

### 6.2.1 *Learning from correlational literature*

The correlational literature is usually engaged in ascertaining capacities, that is, in responding to the question of whether a relationship exists between the single factor  $X$  and a purported effect  $Y$ . The choice of  $X$  and  $Y$  may depend on the discipline and the researcher's interest, or the pragmatic consideration that, were the relationship true,  $X$  could be easily manipulated to improve  $Y$ <sup>5</sup>.

The protocol, however, requires that the relationship between  $X$  and  $Y$  is treated as a hypothesis that "elaborations" with further factors can refute, corroborate, or refine. A robust stream in the correlational tradition identifies three possible elaborations of the original relationship – respectively aimed to explain it away, to interpret it, and to specify it<sup>6</sup>. Of the three, correlational studies have usually narrowed on explanation as a threat to the standing of the relationship, and on interpretation as a guarantee<sup>7</sup>.

A relationship is explained away as spurious when a confounding factor  $C$  exists, on which both  $X$  and  $Y$  depend. The confounding factor dismisses the original relationship  $X \rightarrow Y$  for a truer one  $X \leftarrow C \rightarrow Y$ . However, the substitution may neither happen – when no credible confounders are identified – nor be usable – i.e., when  $C$  is unobserved – nor be complete – i.e., when  $X$  alone still accounts for part of  $Y$ . In all

---

<sup>3</sup> George and Bennett, 2005.

<sup>4</sup> King *et al.*, 1994.

<sup>5</sup> Woodward, 2003.

<sup>6</sup> Kendall and Lazarsfeld, 1950; Lazarsfeld, 1955.

<sup>7</sup> Morgan and Winship, 2015:330 ff.; Little, 1996.

these cases, the relationship stands, although less credibly so. An interpretation can then be added that corroborates the original relationship by probing *how*  $X$  led to  $Y$  in selected cases.

An interpretation finds the intervening or mediating or channeling factors  $M$  that chain  $X$  to  $Y$  so that the original relationship is preserved and detailed as  $X \rightarrow M \rightarrow Y$ . When the mediating factors capture all of the effect of  $X$  on  $Y$ , and their chain is independent on alternative paths to the effect, they provide an identifying mechanism for the effect of  $X$  on  $Y$ <sup>8</sup>. The identifying mechanism renders the chain of events as the stages of the process through which the effect unfolds. In so doing, it supports and illuminates the original relationship, making it credible<sup>9</sup>.

The models from the correlational literature provide predictors, determinants, confounders, and mediators from an array of theoretical approaches to the effect. The suggestions can be overly rich, requiring a selection among factors. A survey of scholars' practices<sup>10</sup> pinpointed four basic strategies. The "comprehensive approach" builds hypotheses that include all the factors from all the relevant theories. The "perspective approach" selects single variables to represent major theories. The "significance approach" narrows on statistically proven variables only. The "second look" approach mixes statistically significant variables with theoretically meaningful factors that did not survive those same tests.

None of these strategies, it has been argued, yields proper configurational hypotheses. *Inus* factors and statistically significant determinants may not overlap. Collections of proven determinants do not support expectations about any particular nomological machine. Mediators leave the conditions unspoken under which their mechanism certainly holds in actual cases<sup>11</sup>. Identifying mechanisms-as-processes provide configurational analysis with at least a valuable criterion to select pertinent factors, however. Knowing the stages of combustion does not tell what made a nomological machine in the warehouse but allows us assuming that certain functions had to be performed for the fire to occur. So, it reduces the range of the local factors we should consider for explaining a particular occurrence of the effect.

---

<sup>8</sup> Pearl, 2009; Gerring, 2011.

<sup>9</sup> Little, 1996.

<sup>10</sup> Amenta and Poulsen, 1994; Berg-Schlusser and De Meur, 2009.

<sup>11</sup> Falletti and Lynch, 2009.



This literature provides vital suggestions about the dynamic underlying the scheme. A not-so-good idea, however, may be that of “currying” chains of mediators into a conjunction, as the dependencies to an effect do not amount to a nomological machine<sup>12</sup>. A complete nomological arrangement, instead, is that which provides the homogeneous partition as the “scope condition” under which the correlation holds<sup>13</sup>.

A factor from the correlational literature hence qualifies as a possible component in a mechanism if it can develop directional expectations about its contribution as a triggering, supporting, or shielding factor that jointly makes the identifying mechanism obtain.

### 6.2.2 *Learning from case-oriented literature*

QCA is closer to case-oriented approaches due to the similar engagement with complexity. The question often drives the case-oriented research asking whether the cases at hand are instances of a theory of interest, and learn from the gap between the set of features that cases display and the set of features that the theory envisages<sup>14</sup>.

The usual starting point of a case-oriented analysis is a concept, i.e., a label endowed with meaning. Any concept has an intension decided by the properties that together define the meaning and an extension as the range of entities that share those properties. Intension and extension are related, as the properties under a label change the width of its applicability – the less and more abstract the properties, the more extensive their coverage. Fruitful concepts in comparative empirical analysis are associated with a stable set of unambiguous attributes, which preserves the identity of the concept across time and contexts and allows deciding on the nature of actual cases properties<sup>15</sup>. When a theory exists that attaches particular expectations to a concept, the evidence from the case establishes whether the case is an instance of the concept. The marks that causation leaves in the case, or its changes over time, provide evidence of the empirical standing of the theoretical expectations<sup>16</sup>.

---

<sup>12</sup> Baumgartner, 2013; Rihoux and De Meur, 2009; Damonte, 2018.

<sup>13</sup> Walker and Cohen, 1985.

<sup>14</sup> Ragin, 2000; Ragin and Becker, 1992; George and Bennett, 2005.

<sup>15</sup> Sartori, 1984; Collier and Mahon, 1993; Goertz, 2006.

<sup>16</sup> Elman, 2005, Fairfield and Charman, 2017, Bennett and Elman, 2006; Rohlfing, 2012; Beach and Pedersen, 2016.

The strategy of ascription may be different, however, depending on how the set of core properties is defined. The methodological scholarship discriminates at least between cores as ideal types and as classification types<sup>17</sup>. The ideal-type follows from a purposeful selection of properties that may have been abstracted from exemplary instances yet may seldom, if ever, be observed jointly, or with the intensity entailed by the concept. Adjudication against an ideal type consists in establishing the relative distance, or deviance, of a case from such an ideal pole.

In the Weberian tradition, ideal types may also portray a fictional twin world in which a property of the concept is given in the wrong state, to serve as a counterfactual mind experiment that emphasizes the relevance of that property to the meaning of the concept. Thus, we may prove irrationality is constitutive of the concept of financial panic by imagining what would have happened in 1929 had people reacted rationally.

The counterfactual application of ideal-types leads us to classification types. Classification types follow from the sociological tradition of the “theories of the middle range” – images developed for thinking about some behavior of particular classes of entities, and about the mechanisms it entails<sup>18</sup>. The theory specifies the attributes that identify the class of entities where the mechanism is expected to obtain. The attributes define a “property space,” originally visualized as multi-story contingency tables; the type arises by their reduction – i.e., the dismissal of some attribute based on pragmatic, statistical, or logical consideration<sup>19</sup>. In contrast to ideal types, cases “belong to” the types that result from such a reduction.

The typological literature pinpoints attributes and properties that naturally suit the configurational mindset. The issue rather arises when instances are attributed *ad hoc* features, as it may happen in the exceptionalism literature, or with any thick descriptions of processes and situations in single cases. In this regard, Verba<sup>20</sup> suggests three operations to learn systematically from these portrayals, which also apply to the construction of configurational hypotheses – as follows.

1. When a single factor is invoked as a relevant condition to an effect in a particular case, the factor should be included in the hypothesis.

---

<sup>17</sup> Collier and Mahon, 1993; Elman, 2005; McKinney, 1969; Capecchi, 1968.

<sup>18</sup> Merton, 1968: 43.

<sup>19</sup> Barton, 1955, Capecchi, 1968, Elman, 2005, Lazarfeld and Barton, 1951.

<sup>20</sup> Verba, 1967.

That is, if the eruption of Mount Vesuvius is reported to account for the disruption of the Pompeiian administration, the configurational hypothesis that learns from the case may systematically consider the presence or the absence of a similar event across cases. In so doing, the hypothesis allows adjudicating on the *inus* status of absent factors, too.

2. When the limited scope of some local factor prevents its portability, the factor can be redefined as the particular manifestation of a more abstract attribute that preserves its meaning beyond the instance that motivated it. In learning from the Pompeiian case, a configurational hypothesis may more fruitfully model a property about “natural calamities” instead of one about “volcanic eruptions,” as the latter builds a property that is meaningful within volcanic regions only.
3. When the purported explanation includes an exogenous factor, better hypotheses follow if the exogenous factor is endogenized as some particular attribute that the unit of analyses reveals when exposed to the factor, and that is relevant to support or obstruct the mechanism of interest. That is, we may develop a better hypothesis if, to explain administrative breakdown, the exogenous condition “natural calamity” is referred to the attributes that theories would suggest a successful administrative system displays when faced with it – such as ⟨resourceful⟩, or ⟨anticipatory⟩ – and which possibly makes the difference with failing systems.

So shaped, the case-oriented literature provides a useful complement to what we learn from correlational research as far as it embodies the third kind of elaboration of a purported causal relationship – the specification of the attributes as local scope conditions under which an interpretable mechanism inevitably unfolds<sup>21</sup>.

### 6.3 Summing up

A Salmon-complete explanatory configurational hypothesis is that which specifies the relevant conditions endogenizing the mechanistic process to an effect – or which identifies the attributes of the explanatory type from a mid-range theory about a mechanism – so that the

---

<sup>21</sup> Walker and Cohen, 1985, George and Bennett, 2005.

mechanism would surely obtain, were our understanding of it true in the cases at hand.

The correlational literature contributes to configurational hypotheses as it can test the relationship between single components of the mechanistic process and the outcome out of several instances, net of idiosyncrasies, although abstracted from contexts. The case-oriented literature contributes to configurational hypotheses by providing detailed, although possibly idiosyncratic portrayals of actual generative processes, and by pointing at the attributes that qualify real cases as instances of types of mechanistic processes.

Mechanisms, therefore, emerge as the common ground around which a configurational hypothesis can take its shape about the circumstances, or the set of attributes, that make an effect “unavoidable” in the field<sup>22</sup>.

The contribution of the typological literature to explanatory QCA, however, goes further than that. Lazarsfeld and Barton<sup>23</sup> indicate the reduction of the property space as a suitable analytical strategy to assess the import of the configurational hypotheses while identifying the explanatory core. Their proposal, however, left the researchers free to choose among alternative strategies of reduction – hence, remained agnostic about the proper algebra of composition and the algorithm of elimination.

Ragin’s Standard QCA reinvents the tradition of classification typologies by rendering and treating the property space with Boolean algebra. Thanks to its unique nature, this language ensures the property space is faithful to the assumptions of configurational causation, while its operations ensure that solutions find the *inus* compound accounting for the cases at hand. Moreover, the language has the convenient characteristic of connecting QCA to the longstanding logical tradition of explanation – especially, with the rationale of reduction sentences and observation reports.

---

<sup>22</sup> Little 1996.

<sup>23</sup> Lazarsfeld and Barton, 1951. See also Barton, 1955.

## 7. A language to render *inus* causation

An algebra is a language of meaningful symbols and rules. An algebra that suits *inus* causation and the typological mindset should allow properties:

1. have basic observable states as presence or absence of attributes;
2. form compounds;
3. have substitutes;
4. establish relations of dependency.

To render them all, QCA resorts to Boolean operators. Introduced as primary devices to analyze human reasoning about the world<sup>1</sup>, these operators have the convenient feature of supporting a twofold reading – a logical one and a set-theoretical one<sup>2</sup>.

### 7.1 Presence and absence of a property

Boolean algebras use “literal symbols” to indicate factors. QCA borrows the convention and indicates:

- the presence of an attribute, with an uppercase letter, such as  $A$ ,  $B$ ,  $C$ ,...
- the absence of the attribute, with the same letter introduced by a curl  $\sim$  reading ‘not’.

Thus,  $A$  reads ‘the presence of attribute  $A$ ’ or ‘ $A$  present’ or ‘ $A$  positive’ or simply ‘ $A$ ’; while  $\sim A$  reads ‘the absence of attribute  $A$ ’ or ‘ $A$

---

<sup>1</sup> De Morgan, 1847; Boole, 1853.

<sup>2</sup> Stone, 1936.

absent' or 'A negative' or 'not A'. Absence can also be indicated with the lowercase, or with an overbar to omit the curl:  $\sim A, a, \bar{A}$  all have the same meaning. Often in empirical QCA, the convention in force is left unspoken, although the reader can infer it from context.

In the following sections, as in the original Boolean notation, the negated set is denoted with the overbar. When a literal indicates more than one attribute, each attribute is denoted by a subscript ranging from 1 to  $k$ . Thus, the set of positive attributes can be  $\{A_1, A_2, \dots, A_k\}$  or  $\{A, B, C, \dots\}$ ; the corresponding set of their negations will be  $\{\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k\}$  or  $\{\bar{A}, \bar{B}, \bar{C}, \dots\}$ .

Literals are freely chosen to stand for a name or an adjective denoting "either a thing, or some quality or circumstance belonging to it"<sup>3</sup>. Hence,  $A$  may mean "inflammable" of a material, "defined" of a jurisdiction, "affluent" of a society, "Independent" of a voter, "adult" of a citizen – or whatsoever attributes we consider as the purported *in* condition for the effect to occur.

From the logical perspective, letters represent propositions about the entities in the world – they establish that the entities of interest 'are  $A$ '. As propositions, they require a universe of discourse to back and define their meaning<sup>4</sup>. Plainly, the universe consists of the entities about which we talk, and to which our propositions refer. Within a universe, each proposition also becomes a set, that is, the extension of the attribute, or the collection of entities to which it applies. Hence, any set  $A_k$  is the collection of the  $i$  entities  $u_i$  that belong to the universe  $\mathbb{U}$  such that each entity shares the  $k$ th attribute  $A_k$ .

In short, a set arises from an operation of classification. In the Boolean logical world, a set partitions automatically the universe in two classes: that of the entities that manifest the property as the select attribute, and that of the entities that do not. Logic clarifies the relationship that the two partitions maintain with themselves and to each other and that operations cannot violate.

The primary relationship is of identity, stating that a set is equal to itself. In Boole's original proposal and all the basic operations of QCA, the set also is idempotent, meaning it satisfies the classical *dictum de omni et nullo*. Recall the principle is met when what can be affirmed or denied of the whole class also holds for each of its members. Therefore,

---

<sup>3</sup> Boole, 1853:27.

<sup>4</sup> De Morgan, 1847.

the principle defines *homogeneous classes of individuals* to the attribute grouping them together. Boole renders idempotent identity as in [18]

$$A^2 \stackrel{\text{def}}{=} A \quad [18]$$

The only two numerical values that satisfy it are 1 and 0. Therefore, Boole's literals can only take these two values – and the basic operations in QCA share the bivalent assumption, too. In its original version, bivalence entails another classical principle, that of the excluded middle. Recall the principle states that an entity can only be  $A$  or  $\bar{A}$  – disallowing borderline cases. The commitment to this principle has significant consequences on the images of the world that we can construe with this algebra.

Set-theoretically, 1 is the conventional membership-value for the inclusion of an entity into a property-set and can be read as 'belongs to', while 0 is the membership value for the exclusion from the same set and can be read as 'does not belong to'. The values allow synthesizing the set-theoretical consequence of the excluded middle as equation [19]:

$$\bar{A}_i \stackrel{\text{def}}{=} 1 - A_i \quad [19]$$

indicating the negation of the attribute is its complement.

Set-theoretically, 0 can also stand for the empty set, that is, the partition containing no elements, and 1 for the tautological set, that is, the partition containing all the entities in a universe. In this sense, however, sets are fractions of the universe, and their value refers to the cardinality of the partition. The cardinality, indicated by a vertical bar on each side of a literal, is the number of the entities that in a universe qualify as instances of the partition, calculated as in [20]:

$$|A| = |\{A_1, A_2, \dots, A_i\}| = \sum_{i=1}^n A_i \quad [20]$$

When the cardinality is rendered as the fraction of the  $n$  entities in the partition to the  $N$  entities in the universe, the empty set yields  $n / N = 0$ ; the tautological set instead yields  $n / N = 1$ .

To avoid ambiguities, in the following we will indicate the empty set as  $\emptyset$ , and the tautological set as  $\mathbb{U}$ . If we denote the difference between sets with a backslash  $\setminus$ , reading 'minus', equation [19] applied to the relationship of  $A$  and  $\mathbb{U}$  reads as equation [21]:

$$|\bar{A}| = |\mathbb{U} \setminus A| \quad [21]$$

meaning the negative set,  $\bar{A}$ , collects those instances in  $\mathbb{U}$  that are excluded from the positive set  $A$ .

Logically, the values convey the truth of the underlying propositions. 1 means *true*, 0 means *false*, and together provide the only allowed answers to the second-order question 'is it true that this entity is  $A$ ?'. Idempotency and the principle of the excluded middle together warrant that:

$$(A_i = 1) \equiv (\bar{A}_i = 0) \quad [22.a]$$

$$(A_i = 0) \equiv (\bar{A}_i = 1) \quad [22.b]$$

Identity [22.a] means that when it is true that the instance is  $A$ , then it is false that the instance is not  $A$ ; identity [22.b] states that when it is false that the instance is  $A$ , then it is true that the instance is not  $A$ . These identities turn the principle of the excluded middle into the principle of non-contradiction, establishing the truth of an attribute and its negation are mutually exclusive.

The truth-value of a contrary statement follows from the application of the connective 'not' to the truth-values of the original statement. The operation can be represented as a truth table<sup>5</sup>.

#### a) *The truth table*

A truth table is the universe specified, plus a connective.

The specification of the universe follows from the logical expansion of the  $k$  elementary propositions – crudely, the exhaustive list of the combinations of their truth-values, which also specifies all the possible alternative states of an instance in the universe independent on their observability. As the expansion is the fundamental ordering operation of a universe<sup>6</sup>, the combinations in a truth table are “primitive”. As each Boolean proposition can only have two values, the total number of primitives in a truth table is  $2^k$ . Conventionally, the first primitive in the truth table reports the combination of all true affirmative propositions, while the last primitive displays the combination of all false-positive propositions.

---

<sup>5</sup> Wittgenstein, 1922.

<sup>6</sup> Quine, 1982.



Each column of a truth table displays a truth function, i.e., the sequence of the truth-values of each elementary proposition, and the connective. The length and structure of the truth function of an elementary proposition are constant and independent on the connective; instead, they depend on the number of elementary propositions considered.

Thus, the expansion of  $\{A\}$  corresponds to the truth function of the single proposition  $A$ , that is, (1 0). In the expansion of two propositions  $\{A, B\}$ , the truth function of  $A$  conventionally is (11 00), while the truth function of  $B$  is (10 10). In the expansion of  $\{A, B, C\}$ , the truth function of  $A$  is (11 11 00 00), that of  $B$  is (11 00 11 00), and the truth function of  $C$  is (10 10 10 10) – and so on.

The last column of the truth table associates each primitive from the expansion to the truth-value of the connective under consideration.

As portrayed in Table 3, that of the negation connective is the most straightforward truth table, as it can apply to a single proposition. The truth function of the connective ‘not’ is (0 1) when applied to the truth function of a single factor, and summarizes the identities in [22.a] and [22.b].

**Table 3. The truth table for the connective  $\bar{\phantom{A}}$ .**

ID	A	$\bar{\phantom{A}}$
(1)	1	0
(2)	0	1

*Keys:* the column “id” attaches an identification number to each of the primitives from the expansion of the elementary proposition.

Column A lists the truth function of the elementary proposition in the expansion.

Column  $\bar{\phantom{A}}$  reports the truth function resulting from the application of the negation connective to each primitive in the expansion.

The truth functions are of use to explore the formal equivalence of complex statements. On this basis, for instance, it is possible to derive De Morgan’s laws on the negation of complex statements. However, as the pointed philosophical literature on paradoxes has long clarified<sup>7</sup>, the formal equivalence of the truth function of statements about phenomena may not warrant the substantial equivalence of actual phenomena – especially when inference is involved.

---

<sup>7</sup> Good, 1967; Salmon, 1989.

## 7.2 Compounds

Compounds arise from trains of letters connected by the ‘and’ operator. In QCA, the operator is indicated by a dot  $\cdot$  or by a star  $*$ , although the connecting symbol may be omitted. Hence,  $ABC$ ,  $A * B * C$ ,  $A \cdot B \cdot C$  all read ‘ $A$  and  $B$  and not- $C$ ’.

Two facts are worth noting about the ‘and’ operator. Permutation and grouping are irrelevant to a compound:  $ABC$  means the same as  $BAC$  and as  $A(BC)$  – the resulting class groups in the same instances. In short, the Boolean ‘and’ supports the commutative rule and the associative rule, respectively. This entails that compounds are insensitive to the time dimension and sequences. Instead, they emphasize space – the joint occurrence of attributes in an entity, which also defines the class of the entities sharing the same compound.

The ‘and’ operator again supports the twofold interpretation – logical, and set-theoretical.

**Table 4. The truth table for the conjunction  $A \wedge B \wedge \bar{C}$ .**

ID	$A$	$B$	$C$	$\bar{C}$	$A \wedge B \wedge \bar{C}$
(1)	1	1	1	0	0
(2)	1	1	0	1	1
(3)	1	0	1	0	0
(4)	1	0	0	1	0
(5)	0	1	1	0	0
(6)	0	1	0	1	0
(7)	0	0	1	0	0
(8)	0	0	0	1	0

*Keys:* the column “id” attaches an identification number to each of the primitives from the expansion of the elementary propositions.

Columns  $A$ ,  $B$ ,  $C$  list the truth function of the elementary proposition in the expansion.

Column  $\bar{C}$  reports the truth function of the negation of  $C$ .

Column  $A \wedge B \wedge \bar{C}$  reports the truth function of the corresponding compound, and draws attention to its only true point – i.e., primitive (2).

Set theoretically, ‘and’ corresponds to the intersection among property-sets, written  $\cap$ .  $ABC$  has the same meaning as  $A \cap B \cap C$ , and indicates the subset of the entities that display all the attributes in the right state, hence lay in the intersection of the corresponding sets.

An intersection is one of the partitions in which a universe can be carved up, given several property-sets. As each set generates 2 partitions, any further property-set doubles the number of partitions of the universe, and the number of partitions from  $k$  sets is  $2^k$ . A partition becomes observed or empirically true when at least one entity from the universe qualifies as its instance.

The one in an intersection is the most demanding set-membership. To qualify as an instance of  $ABC$ , it has to be true at once that an entity is A while being B and not being C. This leads us on the logical ground, where the 'and' operator is a connective, usually indicated with the wedge  $\wedge$ . The train of literals forms a conjunction of elementary propositions.

A conjunction is logically true when each conjunct is true. Hence, in the truth table of the 'and' connective, there is a single point where the truth function takes the value of 1 – corresponding to its identical primitive, as portrayed in Table 4.

As truth-values and membership-values coincide, the shape of the truth function of the conjunction suggests that the rule for establishing whether an entity is an instance of an intersection of  $k$  attributes  $A$  can be based on the minimum of the entity's values in each attribute, as summarized by equation [23]:

$$\cap A_k = \wedge A_k = \min(A_k) \quad [23]$$

When equation [23] yields the value of 1, it is true that the entity is an instance of the compound; else, it is not. If no entity qualifies, the intersection is logically possible yet unobserved, hence remains an empty set.

The application of the 'and' operator to single properties further clarifies the meaning of the principle of the excluded middle as the consequence of the principle of idempotency. Following equations [18] and equation [20], the intersection of the positive and the negative set must be empty, as in equation [24]:

$$A^2 - A = A(1 - A) = A \cdot \bar{A} = \emptyset \quad [24]$$

Any violation of equation [24] arises a logical contradiction, which can be symbolized by an uptack  $\perp$  or by a  $\mathcal{D}$ , reading contradictory, or, again, false.

### 7.3 Substitutes

Substitutes arise when terms – either single or compound – are connected by the operator ‘or’. In QCA, the operator is indicated by a plus symbol + and is never omitted. Hence,  $A + B + \bar{C}$  reads ‘ $A$  or  $B$  or not- $C$ ’.

Idempotency makes permutation and grouping irrelevant to substitutes, too:  $A + B + \bar{C}$  means the same as  $B + A + \bar{C}$ , and as  $A + (B + \bar{C})$ ; thus, the Boolean ‘or’, too, supports the commutative rule and the associative rule.

Set-theoretically, this ‘or’ corresponds to the union of property-sets, usually indicated by  $\cup$ .  $A + B + \bar{C}$  has the same meaning as  $A \cup B \cup \bar{C}$  and indicates the superset of the entities that belong at least to one of the property sets in the right state. Logically, this set union is a nonexclusive disjunction of elementary propositions, usually denoted by a vee  $\vee$ . The disjunction is true for any primitive that includes at least one of the elementary propositions with the right truth-value. The truth function of any disjunction, therefore, is always true except for the primitive where all the propositions jointly take the opposite truth-values. Thus, the disjunction  $A \vee B \vee \bar{C}$  is only false at the primitive  $\bar{A} \wedge \bar{B} \wedge C$  – as shown in Table 5.

**Table 5. The truth table for the disjunction  $A \vee B \vee \bar{C}$ .**

ID	$A$	$B$	$C$	$\bar{C}$	$A \vee B \vee \bar{C}$
(1)	1	1	1	0	1
(2)	1	1	0	1	1
(3)	1	0	1	0	1
(4)	1	0	0	1	1
(5)	0	1	1	0	1
(6)	0	1	0	1	1
(7)	0	0	1	0	0
(8)	0	0	0	1	1

*Keys:* columns are the same as in Table 4, except for the last one.

Column  $A \vee B \vee \bar{C}$  reports the truth function of the corresponding disjunction and draws attention to the only false point – i.e., primitive (7).

