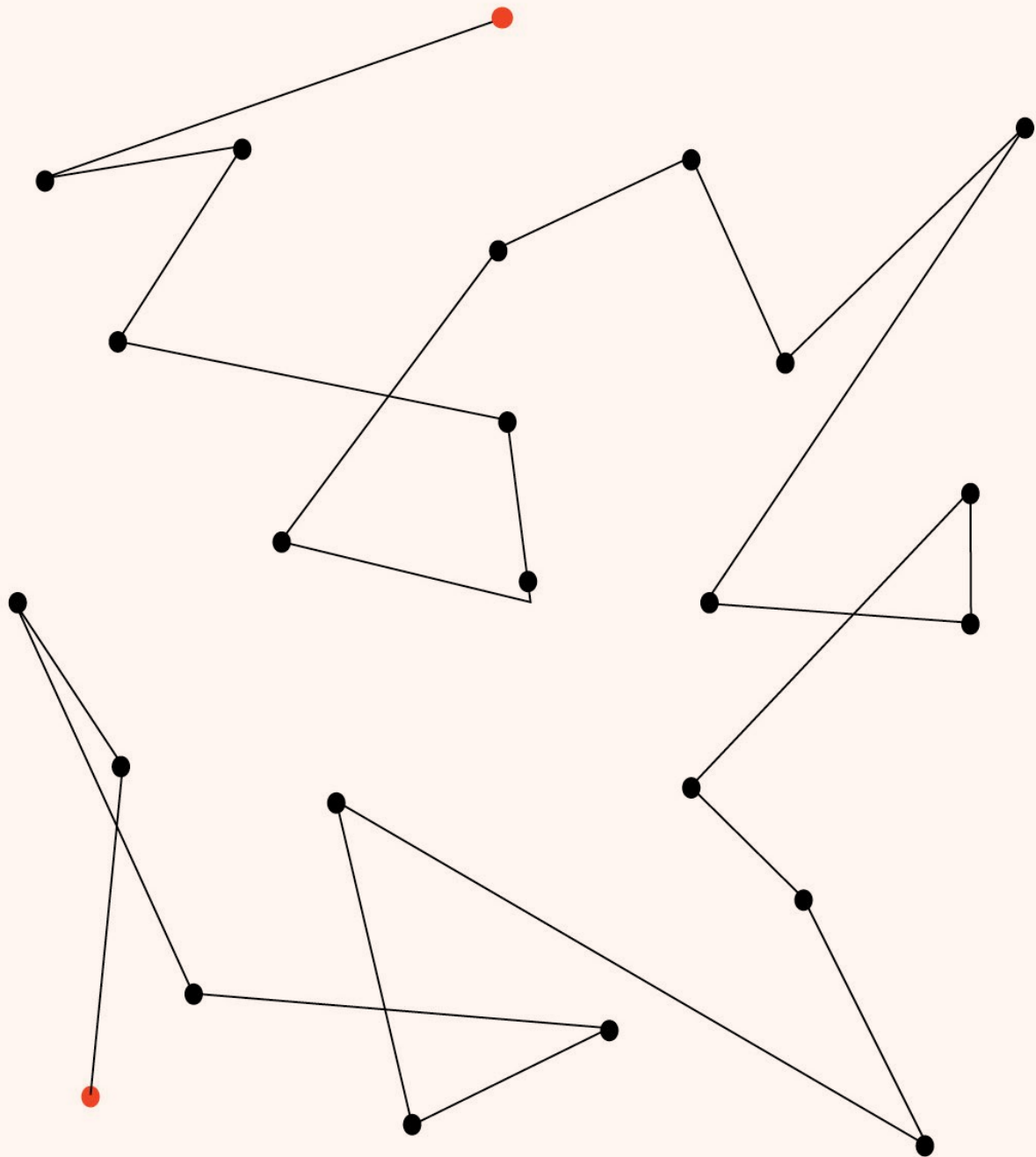


# COGENCY

JOURNAL OF REASONING AND ARGUMENTATION

---





# COGENCY

JOURNAL OF REASONING AND ARGUMENTATION

**COGENCY**  
**Journal of Reasoning and Argumentation**  
**Vol. 13, No. 2, Summer 2021**

Cogency is a publication edited by the Faculty of Psychology, Diego Portales University, Vergara 275, Santiago, Región Metropolitana, Chile.

Email: [cogency.journal@mail.udp.cl](mailto:cogency.journal@mail.udp.cl); web site: [www.cogency.udp.cl](http://www.cogency.udp.cl)

The papers of Cogency are indexed or summarized in: Scopus; Latindex; Dialnet; The Philosopher's Index; International Directory of Philosophy (Philosophy Documentation Center); Elektronische Zeitschriftenbibliothek EZB (Electronic Journals Library); Genamics Journal Seek; NewJour. Electronic Journals & Newsletters; Philosophical Journals on the web; WorldCat OCLC.