The disjunction is false only when its complement is true – in Table 5, primitive (7). Its truth function suggests the rule for establishing whether an entity is an instance of the union of  $k$  attributes  $A$  can be

based on the maximal value that the entity gets in each of the attributes, as summarized by equation [25]:

$$\bigcup A_k = \bigvee A_k = \max(A_k) \quad [25]$$

Lastly: from equation [19], it follows the disjunction of a proposition, and its complement yields the universe or all the instances, as in equation [26].

$$A_i + \bar{A}_i = \mathbb{U} \quad [26]$$

Logically, such a disjunction generates a tautology, and its result is also denoted by a downtack  $\top$ .

## 7.4 Necessity and sufficiency

In QCA, necessity is indicated with an arrow running from the effect to the causes, and sufficiency with an arrow running from the causes to the effect. Conventionally, expressions are written from left to right, and the effect of interest appears at the end of the expression; under such a convention, sufficiency is a right-headed arrow, and necessity is left-headed. For instance,  $A \rightarrow Y$  reads 'A is sufficient to Y';  $\bar{A} \leftarrow Y$  reads 'not-A is necessary to Y'.

These two relationships have long been given a set-theoretical and a logical rendering.

### 7.4.1 Necessity

Set-theoretically, the necessity of  $A$  to  $Y$  corresponds to  $A$  being a superset of  $Y$ . The relationship is satisfied when all the  $Y$  also are  $A$ , although the reverse may not be true, and, in the universe, there may be instances of  $A$  that do not display  $Y$ . The hallmark of necessity is the impossibility of the effect in the absence of the factor, as in [27]:

$$\bar{A} \cap Y = \emptyset \quad [27]$$

The requisite makes better sense in the logical reading of the relationship.

In logic, the superset distribution corresponds to the truth table for the implication connective  $\leftarrow$ , reading ‘implies’, as portrayed in Table 6.

Requisite [27] corresponds to Primitive (3) and embodies the empirical proof of the counterfactual argument that factor  $A$  is required by the effect, although it does not compel it. Regardless of their complexity, hence, necessary attributes cannot sort instances of the effect from instances of its failure – hence, they cannot account for local differences in the effect. Sufficiency can do instead.

**Table 6. The truth table for the implication  $A \leftarrow Y$ .**

ID	$A$	$Y$	$\leftarrow$
(1)	1	1	1
(2)	1	0	1
(3)	0	1	0
(4)	0	0	1

*Keys:* Column  $ID$  assigns a progressive number to the possible conjuncts of truth-values of the factor  $A$  and the effect  $Y$ ;

Column  $A$  reports the truth function of the allegedly causal factor  $A$ ;

Column  $Y$  reports the truth function of the effect  $Y$ ;

Column  $\leftarrow$  reports the truth function of the implication  $A \leftarrow Y$  and draws attention to the only false point – i.e., primitive (3).

#### 7.4.2 Sufficiency

Set-theoretically, the sufficiency of a factor or a compound  $W$  to  $Y$  corresponds to  $W$  being a subset of  $Y$ . The relationship is satisfied when all the  $W$  are also  $Y$  although the reverse may not be true, and instances of  $Y$  may exist in  $\mathbb{U}$  that do not display  $W$ . In short, a sufficient factor compels the effect, although the effect can occur otherwise.

The hallmark of sufficiency is the impossibility that the effect fails when the factor is present, summarized by requisite [28]:

$$W \cap \bar{Y} = \emptyset \quad [28]$$

Again, the requisite makes better sense from a logical perspective.

In logic, the subset distribution corresponds to the truth table for the conditional connective  $\rightarrow$ , reading ‘if – then’, as portrayed in Table 7. The ‘if... then’ connective suits a discursive situation in which the compound proposition on the left is a premiss compelling the proposition

on the right as its logical conclusion. The unobserved primitive (2) ensures the factual statement is not contradictory. Were it true of the entities in  $\mathbb{U}$  that ‘if they are  $W$ ; then they are  $Y$ ’, and ‘if they are  $W$ ; then, they are not  $Y$ ’, we should get to the conclusion that, in  $\mathbb{U}$ , being  $W$  can lead to both  $Y \cdot \bar{Y}$  – that is, to a contradiction. Observing primitive (2), hence, logically refutes the claim that  $W$  compels  $Y$ .

The claim of sufficiency instead logically holds despite primitive (3) is observed, due to the classical principle *ex falso quodlibet sequitur*, establishing that anything can follow from a false premiss. Indeed, we would incur in a fallacy, were we accepting a false premiss invalidates a true conclusion. Instances of primitive (3) are logically valid although not empirically sound – they are vacuously true.

**Table 7. The truth table for the conditional  $W \rightarrow Y$ .**

ID	$A$	$Y$	$\rightarrow$
(1)	1	1	1
(2)	1	0	1
(3)	0	1	0
(4)	0	0	1

*Keys:* Column *id* assigns a progressive number to the possible conjuncts of truth-values of the compound  $W$  and the effect  $Y$ ;  
 Column  $W$  reports the truth function of the allegedly nomological compound  $W$ ;  
 Column  $Y$  reports the truth function of the effect  $Y$ ;  
 Column  $\rightarrow$  reports the truth function of the logical conditional  $W \rightarrow Y$  and draws attention to the only false point – primitive (2) – and the vacuously true point – primitive (3).

From Mackie’s perspective of incomplete explanations, like cell (3) of Figure 5, primitive (3) accommodates the instances where the effect arises due to different compounds than originally hypothesized.

## 7.5 Putting it all together for the analysis

The QCA scholarship agrees<sup>8</sup> Boolean algebra allows expecting that a complete explanation of the occurrence of an effect reads as expression [29.a]:

$$I_1 + \dots + I_m \rightarrow Y \quad [29.a]$$

<sup>8</sup> Duşa 2018, Mahoney *et al.* 2013.

where  $I_m$  are prime implicants and indicate one of the  $m$  alternative complete conjunctions of the  $l \leq k$  *inus* attributes implying one or more primitives from our starting hypothesis, as in [29.b]:

$$A_1 \cdot \dots \cdot A_l \leftarrow I_m \quad [29.b]$$

In short, the general QCA solution is a disjunction of sufficient conjunctions  $I_m$  of proven *inus* attributes  $A_k$  – or, the disjunction of  $m$  K-machines.

The scholarship is less univocal about the shape of the starting hypothesis. Those who maintain QCA is for retrieving articulated theoretical models<sup>9</sup> consider that [29.a] should provide both the starting and the endpoint of the analysis, as QCA is for retrieving the true structure of causal dependencies. Schneider and Wagemann<sup>10</sup> allow for results differing from ex-ante theoretical expression instead, and maintain the empirical holding of specific set-theoretical propositions requires a different protocol without minimizations.

As a reduction technique that “starts by assuming maximum causal complexity and then mounts an assault on that complexity”<sup>11</sup>, however, the starting point may rather resemble a theoretically meaningful yet unwarranted claim about a redundant compound of  $k$  possibly-*inus* attributes to the effect. The claim may assume that were it true that the  $k$  selected attributes together make one or more “nomological” arrangements, then the effect should inevitably occur in a theoretical instance displaying them all. The starting hypothesis that warrants this reading is an overdetermined claim of sufficiency about a theoretical instance, as in expression [29].

$$A_1 \cdot \dots \cdot A_k \rightarrow Y \quad [29]$$

Boolean operations turn the hypothesis [29] into solution [29.a] in three steps. They:

1. expand the hypothesis into a truth table;

---

<sup>9</sup> Baumgartner and Thiem 2017.

<sup>10</sup> Schneider and Wagemann, 2012; see also Oana and Schneider, 2018.

<sup>11</sup> Ragin, 1987: x.



2. assess the consistency of each primitive with the claim of sufficiency in the hypothesis based on the distribution of the instances from the universe across the primitives;
3. reduce the consistent primitives to identify the implicants  $I_m$  – originally, with the Quine-McCluskey algorithm.

As a preliminary operation, however, the protocol prescribes that we assess the empirical standing and direction of the expectations we associate with each attribute in the hypothesis.

### 7.5.1 *Checking directional expectations*

We include an attribute in the starting hypothesis under the directional expectation that, according to the existing literature, in the right state, it contributes to a nomological compound. The expectation holds when the attribute in the wrong state always impedes the effect and  $|\bar{A} \cap Y| = \emptyset$ . The consistency of necessity (N.cons for short) controls for the threats to this claim that can arise from “miracle” instances of an M-machine. The index is calculated as in [30]:

$$\text{N.cons} = \frac{|A \cap Y|}{|Y|} = \frac{\sum_{i=1}^n \min(A_i, Y_i)}{\sum_{i=1}^n Y_i} \quad [30]$$

When  $Y$  only occurs with  $A$ , the numerator equals the denominator, and the index takes its higher value of 1.00, supporting the claim that the attribute is necessary to the effect in the whole universe. Since the compound is that which mediates the necessity of an attribute to the effect, a full N.cons value means the attribute is like oxygen to combustion – a necessary component to any alternative compound, without substitutes in the field. Lower N.cons values weaken the claim instead.

A second index captures the information about the empirical relevance of the necessary attribute: the coverage of necessity (N.cov for short), calculated as in [31]:

$$\text{N.cov} = \frac{|A \cap Y|}{|A|} = \frac{\sum_{i=1}^n \min(A_i, Y_i)}{\sum_{i=1}^n A_i} \quad [31]$$

The N.cov renders the share of the instances with the necessary factor in which the factor also enters a known effective compound. Usually, the index is only reported for those factors with a high N.cons

score. When all the instances of the effect occur under the necessary factor, the numerator equals the denominator, and the index takes its higher value of 1.00. The lower the N.cov index, the lower the extension of the claim that the necessary factors always come with an effective compound.

A further index assesses the truth of the statement that the absence of the necessary factor and the failing of the effect coincide. The Relevance of Necessity (RoN<sup>12</sup>) establishes to which extent the instances where the effect does not occur also are the instances where the factor is in the wrong state, as in [32].

$$\text{RoN} = \frac{|\bar{A}|}{|\bar{A} \cup \bar{Y}|} = \frac{\sum_{i=1}^n (1 - A_i)}{\sum_{i=1}^n (1 - \min(A_i, Y_i))} \quad [32]$$

The index takes its higher value of 1.00 when the failing effect always occurs under the wrong state of the factor; the lower the RoN, the more trivial the condition – to the point of becoming a meaningless constant across the universe.

Calculated on each state of an attribute against both the effect and its failure, and supported by the information from the N-cov and the RoN, the N.cons allows assessing whether our theoretical expectations about the single factors hold in the universe of reference for the analysis. Especially in the Enhanced versions of the Standard Analysis (ESA<sup>13</sup>), the analysis of the necessity of individual attributes provides significant indications for refining the directional expectations to be later used in minimizations.

### 7.5.2 *Building the truth table*

A crucial step of the procedure consists in construing the truth table as the full specification of the universe by the factors in the starting hypothesis.

After expanding the attributes, each of the instances under analysis is classified as a member of the corresponding primitive. The classification is exclusive and follows from the calculation of the membership score of each instance in each primitive as in equation [23]. The instance is then assigned to the single primitive in which it scores higher than 0.5.

---

<sup>12</sup> Schneider and Wagemann 2012, Duşa 2018.

<sup>13</sup> Schneider and Wagemann, 2012, Oana and Schneider, 2018.

At the end of the operation, the truth table displays two kinds of primitives – observed and unobserved. The latter is also known as logical remainders and constitutes a common occurrence. Although the ratio of cases to attributes plays a role in generating the remainders<sup>14</sup>, the phenomenon is relatively independent of the richness of the hypothesis. The logical remainders mainly mirror the limited diversity of the actual nomological arrangements, indicating some combination of attributes remains unrealized in the cases at hand<sup>15</sup>. In the Weberian tradition of typological analysis, these unrealized combinations can nevertheless serve minimizations as counterfactual mind experiments.

So settled, the truth table allows for testing whether the claim of the sufficiency of the starting hypothesis holds in  $\mathbb{U}$ .

### 7.5.3 Assessing the sufficiency of the hypothesis

The truth table provides the first proof of the claim of *inus* causation. If the attributes in the starting hypothesis capture at least a nomological compound, then each observed primitive from the expansion of the hypothesis distributes the entities in the shape of sufficiency.

Standard QCA<sup>16</sup> considers that two indexes can assess the fit of the distribution to the shape of sufficiency.

The coverage of sufficiency (S.cov for short) controls for the threats that arise from vacuous miracle instances, and is calculated as in [33]:

$$S.cov = \frac{|W \cap Y|}{|Y|} = \frac{\sum_{i=1}^n \min(W_i, Y_i)}{\sum_{i=1}^n Y_i} \quad [33]$$

When all the instances of a primitive  $W_i$  display the effect, the numerator equals the denominator, and the index takes its higher value of 1.00, supporting the claim that the compound is sufficient to the effect in the whole universe. The more the exceptions, the lower the S.cons and the credibility of the extension of the claim.

The consistency of sufficiency (S.cons for short) instead controls for the threats that arise from contradictions to the claim of sufficiency, that is, from instances of the same primitive displaying opposite effects. The S.cons index is calculated as in [34]:

---

<sup>14</sup> Marx and Duša, 2011.

<sup>15</sup> Ragin, 2008.

<sup>16</sup> Ragin, 2008; Schneider and Wagemann, 2012; Duša, 2018.

$$\text{S.cons} = \frac{|W \cap Y|}{|W|} = \frac{\sum_{i=1}^n \min(W_i, Y_i)}{\sum_{i=1}^n W_i} \quad [34]$$

When all the instances of  $W_i$  agree on the effect, the numerator equals the denominator, and the index takes its higher value of 1.00, supporting the claim that the compound is sufficient to the effect in the universe. The higher the disagreement, the lower the S.cons index and the weaker the standing of the claim.

To assess the claim of the sufficiency of the starting hypothesis, the S.cons again bears higher relevance than the S.cov. Calculated on each of the primitives in the truth table to the positive and to the negative outcome, the S.cons suggests the hypothesis is well specified when all the observed primitives display high S.cons to the effect and low S.cons to its negation, or vice versa.

In any case, the S.cons values identify which primitives can be minimized to find the compound of *inus* factors explaining the effect and its negation.

#### 7.5.4 Minimizing the truth table

The minimization applies to consistent primitives, compared pairwise, according to the “single difference rule” as implemented by the Quine-McCluskey algorithm. The rule allows dismissing a conjunct as redundant when two primitives associated with the same outcome have the same number of literals and all in the same status except one.

For instance, we can minimize the primitive  $ABCD$  and the primitive  $ABC\bar{D}$  if both display high S.cons values to the same effect. The formal reason is, the two allow the factorization  $ABC(D + \bar{D})$ . Recall from equation [26] and [24] that the disjunction of an attribute and its negation is a tautology, as the union of a partition and its complement corresponds to the universe:  $D + \bar{D} = U = \perp$ . Therefore, the comparison highlights  $ABC$  yields  $Y$  regardless of the state of  $D$ , which does not qualify as an *inus* component of that compound to  $Y$ .

The decision on the *inus* nature of single components, however, may change depending on how minimizations consider the unobserved primitives. The Standard Analysis contemplates three alternative assumptions about them:

- i. that none could have been sufficient to the effect, if observed;
- ii. that any could have been sufficient, if observed and useful to drop a further literal from a compound;

- iii. that only those qualifying as easy counterfactuals could have been sufficient if observed.

a) *Hard and easy counterfactuals*

The difference between hard and easy counterfactuals is easier to grasp with formalization. Then, let us imagine the following.

We include the attribute  $A$  to the starting hypothesis under the theoretical expectation that it is an *inus* factor, and that it contributes to the machine of the effect  $Y$  when present. This means that  $A$  should be present with other conjuncts  $\Phi$  for the outcome to occur; *vice-versa*, the absence of  $A$  should make  $\Phi$  fail. Thus, our directional expectations are (i)  $A\Phi \rightarrow Y$  and (ii)  $\bar{A}\Phi \rightarrow \bar{Y}$ .

After we build and populate the truth table, we find that the primitive  $W_1 = ABCD$  is observed with an S.cons of 1.00 to  $Y$ , while we do not observe (hence we star) the primitive  $W_9^* = \bar{A}BCD$ .

According to the single difference rule,  $W_1$  and  $W_9^*$  can be minimized. The minimization dismisses  $A$  as irrelevant, ascribing sufficiency to the compound  $BCD$ . However, the minimization entails  $W_9^*$  is consistent with  $Y$ , hence  $\bar{A}BCD$  potentially is a nomological arrangement to the effect. This, however, goes against our directional expectation (ii) and makes a hard counterfactual of  $W_5^*$ .

Let us now imagine the opposite.

We find the primitive  $W_{13} = \bar{A}\bar{B}CD$  is observed with an S.cons of 0.98 to  $Y$  while we do not observe (hence we star) the primitive  $W_5^* = A\bar{B}CD$ . Again, according to the single difference rule,  $W_{13}$  and  $W_5^*$  can be minimized, and the minimization dismisses  $A$  as irrelevant, ascribing sufficiency to the compound  $\bar{B}CD$ . Again, the minimization entails  $W_5^*$  potentially is a consistent compound with  $Y$ . In this case, however, the assumption is  $A\bar{B}CD$  would portray a nomological compound – i.e., the presence of  $A$  would contribute to make  $\bar{B}CD$  obtain the effect. As this is consistent with our directional expectation (i),  $W_5^*$  is an easy counterfactual.

Minimizations first operate under the prudential assumption that none of the unobserved primitives would have been sufficient. This drops redundant factors from observed primitives and yields the *conservative* or *complex* solution.

The second round employs any matching counterfactual. The result is a superset of the conservative, known as the parsimonious solution. The surviving factors are the real *inus* components in the hypothesis, as no solution can prove consistent without them. However, minimizations can use hard counterfactuals that defy directional expectations. Otherwise said, *inus* solutions – made of essential components – are not necessarily Salmon-complete.

In the third round, the hard counterfactuals are barred from minimizations. Their implicants are richer than in the parsimonious, as the minimizations restore those further attributes from the starting hypothesis that “co-unvary” with the prime of the parsimonious<sup>17</sup>. Therefore, these results are the *intermediate* or *plausible* ones. The additional attributes may not be essential to tell the positive from the negative instances of the outcome. Nevertheless, they deserve to be discussed as they improve the consistency of the overall solution with the outcome<sup>18</sup>. In a nutshell, higher parameters of fit may prove plausible solutions are Salmon-superior to parsimonious solutions.

## 7.6 Wrapping up

Boolean algebra provides an especially suitable language to render and test individual necessity and joint sufficiency of *inus* factors to an outcome. Attributes are elementary propositions and sets, and their relationship is defined by logical connectives, each corresponding to a particular set relationship.

In this language, the starting hypothesis can be conceived of as the ideal conjunction of all the attributes that jointly bear a claim of sufficiency.

The hypothesis holds to the cases in a universe when the expansion of the attributes orders the universe in homogeneous partitions, each bearing the same claim – i.e., when all the observed primitives in the truth table have high S.cons values to one effect and low S.cons values to its opposite, not both or none.

The Quine-McCluskey algorithm then applies to observed primitives with only high or only low consistency to an outcome. It identifies prime implicants under three assumptions about the unobserved primitives – that they would never obtain; that could obtain if close enough

---

<sup>17</sup> Damonte, 2018.

<sup>18</sup> Duşa, 2019.

to an observed configuration; that would obtain if close enough and preserving the directional expectations.

The prime implicants from the latter minimizations are the reduced types providing the plausible explanation of why an effect occurred in some cases and why it failed in some others. The types are both interpretable and logically valid when they render a configurational hypothesis about a constitutive mechanism. Nevertheless, they still could be unsound, were they based on unreliable measures.

## 8. Modeling membership scores

The actual link between sets, propositions, and the real world is decided by how membership scores are assigned – that is, by measurement.

The standard assumption in measurement theory maintains real-world properties depend on some entities' deep structure that we can know indirectly only, as meaningful variations in their empirical manifestation as attributes<sup>1</sup>. Measurement theory assumes we can represent these attributes as numerical images, and capture their variation as a relationship among numbers through proper scales. Scales are valid when they warrant that, for any manifestation  $A_i$  of the property  $P$ , there is a measure  $q_i$  of the image  $Q$  such that the functional relationship between measures preserves the fundamental relationship between manifestations.

The seminal work of Stevens<sup>2</sup> narrowed on four of such fundamental relationships: sameness, rank, distance, and proportion. Thus, sameness is preserved, and the nominal scale is valid, when  $q_1 = q_2$  if  $A_1 = A_2$ ; rank is preserved, and the ordinal scale is valid, when  $q_1 < q_2$  if  $A_1 < A_2$  equality of differences is preserved, and the interval scale is valid, when  $(q_1 - q_2)/(q_3 - q_4)$  mirrors  $(A_1 - A_2) / (A_3 - A_4)$ ; equality of proportions is preserved, and the ratio scale is valid when  $q_1/q_2$  renders  $A_1/A_2$ . Classical methodology textbooks have long taught that a hierarchy of scope exists among measurements with the ratio scale at the top, as the latter is the more robust – i.e., abstracted from actual contexts – and allows for the broadest array of transformations.

Meant as a prudential guideline for preventing nonsensical claims from naïve statisticians<sup>3</sup>, “permissible transformations” have become a

---

<sup>1</sup> Campbell, 1928.

<sup>2</sup> Stevens, 1946.

<sup>3</sup> Luce, 1959.



canon of statistical procedures. As such, it has been disputed since its introduction. Skeptics argue that data are not prior to scales and that Stevens' four types are far from exhaustive or mutually exclusive. Moreover, any operation – including that of gauging properties – entails a loss function, and is admissible to the extent that it minimizes the loss of information of interest to the researcher<sup>4</sup>. Along this line, comparatists recognize entities can manifest properties with different intensity yet contend that any variation in intensity should be recorded as equally meaningful as the ratio scales assume. Classes and nominal scales arise when some select point of variation is emphasized that corresponds to a meaningful change in the state of the entities bearing the property, while any other differences are dismissed as irrelevant<sup>5</sup>.

In short, scales entail a trade-off between precision and meaning. The two perspectives become compatible, though, when metric variables are remapped onto properties as fuzzy sets.

## 8.1 Giving variation a meaning

Lofti Zadeh introduced fuzzy sets in the 1960s, motivated from dissatisfaction with metric gauges germane yet slightly different from that of comparatists. He noted how the problem-solving embodied by machine algorithms was more accurate although of a narrower scope than the human, as “the class of nontrivial problems for which one can find precise algorithmic solutions is quite limited”<sup>6</sup>. He then identified the unique decision-making capacity of the human brain in its reliance on approximate characterizations of a collection of data as conveyed by language structures. To the question whether a given unit is  $A$ , only simple properties admit a crisp ‘yes’ or ‘no’ answer that Boolean algebra can render without loss. Complex properties, instead, allow for nuanced answers, signaled by hedges such as ‘very’, ‘somewhat’, ‘almost’. They still can be rendered as classes and sets, but indicate fuzzy transitions instead of crisp boundaries between membership and non-membership.

Fuzzy scores translate hedges into weights  $\mu$  from 0.00 to 1.00, conveying the degrees of fit or misfit of any  $i$ th unit  $u_i$  to the concept entailed by the attribute  $A_k$ .

---

<sup>4</sup> Lord, 1946; Guttman, 1977; Mosteller and Tukey, 1977.

<sup>5</sup> Collier *et al.* 2012, Collier and Mahon, 1993; Sartori, 1970, 1984.

<sup>6</sup> Zadeh, 1968:99, 1978.

Set theoretically,  $\mu_A = 1.00$  indicates the full membership in the set of attribute  $A$ , and  $\mu_{\bar{A}} = 0.00$  indicates the full non-membership in the same attribute. With fuzzy scores, the opposite of a set is still its complement:  $\mu_{\bar{A}} = 1.00 - \mu_A$ . Therefore, an entity's memberships in a set and its complement still add up to 1.00:  $\mu_{\bar{A}} + \mu_A = 1.00$ .

The relation between opposites is warranted by a third relevant value, namely, the *crossover*. Conventionally weighing 0.50, this is the point of neutrality and signals a membership neither in the attribute set nor in its complement. Due to it, fuzzy scores defy a strictly bivalent logic<sup>7</sup>. Logically, fuzzy scores capture the possibility that the statement 'u is A' is true for the actual entity  $u_i$ . 1.00 indicates the statement is undoubtedly true; 0.00 indicates the statement is certainly not true; 0.50 indicates the information about  $u$  is so uncertain that the adjudication on its nature is impossible.

In the original proposal<sup>8</sup>, the three points provide a strategy to align linguistic hedges, sets, and metric variables, under the assumption that the shape of the function that filters the metric values  $v_A$  – for instance, age in years – into fuzzy membership scores  $\mu_A$  – for instance, in the fuzzy set <teenager> – is triangular, trapezoidal, or bell-shaped. Given a shape, the researcher can establish:

1. the value of inclusion on the metric measure of the attribute, or  $\alpha_A$ . The inclusion threshold censors any variation above  $\alpha$  as irrelevant, as for any higher values entity  $u_i$  does qualify as an instance of the attribute, and its corresponding fuzzy score is 1.00.
2. the value of exclusion on the metric measure of the attribute, or  $\beta_A$ . The exclusion threshold censors any variation below  $\beta$  as irrelevant, as for any lower values  $u_i$  does not qualify as an instance of the attribute, and its fuzzy score is 0.00;
3. the metric value of the crossover  $\gamma_A$ , at which the classification of  $u_i$  is uncertain so that its fuzzy score is 0.50. By Zadeh's definition, the value of the crossover on the metric variable coincides with their arithmetic mean of  $\alpha_A$  and  $\beta_A$ .

The shape of the relationship between metric values and fuzzy values define the correspondence of some linguistic hedges with set-membership and metric values: 'almost' falls between  $\alpha_A$  and  $\gamma_A$ , 'quite' falls

---

<sup>7</sup> Lee and Zadeh, 1969.

<sup>8</sup> Zadeh, 1978:404.

between  $\gamma_A$  and  $\beta_A$ . Other hedges instead operate transformations on fuzzy scores: 'very' is understood as the stricter scores from the quadratic concentration of the original scores, while 'slightly' yields the more generous square root dilation of the original scores.

## 8.2 Ragin's improvements to fuzzy scores, and some concerns

Zadeh's proposal is affected by a twofold ambiguity problem. First, linguistic hedges are seldom clearly ordered, so that a straightforward correspondence with particular fuzzy scores can prove arbitrary. Second, triangular, trapezoid or bell-shaped relations can make each fuzzy score  $\mu_A$  correspond to more than one raw score on  $v_A$ . For instance, 'almost' and 'no longer' can both weigh  $\mu_A = 0.67$  and, when applied to the attribute <teenager>, indicate either someone of about ten or in their 20s. Thus, his map is not isomorphic: if we only know the thresholds and the fuzzy score, we can retrieve neither the associated hedge nor the raw value.

**Table 8. Positions that any  $i$ th entity  $u_i$  can take to any set  $A$  and the corresponding membership values  $\mu_A$ .**

Position	$\mu_A$						
	(a)	(b)	(c)	(d)	(e)	(f)	(g)
<b>fully in</b>		<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
mostly in					$4/5$	$5/6$	
more in than out	<b>1</b>			$3/4$		$4/6$	$1/2 < \mu_A < 1$
more or less in			$2/3$		$3/5$		
<b>neither in nor out</b>		$1/2$		$1/2$		$1/2$	<b>0.5</b>
more or less out					$2/5$	$2/6$	
more out than in			$1/3$	$1/4$			$0 < \mu_A < 1/2$
mostly out	<b>0</b>				$1/5$	$1/6$	
<b>fully out</b>		<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>

Source: Ragin (2000:156, 2008: 31).

Ragin's fuzzy sets<sup>9</sup> avoid these issues as they serve a different purpose than Zadeh's. Instead of quantifying a natural language, they aim to provide a gauge that includes both types of information of interest to comparatists – of difference in degree and in kind among units to an attribute. Here, membership values are monotonic non-decreasing functions, which re-establishes the isomorphism between raw values, fuzzy membership scores, and selected hedges – as in Table 8.

The mapping of continuous raw variables  $v_A$  into fuzzy scores  $\mu_A$  is especially illuminating of Ragin's rationale of conversion. He understands it as an operation of calibration – that is, like the fine-tuning of an instrument to improve the validity of measurement. The instrument is the filter function, shaped and applied following two different methods<sup>10</sup>.

The less common choice is the indirect method of calibration and proceeds in two steps. In the first step, the filter function is established as the “qualitative score” from a scale – such as (c) or (f) from Table 8 – that the researcher assigns to groups of similar cases on the same raw variable. In the second step, the cases' raw values are turned into predicted fuzzy scores through the qualitative scores by a fractional polynomial regression.

The method of reference, instead, is the direct method of calibration, stipulating the filter function is a growth curve of odds. The different smoothness of the slopes is decided every time by the actual raw values the researcher establishes as  $\alpha_A, \beta_A, \gamma_A$ , whose corresponding membership scores are conventional and fixed. As in Zadeh, Ragin maintains the membership score of any crossover is  $\mu_\gamma = 0.5$ , and fixes the two membership values of inclusion  $\mu_\alpha$  and of exclusion  $\mu_\beta$  as in [35.a] and [35.b], respectively:

$$\mu_\alpha \stackrel{\text{def}}{=} 0.953 \quad [35.a]$$

$$\mu_\beta \stackrel{\text{def}}{=} 0.047 \quad [35.b]$$

The log-odds associated with  $\mu_\alpha$  equals to  $\ln(0.953/(1 - 0.953)) = 3$ , while the log-odds associated with  $\mu_\beta$  equals to  $\ln(0.047/(1 - 0.047)) = -3$ .

---

<sup>9</sup> Ragin, 2000, 2008.

<sup>10</sup> Ragin 2008:96, 2000; Duşa, 2018.

The scalar that filters the raw inclusion value  $\alpha$  into  $\mu_\alpha$  is  $\kappa_\alpha$  as in [36.a], whereas the scalar that filters the raw exclusion value  $\beta$  into  $\mu_\beta$  is  $\kappa_\beta$  as in [36.b]:

$$\kappa_\alpha = 3/\alpha \quad [36.a]$$

$$\kappa_\beta = -3/\beta \quad [36.b]$$

To retrieve the fuzzy membership of the  $i$ th entity with the raw value  $v_i$ ,  $\kappa_\alpha$  or  $\kappa_\beta$  is applied to the distance of  $v_i$  from the crossover  $\gamma$  depending on whether the entity is above or below  $\gamma$ ; then the transformation is inverted as in [37]:

$$\mu_i = \begin{cases} \frac{e^{\kappa_\alpha(v_i-\gamma)}}{1 + e^{\kappa_\alpha(v_i-\gamma)}}, & v_i > \gamma \\ 0.5, & v_i = \gamma \\ \frac{e^{\kappa_\beta(v_i-\gamma)}}{1 + e^{\kappa_\beta(v_i-\gamma)}}, & v_i < \gamma \end{cases} \quad [37]$$

Ragin's calibration thus embeds the idea that the membership scores are the odds that an entity is a member of the set given its raw value. It then provides a correction to the crisp classification of instances as merely in or out. The same rationale applies to the fractional values of the scales (b) to (f) in Table 8, which can be directly assigned to ranked instances or clusters of instances when the uncertainty in membership depends on evidence without a continuous numeric nature.

#### a) *Measurement error*

A common source of concern about calibration stems from the fact that it yields values without measurement error<sup>11</sup>.

The concern entails the assumption that a *mesurand* entity has a correct score in the property set, and that a biased instrument can mistake it. Under such an assumption, the log-odds transformation makes the measurement errors on  $v_A$  interact with the researcher's classification error as embodied by choice of the raw values of  $\alpha_A$ ,  $\beta_A$ ,  $\gamma_A$ , and in ways that may or may not be systematic.

---

<sup>11</sup> Hug, 2013.

The concern is more severe in large-N QCA, where the entities hardly maintain their qualitative identity of cases. Here, it has promoted frontier literature on false negatives and false positives in solutions<sup>12</sup>, as well as on the opportunity of changing the shape of the filtering function<sup>13</sup>.

Especially when the number of cases is limited, the standard strategy may anticipate and discount such consequences by running parallel analyses with different membership scores from purposeful changes in the three calibration thresholds<sup>14</sup>.

A corresponding, straightforward strategy narrows on the researcher's bias and treats them as a more conventional coder bias<sup>15</sup>. The reliability of fuzzy scores can improve by multiplying the coders and their heterogeneity, then solving discrepancies in assigned values or thresholds by a collective agreement on weights.

As an alternative, the systematic tendency to strictness, generosity, certainty, or uncertainty that may affect a single coder's calibration can be simulated by calculating the concentration [38], dilation [39], intensification [40] or moderation [41] of the original  $\mu_{ik}$ .

$$\textit{concentrated } \mu_{iA} = (\mu_{iA})^{2.0} \quad [38]$$

$$\textit{dilated } \mu_{iA} = (\mu_{iA})^{0.5} \quad [39]$$

$$\textit{intensified } \mu_{iA} = \begin{cases} (\mu_{iA})^{0.5}, & \mu_{iA} > 0.5 \\ (\mu_{iA})^{2.0}, & \mu_{iA} < 0.5 \end{cases} \quad [40]$$

$$\textit{moderated } \mu_{iA} = \begin{cases} (\mu_{iA})^{2.0}, & \mu_{iA} > 0.5 \\ (\mu_{iA})^{0.5}, & \mu_{iA} < 0.5 \end{cases} \quad [41]$$

As these transformations render the perturbation of all the calibration decisions in known directions, parallel analyses on the resulting datasets provide an assessment of the robustness of the solutions. From

---

<sup>12</sup> Bramoeller, 2015; Rohlfing, 2018.

<sup>13</sup> Thiem, 2010.

<sup>14</sup> Maggetti and Levi-Faur, 2013; Skaanig, 2011.

<sup>15</sup> Smithson and Verkuilen, 2006.

a different perspective, the comparison among solutions provides a validation of the main threshold-setting decisions when they ensure higher parameters of fit.

### 8.2.1 *The relationship between fuzzy and crisp scores*

The capacity of fuzzy scores to render classification uncertainty does not mean they lose grasp of the qualitative difference among instances. The crisp classification still holds with fuzzy scores, following the rule of conversion in [42]:

$$A_i = \begin{cases} 1, & \mu_{iA} > 0.50 \\ 0, & \mu_{iA} < 0.50 \end{cases} \quad [42]$$

where  $A_i$  is the crisp membership of the  $i$ th entity in the set of the attribute  $A$ , while  $\mu_{iA}$  is the fuzzy membership of the same  $i$ th entity in the same set. Ragin's fuzzy sets can be better conceived of as crisp sets with internal variation – or crisp sets for which the uncertainty in membership is reported for each unit.

The conceptual correspondence of fuzzy and crisp sets is emphasized by the absence of conversion rules for the fuzzy value of 0.5 in [42]. The absence follows from the later convention that the crossover shall not be assigned to any actual unit of analysis, or that the instances on the crossover shall be dropped as they carry useless information to assess set relationships<sup>16</sup>.

When designed to satisfy [26], fuzzy sets allow truth tables where the distinction between observed and unobserved primitives still holds in the universe based on information about the entities' differences in kind. The consistency of sufficiency, instead, is based on the degrees of membership that any unit maintains into any primitive and any solution – which requires some additional caution.

Fuzzy sets preserve full idempotency for certainty instances only. Those with partial membership instead strain the principle of the excluded middle, although not the principle of non-contradiction. For example any  $u_i$  with  $\mu_{iA} = 0.4$  has a  $\mu_{i\bar{A}} = 0.6$ , so that  $\mu_{iA} \cap \mu_{i\bar{A}} = \min(0.4, 0.6) = 0.4$ . It is easy to generalize that, for any  $u_i$ , its membership in the intersection  $\mu_{iA} \cap \mu_{i\bar{A}}$  is always lower than 0.5 so that the entity never belongs to it. The statement 'u is A' and its opposite, then,

---

<sup>16</sup> Ragin, 2008; Duşa, 2018.

again are not true at the same time of the same instance as in classical logic and in the original Boolean algebra; nevertheless, numerically the intersection of a fuzzy set and its complement is not empty, which affects the parameters of fit.

### 8.2.2 *Necessity and sufficiency with fuzzy scores*

With fuzzy scores, sufficiency as a subset relationship is established as the *containment*<sup>17</sup> of the membership function of factor  $W$  by the membership function of effect  $Y$ , as in [43]:

$$\mu_{iW} < \mu_{iY} \quad [43]$$

where  $\mu_{iW}$  is the fuzzy membership of  $u_i$  in set  $W$ ;  $\mu_{iY}$  is the fuzzy membership of the same instances in set  $Y$ , and  $<$  is the relation of containment, reading 'lower than'.

Thus, fuzzy sufficiency is proved when the memberships in  $W$  are systematically lower than the memberships in  $Y$ . Similarly, necessity is defined as the containment of the membership function of  $Y$  by that of  $A$ , so that the fuzzy set  $A$  is necessary to the fuzzy set  $Y$  when the instances' memberships in  $A$  are systematically higher than the memberships in  $Y$ , as in [44] below:

$$\mu_{iA} > \mu_{iY} \quad [44]$$

It follows that the relationship of necessity and sufficiency eventually arises when the memberships in a factor equal those in the effect and  $\mu_{iA} = \mu_{iY}$ .

In short, with fuzzy sets, a necessary factor is that which distributes the instances of the universe below the bisector in a lower-triangular shape, while a sufficient factor is that which distributes the instances of the effect above the bisector in an upper-triangular shape, and a necessary and sufficient one would lay instances on the bisector.

These shapes do not undermine the capacity of the parameters of consistency [30] and [31], and of coverage [33] and [34], to assess the fit of actual distributions. However, simultaneous residual memberships in  $Y$  and  $\bar{Y}$  inflates the calculation of consistency and may affect the identification of solutions.

---

<sup>17</sup> Ragin, 2000, *cfr.* Zadeh, 1978.



A further parameter, the Proportional Reduction of Inconsistency (PRI<sup>18</sup>), has hence been devised to complement the information from the S.cons and to discount the possible inflation of the score due to residual memberships.

$$PRI = \frac{|W \cap Y| \setminus |W \cap Y \cap \bar{Y}|}{|W| \setminus |W \cap Y \cap \bar{Y}|} = \frac{\sum_{i=1}^n \min(W_i, Y_i) - \sum_{i=1}^n \min(W_i, Y_i, \bar{Y}_i)}{\sum_{i=1}^n W_i - \sum_{i=1}^n (W_i, Y_i, \bar{Y}_i)} \quad [45]$$

Calculated as in [45], the PRI index again spans between 0.00 and 1.00, and its higher values indicate that the claim  $W \rightarrow Y$  is fuzzy consistent. Its values are less generous than those of the S.cons. They are commonly used to refine the decisions on which configurations shall enter the minimization as sufficient to the effect under analysis.

### 8.3 Designing membership scores

Whether fine-grained membership better renders the observed variation of the manifestations of a concept, hence, only depends on how we construe them – that is, on how we establish our thresholds. Setting thresholds is the most delicate of the operations. It elicits a solution to the thorny problem of establishing the extension it has in the universe of analysis so that it faithfully mirrors its intension.

The deductive strategy of clarifying the intension first prevents the risk of stretching attributes beyond their meaning, which would embark far more hidden heterogeneity than desirable for the analysis<sup>19</sup>. At the same time, thresholds are meant to shape fuzzy scores that capture the relevant variation in the attribute across the universe of reference. As such, they may preclude or flaw the analysis when they enforce an ideal yardstick that none of the entities in a universe displays. In short, thresholds can become useless when decisions are not fine-tuned to the universe of discourse.

The scholarship has circulated several recommendations to prevent these risks, which together compose a research design to ensure the claim of the sufficiency of the starting model is empirically sound, amenable to interpretation, and treatable with Boolean algebra. The recommendations help the researcher to tackle three intertwined problems –

---

<sup>18</sup> Ragin, 2006; Schneider and Wagemann, 2012.

<sup>19</sup> Sartori, 1970.

namely, to (a) establish the empirical domain of reference; (b) define the operationalization of the attributes; (c) identify thresholds that align attributes and evidence. In actual research, the point of attack may change; however, the resulting membership scores shall provide a single solution to these three issues at once – likely, after some iteration.

### 8.3.1 *Establish the universe of reference*

As in any technique, cases provide an empirical ground to the analysis as solid as the criterion for their selection. The selection of cases for QCA, then, must ensure as unbiased minimizations as possible.

Standard minimizations compare alike configurations with the same outcome and drop their unique changing condition under the further logical constraint that the solutions to the effect and those to the absence of the effect do not overlap. In so doing, as in most dissimilar systems with similar outcome designs<sup>20</sup>, they illuminate meaningful constant associations against a background of theoretically and logically irrelevant variations. Cases' *heterogeneity to the background factors* is the first criterion that case selection shall meet.

Cases allow for meaningful set relationships and the possibility that the hypothesis is logically wanting when they are selected to ensure enough variation on both the outcome and each explanatory factor<sup>21</sup>. Were the cases only selected on the presence of some explanatory factor of interest, the attribute would turn into a constant and easily disappear from the parsimonious solutions simply due to distributional reasons. Were the cases only selected on the presence of the effect of interest, evidence would artificially prevent inconsistencies – thus making the assessment impossible of the logical validity of results. Hence, the second criterion that cases selection shall meet is to *ensure the variability of all the factors*.

The negative existential claim embodied by [27] and [28] only stands if we do not cherry-pick our cases but also if we refer the analysis to a closed universe of reference. The conclusion that the starting model generates a non-contradictory truth table, hence that its implicants do explain the effect, would prove hasty unless we consider all the actual diversity, which in turn is humanly impossible within an open universe. Besides, these boundaries provide a geographical, historical, cultural limit to the heterogeneity of the instances – thus supporting meaningful

---

<sup>20</sup> Przeworski and Teune, 1970; Little, 1996; Rihoux and Grimm, 2006.

<sup>21</sup> Collier and Mahoney, 1996.

decisions about the inclusion and exclusion of each entity from the set of an attribute. Indeed, whether a country qualifies as ⟨rich⟩, ⟨democratic⟩, or ⟨equal⟩ may require different  $\alpha$ ,  $\beta$ ,  $\gamma$  depending on the world region and the timeframe of the analysis. The last criterion for case selection, hence, consists of finding the *ultimate scope condition* that enables meaningful calibration while setting the limits of the validity of results.

### 8.3.2 *Operationalize attributes' intension*

The operationalization of an attribute consists of finding the pieces of evidence that count as its local manifestation across contexts – i.e., the evidence related to an observation sentence. Attributes so become the features that we recognize to an entity in force of some property, and that makes it indiscernible from others manifesting the same or equivalent features. The scholarship emphasizes that evidence can come in any shape – from official statistics to interview responses and expert scorings or portrayals<sup>22</sup>. The operation of aligning evidence to observation sentences in a meaningful way is seldom straightforward, however.

A first problem is the one Verba<sup>23</sup> pinpointed – of treating pieces of evidence as they were attributes, leading to narrow observation sentences. Following Verba's volcanic suggestions, the problem can be solved by scaling up the ladder of abstraction, and by establishing functional equivalences in a catalog of items equally suitable to fill a role in the explanatory schema. If required, moreover, the broader class can be turned into the attribute that the theory expects successful units would display, and failing units would lack.

The second and opposite problem arises when types are far too encompassing to be meaningful. This mainly affects ambiguous or contested attributes for which no precise opposite exists. Such openness can serve the purpose of defining attributes whose intension does not follow the hierarchic structure entailed by the ladder of abstraction<sup>24</sup>. Nevertheless, without clear opposites, we risk “theoretically, a ‘nullification of the problem’ and, empirically, what may be called an ‘empirical vaporization’”<sup>25</sup>.

---

<sup>22</sup> Basurto and Speer, 2012.

<sup>23</sup> Verba, 1967.

<sup>24</sup> Goertz, 2017; Collier and Mahon, 1993.

<sup>25</sup> Sartori, 1970:68.

The solution follows from the consideration that, usually, concepts are left open when their intension is multidimensional, and the rule of composition is unclear. We can address such an uncertainty within a given research question, by selecting those dimensions of the concept that better fill the role in our nomological arrangements. We can use them as separate factors, and let their contribution and relationships be decided empirically, by minimizations: when the factors in the model are too many to the cases at hand, however, the findings multiply, and their interpretation can prove daunting. Alternatively, we may establish suitable composition rules under some theoretical assumption, and “compress” them into “superconditions”<sup>26</sup>. The compression reduces the primitives in the truth table and, as a consequence, the number of solutions from minimizations. Besides, configurational findings from compressed factors do preserve their interpretability.

This, in the happy case that raw measures render each dimension of the attribute faithfully for all the cases. Often, however, some measures will prove unsuited to a QCA due to missing values. The technique cannot handle them, as the units for which the value is missing would have an open membership in the attribute – that is, should be given the score of highest uncertainty. Missing values require some credible raw estimation to resolve the dilemma of adjudication – else, the measure can be substituted with a complete gauge of the same intension, or the uninformative unit can be dropped from the analysis<sup>27</sup>.

### 8.3.3 *Identify the thresholds for calibration*

Thresholds explicate the rule the researchers follow to establish a unit is an instance of the attribute.

The default recommendation is to base these decisions on external anchors<sup>28</sup>. The advice is especially compelling about the crossover, as it determines which configurations will be observed and prove consistent with which effect. At the same time, the advice can result in misleading measures when applied automatically.

In policy studies, particularly for indexes of extensive usage of national systems performance, recognized tipping points exist that can be employed as external anchors. Among the many possible examples are

---

<sup>26</sup> Berg-Schlosser and De Meur, 2009; Elman, 2005; Damonte *et al.*, 2014; Damonte and Negri, 2019.

<sup>27</sup> Ragin, 2008; Basurto and Speer, 2012; Duşa, 2018.

<sup>28</sup> Ragin, 2000, 2008.

interest rates around 4%, taken as a neutral signal about the overall performance of the economy; a Gini index equal or above 30, sorting high from low economic inequality; a rate of 60% of citizens in the right age cohort enrolled in higher education programs, marking a watershed in the expected quality of societies. These are all critical points as they flag a shift in the state of the underlying system. Strategies and interactions of the key actors in the field change depending on which side of the gauge the system lays at a particular time. Nevertheless, they may also prove misleading or useless – especially when no natural zero is involved. Periods of very low-interest rates may lower the tipping point; outside postwar Europe, the Gini index may never yield low scores; the enrollment rates of teenagers from low earning families, or some minority, may systematically remain below the chosen threshold. In short, under special contingencies or within particular regions, a strict reliance on common knowledge may generate skewed distributions – which, in QCA terms, will curtail the empirical diversity by calibration design, and create quasi-constants that, for this reason, may disappear from parsimonious solutions. Moreover, a conventional tipping point may happen to coincide with some cases, making them uninformative in spite that all their raw information is available.

To avoid the side issues from blind reliance on common knowledge, a researcher can correct conventions with distributional considerations. The practical suggestions mainly concern distributions generated by continuous raw measures of single attributes<sup>29</sup>. Although standard descriptive statistics are usually deplored as a significant guide for threshold setting because of their lack of qualitative meaning, considerations about quintiles seem unavoidable in large-N designs<sup>30</sup>. A complementary strategy, and more sensitive to the qualitative arguments of homogeneity, prescribes the reliance on cluster analysis to identify possible positions for thresholds, following the assumption that cases closer to each other along the single dimension should belong to the same side of the property – hence, that threshold should fall in the “natural gaps” between two clusters. The cluster analysis to find thresholds has long been offered as a standard function by many software packages<sup>31</sup>, although it can also be performed more intuitively. This support, however, only partially relieves the researcher from the responsibility of choosing

---

<sup>29</sup> Ragin, 2000, 2008.

<sup>30</sup> Ragin and Fiss, 2017.

<sup>31</sup> Duşa, 2018; Cromqvist, 2011.

the threshold, giving that the natural gaps identified with cluster analysis change depending on the number of clusters the researcher asks the algorithm to retrieve. In such cases, multiple thresholds can be calculated, to opt for the two robust ones that censor the tails without over-emphasizing certainty, and for the mid one closer to statistical as well as conventional knowledge.

A qualitative check can be performed on the cases that would lay above and below the crossover, to dispel the last doubts and verify whether the decision chimes with the available qualitative knowledge<sup>32</sup>. As instances with partial membership between 0.51 and 0.70 will shift below the crossover under concentration, while those with partial membership between 0.26 and 0.49 will shift above the crossover under dilation, the test of robustness with fuzzy transformations will nevertheless provide the opportunity to consider the effect that residual biases in calibration may have on contradictions in primitives. Robustness tests are even more justified when fuzzy scores are directly assigned based on qualitative evidence<sup>33</sup>.

---

<sup>32</sup> Ragin, 2000, 2008.

<sup>33</sup> de Block and Vis, 2018.

## 9. The protocol

To run an explanatory QCA, a researcher may want to:

### **a) specify an interpretable statement of sufficiency –**

1. *Define the effect of interest, and learn about its generation.*  
Review the related theoretical and empirical literature to find testable definitions of the effect of interest, and identify a convincing underlying process to it. Learn how the process and its conditions have been understood and measured. When the literature reports predictors without clear causal channels, you may still ask yourself what the predictor captures of the process.
2. *Formalize a nomological machine.*  
Identify a plausible nomological arrangement. Recall QCA is committed to the assumption that causation is ontologically complex, and that mechanisms only obtain under the conjunction of special attributes related to each other and the effect by *inus* composition rules. The starting model can be conceived of as an overdetermined nomological arrangement under which the process to the effect would not fail, was it observed. Hence, it has the shape of the sufficiency statement  $\bigcap_{k=1}^m A_k \rightarrow Y$ .  
Elucidate the directional expectation from theory and previous literature that the model embarks about each of the  $k$  attributes.

**b) operationalize it –**3. *Select the proper population and gauge.*

Identify the universe of reference in which to assess the statement and the raw variables that render its components in that universe. Please note that data availability may affect the specification and the operationalization of the attributes in the statement, as well as the width of the population. Within these constraints, choose the raw measures that render each attribute as faithfully as possible across the universe without missing values and at the proper level of abstraction.

4. *Turn the measures of each inus component and the effect into set membership scores.*

Discount uncertainty in classification by identifying the degree of variation in membership that the raw measures allow; identify the required thresholds; assign membership scores to instances with proper operations.

As different scales may affect the assessment of set-relationships, apply the same scale to all the attributes.

Consider whether the specification of the model may benefit from the compression of some components; in case, add the compressed attributes to the dataset.

**c) analyze it –**5. *Assess the claim of the individual necessity of the components to the occurrence of the effect.*

Conventionally, the claim is supported when the attribute in any state gets values of N.consistency higher than 0.95 to either the effect or its negation – although an attribute that is necessary to a single compound can have lower scores, down to 0.80. Consider the N.coverage and the Relevance of Necessity to exclude the attributes are trivial, as trivial necessary conditions – i.e., constants – can be dropped.

Consider whether compressed conditions obtain better or worse scores on N.consistency and N.coverage than the original ones; in case of better scores, consider substituting the initial attributes with their compression.



Consider N.consistency and N.coverage to appraise the empirical standing of the directional expectations established in stage *a*), step 2.

6. *Assess the claim of sufficiency.*

Build a truth table from the expansion of the statement, assign instances to primitives, and calculate the S.consistency and the PRI of the observed primitives to both the effect and its negation.

As the claim of sufficiency is weakened when observed configurations cluster instances with opposite effects, check for contradictory configurations. If found, check whether the contradiction arises because of:

- the misclassification of a case to the crossover in the calibration of the attributes or the effect. In such a case, change the threshold;
- the improper operationalization of one or more attributes. In such a case, change the gauge;
- the misspecification of the statement. In such a case, add a further attribute consistent with the starting hypothesis to the statement and capable of telling the positive from the negative instances in the contradictory configurations.

Then, re-run the analysis.

7. *Minimize.*

Establish the cutoff in the values of S.consistency below which the observed primitives will not be deemed consistent with the claim of sufficiency to both the positive and the negative outcome. The original software for the Standard Analysis implements a default cutoff of the S.consistency to the effect of 0.85, although lower values can be justified if the PRI scores preserve high values. Please note the R package instead performs this operation while building the truth table.

Find the conservative, the parsimonious, and the plausible solutions.

Consider the difference in the composition of each prime impli- cant from the parsimonious and the plausible solution. If new conditions appear in the latter, check whether the S.consistency is higher in the plausible than in the parsimonious – whether the plausible solution is Salmon-superior to the parsimonious. If so, the addition is detectably meaningful, and the plausible solution is

that which deserves full discussion. If not, consider re-running the analysis from step 6 on a hypothesis without these additional conditions.