ISSN 0718-8285

**Director:**

Francisco Pizarro, Diego Portales University, Santiago, Chile

**Assistant Director:**

Julián "Iñaki" Goñi, Pontificia Universidad Católica, Santiago, Chile

**Editorial Board:**

Margaret Cuonzo, Long Island University, New York, United States  
Frans van Eemeren, University of Amsterdam, Amsterdam, The Netherlands  
Selma Leitao, University Federal de Pernambuco, Pernambuco, Brazil  
Christian Kock, University of Copenhagen, Copenhagen, Denmark  
Christopher Tindale, University of Windsor, Windsor, Canada  
Luis Vega Reñón, Nacional de Educación a Distancia University, Madrid, Spain  
Dima Mahommed, Universidade Nova de Lisboa, Lisboa, Portugal  
Steven Patterson, Marygrove College, Detroit, United States  
Sofía Inés Albornoz, University of do Vale do Rio dos Sinos, Sao Leopoldo, Brasil  
Katarzyna Budzynska, Cardinal Stefan Wyszyński University, Warsaw, Poland  
Peter Cramer, Simon Fraser University, Burnaby, Canada  
Eduardo Fermandois, Catholic University, Santiago, Chile  
María Marta García Negroni, Buenos Aires University, Buenos Aires, Argentina  
Michael Gilbert, York University, Toronto, Canada  
David Godden, Old Dominion University, Norfolk, United States  
Leo Groarke, Trent University, Canada  
Cornelia Ilie, Strömstad Academy, Strömstad, Sweden  
José Julio León, Universidad Santo Tomás, Santiago, Chile  
Roberto Marafioti, University of Buenos Aires, Buenos Aires, Argentina  
Andrei Moldvan, University of Barcelona, Barcelona, Spain  
Rudi Palmieri, University of Lugano, Lugano, Switzerland  
Francisca Snoeck-Henkemans, University of Amsterdam, Amsterdam, The Netherlands

**Correspondence, subscription and sales**

Cogency Journal  
Vergara 275, Santiago, Chile  
E-mail: [cogency.journal@mail.udp.cl](mailto:cogency.journal@mail.udp.cl)

**Production design:**

Camila González S.

© 2022 by Universidad Diego Portales. All rights reserved.

*No part of this publication may be reproduced, displayed, or transmitted in any form or by any means, electronic or mechanical, including photocopying or by any information storage or retrieval system, without the prior written permission from the Director of the journal.*

# Contenidos

7 TOWARDS THE DEFINITION OF LOGICAL  
COMPETENCE *A. Schumann*

41 SÀÁGUA AND BAUMTROG ON PRACTICAL  
RATIONALITY *D. Botting*

79 PRAGMÁTICA Y ARQUITECTURA MENTAL:  
UNA PROPUESTA DESDE LA MODULARIDAD  
MASIVA DE LA MENTE *N. Acuña Luongo*

103 EL ROL DEL RAZONAMIENTO ABDUCTIVO EN EL  
RAZONAMIENTO PROBATORIO DEL DERECHO:  
UNA PERSPECTIVA ORIENTADA HACIA LA PRÁCTICA  
*J. E. Sotomayor Trelles, G. E. Uscamayta & N. J. Apaza Jallo*

139 HOW UNIVERSAL GENERALIZATION WORKS  
ACCORDING TO NATURAL REASON *K. S. Hodge*



# TOWARDS THE DEFINITION OF LOGICAL COMPETENCE

\*

ANDREW SCHUMANN

Head of the Department of Cognitive Science and Mathematical Modelling,  
University of Information Technology and Management in Rzeszow, Sucharskiego 2,  
35-225 Rzeszów, Poland  
andrew.schumann@gmail.com

## Abstract

*According to logical psychologism that was popular in the nineteenth century, logic was regarded as a natural ability of human psychology. Consequently, the logical competence as a realization of logic knowledge was treated as one of the innate features of human thinking. Nevertheless, within cognitive science it was experimentally proved that our thinking is not free from cognitive biases and to the same extent our reasoning is not free from logical fallacies. Hence, we are forced to consciously clear the thinking of possible distortions to follow logical norms and to realize the logical competence thereby. In the paper, it is shown that some cognitive biases are observed even at the level of cellular reactions to the environment and then at the level of animal behavior. Therefore not logic, but cognitive biases are a natural (biological) mechanism of human thinking. So, the problem of defining the logical competence arises. In this paper, there are some arguments that the logical competence has once appeared as an especial social practice and it has then been developed for a long time before the first treatises on logic.*

**Keywords:** logic, cognitive bias, optical illusion, logical competence, legal hermeneutics

## 1. Introduction

There may be two general approaches to an explication of logical knowledge: normative and phenomenological. In the normative approach, logical concepts and relationships are regarded as established through a creation of logical norms. This approach began with introducing some logical laws (by Aristotle and Gottfried Wilhelm Leibniz) such as *tertium non datur* and the principle of sufficient reason and then it was transferred into a qualitatively new level with creating the first axiomatic systems of logic (by Bertrand Russell, David Hilbert and many others logicians of the beginning of the 20<sup>th</sup> century). The second approach emphasizes descriptions, identifications, and developments of logical structures in existing knowledge systems, for example, in jurisprudence. Nevertheless, even an explication of logic in the light of the phenomenological approach may have a normative character, too – such a way was implemented, for example, by E.V. Bulygin in his analysis of logical relations in legal discourse (Alchurron & Bulygin, 2013).