8. *Plot the solutions to the effect and its negation.*

Draw the XY-plots to check the fitting of the instances to the upper triangular shape, assuming the fit is good when the solution lays all the cases above the  $y = x + 0.1$  line.

Discuss which implicants explain which instances of the effect. Consider unexplained instances, if any.

**d) check robustness, and go back to the theory**

9. *Check the robustness of results.*

Calculate different datasets with perturbations – diluted, concentrated, moderated, and intensified scores. Delete some cases to make some observed configuration unobserved. Run parallel analyses on these datasets.

Compare the insurgence of contradictory configurations, the change in consistency and coverage of individual attributes, the configurations that survive, and those that do not, the differences in solutions.

Draw conclusions about the standing of your results, and the assumptions that support it.

Please note these results usually go into the limitation section and in the appendix<sup>1</sup>.

10. *Return to theory.*

Consider what the findings of the nomological compound add to the previous empirical and theoretical knowledge about explanatory types and whether they refine the expectations about the capacity of single attributes to yield the effect.

Consider what the cases in failing configurations can learn from those in close successful configurations to improve their effectiveness. Please do not overload the results with meaning about the nature of the cases. Keep in mind that the findings are valid within

---

<sup>1</sup> For further details on the current standards of transparency, please refer to Schneider *et al.* (2019).

the space-time region where the gauges and setting of the operational reports make sense and that cases can change their membership into configurations over time.

Consider limitations and further extensions.

The last part illustrates the protocol with the support of an actual example. Please note that the example is not to be taken as the blueprint of an ideal “research product” – be it an article, a policy report, or a book. Instead, it has to be taken as an example of how to develop explanations with a configurational mindset and get results in line with fundamental criteria and assumptions about configurational causation. The presentation of the findings and the causal story that supports them have to be fine-tuned to the need of the addressees – a topic beyond the scope of this work.

*intentionally blank page*

Part III

THE EXAMPLE

*intentionally blank page*

## 10. What causes corruption, and what can hinder it

Even a cursory look at the literature reveals that, first, not everywhere, and not always, corruption has been understood as a problem. For decades, it instead had been justified for its capacity of greasing the wheels of oppressive bureaucracies, easing transactions, and improving allocation. It had been seen as the strategy available to the most entrepreneurial players to bypass “red tape” and speed up procedures and as a tool for keeping wages low and thus contributing to national competitiveness. Corruption had even been ascribed the capacity of cementing fragmented countries together, as it would open informal access to the policymaking process to groups and interests that would otherwise be excluded.

Corruption was thus acknowledged and justified in the light of a particular belief. It was a byproduct of “second-best worlds,” such as weak states, authoritarian systems, or democracies in their early stages of development, and the “modernization” of the political system would have made it disappear<sup>1</sup>. A rich empirical literature came to challenge such a “romantic” view<sup>2</sup>. Evidence heaped up that corruption spoils national economies instead of contributing to their innovation<sup>3</sup>. Corruption may strengthen large enterprises in the short run, allowing them to build monopoly rents and scale economies, but, in the long run, it takes a toll on their competitiveness. To small and medium enterprises, corruption is of little benefit as it always reduces their rate of return to

---

<sup>1</sup> Huntington, 1968; Nye, 1967; Leff, 1964.

<sup>2</sup> Tanzi and Daavodi, 2001.

<sup>3</sup> Gardiner, 1993; Kitschelt and Wilkinson, 2007; Klitgaard, 1988; Scott, 1972; Tanzi, 1998; della Porta and Vannucci, 2013.

investment<sup>4</sup>. An inverse relation has also been detected between corruption and direct foreign investment, with a marginal effect equal to an increase of more than three times in the tax rate<sup>5</sup>. Furthermore, corruption affects the productivity and composition of public investments and the national allocation of talent in ways that undermine growth. It widens the share of the informal sector, undermines the progressivity of taxation systems, decreases the rate of revenues, and fuels national fiscal deficits<sup>6</sup>.

Although such empirical results may be far from conclusive, gains from corruption can no longer be assumed. More importantly, the geography of scandals proven mature democracies are not immune to malpractices. Instead of vanishing, corruption has been often observed to adapt, survive reforms, and spread around<sup>7</sup>. The typical bottom line to these research streams is, everywhere the “quality of government” can decline – corruption being a significant indicator, catalyst, or cause of such deterioration<sup>8</sup>. However, corruption metrics display significant variations across political systems with similar degrees of maturity. Thus, our guiding question asks: what can explain these differences across cases?

## 10.1 Define: how do we recognize corruption?

A usual starting point in corruption studies is the remark that the phenomenon lacks a precise and shared definition. For instance, corruption is commonly understood as office-holders’ “misuse” of their power at their advantage<sup>9</sup>. The problem here is, which criteria should identify “misuse.” The literature resorts to three renowned criteria to draw this line<sup>10</sup> – none of which has escaped criticisms.

The first criterion has a legal nature and refers either to some meta-historical principles or to the rules in force in the given context. However, the former could prove culturally biased; the latter, too relativistic and lax as, in some context, certain practices may not be perceived as

---

<sup>4</sup> Isham *et al.*, 1995.

<sup>5</sup> Wei, 1997.

<sup>6</sup> Mauro, 1998; Tanzi and Davoodi, 2001.

<sup>7</sup> della Porta and Vannucci, 2016; Rothstein, 2011; Rose-Ackerman, 2001; Piattoni, 2001; Elliott, 1997.

<sup>8</sup> Rothstein, 2011.

<sup>9</sup> Rose-Ackerman, 1999.

<sup>10</sup> Heidenheimer and Johnston, 2007.



corruption. Article 1 of the United Nations Convention against Corruption has attempted to circumvent the issue and defined the phenomenon by antithesis. Thus, corruption is discussed as the opposite of “integrity, accountability, and proper management of public affairs and public properties.” Since 2003, further standards have been specified by the United Nations’ Convention against Corruption, based on inventories of offenses. Its article 1 posits integrity, accountability, and proper management of public affairs and public property as the purpose of anti-corruption. In Chapter III, it associates corruption with bribery, embezzlement, and diversion, trading in influence, abuse of functions, illicit enrichment, laundering of proceeds of crime, concealment, obstruction to justice. The Convention, however, leaves the state parties the freedom to detail the meaning of “undue advantage” or of “lawful income,” and the right to consider themselves unbound by special provisions. Interestingly enough, state parties’ reservations mostly concern Article 66.2, establishing the International Court of Justice as the default alternative to arbitration in settlement of international disputes on the application of the Convention. Scattered national resistances to the definition of a joint international jurisdiction signal that important differences remain among legal systems in their approach to corruption and leveling differences constitutes a sensitive political issue.

The second criterion understands corruption as the influence that bends public decisions in directions contrary to the public interest. However, different understandings of the public interest may coexist in the same society. The problem does not disappear if we rely on some consolidated economic yardstick to define it. The definition of winners, losers, and possible compensations may draw the issue back to the matter of standards. If “public” means “general” or “universal” policies, the criterion again may prove demanding, as policies often benefit particular constituencies that “deserve” it<sup>11</sup>.

The third criterion recognizes corruption in the improper application of market and patrimonial rationales in the domain of legality. Along this line, it occurs when the power-holders maximize the benefits from their positions by giving higher bidders a disproportionate influence on the policymaking process. This would make corruption overlap policy capture by lobbyists: but while the latter can ideally be reversed by a policy change<sup>12</sup>, the hallmark of corruption is its capacity to survive

---

<sup>11</sup> Kitschelt and Wilkinson, 2007.

<sup>12</sup> Hall, 1993.

regulatory changes – even of institutional regimes<sup>13</sup>. Moreover, an inverse relationship has been detected between contexts where legitimate pressures are allowed and those where illicit channels provide the only access to the policymaking. Besides, from the perspective of the private agent in search of a favorable policy decision, lobbying has been proven a more effective strategy than bribing to gain and retain an effective influence on the content of policy decisions<sup>14</sup>.

Further analyses agree that, beyond their differences, all these definitions revolve around a core meaning, that of corruption as *unfairness*<sup>15</sup>. Fairness consists of the enduring commitment to the unbiased treatment of any party<sup>16</sup>. Thus, office-holders qualify as corrupt when they trade their discretion on the public resources that should serve whatsoever policy purpose and instead deliver individualized benefits beyond or against that purpose. This definition has the advantage of applying to the petty corruption of “grease money” as much as to the grand corruption of grafts and understands both in the light of a basic principle of the rule of law. Besides, fairness suggests a useful ground to grasp the mechanism that makes corruption toxic and hard to eradicate.

## 10.2 Learn: how does corruption come about?

The literature suggests that corruption hinges on a self-standing social mechanism and that the effort of curbing it consists of locating the levers and pulleys to block it<sup>17</sup>. Moreover, the literature emphasizes that perceptions as beliefs and personal experience play an active role in maintaining and institutionalize the social mechanism of corruption, as they affect people’s expectations about the actual working of the system and the individual strategies of access to the policymaking system. In short, perceptions do change the functioning of the policy system. The causation stems from the connection that fairness and generalized trust maintain in democracies through the mediation of institutions, and the perceived corruption tarns the relationship between fairness and trust

---

<sup>13</sup> della Porta and Vannucci, 2006, 2012.

<sup>14</sup> Campos and Giovannoni, 2007.

<sup>15</sup> Rothstein and Teorell, 2008.

<sup>16</sup> Lind and Tyler, 1992.

<sup>17</sup> della Porta and Vannucci, 2006, 2012.

against the backdrop of the assumption of actors with bounded rationality.

### *10.2.1 Cogs and wheels*

Experimental evidence has long suggested that, before knowing the gainers and the losers, fairness is people's preferred ruling principle. Individuals can agree on restraining themselves to the benefit of others if the restrictions would have applied to anybody in their same position and situation. Moreover, they can stick to the principle in their decisions over time until they can assume that everybody does the same or, at least, that opportunists are sanctioned. When people understand that abuses are seldom prosecuted, opportunism becomes a dangerous temptation. It paves the way to a tragedy of the commons in which even basic provisions are primarily delivered after unfair access.

The standard solution to a tragedy of the commons is regulatory and includes a capable pool monitor, and certainty of sanctions on transgressors as deterrents<sup>18</sup>. However, monitors can be corrupted, or violations meet forbearance, or whistleblowers be harassed. These possibilities trigger a "second-order trap." In such a situation, people can find it rational to not report violations and let the controls fail, too<sup>19</sup>. The second-order trap has especially harmful effects as it institutionalizes corruption as the way to access a provision regardless of whether one is entitled to it. As a consequence, even the best-designed policies – included those against corruption – turn into distribution machines and means of political competition<sup>20</sup>, while the principle of fairness associated with formal subjective positions weakens.

Unsurprisingly, then, perceived fairness has been found to affect people's support for authorities and their decisions, even regardless of the policy delivered<sup>21</sup>. This effect has been robustly detected in local contexts such as workplaces, classrooms, police stations, and courts<sup>22</sup>.

Extended to the wider political context, the perception of fairness has been found to moderate the relationship between voters' perceptions of economic performance and their approval of the government, then the relationship between economic outcomes and electoral support

---

<sup>18</sup> Ostrom, 2004; Frey and Bohnet, 1995.

<sup>19</sup> Rothstein and Teorell, 2015; Mungiu-Pippidi, 2013; Ostrom, 1998.

<sup>20</sup> Zhu and Zhang, 2017.

<sup>21</sup> Lind and Tyler, 1988; Brockner, 2011.

<sup>22</sup> Colquitt *et al.*, 2001, 2013.

for the incumbent parties, regardless of voters' political sophistication<sup>23</sup>. The generalization yields less convincing results when tested within a country<sup>24</sup> than in cross country comparisons<sup>25</sup> – which points to the contribution of the national institutional framework on these perceptions. Thus, we can say that inequality is accepted when it follows fair policies.

The effect of fairness on the acceptance of corruption has been emphasized by further experiments proving that the time spent in a less corrupt context makes the country of origin lose its power to predict individual propensity to enter bribes<sup>26</sup>. Evidence corroborates the theoretical tenet that free-riding mainly is a local and conditional result – some myopic rational response to institutional incentives, which makes them insensitive to the prospect of harmful collective consequences in the long run<sup>27</sup>.

Although elicited by local experience, the belief of system unfairness has lasting systemic effects as it undermines individuals' predisposition to *trust*.

Trust arises when “the probability” we entertain that someone “will perform an action that is beneficial or at least not detrimental to us is high enough for us to consider engaging in some form of cooperation with him”<sup>28</sup>. As such, trust has been associated with lively economies and robust democratic compacts, although the literature contends that the relationship is straightforward.

Ariely and Uslaner<sup>29</sup> establish a difference between the trust that one places in particular groups, and the trust that one reveals as a positive general disposition toward cooperation. They consider the two forms are at odds with each other and see both as the byproduct of economic inequality. From this perspective, inequality triggers the first order trap as it sprawls the belief that the institutional system is stacked in favor of the few. Therefore, people come to distrust other people outside the strict boundaries of kinship, which becomes the reason for demands for less redistributive policies as well as a cooperative basis to gain corrupt access to provisions in a self-reinforcing dynamics, as the diffusion of

---

<sup>23</sup> Magalhães and Aguiar-Conraria, 2019.

<sup>24</sup> Bøggild, 2016; Kumlin, 2004.

<sup>25</sup> Dahlberg and Linde, 2016.

<sup>26</sup> Barr and Serra, 2010.

<sup>27</sup> Ostrom, 2004.

<sup>28</sup> Gambetta, 1988: 2.

<sup>29</sup> Ariely and Uslaner, 2016.

opportunism makes control ineffective. The other way round, when inequality is low, people are better disposed to trust, easily agree on redistributive policies, and stick to fairness in relationships, which makes opportunism a rare occurrence and easily uncovered and punished<sup>30</sup>.

In short, on this account, the causal chain goes from (in)equality to corruption via perceived (un)fairness, generalized (dis)trust, and (non-)acceptance of opportunism as a rule, which reinforces (in)equality. So shaped, the hypothesis is that there is no endogenous solution to the problem. The only route over the inequality trap passes through effective redistributive policies, which in corrupt contexts are easily captured and doomed to fail. Such a discouraging portrayal, however, disregards the experimental evidence that restoring fairness can change perceptions and the inclination to trust. Generalized trust here becomes closer to an imprinted personal trait than to an adaptive response. Rose-Hackerman<sup>31</sup> contends that trusting others can be an imprinted property that pits kinship against universal fairness through the experience of inequality. Instead, she maintains the crucial trade-off arises between two kinds of trust – “one-sided,” and “two-sided” – because of their different rationale.

One-sided trust characterizes the ordinary principal-agent relations that political science posits as constitutive of modern representative governments. It supports the mandate from voters to elected office-holders, as well as from elected to unelected office-holders. It arises when the principal *decides* that the agent can be trusted to act on their behalf, regardless of whether the agent reciprocates. The principal’s decision to trust depends on their consideration that the agent will use their discretion to make choices beneficial to the principal – or, at least, that the agent’s decisions will not harm the principal. The source of one-sided trust, hence, rests on the alignment of the trustee’s and the principal’s interest.

The broad neo-institutional wisdom<sup>32</sup> considers that the principal can expect such an alignment when the trustee’s intentions and behavior are assessed against a standard, and the assessment provides the agent with incentives. It can provide a *social reputational* incentive when it produces information to be circulated for praise or blame. A different reputational incentive comes from *professional* membership, as failing the professional standards of conduct may tarnish the agent’s standing

---

<sup>30</sup> Uslaner, 2004.

<sup>31</sup> Rose-Ackermann, 2001:532 ff.

<sup>32</sup> e.g., Majone, 2001; Bovens, 2007.

among peers. Third, *procedural* standards can justify one-sided trust by making the behavior of an agent more predictable.

Instead, two-sided trust *grows* between two parties from their interactions and over time. It can stem from *mutual dependence*: the principal comes to trust the agent as the agent lets the principal hold something that they value as “security” for the agent’s future behavior. Alternatively, it can depend on *kinship*, and rest on some ascriptive traits when it stems from the agent and the principal belonging to the same group, or on chosen traits when the principal and the agent share beliefs and values that tell them apart from the near context.

This theory, too, recognizes a trade-off between two-sided trust and one-side trust. When activated in fair contexts, dependence and kinship can be exploited to open illegitimate channels to access public resources and bend the functioning of the system. When activated in unfair contexts, however, two-sided trust can be employed to defuse the traps and initiate a paradigmatic change toward fairer institutions. The expected contribution of two-sided trust to the functioning of institutions, therefore, is mixed, as it largely depends on the context. The direction of the contribution, within this perspective, can be decided by pool-monitoring bodies that adjudicate on the contribution of different kinds of kinship on fairness in critical circumstances.

On a more general note, Rose-Ackermann expresses reservations about imprinting as a significant explanation of people’s actual predisposition to trust. In her view, trust is the default predisposition, and distrust remains “a function of the particular situation”<sup>33</sup>. In the light of experimental evidence, such an anthropological optimism can be assumed as a necessary and plausible starting point toward practicable policy solutions. Attempts at defusing the mechanism of corruption can prove worth the effort only if people can change their mind about how trustworthy policy-makers are – hence, if dis/trust is contingent on contexts.

### 10.3 Formalize: hindering corruption

Theory prospects the possibility of reverting the mechanism of distrust beneath the corruption trap by providing citizens with reasons to decide that the policy-makers are trustworthy. Neo-institutional theory and

---

<sup>33</sup> Rose-Ackermann, 2001: 538.

corruption studies consider that these effective reasons depend on whether local institutional designs provide policy-makers with systematic incentives to stay clean even when faced with pressures to bend their decisions.

The broad explanatory schema, then, can be summarized as in Hp.0.0, reading:

$$[\text{Hp. 0.0}] \quad I \rightarrow \bar{C}$$

where  $I$  is for institutional incentives,  $C$  is for perceived corruption, the overbar negates, and the arrow reads “is sufficient to.”

Hp. 0.0 allows maintaining that a given set of incentives is sufficient to account for the perception of corruption. Sufficiency is backed by the processes captured by the interlaced concepts of one-sided dis/trust and un/fairness – all of which remain unobserved at a different analytic level. Please note that the mechanism underlying Hp.0.0 entails a continuity between the “black zone” where corruption takes the shape of crime, and the “gray zone” of malfeasance and policy capture. The continuity is justified from the perspective of hindrance, as malfeasance and capture contribute to undermining the perception of fairness that triggers distrust and convinces people to resort to corruption.

Before we can test Hp.0.0, however, we have to establish whose incentive matter – that is, the reference class.

### *10.3.1 Locating levers and pulleys*

Within the broader policy process that connects voters’ preferences to policy delivery, an adequate reversal strategy has to take into account the contribution of administration in setting the trustworthiness of the system. Even when initiated at the political level by some illicit exchange in the shadow of an “electoral machine”<sup>34</sup>, the delivery of corrupt benefits requires at least the compliance of nonelected office-holders to the pressures to force or bypass the rules during implementation. Being this true, the counterfactual argument should also hold that office-holders can defuse corruption by resisting pressures<sup>35</sup>.

The claim of the hypothesis does not change, but now we consider that the role of institutional incentives in hindering corruption is of supporting administrative resistances to pressures. The driving question in

---

<sup>34</sup> Scott, 1969.

<sup>35</sup> Dahlström *et al.*, 2012; Pelizzo and Stapenhurst, 2012.

the specification stage, therefore, asks under which institutional conditions the appointed office-holders are prone to accommodate corruption, and under which they resist instead.

Studies on the developmental role of the state argue that resistance to corruption is a typical trait of Weberian-like bureaucrats displaying ethical specificity and closedness. Good remuneration is also expected to play a role, as it shields bureaucrats from the temptation to complement poor wages with illegal earnings; and in a design that promotes the contrast of interests. In this stream, Dahlström *et al.*<sup>36</sup> develop these features into two indexes.

The first index captures “professionalism” and is based on the presence of meritocratic recruitment, the absence of political recruitment, and the possibility of internal careers. The second index captures “closedness” that arises from a formal examination system, lifelong careers, and special employment laws for public employees.

Tested on the inverse of the World Bank indicator on the “control of corruption” from 2008, and controlled for economic, cultural, and constitutional confounders, they find professionalism meaningful. However, they conclude, merit-based recruitment alone is not enough to ensure the resistance of bureaucrats to corruption. The study concluded, but left untested, the major explanatory claim that resistance unfolds when administrators develop separate interests from politicians. Such a desirable “contrast of interests” becomes a structural feature of the administrative systems when administrators are held accountable on a different ground from that of politicians.

This conclusion leads to the reformulation of the hypothesis as in [Hp. 0.1] below:

$$[\text{Hp. 0.1}] \quad A \rightarrow \bar{C}$$

where  $A$  is for administrative accountability.

Now, administrative accountability is the common name in the literature to indicate those rules designed to steer bureaucratic discretion in the making of policy decisions “from outside”<sup>37</sup>. The accountability perspective emphasizes that corruption is made possible by the bureaucratic monopoly of the rulemaking and the margins of discretion that

---

<sup>36</sup> Dahlsom *et al.*, 2012.

<sup>37</sup> e.g., McCubbins and Schwartz, 1984; McCubbins, 1985; McCubbins *et al.*, 1987, 1989; Lupia and McCubbins, 1994; Huber and Shipan, 2002; Majone, 2001; Gilardi, 2002.



appointed officials enjoy<sup>38</sup>. Unsurprisingly, accountability constraints render the spirit of Rose-Ackermann’s incentives to ensure one-sided trust.

A common tool employed to ensure administrative accountability is the obligation to justify decisions and behavior. These obligations contribute to professionalism when they commit the rulemakers to expose their decisions and behavior to the scrutiny of third parties beyond the line of the political mandate. Following the principle of the contrast of interests, special gains are expected when the scrutiny is carried on by a plurality of account-holders different from the political principal<sup>39</sup>.

The relevant types of third-party account-holders can be reduced to two: expert bodies, exerting a professional control; and the praising and blaming public<sup>40</sup>. The concerns that may be expressed by the expert bodies about the shape of a policy are possibly at odds to those of the general public. Exposition to the public is expected to mobilize affected interests or special constituencies – those unhappy with the distribution of costs and benefits and attentive to the fairness of the criteria that justified it. The expert bodies, instead, may be interested in the sustainability, legality, and cost-efficiency of the policy decisions. Each separately may “capture” the rulemaking to serve a single concern, which opens to systematic unjustified biases toward a special group. When activated jointly, instead, they can compel some balanced decisions.

The further specification of the hypothesis, then, reads as in [Hp.0.2] below:

$$[\text{Hp. 0.2}] \quad E * P \rightarrow \bar{C}$$

The hypothesis maintains that non-corrupt systems are those in which the rulemakers are accountable to expert bodies  $E$  and, together, to the general public  $P$ . The underlying theory maintains that the two constraints endow the bureaucracies with the incentives to resist undue pressures.

The effectiveness of accountability obligations depends on the experts and the public having the right to access administrative information, and the obligation of the rulemakers to release information and justification in response. Rights and obligations, however, may prove

---

<sup>38</sup> Klitgaard, 1998.

<sup>39</sup> Damonte, 2018; Damonte *et al.* 2014; della Porta and Vannucci 2012.

<sup>40</sup> Schedler 1999; Bovens 2007; Hood, 2010.

erratic or void when they are not backed by some enforcing authority<sup>41</sup>. The consideration brings the judiciary into the picture as the authority of last resort that makes accountability obligations credible. In the case of noncompliant administrative bodies, an effective judiciary allows third parties to enforce their rights through the judicial process; ineffective courts, instead, can turn any right and obligation into empty shells. The last theoretical specification from the literature review, then, reads:

$$[\text{Hp. 0.2}] \quad E * P * J \rightarrow \bar{C}$$

where  $J$  indicates the presence of an effective judiciary and renders its contribution to making the obligations credible of being accountable to experts and the public.

### 10.3.2 Setting theoretical expectations

Given the literature, then, we can expect that the three conditions in the antecedent of [Hp.0.2] be *inus* components: no condition can plausibly ensure the effect unless the remaining two are present, too. Together, the three *inus* components may also prove a Salmon-complete compound in a universe of democratic systems – that is, provided that citizens are endowed with effective political, economic, social, and civil citizenship rights.

Of the three conditions, moreover, we can expect that the accountability to the public and the experts provide the channel to the effect according to the *inus* rationale of the “and” connective. The only positive effect is expected under their joined presence; in the absence of one or the other or both, the constraints should weaken and the rulemakers’ capacity to resist pressures with them.

Finally, the judiciary can be expected to enhance the effect of constraints when the accountability rights and obligations are adequate. When accountability is inadequate, instead, legal prosecution without changes in the rules of accountability may unintentionally provide new reasons for distrusting the policymakers.

### 10.3.3 Gauging levers and pulleys

Unless we are in the happy condition of designing and gathering suitable data, the explanatory factors in the starting hypothesis may have to find their operationalization in existing measures. The topic of

---

<sup>41</sup> Ostrom, 2005; Schillemans, 2008.

corruption and the quality of administration is, fortunately, relevant enough for justifying the systematic release of indicators. However, the overlapping between the index and the conditions in the model is often only partial. Thus, we may need more than an index – or than a sub-index – to render the *inus* factor in the model.

This study chooses its sub-indexes from the Rule of Law Index maintained by the World Justice Project (see Table 10 and notes). These gauges follow from systematic surveys and transparent cross-validation, although they are not released with a measurement error, and the series before 2015 cannot be compared cross-time, which defines the time boundaries of the analysis.

From the dataset related to 2017, the following gauges will be used:

- subindex 1.3. “Government powers are effectively limited by independent auditing and review,” for the condition <ATEC>. According to the metadata, the raw variable gauges whether comptrollers or auditors, as well as national human rights ombudsman agencies, have sufficient independence and the ability to exercise adequate checks on and oversight of the government. The condition will be interpreted as the presence of effective accountability constraints from expert bodies, and an *inus* component in the accountability mechanism.
- subindex 1.5 “Government powers are subject to non-governmental checks,” to calibrate the condition <ASOC>. The raw variable gauges if independent media, civil society organizations, political parties, and individuals are free to report and comment on government policies without fear of retaliation. The condition will be interpreted as conditions enabling and protecting societal actors interested in scrutinizing the policymaking activities, and reporting vital information to the public at large.
- subindex 3.1 “Publicized laws and government data” to calibrate the condition <APUB>. The raw measure gauges whether basic laws and information on legal rights are publicly available, presented in everyday language, and made accessible in all languages. It also measures the quality and accessibility of information published by the government in print or online, and whether administrative regulations, drafts of

legislation, and high court decisions are made accessible to the public promptly.

The condition will be interpreted as the availability of clear information about policymaking activities and contents, and as a trigger of the mobilization of general societal interest.

- subindex 3.2 “Right to information” to calibrate the condition <RTA>.

The underlying raw measure gauges whether requests for information held by a government agency are granted, how timely, if the information provided is pertinent and complete, and if requests for information are granted at a reasonable cost and without having to pay a bribe. It also measures whether people are aware of their right to information and whether relevant records are accessible to the public upon request.

The condition will be interpreted as the right to access government information enabling effective scrutiny.

- subindex 7.6 “Civil justice is effectively enforced” as the condition <ENFOR>.

The raw variable gauges the effectiveness and timeliness of the enforcement of civil justice decisions and judgments in practice.

The condition will be interpreted as the primary enabling factor ensuring the effectiveness of any other element in the nomological arrangement.

The operationalization of the effect as the condition <CLEAN>, instead, relies on the Corruption Perception Index maintained by Transparency International. This again provides a suitable gauge of the perceived level of corruption of the administrative bodies, collected from surveys, and, since 2012, validated through a transparent methodology.

The operationalization turns [Hp.0.2] into [Hp.1]:

[Hp.1] *A TEC \* ASOC \* APUB \* RTA \* ENFOR* → *CLEAN*

Following theoretical assumptions and the operationalization, the analysis will maintain the presence of each condition is required for the compound to yield the positive effect, and the absence of any condition sufficient for it to fail – as summarized in Table 9.

**Table 9. Directional expectations**

<i>Contribute to CLEAN:</i>	<i>Contribute to ~CLEAN:</i>
ATEC * $\Phi$	~ATEC * $\Phi$
ASOC * $\Phi$	~ASOC * $\Phi$
APUB * $\Phi$	~APUB * $\Phi$
RTA * $\Phi$	~RTA * $\Phi$
ENFOR * $\Phi$	~ENFOR * $\Phi$

As the measures of the outcome and of the *inus* factors come from surveys and discount the variations in the year of reference, no need for lagging the effect is envisaged. The perception of the accountability of the administration and the perception of public sector corruption count as individual responses to the same state of the policymaking system. The relevant issue for the method is whether the state of accountability yields observation reports that support the claim as in [Hp.1].

The data of the Corruption Perception Index and the World Justice Project are all collected from a variety of world regions, although not from the same countries. When combined, the more comprehensive coverage is of the countries of the European Union and the European Free Trade Area, and the Anglophone countries. Together, the administrative and institutional systems provide enough diversity to support the patterns detected by the observational reports are meaningful. At the same time, they all are uninterrupted democratic systems, although at different degrees of maturity, which ensures the gauges of the conditions in the model can be given unambiguous interpretations.

After dropping the cases with missing values, the population suitable for the analysis includes 26 cases – namely, Austria, Belgium, Bulgaria, Canada, Czechia, Germany, Denmark, Spain, Estonia, Finland, France, Great Britain, Greece, Croatia, Hungary, Italy, the Netherlands, Norway, Poland, Portugal, Romania, Slovenia, Sweden, United States, Australia, New Zealand – labeled after the Alpha-3 country codes. The raw dataset reads as in Table 10.

**Table 10. Raw data**

CODE	CLEAN	A TEC	ASOC	APUB	RTA	ENFOR
AUS	0.77	0.814708	0.829643	0.910391	0.655894	0.743761
NZL	0.89	0.837319	0.832978	0.855778	0.746930	0.734909
AUT	0.75	0.756031	0.851111	0.714408	0.637350	0.835960
BEL	0.75	0.842677	0.800369	0.684961	0.684168	0.783568
BGR	0.43	0.412690	0.619295	0.576252	0.474647	0.593935
CAN	0.82	0.799486	0.846941	0.875665	0.681263	0.765940
CZE	0.57	0.719355	0.719058	0.667046	0.635679	0.678652
DEU	0.81	0.856441	0.860407	0.752887	0.753014	0.897994
DNK	0.88	0.915851	0.973083	0.843298	0.794127	0.914966
ESP	0.57	0.704767	0.738899	0.704143	0.624528	0.527812
EST	0.71	0.857182	0.794909	0.861620	0.730704	0.714395
FIN	0.85	0.941276	0.905502	0.899718	0.811627	0.882495
FRA	0.70	0.815772	0.755039	0.855324	0.767304	0.720260
GBR	0.82	0.820408	0.849469	0.903424	0.694535	0.772559
GRC	0.48	0.729397	0.683528	0.563534	0.616350	0.411051
HRV	0.49	0.554981	0.682452	0.553799	0.672758	0.540608
HUN	0.45	0.418136	0.490233	0.573962	0.451707	0.497701
ITA	0.50	0.760439	0.685054	0.606626	0.651787	0.367565
NLD	0.82	0.870628	0.853806	0.786342	0.803218	0.886952
NOR	0.85	0.939140	0.925695	0.877084	0.880455	0.912023
POL	0.60	0.589044	0.628946	0.618562	0.663268	0.617409
PRT	0.63	0.789964	0.816835	0.541897	0.664296	0.549812
ROU	0.48	0.548285	0.762163	0.610889	0.602299	0.655217
SVN	0.61	0.627736	0.697531	0.692308	0.645424	0.492498
SWE	0.84	0.942595	0.917292	0.792256	0.947463	0.927448
USA	0.75	0.778349	0.808321	0.766653	0.707624	0.728341

*Key:* AUT: Australia; BEL: Belgium; BGR: Bulgaria; CAN: Canada; CZE: Czechia; DEU: Germany; DNK: Denmark; ESP: Spain; EST: Estonia; FIN: Finland; FRA: France; GBR: Great Britain; GRC: Greece; HRV: Croatia; HUN: Hungary; ITA: Italy; NLD: the Netherlands; NOR: Norway; POL: Poland; PRT: Portugal; ROU: Romania; SVN: Slovenia; SWE: Sweden; USA: United States; AUS: Australia; NZL: New Zealand.

*Source:* CLEAN reports the value of the Corruption Perception Index 2017 collected by Transparency International: [https://www.transparency.org/news/feature/corruption\\_perceptions\\_index\\_2017](https://www.transparency.org/news/feature/corruption_perceptions_index_2017). The other raw measures are from the Report 2017 of the World Justice Project, <http://data.worldjusticeproject.org/>, respectively corresponding to Subindexes 1.3 (ATEC), 1.5 (ASOC); 3.1 (APUB); 3.2 (RTA); 7.6 (ENFOR).

# 11. Set-theoretic configurational analysis

## 11.1 Calibrate the membership scores

All the raw variables in our model are continuous. Hence, we can apply the direct method of calibration. With the assistance of a suitable package, we can inspect the single variables with the following procedure: (a) insulate a raw variable, (b) identify the gaps where the thresholds can be positioned meaningfully, (c) calibrate. The crux of the procedure, as already noted, is the identification of the thresholds.

Threshold setting is the solution to a  $k$ -cluster problem – one for each raw measure in the model. The related functions in the packages can apply a variety of clustering algorithms, none of which, so far, is unassisted optimizing. By design, the decision is intentionally left to the researcher about which variation is meaningful and which instead can be censored. The default setting for continuous variables performs a hierarchical clustering based on the Euclidean distance calculated from complete linkage, which maximizes the distance between the elements in different clusters taken pairwise.

Conventionally, the clustering algorithm operates on the raw variable without previous standardization. In the absence of specific knowledge to support the decision, but also as a strategy for confronting possible biases in our previous knowledge, we can iterate the function and call for increasing numbers of thresholds. This provides information about which groups of units are geometrically closer to each other than to the rest of the population under different discrimination criteria. It should be noted that the first threshold to be identified often insulates the outliers. Thus, symmetry in the order of the gaps cannot lead our reading of the function's output – the crossover is not necessarily the third value retrieved by the setting function when we call for five gaps. Symmetry, however, may prove significant given the effect

that skewed distributions may have in construing conditions with almost trivial individual N.cons scores, which affect the PRI values in the truth table. Standard statistics of each raw measure can then be computed to assist the decision on thresholds and avoid unjustified skewness<sup>1</sup>.

The guiding principle remains that the thresholds should assist the researcher's decision on the raw values that separate the irrelevant from the relevant variation. The following section illustrates the craft of the calibration procedure of the outcome CLEAN. As the keystone of the analysis, this operation deserves special attention.

*a) From raw to fuzzy: calibrating the outcome*

From the statistics of the raw condition CLEAN, we can learn that the span extends from a minimum value of 0.43 to a maximum of 0.89; the median (0.73) is higher than the mean (0.68). Hence the distribution is skewed, and the first and the third quartile fall at 0.57 and 0.82, respectively (see Table 11). This, however, conveys just a rough idea of the actual shape of the distribution. Visualization can improve our grasp, and, since long, it has been the recommended first approach to the issue. The intuitive strategy suggests to plot the ordered data and draw the statistics to better put the distribution in perspective, as in Figure 6.

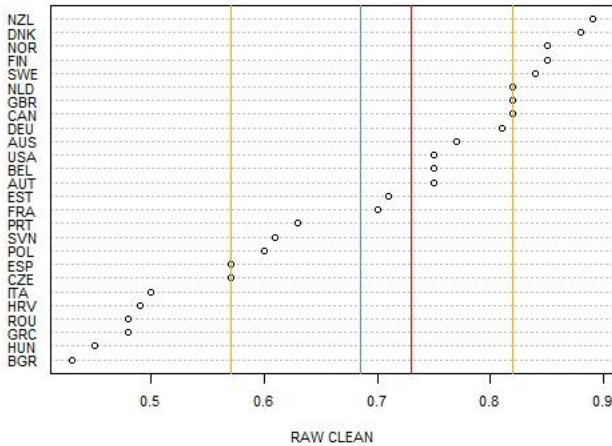
Figure 6 conveys that in 2017, according to the index from Transparency International, the countries from our population with a lower perception of corruption were New Zealand and Denmark, while the perception was higher in Bulgaria and Hungary. The distribution in Figure 6 also shows interesting "natural gaps," where the thresholds could be conveniently placed. The countries below the raw score of 0.5 – Italy, Croatia, Romania, Greece, Hungary, and Bulgaria – seem to be cases with high perceived corruption. Czechia, Spain, Poland, Slovenia, and Portugal seem to cluster together as countries where the perception of corruption is relatively high, to some degree. Then we have two cases in between the distribution, Estonia, and France, which lean toward the almost-certainly clean groups of Austria, Belgium, and the United States, but with a gap in between where the median also dwells. Besides, the distribution of clean countries displays many small groups. From the distribution, we could be quite confident about the positioning of the exclusion threshold; however, we may not be able to dispel the doubts about the inclusion threshold and the crossover.

---

<sup>1</sup> Ragin and Fiss 2017.



**Figure 6. Ranking of cases by their raw scores on <CLEAN>**

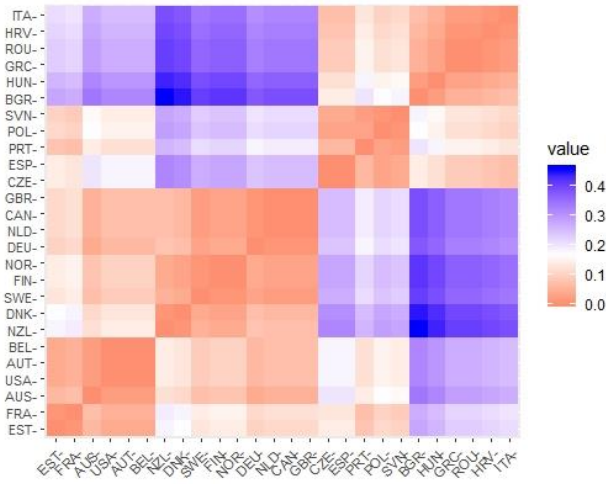


*Note:* the vertical bars in the picture indicate, from left to right, the 1<sup>st</sup> quartile, the mean, the median, and the 4<sup>th</sup> quartile of the distribution.

Then we can resort to the clustering functions to bring further considerations into the picture and hopefully dispel the last doubts. Clustering is based on two types of information, a matrix of distances, and the proximities detected during the process of aggregation.

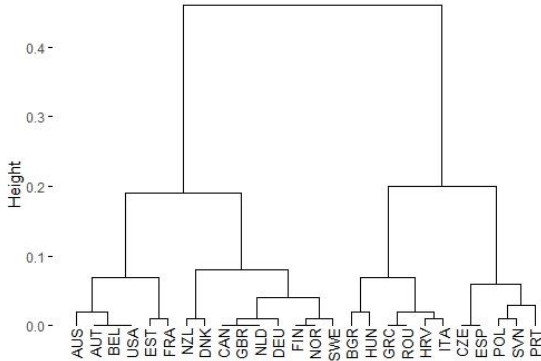
The distances can be displayed as a heatmap, as in Figure 7. The heatmap emphasizes the distances of a case from any other pairwise and renders the information as a different shade – here, of red for those closer to zero, and blue for the farther ones. This information comforts us in our grouping of France and Estonia. The matrix of distances beneath the heatmap says that France has a distance of 0.19 from the cleaner case and of 0.27 from the less clean one; while Estonia is 0.18 from the clean pole and 0.28 from the less clean one, which supports the guess that they shall be above the crossover with the clean countries, despite the uncertainty. The results of the clustering algorithm displayed as a dendrogram in Figure 8 further confirms the intuition. The dendrogram groups France and Estonia with Austria, Belgium, the USA and Australia on the same side of the partition.

**Figure 7. Cases' distances on the raw measure <CLEAN>**



Keys: Each cell displays the Euclidean distance of the case in the row from the case in the column as a different shade. The shades correspond to the values as displayed in the legend on the right side of the map.

**Figure 8. Cases' clusters on the raw measure <CLEAN>**

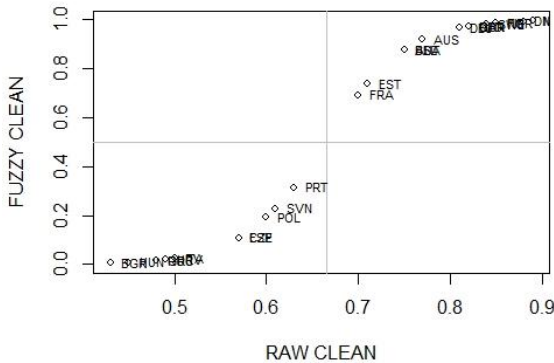


Keys: The dendrogram reports the result of the aggregation of the cases by the maximal distance that they maintain within their cluster from another cluster.

Figure 8 suggests that the inclusion threshold can group as clean all the cases with a raw score of 0.8 or higher. The function for setting the threshold provides us with the values for calibration. The function suggests the exclusion threshold at 0.535, the crossover at 0.665, and the inclusion at 0.790 – as reported in Table 11, together with the calibration decisions of every other condition in the model.

The calibration of the raw variable into the fuzzy condition through the logistic filtering function yields the distribution, as in Figure 9.

**Figure 9. CLEAN: raw by fuzzy scores**



**Table 11. Raw variable statistics**

Raw variable	Min	Q1	Med	Mean	Q3	Max
CLEAN	0.430	0.570	0.730	0.685	0.820	0.890
ATEC	0.413	0.708	0.795	0.756	0.853	0.943
ASOC	0.492	0.703	0.804	0.782	0.851	0.973
APUB	0.542	0.613	0.734	0.734	0.856	0.910
RTA	0.452	0.639	0.678	0.692	0.751	0.947
ENFOR	0.368	0.561	0.724	0.698	0.823	0.947

**Table 12. Thresholds applied for calibration**

Raw variable	Fuzzy label	Exclusion	Crossover	Inclusion
CLEAN	CLEAN.Z	0.535	0.665	0.790
ATEC	ATEC.Z	0.666	0.742	0.807
ASOC	ASOC.Z	0.708	0.778	0.823
APUB	APUB.Z	0.643	0.734	0.818
RTA	RTA.Z	0.630	0.689	0.781
ENFOR	ENFOR.Z	0.572	0.696	0.810

The fuzzy membership values of the outcome are reported in Table 13, together with the result of the same operations on all the other raw measures in the model.

**Table 13. Fuzzy membership scores**

Cases	CLEAN.Z	ATEC.Z	ASOC.Z	APUB.Z	RTA.Z	ENFOR.Z
AUS	0.918	0.962	0.965	0.998	0.166	0.770
NZL	0.994	0.986	0.971	0.985	0.860	0.728
AUT	0.877	0.651	0.991	0.350	0.074	0.972
BEL	0.877	0.989	0.807	0.175	0.441	0.902
BGR	0.005	0.000	0.001	0.007	0.000	0.085
CAN	0.973	0.928	0.988	0.992	0.407	0.854
CZE	0.108	0.298	0.081	0.107	0.069	0.401
DEU	0.966	0.994	0.995	0.657	0.881	0.994
DNK	0.993	1.000	1.000	0.977	0.964	0.996
ESP	0.108	0.196	0.167	0.280	0.041	0.020
EST	0.738	0.994	0.747	0.988	0.787	0.614
FIN	0.986	1.000	1.000	0.997	0.979	0.991
FRA	0.692	0.963	0.280	0.985	0.921	0.649
GBR	0.973	0.970	0.990	0.997	0.543	0.874
GRC	0.016	0.383	0.020	0.004	0.028	0.001
HRV	0.020	0.001	0.019	0.003	0.311	0.026
HUN	0.008	0.000	0.000	0.006	0.000	0.010
ITA	0.025	0.694	0.021	0.017	0.140	0.000
NLD	0.973	0.997	0.992	0.858	0.973	0.992
NOR	0.986	1.000	1.000	0.993	0.998	0.996
POL	0.191	0.003	0.002	0.025	0.221	0.139
PRT	0.315	0.894	0.923	0.002	0.230	0.032
ROU	0.016	0.001	0.342	0.020	0.014	0.279
SVN	0.228	0.013	0.035	0.211	0.106	0.009
SWE	0.983	1.000	1.000	0.881	1.000	0.997
USA	0.877	0.834	0.875	0.754	0.642	0.694

## 11.2 Assess the claim of individual necessity

We can now inspect the set-relationship that the single conditions in the dataset entertain with the outcome. The “pofind” function of the R package yields the value of the N.cons (dubbed “inclN” in the output), the RoN, and the N.cov (dubbed “covN” in the output) calculated for

each condition and its negation against the effect. The operation is repeated against the negation of the effect, too, to be certain that we have considered the four possible combinations of presence and absence. The results are reported in Table 14. Given that the parameters of fit have a double meaning – the consistency of sufficiency and the coverage of necessity are calculated with the same formula, as the coverage of sufficiency and the consistency of necessity – both meanings will be reported in the headings.

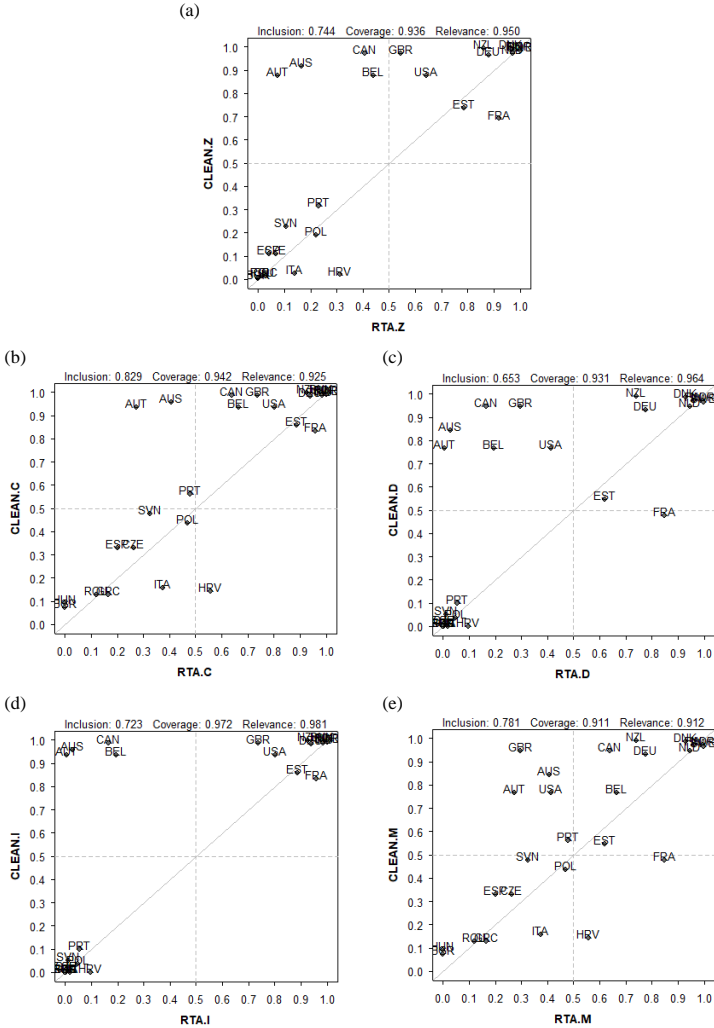
**Table 14. Analysis of individual necessity**

ID	Condition tested	CLEAN.Z			clean.z		
		<i>N.cons</i> <i>S.cov</i>	<i>RoN</i>	<i>N.cov</i> <i>S.cons</i>	<i>N.cons</i> <i>S.cov</i>	<i>RoN</i>	<i>N.cov</i> <i>S.cons</i>
1	<i>atec.z</i>	0.106	0.685	0.170	0.760	<b>0.956</b>	<b>0.916</b>
2	<i>ATEC.Z</i>	<b>0.948</b>	0.776	0.840	0.311	0.411	0.207
3	<i>asoc.z</i>	0.147	0.639	0.202	<b>0.884</b>	<b>0.942</b>	<b>0.914</b>
4	<i>ASOC.Z</i>	<b>0.937</b>	<b>0.893</b>	<b>0.915</b>	0.228	0.460	0.167
5	<i>apub.z</i>	0.233	0.589	0.272	<b>0.923</b>	<b>0.845</b>	0.808
6	<i>APUB.Z</i>	0.836	<b>0.937</b>	<b>0.935</b>	0.168	0.528	0.142
7	<i>rta.z</i>	0.361	0.571	0.377	<b>0.932</b>	0.756	0.732
8	<i>RTA.Z</i>	0.744	<b>0.950</b>	<b>0.936</b>	0.207	0.599	0.196
9	<i>enfor.z</i>	0.203	0.610	0.252	<b>0.924</b>	<b>0.894</b>	<b>0.861</b>
10	<i>ENFOR.Z</i>	<b>0.888</b>	<b>0.934</b>	<b>0.940</b>	0.197	0.503	0.157

This stage of the analysis supports the theoretical intuitions that the selected condition is connected to the outcome by some meaningful set relationships – thus, they are *inus* factors to the effect. Moreover, the conditions display the connection in the expected sign: the presence empirically contributes to low perceptions of corruption, whereas their absence contributes to high perception. Among all the conditions, ENFOR.Z stands out as a rare necessary and sufficient condition to the occurrence of the outcome when present, and of its non-occurrence when absent. The connection weakens but is not lost when the fuzzy scores are transformed to mimic the different possible coder’s biases, as reported in Table 15.

However, as Figure 10 shows with the condition RTA, the fit is unsurprisingly affected by the biases of the coder –simulated as generous in (b), strict in (c), confident in (d), and shy in (e). The comparison suggests that the fit may change remarkably, and the challenging cases increase in number with the coders’ hesitation to decide on the nature of the observations.

**Figure 10. XYplot of RTA and the outcome under different fuzzy transformations**



Keys: in the labels of the conditions, the suffix “.C” is for concentration as in [29], “.D” is for dilation as in [30]; “.I” is for intensification as in [31], while “.M” is for moderation as in [32]. “Inclusion” is for the values of N.cons and S.cov, “Coverage” is for the values of N.cov and S.cons, “Relevance” is for the values of RoN.

Table 15. Fit of transformed conditions

(a)	Conditions	CLEAN.C			clean.c		
		N-con	RoN	S-con	N-con	RoN	S-con
1	<i>atec.c</i>	0.159	0.783	0.352	0.738	0.933	<b>0.832</b>
2	<i>ATEC.C</i>	<b>0.924</b>	0.772	<b>0.874</b>	0.425	0.35	0.205
3	<i>asoc.c</i>	0.184	0.749	0.356	0.874	0.933	<b>0.863</b>
4	<i>ASOC.C</i>	<b>0.929</b>	0.889	<b>0.935</b>	0.347	0.387	0.178
5	<i>apub.c</i>	0.252	0.72	0.418	<b>0.933</b>	0.875	0.786
6	<i>APUB.C</i>	0.871	0.947	<b>0.962</b>	0.309	0.447	0.174
7	<i>rta.c</i>	0.327	0.744	0.518	<b>0.900</b>	0.837	0.728
8	<i>RTA.C</i>	0.829	0.925	<b>0.942</b>	0.404	0.483	0.234
9	<i>enfor.c</i>	0.228	0.747	0.413	0.880	0.902	<b>0.812</b>
10	<i>ENFOR.C</i>	0.896	0.900	<b>0.936</b>	0.363	0.417	0.193
(b)	Conditions	CLEAN.D			clean.d		
		N-con	RoN	S-con	N-con	RoN	S-con
1	<i>atec.d</i>	0.117	0.623	0.142	0.783	0.961	<b>0.943</b>
2	<i>ATEC.D</i>	<b>0.953</b>	0.793	0.815	0.287	0.483	0.244
3	<i>asoc.d</i>	0.148	0.575	0.159	<b>0.887</b>	0.955	<b>0.946</b>
4	<i>ASOC.D</i>	<b>0.950</b>	0.893	<b>0.894</b>	0.211	0.522	0.197
5	<i>apub.d</i>	0.262	0.515	0.238	<b>0.903</b>	0.813	<b>0.814</b>
6	<i>APUB.D</i>	0.794	0.919	<b>0.892</b>	0.154	0.599	0.172
7	<i>rta.d</i>	0.453	0.455	0.351	<b>0.951</b>	0.669	0.732
8	<i>RTA.D</i>	0.653	0.964	<b>0.931</b>	0.156	0.703	0.22
9	<i>enfor.d</i>	0.264	0.526	0.244	<b>0.953</b>	0.868	<b>0.873</b>
10	<i>ENFOR.D</i>	0.862	0.959	<b>0.948</b>	0.174	0.596	0.191
(c)	Conditions	CLEAN.I			clean.i		
		N-con	RoN	S-con	N-con	RoN	S-con
1	<i>atec.i</i>	0.038	0.655	0.060	0.794	0.984	<b>0.971</b>
2	<i>ATEC.I</i>	<b>0.981</b>	0.799	<b>0.859</b>	0.231	0.400	0.158
3	<i>asoc.i</i>	0.092	0.602	0.120	<b>0.899</b>	0.943	<b>0.920</b>
4	<i>ASOC.I</i>	<b>0.939</b>	0.907	<b>0.923</b>	0.14	0.457	0.107
5	<i>apub.i</i>	0.173	0.545	0.191	<b>0.962</b>	0.851	<b>0.831</b>
6	<i>APUB.I</i>	<b>0.847</b>	0.968	<b>0.966</b>	0.064	0.523	0.057
7	<i>rta.i</i>	0.311	0.505	0.299	<b>0.974</b>	0.728	0.733
8	<i>RTA.I</i>	0.723	0.981	<b>0.972</b>	0.070	0.602	0.073
9	<i>enfor.i</i>	0.088	0.574	0.109	<b>0.970</b>	0.952	<b>0.939</b>
10	<i>ENFOR.I</i>	<b>0.950</b>	0.972	<b>0.976</b>	0.079	0.470	0.063
(d)	Conditions	CLEAN.M			clean.m		
		N-con	RoN	S-con	N-con	RoN	S-con
1	<i>atec.m</i>	0.243	0.757	0.414	0.732	0.910	0.820
2	<i>ATEC.M</i>	<b>0.895</b>	0.769	<b>0.835</b>	0.477	0.437	0.293
3	<i>asoc.m</i>	0.265	0.738	0.420	0.858	0.940	<b>0.895</b>
4	<i>ASOC.M</i>	<b>0.934</b>	0.871	<b>0.909</b>	0.444	0.461	0.284
5	<i>apub.m</i>	0.351	0.704	0.476	0.863	0.843	0.769
6	<i>APUB.M</i>	0.829	0.891	<b>0.902</b>	0.412	0.532	0.295
7	<i>rta.m</i>	0.466	0.720	0.583	0.883	0.797	0.727
8	<i>RTA.M</i>	0.781	0.912	<b>0.911</b>	0.493	0.600	0.378
9	<i>enfor.m</i>	0.388	0.710	0.514	0.871	0.831	0.758
10	<i>ENFOR.M</i>	0.817	0.899	<b>0.906</b>	0.441	0.553	0.322

Keys: suffix “.C”, “.D”, “.I”, and “.M” as in Figure 10.

### 11.3 Analysis of sufficiency, first: assess the hypothesis

The analysis of necessity confirms the theoretical expectations set at the theoretical stage about the sign of the contribution of each condition to the occurrence and the nonoccurrence of the outcome. The analysis of sufficiency, instead, probes the explanatory capacity of our hypothesis. However, as Schneider and Wagemann<sup>2</sup> underlined, the analysis of sufficiency has two distinct moments. The first concerns the assessment of the truth table, while the second concludes with the solutions from the three different types of minimizations.

The truth table analysis may include the researcher's direct adjudication on the nature of the logical remainders. In this way, even the parsimonious minimizations yield theoretically plausible solutions. The practice, however, lays the solutions open to the suspicion of confirmatory bias. Besides, directional expectations operate a correction whose effects can be appreciated in the different richness of the prime implicants in the parsimonious solution when compared to the intermediate.

From the explanatory perspective, instead, the truth table is a key step in the analysis of sufficiency as it already tells us whether the observational reports – that is, the configurations that the cases in the population make observed – dis/confirm our starting hypothesis. Bluntly, the criteria R1 to R4 set by Mill, the considerations about the confirmation and disconfirmation of conditional hypothesis with different scope, and the principle of Salmon-completeness require that an explanatory truth table clusters the instances of the effect from the population into homogeneous groups.

Recall that the homogeneity with fuzzy scores is again a matter of degrees and thresholds. It manifests itself at the configuration level as well as at the case level, and special criteria have been developed to pinpoint those “contradictory” cases that challenge the claim of sufficiency of the configuration<sup>3</sup>. However, the real challenge to the claim of sufficiency and the Salmon-completeness of the solution arises from (b)-type of instances, as they violate R4. A naïve yet effective strategy to detect them turns the fuzzy membership scores into crisp-scores, and checks whether and where contradictions arise, if any, in the truth table.

---

<sup>2</sup> Schneider and Wagemann, 2012.

<sup>3</sup> Rubinson, 2013.



**Table 16. Observational reports from crisp configurations**

N	ID	F1	F2	F3	F4	F5	CLEAN
01	Z26	1	1	0	0	1	1
02	Z30	1	1	1	0	1	1
03	Z24	1	0	1	1	1	1
04	Z32	1	1	1	1	1	1
05	Z25	1	1	0	0	0	0
06	Z17	1	0	0	0	0	0
07	Z01	0	0	0	0	0	0
01	C30	1	1	1	0	1	1
02	C28	1	1	0	1	1	1
03	C32	1	1	1	1	1	1
04	C25	1	1	0	0	0	1
05	C05	0	0	1	0	0	0
06	C18	1	0	0	0	1	0
07	C03	0	0	0	1	0	0
08	C10	0	1	0	0	1	0
09	C17	1	0	0	0	0	0
10	C01	0	0	0	0	0	0
01	D30	1	1	1	0	1	1
02	D26	1	1	0	0	1	1
03	D28	1	1	0	1	1	1
04	D32	1	1	1	1	1	1
05	D31	1	1	1	1	0	1
06	D29	1	1	1	0	0	1
07	D23	1	0	1	1	0	0
08	D10	0	1	0	0	1	1
09	D25	1	1	0	0	0	0
10	D01	0	0	0	0	0	0
01	I26	1	1	0	0	1	1
02	I30	1	1	1	0	1	1
03	I24	1	0	1	1	1	1
04	I32	1	1	1	1	1	1
05	I25	1	1	0	0	0	0
06	I17	1	0	0	0	0	0
07	I01	0	0	0	0	0	0
01	M28	1	1	0	1	1	1
02	M30	1	1	1	0	1	1
03	M32	1	1	1	1	1	1
04	M14	0	1	1	0	1	1
05	M31	1	1	1	1	0	C
06	M29	1	1	1	0	0	1
07	M25	1	1	0	0	0	1
08	M05	0	0	1	0	0	0
09	M18	1	0	0	0	1	0
10	M10	0	1	0	0	1	0
11	M03	0	0	0	0	1	0
12	M17	1	0	0	0	0	0
13	M01	0	0	0	0	0	0

*Note:* F1: ATEC, F2:ASOC, F3: APUB, F4: RTA, F5: ENFOR. The type of calibration (.Z, .D, .C, .I, .M) is indicated in the first column, with the identification number (ID) of the report as a row of the truth table

**Table 17. Observations in .Z, .C, .D, .I, and .M reports**

<b>Id</b>	<b>Observations</b>
<b>Z01</b>	<b>BGR,CZE,ESP,GRC,HRV,HUN,POL,ROU,SVN</b>
Z17	ITA
Z24	FRA
<b>Z25</b>	<b>PRT</b>
Z26	AUT,BEL
<b>Z30</b>	<b>AUS,CAN</b>
<b>Z32</b>	<b>NZL,DEU,DNK,EST,FIN,GBR,NLD,NOR,SWE,USA</b>
<b>C01</b>	<b>BGR,HUN,POL,SVN</b>
C03	HRV
C05	ESP
C10	ROU
C17	GRC,ITA
C18	CZE
<b>C25</b>	<b>PRT</b>
C28	BEL
<b>C30</b>	<b>AUS,AUT</b>
<b>C32</b>	<b>NZL,CAN,DEU,DNK,EST,FIN,FRA,GBR,NLD,NOR,SWE,USA</b>
<b>D01</b>	<b>BGR,CZE,ESP,GRC,HRV,HUN,ITA,POL,ROU,SVN</b>
D10	AUT
D23	FRA
<b>D25</b>	<b>PRT</b>
D26	BEL
D28	DEU
D29	USA
<b>D30</b>	<b>AUS,CAN,GBR</b>
D31	EST
<b>D32</b>	<b>NZL,DNK,FIN,NLD,NOR,SWE</b>
<b>I01</b>	<b>BGR,CZE,ESP,GRC,HRV,HUN,POL,ROU,SVN</b>
I17	ITA
I24	FRA
<b>I25</b>	<b>PRT</b>
I26	AUT,BEL
<b>I30</b>	<b>AUS,CAN</b>
<b>I32</b>	<b>NZL,DEU,DNK,EST,FIN,GBR,NLD,NOR,SWE,USA</b>
<b>M01</b>	<b>BGR,HUN,ITA,POL,SVN</b>
M03	HRV
M05	ESP
M10	ROU
M14	AUT
M17	GRC
M18	CZE
<b>M25</b>	<b>PRT</b>
M28	BEL,DEU
M29	USA
<b>M30</b>	<b>AUS,GBR</b>
M31	EST,FRA
<b>M32</b>	<b>NZL,CAN,DNK,FIN,NLD,NOR,SWE</b>

*Note:* configurations ID as in Table 16. In bold, the cases that preserve their association with the same configuration across different coding biases.

Table 18. Truth tables from .Z, .C, .D, .I, and .M conditions

<i>ID</i>	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>	<i>n</i>	<i>CLEAN</i>	<i>PR.I.P</i>	<i>~CLEAN</i>	<i>PR.I.N</i>
Z26	1	1	0	0	1	2	1.000	1.000	0.348	0.000
Z30	1	1	1	0	1	2	1.000	1.000	0.291	0.000
Z24	1	0	1	1	1	1	1.000	1.000	0.706	0.000
Z32	1	1	1	1	1	10	0.999	0.999	0.135	0.000
Z25	1	1	0	0	0	1	0.662	0.204	0.858	0.665
Z17	1	0	0	0	0	1	0.333	0.009	0.994	0.991
Z01	0	0	0	0	0	9	0.137	0.002	0.996	0.996
C30	1	1	1	0	1	2	1.000	1.000	0.348	0.000
C28	1	1	0	1	1	1	1.000	1.000	0.501	0.000
C32	1	1	1	1	1	12	1.000	1.000	0.113	0.000
C25	1	1	0	0	0	1	0.952	0.624	0.921	0.376
C05	0	0	1	0	0	1	0.900	0.024	0.998	0.976
C18	1	0	0	0	1	1	0.818	0.168	0.963	0.832
C03	0	0	0	1	0	1	0.794	0.011	0.998	0.989
C10	0	1	0	0	1	1	0.764	0.282	0.907	0.718
C17	1	0	0	0	0	2	0.551	0.005	0.998	0.995
C01	0	0	0	0	0	4	0.481	0.002	0.995	0.991
D30	1	1	1	0	1	3	1.000	1.000	0.341	0.000
D26	1	1	0	0	1	1	1.000	1.000	0.480	0.000
D28	1	1	0	1	1	1	1.000	1.000	0.389	0.000
D32	1	1	1	1	1	6	0.998	0.998	0.163	0.000
D31	1	1	1	1	0	1	0.994	0.988	0.496	0.000
D29	1	1	1	0	0	1	0.993	0.987	0.491	0.013
D23	1	0	1	1	0	1	0.931	0.326	0.927	0.291
D10	0	1	0	0	1	1	0.930	0.849	0.607	0.151
D25	1	1	0	0	0	1	0.571	0.234	0.869	0.766
D01	0	0	0	0	0	10	0.064	0.002	0.997	0.997
I26	1	1	0	0	1	2	1.000	1.000	0.139	0.000
I30	1	1	1	0	1	2	1.000	1.000	0.162	0.000
I24	1	0	1	1	1	1	1.000	1.000	0.407	0.000
I32	1	1	1	1	1	10	1.000	1.000	0.048	0.000
I25	1	1	0	0	0	1	0.292	0.078	0.904	0.875
I17	1	0	0	0	0	1	0.168	0.005	0.996	0.995
I01	0	0	0	0	0	9	0.030	0.001	0.998	0.998
M28	1	1	0	1	1	2	1.000	1.000	0.543	0.000
M30	1	1	1	0	1	2	1.000	1.000	0.497	0.000
M32	1	1	1	1	1	7	0.999	0.998	0.290	0.000
M14	0	1	1	0	1	1	0.994	0.975	0.762	0.025
M31	1	1	1	1	0	2	0.985	0.956	0.660	0.030
M29	1	1	1	0	0	1	0.979	0.938	0.678	0.062
M25	1	1	0	0	0	1	0.964	0.764	0.882	0.236
M05	0	0	1	0	0	1	0.915	0.086	0.992	0.914
M18	1	0	0	0	1	1	0.881	0.369	0.931	0.631
M10	0	1	0	0	1	1	0.830	0.401	0.887	0.599
M03	0	0	0	1	0	1	0.767	0.029	0.993	0.971
M17	1	0	0	0	0	1	0.662	0.023	0.992	0.977
M01	0	0	0	0	0	5	0.487	0.007	0.993	0.987

Note: Factors and type of calibration, as in Table 16.

PR.I.P: Proportional Reduction of Inconsistency against the positive outcome;  
 PR.I.N: Proportional Reduction of Inconsistency against the negative outcome. The columns CLEAN and ~CLEAN report the S.cons of the configuration to the positive and the negative outcome, respectively.

Table 16 reports the observed configurations from the same truth table under different calibrations of the conditions: the original ones, *.Z*, then their transformation by dilation *.D*, concentration *.C*, intensification *.I*, and moderation *.M*. When read in the light of the cases that make configurations observed, reported in Table 17, we can notice how few cases remain in the same configuration under different coding biases. Bulgaria, Hungary, Poland, and Slovenia always make configuration 01 (0000) observed, while New Zealand, Denmark, Finland, the Netherlands, Norway, and Sweden always populate configuration 32 (1111). Stable membership across different calibrations also characterizes Australia in configuration 30 (1110), and Portugal in configuration 25 (1100).

Table 16 also shows that no truth table displays crisp-contradictory primitive configurations except for *.M*. Recall that moderation is obtained by square-rooting the fuzzy scores lower than 0.5 and by squaring those higher than 0.5, which sorts a swapping effect on observations. The moderate coder adopts a changing crossover – one for the presence of the outcome, and a different one for the absence – that may end up designing unjustified contradictions. Moreover, the *.M* truth table includes a higher number of observed primitive configurations – 13 out of 32 that the five initial conditions make logically possible. In the light of the wisdom that considers the saturated truth table desirable and limited diversity problematic, the *.M* truth table should appear superior to any other. However, if we look at the cases in Table 17 that make the configurations observed, we realize that the higher saturation of the *.M* truth table follows from “pigeonholing” cases in the truth table without improving the consistency of the statements. Let us narrow on the consistency values of the *M* and the *.Z* configurations in Table 18.

The *.Z* truth table displays a promising symmetry: of the seven observed configurations, four (*Z26*, *Z30*, *Z24*, and *Z32*) have full *S*.consistency to the positive outcome with almost identical *PRI.P* values and clear null *S*.consistency to the negative outcome with null *PRI.N*. The remaining three reports (*Z25*, *Z17*, *Z01*) again take high *S*.consistency to the absence of the outcome, and negligible *S*.consistency to the presence of the outcome, all supported by consistent *PRI* values. Together, they suggest homogeneous partitions and, hence, the relevance of the explanatory model.

Of the thirteen reports in the *.M* truth table, instead, consistent to *CLEAN* are the reports *M28*, *M30*, *M32*, *M14*, but also the crisp-contradictory *M31*, and *M29*. Report *M25* has high *S*.consistency to *CLEAN*

(0.964) but displays a drop in the related PRI.P (0.764), and, in fact, it also proves consistent to ~CLEAN (0.882) although the associated PRI.N is low (0.236). The adjudication of M10 is hard, with an S.consistency to CLEAN of 0.881 and a PRI.P of 0.401 but an S.consistency to ~CLEAN of 0.887 and the relative PRI.N of 0.599. Clear reports of the negative outcome are M03, M01, and M17. Thus, we cannot expect neat homogeneous partitions from .M calibration, and the empirical support for the starting hypothesis is inferior to the original one. Although the .I partitions are more homogeneous, we may be wary of such cleanness, as it may come at the further loss of relevant information.

The two remaining transformations, .D and .C, again qualify as worsening versions of the original calibration: .C reports are more diverse, yet generally less consistent claims of sufficiency; .D consistent configurations are more numerous to the positive outcome, while those to the negative are reduced.

The comparison between .Z and its transformations provides further arguments in favor of the original design with the possible exception of .I. The substantive merits of the models can now be adjudicated by minimization.

#### **11.4 Analysis of sufficiency, second: Minimizing solutions**

The truth table exposes the consistency of the observational reports to the claim that the hypothesis holds in the population; however, it does not tell which local conditions explain the occurrence of the outcome in which cases. Ascription is the task that minimizations perform instead, under different assumptions about the unobserved configurations in the truth table. Now, from the previous analysis of the truth table, we have learned that even the same hypothesis in the same population may leave different logical remainders due to calibration. So, deciding their plausibility in the light of theoretical considerations may risk reinforcing some bias. The default strategy manages the logical remainders with three different types of expectations and yields three sets of solutions of different specificity – the conservative, the parsimonious, and the intermediate or plausible.

Applied to the .Z truth table, the minimizations finds that the conservative solution of the positive outcome has two prime implicants, both with consistency and PRI above 0.999. The first reads

ATEC.Z\*ASOC.Z\**rta.z*\*ENFOR.Z and covers the first two consistent configurations Z26 and Z30 by dropping APUB.Z. The second term reads ATEC.Z\*APUB.Z\**RTA.Z*\*ENFOR.Z and covers the two consistent configurations, Z24 and Z32, by dropping ASOC.Z. Together, they cover the whole of the instances of the outcome CLEAN.

The conservative solution to the negative outcome again has two terms. The first term reads ATEC.Z\**apub.z*\**rta.z*\**enfor.z* and follows from the minimization of Z25 and Z17 by dropping ASOC.Z; the second term reads *asoc.z*\**apub.z*\**rta.z*\**enfor.z* and covers configurations Z17 and Z01. The minimization uses Z17 twice, meaning that both implicants will cover the instances of Z17 (in this example, Italy).

The unconstrained minimizations to the parsimonious solutions find a single-factor term: ENFOR.Z to CLEAN, and *enfor.z* to *clean.z*. Indeed, the population can be sorted into two homogeneous partitions by only considering whether the factor is present or absent. This does justice to the consistency scores in the analysis of necessity reported in Table 14. The minimization to the positive outcome employs twelve logical remainders (Z2, Z4, Z6, Z8, Z10, Z12, Z14, Z16, Z18, Z20, Z22, and Z28) under the assumption that they would lead to the outcome if observed. The minimization to the negative instead employs thirteen logical remainders (Z3, Z5, Z7, Z9, Z11, Z13, Z15, Z19, Z21, Z23, Z27, Z29, Z31) to reach its result. Thus, the minimization to the opposite outcome does not make opposite counterfactual assumptions about the same remainders, which would make the solution contradictory. All in all, the finding is formally sound, although essential. With an S.cons of 0.940 and a PRI of 0.932, its fit proves lower than the worse solution term in the conservative solution.

**Table 19. Solution terms from the .Z truth table**

<i>Outcome</i>	<i>Type</i>	<i>Id</i>	<i>Term</i>	<i>S-cons</i>	<i>PRI</i>
Positive	Cons Int	1	ATEC.Z*ASOC.Z* <i>rta.z</i> *ENFOR.Z	1.000	1.000
		2	ATEC.Z*APUB.Z* <i>RTA.Z</i> *ENFOR.Z	0.999	0.999
	Pars	1	ENFOR.Z	0.940	0.932
Negative	Cons Int	1	ATEC.Z* <i>apub.z</i> * <i>rta.z</i> * <i>enfor.z</i>	0.923	0.886
		2	<i>asoc.z</i> * <i>apub.z</i> * <i>rta.z</i> * <i>enfor.z</i>	0.997	0.996
	Pars	1	<i>enfor.z</i>	0.861	0.832

Key: Cons: conservative; Int: intermediate; Pars: parsimonious. Cut set at 0.85.

The intermediate solutions are obtained by preventing the minimization with hard counterfactuals. Those to the positive outcome maintain that each condition contributes to the outcome when present; those

to the negative outcome, when absent. Under these constraints, in this analysis, the minimizations find the conservative solutions.

The consistency scores also indicate that the plausible solution ensures higher homogeneous partitions than the intermediate. Hence, in this population, the plausible solution is more Salmon-complete than the parsimonious one. On a more general note, the minimizations confirm that the hypothesis holds in the population as a reduction statement that the observation reports from its development confirm and do not falsify.

The standing of the hypothesis before evidence, however, is better conveyed by a graphical representation of the solutions.

## 11.5 Plot solutions

Many graphical solutions suit the representation of the findings from QCA, depending on the intent of the analysis<sup>4</sup>. An adequate rendering to the explanatory use of the technique is the XY-plot. This special Cartesian space emphasizes the ordering power of *inus* solutions. Its “zones” allow identifying the instances that violate Mill’s requirements at a glance, with further attention for the effect that fuzzy scores impress to the analytic space<sup>5</sup>. Recall that, with fuzzy scores, the sufficiency of a term to the outcome is understood as a lower-than relationship. This rotates the analytic space and establishes the bisector as the further relevant axis for assessing the claim of sufficiency, with a slack of +0.1 to accommodate measurement issues<sup>6</sup>. The XY-plot, in short, provides a useful diagnostic of the fit of the solutions to the claim of sufficiency, although it does not tell whether the possible misfit is due to a weak hypothesis, mistakes in the selection of the gauges, or poor calibration decisions.

Figure 11 uses the XY-plot to compare the distributions generated by the parsimonious and intermediate/complex solutions from the .Z data. Indeed, the graphs suggest that the indications from the parameters of fit are empirically sound. There is no instance in the top left corner to indicate that the compound has substitutes: hence, all the channels to the effect are captured by the conditions in the solutions. More important, there is no instance in the bottom right corner to falsify the

---

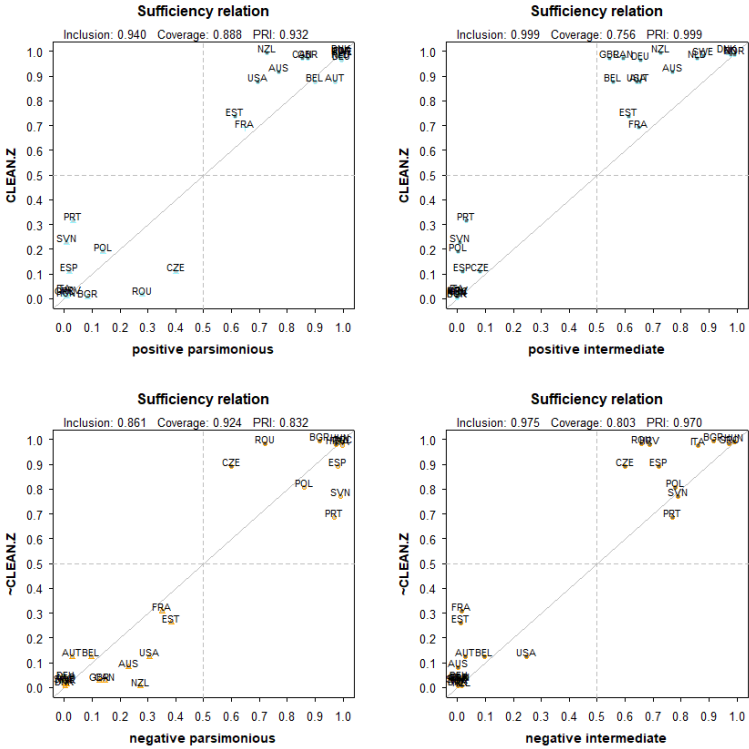
<sup>4</sup> Rubinson, 2019.

<sup>5</sup> Schneider and Rohlffing, 2016.

<sup>6</sup> Ragin, 2000.

claim of sufficiency: the solutions are Salmon-complete. However, the intermediate terms display higher relevance than the parsimonious ones. The effect is especially visible on the instances of the negative outcome, which are pushed above the bisector when the solution term is better specified. This effect may be noted for later theoretical considerations.

**Figure 11. XY-plots from .Z calibration**



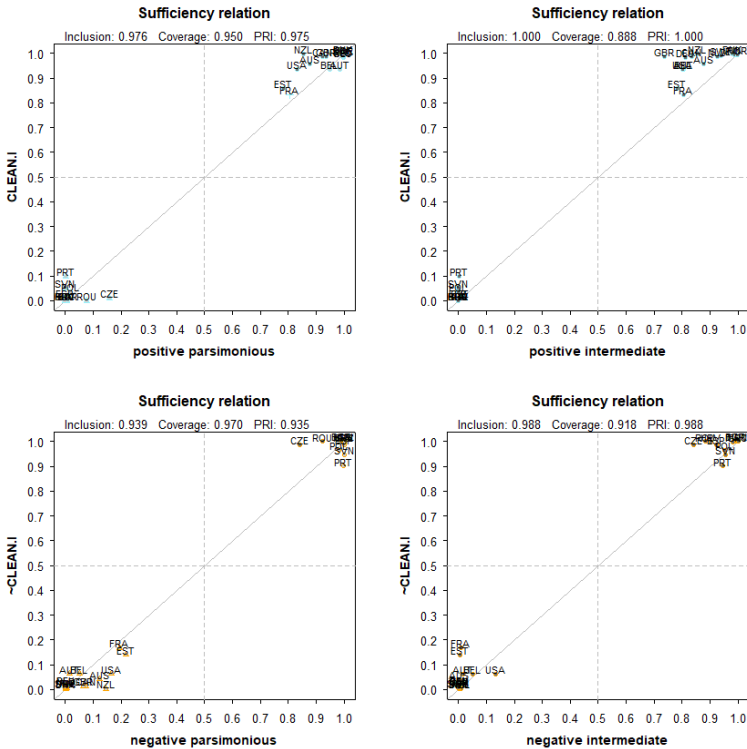
*Keys:* *Inclusion* is for the S.consistency of the overall solution, calculated as the disjunction of the prime implicants. *Coverage* is for the S.coverage. The *PRI* again is calculated on the disjunction of the prime implicants.

The XY-plots allow immediate comparisons with the solutions from alternative calibrations. For instance, from Figure 12, we can better appreciate the effects that the intensification of the fuzzy scores exert on



the distribution. Intensification does not change the findings but heightens their consistencies by shrinking the distances of each observation from the corners of sure exclusion (0,0) and of sure inclusion (1,1). The effect of higher certainty in classification is the reduction of the inconsistencies: the observations that still fall below the bisector are moved in the +0.1 area of tolerance.

**Figure 12. XY-plots from .I calibration**

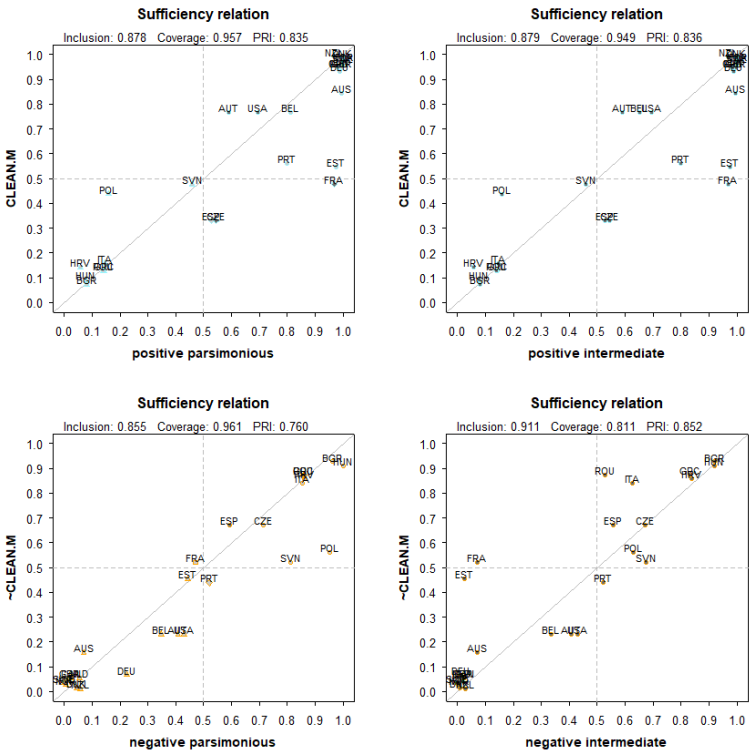


Keys: same as above.

Just the opposite happens to the analysis with moderated fuzzy scores. Recall these membership scores pigeonhole the observations and make thirteen configurations observed out of thirty-two. The conservative solution to the positive outcome reports six terms, four more

than the same solution with .Z scores. The parsimonious positive solution has three terms, arising from multiple usages of the same configuration for minimization. So, 14 instances are covered by different solutions, and only in one of them, the single parsimonious factor from .Z minimizations appears. However, if we had to rely on the parameters of fit only, we could accept the solutions as sound: each term's S.consistency is higher than the conventional thresholds, and even their PRI values do not raise particular concerns. Full evidence of the little credibility of the solution only comes from the actual distribution, as displayed in Figure 13.

**Figure 13. XY-plots from .M calibration**



Keys: same as above.

Figure 13 shows the solutions from .M calibration to both the positive and the negative outcome, unsurprisingly, distribute instances in a shape that misfits the claim of the sufficiency of the hypothesis. Here, we know the falsification arises from the limited capacity of the moderate coder to “carve nature at its joints”; beyond the particular circumstances, however, we might be wary of solutions with contradicting instances as a severe warning about our research design first, and ultimately about our hypothesis.

## 11.6 Return to the configurational hypothesis

The lesson from the comparison of the distribution under different calibration decisions is that, regardless of the decisions that we can make about membership, the conditions in the model do capture some relevant facts of the causal link between institutional designs and perception of corruption. The operationalization of the starting hypothesis, therefore, stands in our selection of cases. The credibility of these results depends on raw measures. Despite our choice of sub-indexes, the gauges contain heterogeneous information about the same dimension. For instance, RTA scores are based on the fact that a right is established to access information, and people are aware of it, then its timeliness, the completeness of the information obtained, its cost. Thus, the score provides an overall compensation of many facets, and the same score may be due to different compensations. Absent finer measures, we have to stay content with the existing gauges. In their heterogeneity, the gauges still provide indications about which (aggregate) conditions prove locally capable of motivating, sustaining, and protecting the people’s perceptions of low corruption of the public sector.

The more credible solution comes from .Z observation reports. The parsimonious solution to the positive and the negative outcome agree that the single factor underlying the high and low perception of corruption is the quality of civil justice. The factor is linked to the outcome by a biconditional relation, such that the parsimonious solution reads as in [PS] below:

[PS]     *ENFOR*  $\equiv$  *CLEAN*

The distribution from the single factor, however, shows that the fitting improves when the solutions are built on those counterfactuals that

meet theoretical expectations. The constraint indicates that the lock-in effect on the positive outcome has a common part and two variants. The intermediate solution reported in Table 19 can be written as [IS.pos]

[IS.pos]  $A TEC * ENFOR * (ASOC * rta + APUB * RTA) \rightarrow CLEAN$

The common part indicates effective constraints on the public sector's operation by audit and review activities of independent authorities (ATEC), together with the effectiveness and timeliness of civil justice enforcement (ENFOR). These provide the enabling and shielding unreplaceable conditions for the two replaceable triggers. The one that can be dubbed "lively societal pulleys" follows from the existence of societal checks as free media, opposition parties, and civil organizations, all supported by the freedom of speech (ASOC), even when the right of information is weak (rta). This is the relevant trigger in Austria, Belgium, Australia, and Canada. The other emphasizes "legal levers" and passes through the transparency of the administration, as ensured by the right of information (RTA) and by the obligations that the government has to circulate clear and accessible information widely (APUB). The second trigger operates in cases as different as France, New Zealand, Germany, Denmark, Estonia, Finland, Great Britain, the Netherland, Norway, Sweden, and the United States.

The return to the configurational theory is incomplete without the recipes of failure. Here again, we find complex machinery with two parts. The solution can be written as [IS.neg]

[IS.neg]  $apub * rta * enfor * (ATEC + asoc) \rightarrow clean$

The solution allows the interpretation of high levels of perceived corruption as engrained in weak legal constraints on the public sector – low obligation to circulate information (apub), weak right of access to information (rta), and ineffective enforcement of civil justice (enfor). Within such a dimmed rule of law, the triggers of perceiving corruption depend on the constraints exerted by the oversight bodies in Italy and Portugal; and on weak societal checks in Bulgaria, Czechia, Spain, Greece, Croatia, Hungary, Poland, Romania, Slovenia, and Italy.

Together, the two halves of the causal story suggest that the effectiveness of civil justice is a crucial element to enable virtuous mechanisms. However, its positive effects only follow when oversight rights are distributed to institutional actors, and societal actors are either lively

engaged or kept updated on government activities. Despite their generality, these findings provide a useful suggestion on how to intervene on the perception of corruption. These findings may open to further research on the actual processes that occur under the different configurations in single selected cases in nested designs. However, adding details may enrich a causal story that is already a self-standing credible *in us* “reduction” of a generative process.

## 12. Bottom lines

The current discourse on scientific explanation has shifted the ground on which we can develop credible accounts from law-like generalizations to mechanisms. The definitional debate on what precisely a mechanism is, where it is located, and how it can be captured far exceeds the purpose of this work. Nevertheless, the representations from the strand of the modelers in the philosophy of science converge on the picture of a flow of activities and interactions that occur under the right conditions and that, in the absence of obstructions, bring about the outcome.

The picture is Aristotelean in its underlying assumption that special entities express the power to make the outcome happen when properly arranged. At the same time, the picture is not Aristotelean in its assumption that the “mechanism” – understood as the organization of the right entities under the right conditions – is self-standing. As Mill long established with his modes of explanation, we can only resolve one mechanism in its components, or abstract many specific mechanisms into a more general one, or specify which components are relevant to a particular effect. Each of these operations, however, deepens our knowledge of a situation, bridges sparse notions, and possibly improves our understanding as much as the effectiveness of our interventions. In short, the mechanistic perspective allows resolving causation into what happens under the right configuration; the right configuration is that which responds to our why questions. The analytic challenge consists of filling the picture with the relevant details that turn the picture into a usable yet reliable model of the occurrence of an outcome of interest.

The possible points of attack to filling the sketch are many but can be reduced to a simple alternative. Models can capture the changes in the picture under the assumption that every variation counts. Alternatively, they can be developed under the tenet that only a few critical points matter, as any of the remaining variations will keep the overall system in a state associated with approximately similar levels of the

outcome. In the latter mindset, the question of interest asks which conditions – if any – keep the system in a state – either desirable or not. This is the particular question to which QCA can respond, mainly due to its language. Boolean algebra allows modeling the presence and absence of the features of the system, or its component, and developing hypotheses about the right configurations. The language also comes with principles of combinations, criteria to establish whether the hypothesis holds or is wanting, and diagnostics that indicate the direction for improving it – undetected plurality, or omitted *inus* conditions. Moreover, the language allows for “observation reports” that ensure the testability of the logical hypotheses on populations of cases, as configurations can be rendered as intersections to which the cases belong. The technique also comes with an algorithm that applies a long-honored set of rules to dismiss those components in the hypothesis that prove irrelevant to explain the differences in the outcome across groups and instances.

On the whole, then, the analytic machinery of QCA is adequate for ascertaining *inus* causation. However, its results are as good as our starting hypothesis. The starting hypothesis, in turn, has two weak points – issues that may invite to faulty interpretations.

The first issue, and possibly the less serious at this stage, concerns the causal interpretability of QCA results. Of course, a non-contradictory truth table and consistent solutions do not automatically endow a sparse bunch of conditions with any explanatory power. QCA is “explanatory” as the old syllogistic machinery was: it is “truth-preserving” – yet not “truth-generating.” The outputs of its algorithms are as good as its inputs, and the claim of sufficiency stands when the hypothesis already offers a plausible account of the outcome. This requirement posits the question of what makes a hypothesis credible. As argued in the previous pages, mechanisms provide the mindset to model a plausible hypothetical account, and the language of logic supports our modeling efforts straightforwardly. Logic has first been developed to assist our reasoning about the world: roughly, if you can say it, you can also model it – with literals, overbars, wedges, wees, and horseshoes. Indeed, even a very general definition of a mechanism as the interaction of given entities with special attributes *As*, in contexts with features *Cs* and in the absence of obstructions *Os*, can be turned into a formal expression like [M] below:

$$[M] \quad (\wedge A \wedge \wedge C) \wedge \wedge \bar{O} \rightarrow outcome$$

The reason that qualifies [M] as a plausible explanation, in turn, follows from our confidence that, at the best of our knowledge, these three types of components, together, suffice to capture the unobserved dynamic to the outcome. They include the entities with the capacity or potential to yield an effect, plus the triggering factors that actualize the capacity and make the process to the outcome as inevitable as a flying bullet – that is, inevitable unless obstacles arise. Hence, in modeling a plausible hypothesis, we may want to select those *As* that signal a capacity, those *Cs* that trigger that capacity and sustain its unobserved unfolding, and those *Os* that defuse the triggers or weaken the support. The logical rendering also suggests narrower models. As long as we can assume that a good model of the triggers and supports dovetails the model of the capacity to “make” the effect, we can assume that  $\wedge C \rightarrow \wedge A$ . An abridged explanatory model [M’], then, can explain by taking into account contextual features and obstacles only:

$$[M'] \quad \wedge C \wedge \wedge \bar{O} \rightarrow \text{outcome}$$

Thus, for instance, we can model the institutional constraints and their enforcement as in Part III, and assume that they interact with a particular type of opportunistic behavior, although we may not be able to observe actual opportunism in the making. Then, again as in Part III, we may discover that relevant *Os* capture much yet not all of the effects of the other components. This comes as little surprise, as a good model of *Os* dovetails the model of *Cs*, so that the logical relationships among the component of a mechanism are as in [LR]:

$$[LR] \quad \wedge \bar{O} \rightarrow \wedge C \rightarrow \wedge A \rightarrow \text{outcome}$$

It is worth noting that the chain of conditionalities in [LR] may or may not have the temporal dimension of flows – it only depicts logical relationships as in an Aristotelean syllogism. The message conveyed by [LR] is, we should be able to operate on the whole mechanism by removing or adding obstacles. Which, of course, makes sense – especially from the perspective of policy design and choice architecture. However, here a second issue arises about the sharpness of configurational knowledge.



Indeed, these are *models* – useful reductions of real complex situations. The relationship with the actual world is established by the quality of the data on which the truth tables rely as the “observational reports” of our starting hypothesis. Their quality, in turn, depends on two kinds of decisions: those to calibrate the raw data into membership scores, and those to select the raw data that render our *As*, *Cs*, *Os* to the outcome. Again, the former may raise fewer concerns than the latter.

Calibration is a delicate operation that affects our findings; however, identifying suitable thresholds may itself be addressed as an empirical puzzle that we can tackle with the aid of proper techniques to keep our biases at bay. Moreover, if a reasonable calibration setting – one that pragmatically affords “certainty” and “relative uncertainty” in classification – yields a “neater” truth table than others, we may want to draw conclusions from it, as worse calibrations may entail we are disregarding tipping points and ill-carving the analytic partitions of our universe. Thus, carving does remain a core operation of a credible explanation that requires substantive knowledge. However, it can become a vain exercise without a suitable matter – that is, without adequate raw measures of our *As*, *Cs*, *Os*, and outcome.

The selection of the raw measures is often guided by convention or constrained by the scarcity of reliable information. Even in the age of data deluge, the data about, for instance, the average government, its policy decisions, and their implementation can, unfortunately, be erratic. Reliable series, as we have seen, may be intended to serve many purposes and include high degrees of heterogeneity. The heterogeneity engrained in the raw measures may then boost or spoil the set relationships in a QCA, depending on the relationships between the raw data and the fuzzy construct. The ultimate bottom line, then, suggests an accurate selection of the raw data, then the usual caution when interpreting configurational findings.

As a technique, QCA does afford explanations in the long-honored sense that its machinery exposes the weaknesses of our hypothetical because-answer and isolates those of its parts that prove relevant to the account of interest. This is not the only usage of QCA<sup>1</sup>, neither, of course, is the only kind of knowledge that we need to grasp causation<sup>2</sup>. Nevertheless, its explanations are one of a kind and useful to pinpoint

---

<sup>1</sup> Thomann and Maggetti, 2017; Hino, 2009.

<sup>2</sup> King *et al.*, 1994; George and Bennett, 2005; Brady and Collier, 2010.

the constraints that shape games, the right context of effective interactions, the starting point of Markovian chains to the outcome – or, more simply, realistic risk models and ex-post evaluations of policy interventions. Despite its popularity, its full potential may still have to be unlocked.

Plus, configurational thinking about generative mechanisms makes a great pastime.

Please, try it at home.

## References

- AMENTA, Edwin, and Jane D. POULSEN. 1994. Where to begin. A survey of five approaches to selecting independent variables for Qualitative Comparative Analysis. *Sociological Methods and Research* 23.1: 22-53.
- ANNAS, Julia, and Jonathan BARNES. 2000. *Sextus Empiricus. Outlines of Scepticism*. Cambridge, UK: Cambridge University Press.
- ANNAS, Julia. 1982. Aristotle on inefficient causes. *The Philosophical Quarterly* 32.129:311-326.
- BACON, Francis. 1620. *The New Organon*. Edited by Jardine, Lisa, and Michael Silverthorne. Cambridge, UK: Cambridge University Press.
- BARNES, Jonathan, Ed. 1991. *Complete Works (Aristotle)*. Princeton, N.J.: Princeton University Press.
- BARR, Abigail, and Danila SERRA. 2010. Corruption and culture: An experimental analysis. *Journal of Public Economics* 94.11-12: 862-869.
- BARTON, Allen H. 1955. The concept of property-space in social research. In Lazarsfeld, Paul F. and Morris Rosenberg, Eds. *The Language of Social Research*. New York, NY: The Free Press, 40-62.
- BASURTO, Xavier, and Johanna SPEER. 2012. Structuring the calibration of qualitative data as sets for Qualitative Comparative Analysis (QCA). *Field Methods* 24.2:155-174.
- BAUHR, Monika. 2017. Need or greed? Conditions for collective action against corruption. *Governance* 30.4: 561-581.
- BAUMGARTNER, Michael, and Alrik THIEM. 2015. Identifying complex causal dependencies in configurational data with coincidence analysis. *The R Journal* 7.1: 176-184.
- BAUMGARTNER, Michael, and Alrik THIEM. 2017. Model ambiguities in configurational comparative research. *Sociological Methods & Research* 46.4: 954-987.
- BAUMGARTNER, Michael. 2009. Inferring causal complexity. *Sociological Methods & Research* 38:71-101.
- BAUMGARTNER, Michael. 2013. Detecting causal chains in small-n data. *Field Methods* 25.1: 3-24.
- BEACH, Derek, and Rasmus B. PEDERSEN. 2016. *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*. Ann Arbor, MI: University of Michigan Press.
- BECHTEL, William, and Adele ABRAHAMSEN. 2005. Explanation: a mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36.2: 421-441.
- BECHTEL, William, and Robert C. RICHARDSON. 1993. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, NJ: Princeton University Press.

- BEN-ELIYAHU, Rachel, and Rina DECHTER. 1994. Propositional semantics for disjunctive logic programs. *Annals of Mathematics and Artificial Intelligence* 12.1-2: 53-87.
- BENNETT, Andrew, and Colin ELMAN. 2006. Qualitative research: Recent developments in case study methods. *Annual Review of Political Science* 9: 455-476.
- BENNETT, Andrew, and Jeffrey T. CHECKEL, 2015. Process tracing: From philosophical roots to best practices. In Id., Eds. *Process tracing: From Metaphor to Analytical Tool*. Cambridge, UK: Cambridge University Press, 3-37.
- BENNETT, Jonathan. 1974. Counterfactuals and possible worlds. *Canadian Journal of Philosophy* 4.2: 381-402.
- BERG SCHLOSSER, Dirk, and Gisèle DE MEUR. 2009. Comparative research design: case and variable selection. In Benoit Rihoux and Charles C. Ragin, Eds. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. London, Sage, 19-32.
- BERG SCHLOSSER, Dirk, DE MEUR, Gisèle, RIHOUX, Benoit, and Charles C. RAGIN. 2009. Qualitative Comparative Analysis (QCA) as an approach. In Benoit Rihoux and Charles C. Ragin, Eds. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. London, Sage, 1-18.
- BETT, Richard. 2005. *Sextus Empiricus. Against the Logicians*. Cambridge, UK: Cambridge University Press.
- BLAIR, Graeme, Jasper COOPER, Alexander COPPOCK, and Macartan HUMPHREYS. 2019. Declaring and diagnosing research designs. *American Political Science Review* 113.3: 838-859.
- BLATTER, Joachim, and Markus HAVERLAND. 2012. *Designing Case Studies: Explanatory Approaches in Small-N Research*. New York: Palgrave MacMillan.
- BOCHENSKI, Joseph M., 1951. *Ancient Formal Logic*. Amsterdam, NL: North-Holland.
- BØGGILD, Troels. 2016. How politicians' reelection efforts can reduce public trust, electoral support, and policy approval. *Political Psychology*, 37.6, 901-919.
- BOOLE, George. 1853. *An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities*. London, UK: Walton and Maberly.
- BOVENS, Mark. 2007. Public accountability. In Ferlie, Ewan, Laurence E. Lynn, Jr., and Christopher Pollitt, Eds. *The Oxford Handbook of Public Management*. New York, NY: Oxford University Press, 182–208.
- BRADY, Henry E. 2008. Causation and explanation in social science. In Robert E. Goodwin, Ed. *The Oxford Handbook of Political Science*. Oxford, UK: Oxford University Press, 1054-1108.
- BRADY, Henry E., and David COLLIER, eds. 2010. *Rethinking social inquiry: Diverse tools, shared standards*. Rowman & Littlefield Publishers.
- BRAUMOELLER, Bear F. 2015. Guarding against false positives in qualitative comparative analysis. *Political Analysis* 23.4: 471-487.
- CAMPBELL, Donald T., and J. C. STANLEY. 1963. *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally and Company.
- CAMPBELL, Norman R. 1928. *An Account of the Principles of Measurement and Calculation*. London, UK: Longmans and Green.
- CAMPOS, Mauro F., and Francesco GIOVANNONI. 2007. Lobbying, corruption and political influence. *Public Choice* 131:1-21.
- CAPECCHI, Vittorio. 1968. On the definition of typology and classification in sociology. *Quality & Quantity* 2.1: 9-30.
- CARNAP, Rudolf. 1936. Testability and meaning. *Philosophy of Science* 3.4: 419-471.

- CARNAP, Rudolf. 1937. Testability and meaning – Continued. *Philosophy of Science* 4.1: 1-40.
- CARNAP, Rudolf. 1946. Modalities and quantification. *Journal of Symbolic Logic*, 11.2: 33–64.
- CARNAP, Rudolf. 1949. Truth and confirmation. In Feigl, Herbert, and Wilfrid Sellars. *Readings In Philosophical Analysis*. New York, NY: Appleton - Century. Crofts, 119-127.
- CARNIELLI, Walter A. 1987. Systematization of finite many-valued logics through the method of tableaux. *The Journal of Symbolic Logic* 52.2: 473-493.
- CARTWRIGHT, Nancy. 1979. Causal laws and effective strategies. *Noûs* 13.4: 419-437.
- CARTWRIGHT, Nancy. 1989. *Nature's Capacities and Their Measurement*. London, UK: Clarendon Press.
- CARTWRIGHT, Nancy. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge, UK: Cambridge University Press.
- CARTWRIGHT, Nancy. 2017. Causal powers. Why Humeans can't even be instrumentalist. In Jacobs, Jonathan D., ed. *Causal Powers*. Oxford, UK: Oxford University Press, 9-23.
- COLLIER, David, and James E. MAHON. 1993. Conceptual stretching revisited: adapting categories in comparative analysis. *American Political Science Review* 87.4: 845-855.
- COLLIER, David, and James MAHONEY. 1996. Insights and pitfalls: Selection bias in qualitative research. *World Politics* 49.1: 56-91.
- COLLIER, David, Jody LAPORTE, and Jason SEAWRIGHT. 2012. Putting typologies to work: concept formation, measurement, and analytic rigor. *Political Research Quarterly* 65.1: 217–232.
- COLQUITT, Jason A., Brent A. SCOTT, Jessica B. RODELL, David M. LONG, Cindy P. ZAPATA, Donald E. CONLON, and Michael J. WESSON. 2013. Justice at the millennium, a decade later: A meta-analytic test of social exchange and affect-based perspectives. *Journal of Applied Psychology* 98. 2: 199-216.
- COLQUITT, Jason A., Donald E. CONLON, Michael J. WESSON, Christopher O.L.H. PORTER, and K. YEE NG. 2001. Justice at the millennium: a meta-analytic review of 25 years of organizational justice research. *Journal of applied psychology* 86. 3: 425-445.
- COMPTON, Mallory E., Joannah LUETJENS and Paul 'T HART. 2019. Designing for Policy Success. *International Review of Public Policy* 1:2: <http://journals.openedition.org/irpp/514>
- COOK, T. D., and Donald T. CAMPBELL. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin.
- CORCORAN, John. 1972. Completeness of an ancient logic, *Journal of Symbolic Logic* 37: 696–702.
- CORCORAN, John. 1974. Aristotle's natural deduction system, in Corcoran, J. (ed.) *Ancient Logic and its Modern Interpretation*, Dordrecht, NL: Kluwer Academic Publishers, 85–131.
- CRAVER, Carl F. 2006. When mechanistic models explain. *Synthese*, 153, 355–76.
- CRAVER, Carl F. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford, UK: Oxford University Press.
- CRAVER, Carl F., and David M. KAPLAN. 2018. Are more details better? On the norms of completeness for mechanistic explanations. *The British Journal for the Philosophy of Science*, <https://doi.org/10.1093/bjps/axy015>.

- CRESSWELL, Max J. 2017. Modal Logic. In Goble, Lou, ed. *The Blackwell Guide to Philosophical Logic*, Oxford, UK: Blackwell, 136–158.
- CRONQVIST, Lasse. 2011. *Tosmana: Tool for Small-n Analysis*, Version 1.3.2.0 [Computer Program].
- DAHLBERG, Stefan, and Jonas LINDE. 2016. Losing happily? The mitigating effect of democracy and quality of government on the winner-loser gap in political support. *International Journal of Public Administration*, 39.9: 652–664.
- DAMONTE, Alessia, and Fedra NEGRI. 2019. Gauging fiscal worlds: how the EU countries balanced equality and wealth between 2007 and 2016. *Quality and Quantity* <https://link.springer.com/article/10.1007/s11135-018-00833-x>
- DAMONTE, Alessia, Claire A. DUNLOP, and Claudio M. RADAELLI. 2014. Controlling bureaucracies with fire alarms: policy instruments and cross-country patterns. *Journal of European Public Policy* 21.9: 1330–1349.
- DAMONTE, Alessia. 2018. Gauging the Import and Essentiality of Single Conditions in Standard Configurational Solutions. *Sociological Methods & Research* <https://doi.org/10.1177/0049124118794678>.
- 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*, OnlineFirst, doi:10.1177/1558689818770061
- DE MORGAN, Augustus. 1847. *Formal Logic: Or, The Calculus of Inference, Necessary and Probable*. London, UK: Taylor and Walton.
- DE SWART, Harrie. 2018. *Philosophical and Mathematical Logic*. Springer.
- DELLA PORTA, Donatella, and Alberto VANNUCCI. 2006. A typology of corrupt networks. In Junichi Kawata, Ed. *Comparing Political Corruption and Clientelism*. London, Ashgate, 23–44.
- DELLA PORTA, Donatella, and Alberto VANNUCCI. 2012. *The Hidden Order of Corruption: An Institutional Approach*. London, UK: Routledge.
- DOWE, Phil. 1992. Process causality and asymmetry. *Erkenntnis* 37: 179–196.
- DUNN, J. Michael, and Greg RESTALL. 2002. Relevance logic. In Gabbay Dov M., and Franz Guenther, Eds. *Handbook of Philosophical Logic 2<sup>nd</sup> Edition Volume 6*. Dordrecht, NL: Springer, 1–128.
- DUŞA, Adrian. 2018. *QCA with R: A Comprehensive Resource*. Cham, CH: Springer.
- DUŞA, Adrian. 2019. Critical Tension: Sufficiency and parsimony in QCA. *Sociological Methods & Research*: <https://doi.org/0049124119882456>.
- ELMAN, Colin. 2005. Explanatory typologies in qualitative studies of international politics. *International Organization* 59.2: 293–326.
- FAIRFIELD, Tasha, and Andrew E. CHARMAN. 2017. Explicit Bayesian analysis for process tracing: guidelines, opportunities, and caveats. *Political Analysis* 25.3: 363–380.
- FALLETI, Tulia T., and Julia F LYNCH. 2009. Context and causal mechanisms in political analysis. *Comparative Political Studies* 42.9: 1143–1166.
- FETZER, James H. 2000. The paradoxes of Hempelian explanation. In Id., Ed. *Science, Explanation, and Rationality: Aspects of the Philosophy of Carl Gustav Hempel*. Oxford, UK: Oxford University Press, 111–137.
- FREDE, Michael. 1974. Stoic vs. Aristotelian syllogistic. *Archiv für Geschichte der Philosophie* 56.1: 1–32.
- FREDE, Michael. 1980. The original notion of cause. In Schofield, M., Burnyeat, M. and J. Barnes (eds.). *Doubt and Dogmatism. Studies in Hellenistic Epistemology*. Oxford, UK: Oxford University Press, pp. 217–49.

- GARCÍA-CASTRO, Roberto, and Miguel A. ARINO. 2016. A general approach to panel data set-theoretic research. *Journal of Advances in Management Sciences & Information Systems* 2:63–76.
- GASKIN, Richard. 1995. *The Sea Battle and the Master Argument*. Berlin, DE: De Gruyter.
- GEORGE, Alexander L., and Andrew BENNETT. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: The MIT Press.
- GERRING, John. 2011. *Social Science Methodology: A Unified Framework*. Cambridge, UK: Cambridge University Press.
- GILARDI, Fabrizio. 2002. Policy credibility and delegation to independent regulatory agencies: a comparative empirical analysis. *Journal of European Public Policy* 9.6: 873-893.
- GOERTZ, Gary, and Harvey STARR. 2002. *Necessary Conditions: Theory, Methodology, and Applications*. Lanham: Rowman and Littlefield.
- GOERTZ, Gary, and James MAHONEY. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton University Press.
- GOERTZ, Gary. 2002. The substantive importance of necessary condition hypotheses. In Goertz, Gary, and Harvey Starr. *Necessary Conditions: Theory, Methodology, and Applications*. Princeton, NJ: Princeton University Press, 65-94.
- GOERTZ, Gary. 2006. *Social Science Concepts: A User's Guide*. Princeton, NJ: Princeton University Press.
- GOERTZ, Gary. 2017. *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach*. Princeton, NJ: Princeton University Press.
- GOOD, Irving J. 1967. The white shoe is a red herring. *The British Journal for the Philosophy of Science* 17.4, 322.
- GOODMAN, Nelson. 1947. The problem of counterfactual conditionals. *The Journal of Philosophy* 44.5: 113-128.
- GOODMAN, Nelson. 1955. *Fact, Fiction, Forecast*. London, UK: University of London Press.
- GRANT, Edward. 1978. Aristotelianism and the longevity of the medieval world view. *History of Science* 16.2: 93-106.
- GROFMAN, Bernard, and Carsten Q. SCHNEIDER. 2009. An Introduction to Crisp Set QCA, with a Comparison to Binary Logistic Regression. *Political Research Quarterly* 62.4: 662-672.
- GUTTMAN, Louis. 1977. What is not what in statistics. *The Statistician*, 26.2: 81-107.
- GUYER, Paul, and Allen W. WOOD. 1998. *Immanuel Kant: Critique of Pure Reason*. Cambridge, UK: Cambridge University Press.
- HACKETT, Ursula. 2015. But not both: the exclusive disjunction in qualitative comparative analysis (QCA). *Quality & Quantity* 49.1: 75-92.
- HAESEBROUCK, Tim. 2016. The added value of Multi-Value Qualitative Comparative Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 17(1), Art. 12, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1601129>.
- HAESEBROUCK, Tim. 2019. An alternative update of the two-step QCA procedure. *Quality & Quantity*: 1-16.
- HEDSTRÖM, Peter, and Petri YLIKOSKI. 2010. Causal mechanisms in the social sciences. *Annual Review of Sociology* 36: 49-67.
- HEMPEL, Carl G. 1942. The function of general law in history. *Journal of Philosophy* 39:35-48.
- HEMPEL, Carl G. 1942. The function of general laws in history. Reprinted in Gardiner, Patrick, Ed. 1959. *Theories of History*. Glencoe, Ill.: The Free Press.

- HEMPEL, Carl G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York, NY: The Free Press.
- HINO, Airo. 2009. Time-series QCA. *Sociological Theory and Methods* 24.2: 247-265.
- HINTIKKA, Jaakko. 1969. Modality and quantification. In *Id.*, *Models for Modalities*. Dordrecht, NL: D. Reidel Publishing Company.
- HITCHCOCK, Christopher. 1999. Contrastive Explanation and the Demons of Determinism. *British Journal for the Philosophy of Science* 50, 585-612.
- HOOD, Christopher. 2010. *The Blame Game: Spin, Bureaucracy, and Self-Preservation in Government*. Princeton, NJ: Princeton University Press.
- HORROCKS, Ian Robert. 1997. *Optimising Tableaux Decision Procedures For Description Logics*. Manchester, UK: University of Manchester.
- HUBER, John D., and Charles R. SHIPAN. 2002. *Deliberate Discretion? The Institutional Foundations of Bureaucratic Autonomy*. Cambridge, UK: Cambridge University Press, 2002.
- HUG, Simon. 2013. Qualitative comparative analysis: How inductive use and measurement error lead to problematic inference. *Political Analysis* 21.2: 252-265.
- HUME, David. 1748. An inquiry concerning human understanding. In *The Philosophical Works of David Hume Including All the Essays, And Exhibiting the More Important Alterations and Corrections in the Successive Editions Published By The Author*, Vol. IV. London and UK: Black, Tait, and Tait, 1826.
- HUMPHREYS, Macartan, and Alan M. JACOBS. 2015. Mixing methods: A Bayesian approach. *American Political Science Review* 109.4: 653-673.
- ISRAEL, Jonathan I. 2001. *Radical Enlightenment. Philosophy and the Making of Modernity 1650-1750*. Oxford, UK: Oxford University Press.
- KAISER, Marie L., and Beate KRICKEL. 2016. The metaphysics of constitutive mechanistic phenomena. *The British Journal for the Philosophy of Science* 68.3: 745-779.
- KAPLAN, David M., and Carl F. CRAVER. 2011. The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective. *Philosophy of Science* 78: 601-27.
- KENDALL, Patricia L., and Paul F. LAZARSELD. 1950. Problems of survey analysis. In Merton, Robert K., and Paul F. Lazarsfeld, Eds. *Continuities in Social Research: Studies in the Scope and Method of 'The American soldier.'* New York, NY: The Free Press, 133-196.
- KING, Gary, Robert O. KEOHANE, and Sidney VERBA. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- KITCHER, Paul. 1981. Explanatory Unification. *Philosophy of Science* 48: 507-531.
- KITCHER, Paul. 1989. Explanatory unification and the causal structure of the world. In Kitcher, Paul and Wesley Salmon, eds. *Scientific Explanation*. Minneapolis: University of Minnesota Press, 410-505.
- KNEALE, William, and Martha KNEALE. 1962. *The Development of Logic*. Oxford, UK: Oxford University Press.
- KRIPKE, Saul A. 1959. A completeness theorem in modal logic. *The Journal of Symbolic Logic* 24.1: 1-14.
- KRIPKE, Saul A. 1963. Semantical considerations on modal logic. *Acta Philosophica Fennica* 16: 83-94.
- KUHN, Thomas. 1970. *The Structure of Scientific Explanation*, 2nd edition. Chicago, IL: University of Chicago Press.
- KUMLIN, S. 2004. *The Personal and the Political: How Personal Welfare State Experiences Affect Political Trust and Ideology*. New York, NY: Palgrave Macmillan US.



- KYBURG Jr, Henry E. 1965. Salmon's paper. *Philosophy of Science* 32.2: 147-151.
- The Language of Social Research. New York, NY: The Free Press, 115-125.
- The Policy Sciences. Stanford, CA: Stanford University Press, 155-192.
- Information Sciences 1.4:421-434.
- A Survey of Symbolic Logic. New York, NY: Dover.
- An Analysis of Knowledge and Valuation. La Salle, ILL: The Open Court.
- Symbolic Logic, 2nd Edition. New York, NY: Dover Publications.
- The Social Psychology of Procedural Justice. New York, NY: Plenum Press.
- Varieties of Social Explanation: An Introduction to the Philosophy of Social Science. Boulder, CO: Westview Press.
- The Hellenistic Philosophers Volume I. Translations of the Principal Sources, with Philosophical Commentary. Cambridge, UK: Cambridge University Press.
- American Psychologist 8.12: 750-751.
- Psychological Review 66, 81-95.
- Revue de Métaphysique et de Morale 57.4:456-458.
- Journal of Law, Economics, & Organization 10.1: 96-125.
- Philosophy of Science 67.1: 1-25.
- American Philosophical Quarterly 2.4: 245-264.
- The Cement of the Universe. A Study of Causation. Oxford, UK: Clarendon Press.
- Synthese 34, 361-396.
- Political Psychology 40.1: 165-181.
- Political Research Quarterly 66.1: 198-204.
- Philosophy of Science 66.1: 50-70.
- Handbook of Causal Analysis for Social Research. Dordrecht, NL: Springer, 75-90.
- Comparative Political Studies 41.4-5:412-436.
- European Union Politics 2.1: 103-122.

- MARX, Axel, and Adrian DUŞA. 2011. Crisp-set qualitative comparative analysis (csQCA) and model specification: Benchmarks for future csQCA applications. *International Methodological Innovations Online*, 6:2, 103-148.
- MATES, Benson. 1949. Stoic logic and the text of Sextus Empiricus. *The American Journal of Philology* 70.3: 290-298.
- MATES, Benson. 1961. *Stoic Logic*. Berkeley, CA: University of California Press.
- MATES, Benson. 1968. Leibniz on possible worlds. *Studies in Logic and the Foundations of Mathematics*. 52: 507-529.
- MCCUBBINS, Mathew D. 1985. The legislative design of regulatory structure. *American Journal of Political Science* 29.4: 721-748.
- MCCUBBINS, Mathew D., and Thomas SCHWARTZ. 1984. Congressional oversight overlooked: Police patrols versus fire alarms. *American Journal of Political Science* 28:1: 165-179.
- MCCUBBINS, Mathew D., Roger G. NOLL, and Barry R. WEINGAST. 1987. Administrative procedures as instruments of political control. *Journal of Law, Economics, & Organization* 3.2: 243-277.
- MCCUBBINS, Mathew D., Roger G. NOLL, and Barry R. WEINGAST. 1989. Structure and process, politics and policy: Administrative arrangements and the political control of agencies. *Virginia Law Review*. 75.2: 431-482.
- MCKINNEY, John C. 1969. Typification, typologies, and sociological theory. *Social Forces* 48:1, 1-12.
- MERTON, Robert K. 1968. *Social Theory and Social Structure*. New York, NY: Simon & Schuster.
- MICHAEL, Frederick S. 1976. What is the Master Argument of Diodorus Cronus? *American Philosophical Quarterly* 13.3:229-235.
- MILL, John Stuart. 1843. *A System of Logic, Ratiocinative and Inductive*, Books I-III. In John M. Robson (Ed.) 1974. *Collected works of John Stuart Mill*, Vol. VII. Toronto, CAN: University of Toronto Press.
- MORGAN, Stephen L., and Christopher WINSHIP. 2015. *Counterfactuals and Causal Inference*, 2<sup>nd</sup> ed. Cambridge, UK: Cambridge University Press.
- MOSTELLER, Frederick, and TUKEY, John W. 1977. *Data analysis and regression. A second course in statistics*. Reading, MA: Addison-Wesley.
- MUNGIU-PIPPIDI, Alina, Ed. 2013. *Controlling Corruption in Europe: The Anticorruption Report*. Verlag Barbara Budrich.
- NASSIM, Nicholas T. 2008. *The Black Swan: The Impact Of The Highly Improbable* London: Penguin.
- NEWMAN, Maxwell H.A. 1928. Mr. Russell's "Causal Theory of Perception." *Mind* 37:146, 137-148.
- OANA, Ioana-Elena, and Carsten Q. SCHNEIDER. 2018. SetMethods: an add-on R package for advanced QCA. *The R Journal* 10.1: 507-533.
- OSTROM, Elinor. 1998. A behavioral approach to the rational choice theory of collective action: Presidential address, American Political Science Association, 1997. *American Political Science Review* 92.1: 1-22.
- OSTROM, Elinor. 2005. *Understanding Institutional Diversity*. Princeton, NJ: Princeton University Press.
- PARSONS, Charles. 1982. Intensional logic in extensional language. *The Journal of Symbolic Logic* 47.2: 289-328.
- PEARL, Judea. 2009. *Causality. Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press.

- PEARL, Judea. 2015. Causes of effects and effects of causes. *Sociological Methods & Research* 44.1: 149-164.
- PELIZZO, Riccardo, and Frederick STAPENHURST. 2013. *Parliamentary Oversight Tools: A Comparative Analysis*. London, UK: Routledge.
- PRIEST, Graham. 2002. Paraconsistent logic. In Gabbay Dov M., and Franz Guentner, Eds. *Handbook of Philosophical Logic 2<sup>nd</sup> Edition Volume 6*. Dordrecht: Springer, 287-393.
- PRIOR, Arthur N. 1957. *Time and Modality*, Oxford, UK: Oxford University Press.
- PRZEWORSKI, Adam, and Henry TEUNE. 1970. *The Logic of Comparative Social Inquiry*. New York, NY: Wiley.
- QUINE, Willard V.O. 1951. Main trends in recent philosophy: two dogmas of empiricism. *The Philosophical Review*, 20-43. Reprinted in Id., 1980. *From A Logical Point of View*, 2<sup>nd</sup> Edition. Cambridge, MA: Harvard University Press.
- QUINE, Willard V.O. 1952. The problem of simplifying truth functions. *The American Mathematical Monthly* 59.8: 521-531.
- QUINE, Willard V.O. 1960. *Word and Object*. Cambridge, MA: The MIT Press.
- QUINE, Willard V.O. 1969. Natural kinds. In Rescher, Nicholas, Ed. *Essays in Honor of Carl G. Hempel*. Dordrecht, NL: Springer, 5-24.
- QUINE, Willard. V.O. 1974. *Roots of Reference*. Open Court.
- QUINE, Willard. V.O. 1982. *Methods of Logic*. Cambridge, MA: Harvard University Press.
- RAGIN, Charles C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley, CA: University of California Press.
- RAGIN, Charles C. 1994. *Constructing Social Research*. Thousand Oaks, CA: Sage.
- RAGIN, Charles C. 1998. The logic of qualitative comparative analysis. *International Review of Social History* 43.S6: 105-124.
- RAGIN, Charles C. 2000. *Fuzzy-Set Social Science*. Chicago, IL: University of Chicago Press.
- RAGIN, Charles C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago, IL: University of Chicago Press.
- RAGIN, Charles C., and Howard S. BECKER, Eds. 1992. *What Is A Case? Exploring the Foundations of Social Inquiry*. Cambridge, UK: Cambridge University Press.
- RAGIN, Charles C., and Peer FISS. 2017. *Intersectional Inequality: Race, Class, Test Scores, and Poverty*. Chicago, IL: University of Chicago Press.
- RAGIN, Charles C., and Sarah I. STRAND. 2008. "Using Qualitative Comparative Analysis to Study Causal Order." *Sociological Methods & Research* 36:431-41.
- RAGIN, Charles C., DRASS, K. A., and DAVEY S. 2006. *Fuzzy-Set/Qualitative Comparative Analysis 2.0*. [Computer Program].
- RAMSEY, Frank. 1931. *The Foundations of Mathematics and Other Essays*. Edited by R. B. Braithwaite. London, UK: Routledge and Kegan Paul.
- REESOR, Margaret E. 1965. Fate and possibility in early Stoic philosophy. *Phoenix* 19.4:285-97.
- RESCHER, Nicholas. 2007. *Conditionals*. Cambridge, MA: The MIT Press.
- RIHOUX, Benoit, and Gisèle DE MEUR. 2009. Crisp-set Qualitative Comparative Analysis (csQCA). In Charles C. Ragin and Benoit Rihoux. *Configurational Comparative Methods*. London, UK: Sage, 33-68
- RIHOUX, Benoit, and Heike GRIMM, Eds. 2006. *Innovative Comparative Methods for Policy Analysis. Beyond the Quantitative-Qualitative Divide*. Boston, MA: Springer.

- ROHLFING Ingo, and Carsten Q. SCHNEIDER. 2013. Improving Research On Necessary Conditions: Formalized Case Selection for Process Tracing after QCA. *Political Research Quarterly* 66.1:220–235.
- ROHLFING, Ingo. 2012. *Case Studies and Causal Inference: An Integrative Framework*. London, UK: Palgrave.
- ROHLFING, Ingo. 2018. Power and false negatives in qualitative comparative analysis: foundations, simulation, and estimation for empirical studies. *Political Analysis* 26.1: 72–89.
- ROSSI, Paolo. 1996. Bacon's idea of science. In Peltonen, Markku, Ed. *The Cambridge Companion to Bacon*. Cambridge, UK: Cambridge University Press, 25–46.
- ROTHSTEIN, Bo, and Daniel EEK. 2009. Political corruption and social trust: An experimental approach. *Rationality and Society* 21.1: 81–112.
- ROTHSTEIN, Bo, and Jan TEORELL. 2015. Getting to Sweden, Part II: Breaking with Corruption in the Nineteenth Century. *Scandinavian Political Studies* 38.3: 238–254.
- RUBINSON, Claude. 2013. Contradictions in fsQCA. *Quality & Quantity* 47.5: 2847–2867.
- RUBINSON, Claude. 2019. Presenting qualitative comparative analysis: Notation, tabular layout, and visualization. *Methodological Innovations* 12.2, DOI: 10.1177/2059799119862110.
- RUSSELL, Bertrand, 1927. *The Analysis of Matter*. London, UK: Routledge and Kegan Paul.
- SALMON, Wesley C. 1977. Laws, modalities, and counterfactuals. *Synthese* 35.2: 191–229.
- SALMON, Wesley C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.
- SALMON, Wesley C. 1989. Four decades of scientific explanation. In Kitcher, Philip, and Wesley C. Salmon, eds. *Scientific Explanation*. Minneapolis: University of Minnesota Press, 3–219.
- SALMON, Wesley C. 1998. *Causality and Explanation*. Oxford, UK: Oxford University Press.
- SALMON, Wesley C. 1999. Scientific explanation. In Salmon, Merrilee H., Earman, John, Glymour, Clark, Lennox, James G., Machamer, Peter, McGuire, JE, Norton, John D., Salmon, Wesley C., Schaffer, Kenneth F. *Introduction to the Philosophy of Science*, II ed. Indianapolis, IN: Hackett Publishing, 7–43.
- SARTORI, Giovanni. 1970. Concept misformation in comparative politics. *American Political Science Review* 64.4: 1033–1053.
- SARTORI, Giovanni. 1984. *Social Science Concepts: A Systematic Analysis*. London: Sage.
- SCHEDLER, Andreas. 1999. Conceptualizing accountability. In Schedler, Andreas, Larry Diamond, and Marc F. Plattner, Eds. *Power and Accountability in New Democracies*. Boulder, CO: Lynne Rienner, 13–28.
- SCHILLEMANS, Thomas. 2008. Accountability in the shadow of hierarchy: The horizontal accountability of agencies. *Public Organization Review* 8.2: 175.
- SCHNEIDER, Carsten Q., and Claudius WAGEMANN. 2010. Standards of good practice in qualitative comparative analysis (QCA) and fuzzy-sets. *Comparative Sociology* 9.3: 397–418.
- SCHNEIDER, Carsten Q., and Claudius WAGEMANN. 2012. *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*. Cambridge, UK: Cambridge University Press.

- SCHNEIDER, Carsten Q., and Ingo ROHLFING. 2016. Case studies nested in fuzzy-set QCA on sufficiency: Formalizing case selection and causal inference. *Sociological Methods & Research* 45.3:526-568.
- SCHNEIDER, Carsten Q., VIS, Barbara, and Kendra KOIVU. 2019. Set-Analytic Approaches, Especially Qualitative Comparative Analysis (QCA). *American Political Science Association Organized Section for Qualitative and Multi-Method Research, Qualitative Transparency Deliberations, Working Group Final Reports*, Report III.4, <http://dx.doi.org/10.2139/ssrn.3333474>
- SCOTT, Dana, and Patrick SUPPES. 1958. Foundational aspects of theories of measurement 1. *The Journal of Symbolic Logic* 23.2: 113-128.
- SCRIVEN, Michael. 1958. Explanations, predictions, and laws. In Feigl, Herbert, Michael Scriven, and Grover Maxwell, Eds. *Minnesota Studies in the Philosophy of Science*, vol. 2. Minneapolis, MN: University of Minnesota Press, 99-192.
- SCRIVEN, Michael. 1959. Explanation and prediction in evolutionary theory. *Science* 130. 477-482.
- SKAANING, Svend-Erik. 2011. Assessing the robustness of crisp-set and fuzzy-set QCA results. *Sociological Methods & Research* 40.2: 391-408.
- SMITHSON, Michael, and Jay VERKUILEN. 2006. *Fuzzy Set Theory: Applications in the Social Sciences*. London, UK: Sage.
- STAKELUM, James W. 1940. *Galen and the Logic of Proposition*. Rome, IT: Angelicum.
- STEVENS, Stanley S. 1946. On the theory of scales of measurement. *Science* 103.2684: 677-680.
- STEVENS, Stanley S. 1958. Measurement and man. *Science* 127.3295: 383-389.
- STONE, Marshall H. 1936. The theory of representations of Boolean algebras. *Transactions of the American Mathematical Society* 40:1, 37-111.
- TARSKI, Alfred. 1956. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Oxford, UK: Oxford University Press.
- THIBAUT, John W., and Laurens WALKER. 1975. *Procedural Justice: A Psychological Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.
- THIEM, Alrik. 2010. Set-relational fit and the formulation of transformational rules in fsQCA. *COMPASSSS WP Series* 2010:61, <http://www.compassss.org/wpseries/Thiem2010.pdf>.
- THIEM, Alrik. 2015. Parameters of fit and intermediate solutions in multi-value Qualitative Comparative Analysis. *Quality & Quantity* 49: 657-674.
- THOMANN, Eva, and Martino MAGGETTI. 2017. Designing research with qualitative comparative analysis (QCA) approaches, challenges, and tools. *Sociological Methods & Research*: 0049124117729700.
- TYLER, Tom R. 1990. *Why People Obey The Law*. Princeton, NJ: Princeton University Press.
- Van Evera, Stephen. 1997. *Guide to Methods for Students of Political Science*. Ithaca, NY: Cornell University Press.
- VAN FRAASSEN, Bas C. 2008. *Scientific Representation: Paradoxes of Perspective*. Oxford, UK: Oxford University Press.
- VAN ROOIJ, Robert, and Katrin SCHULZ. 2019. Natural kinds and dispositions: a causal analysis. *Synthese* <https://doi.org/10.1007/s11229-019-02184-y>
- VERBA, Sidney, 1967. Some dilemmas in comparative research. *World Politics*, 20.1: 111-127.
- WAGEMANN, Claudius, Jonas BUCHE, and Markus B. SIEWERT. 2016. QCA and business research: Work in progress or a consolidated agenda?. *Journal of Business Research* 69.7: 2531-2540.

- WALKER, Henry A., and Bernard P. COHEN. 1985. Scope statement: imperatives for evaluating theory. *American Sociological Review* 50.3:288-301.
- WHITEHEAD, Alfred N., and Bertrand RUSSELL, 1927. *Principia Mathematica*, 2<sup>nd</sup> edition. Cambridge, UK: Cambridge University Press.
- WIENER, Philip Paul. 1932. The experimental philosophy of Robert Boyle (1626-91). *The Philosophical Review* 41.6: 594-609.
- WITTGENSTEIN, Ludwig. 1922. *Tractatus Logicus Philosophicus*. London, UK: Kegan Paul, Trench, Trubner & Co.
- WOLNIEWICZ, Boguslaw. 1979. A Wittgensteinian semantics for propositions. In: Anscombe, Gertrude E. M., Cora Diamond, and Jenny Teichman. Eds. *Intention and intentionality. Essays in Honour of G. E. M. Anscombe*. Ithaca, NY: Cornell University Press, 165-178.
- WOLNIEWICZ, Boguslaw. 1999. *Logic and Metaphysics. Studies in Wittgenstein's Ontology of Facts*. Warsaw, PL: Polskie Towarzystwo Semiotyczne.
- WOODWARD, James. 1989. The causal-mechanical model of explanation. In Kitcher, Philip, and Wesley C. Salmon, eds. *Scientific Explanation*. Minneapolis: University of Minnesota Press, 356-383.
- WOODWARD, James. 2002. What is a mechanism? A counterfactual account. *Philosophy of Science*, 69: S366-S377.
- WOODWARD, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford, UK: Oxford University Press.
- ZADEH, Lofti A. 1968. Fuzzy algorithms. *Information and Control* 12.3: 94-102.
- ZADEH, Lofti A. 1978. PRUF a meaning representation language for natural languages. *International Journal of Man-Machine Studies* 10, 395-460.