This paper examines the nature of logical knowledge within the normative approach, namely, logic is presented as a *logical competence* – the normalized ability to build consistent conclusions, applying certain inference rules. As a result of such an analysis, it is possible to determine the very boundaries of the normalization of logical knowledge as such. Within the framework of this consideration, the logical competence is understood, first of all, as a social mechanism of cognition within the framework of common human ways of finding a consensus.

In the every-day life of people, the logical competence is detected only with some biased forms (Tverky & Kahneman, 1974). This raises problems how far logic or logical competence can be considered an innate ability of humans. In the 19<sup>th</sup> century it was maintained (within the logical psychologism), see (Heymans, 1905), that logic gives some limits of our thinking and, therefore, is absolutely natural for us, because it is a part of our psychology. In this paper, I'll show that logic is ever biased not only in our reasoning, but also in biological processes of transmitting signals from the external world, and it can be biased more or less, but its pure (unbiased) forms are almost impossible. So, the logical competence is not natural and it arises as a special social practice of finding different forms of consensus among humans. In Section 2, the logical psychologism presenting logic as an innate human ability is sharply criticized. In Section 3, it is shown that biased and unbiased forms of reasoning are studied together in informal logic or argumentation theory. In Section 4, on the basis of intra- and inter-cellular effects caused by a lateral activation and a lateral inhibition, it is demonstrated that each cell receives signals most often through some “biases”. Let us note that the lateral activation can be treated as a state of satisfaction on the microlevel, while the lateral inhibition as a state of stress. In Section 5, there are regarded some biased ways

of intelligent reactions performed by different animals to show that these reactions repeat the lateral activation and the lateral inhibition, but now on the macrolevel. In Section 6, it is shown that logical competence without some biased manifestations can be realized only as a special social practice among humans and this practice was first formed within the Sumero-Akkadian legal tradition and only then it was reflected within philosophy.

## 2. Logical Psychologism and its Criticism

Due to some experimental data of cognitive science (Tversky & Kahneman, 1974), it became known that logic is not an innate ability of humans and it can be cultivated only in the process of learning and practical applications. Accordingly, the use of logical competence in legal discourse or other human areas does not occur by itself, but it is possible only through a direct constitution of some norms.

In the 19<sup>th</sup> century, there was another general understanding of logical competence, according to which this competence was considered an integral part of human thinking and cognition, and logic itself with all the laws was reduced to psychology and its laws. This approach to the logical competence (in the light of justifying the reducibility of logic to psychology) is called *logical psychologism*. According to it, the logical competence as a whole is natural for human cognition (Heymans, 1905). The development of logic is just the development of reflection on the already existing forms of logical thinking.

The logical psychologism has its roots in traditional definitions of logic as a natural way of thinking, e.g. in this manner logic is defined by Antoine Arnauld (1612 – 1694) and Pierre Nicole (1625 – 1695):

La logique est l'art de bien conduire sa raison dans la connaissance des choses, tant pour s'instruire soi-même que pour en instruire les autres.

Cet art consiste dans les réflexions que les hommes ont faites sur les quatre principales opérations de leur esprit, *concevoir, juger, raisonner et ordonner* (Arnauld & Nicole, 1662/1992, p. 30).

Logic is an art of leading one's reason well to the knowledge of things, both to educate oneself and to instruct others.

This art consists of the reflections that men have made on the four main operations of their mind: *representation, judgement, entailment, and ordering*.

This understanding of logic was common for logic handbooks written from the 19<sup>th</sup> century to the early 20<sup>th</sup> century: logic is “a science of the laws of correct thinking” (Chelpanov, 1917, p. 1). The point is that logic gives general frameworks of our knowledge:

Logic can do nothing but bring to our knowledge the conditions under which this subjective feeling of necessity arises, and express them in a general form (von Sigwart, 1895, p. 15).

Our cognitions are possible only due to logic:

Logic is a science of the methods of thinking, with the help of which the human mind seeks to achieve the knowledge of existence (Povarnin, 1916, p. 3).

Logic is a science that shows how thinking should be done in order for truth to be achieved (Chelpanov, 1917, p. 1).

There are several laws of our cognition (first of all, the principle of identity, the law of contradiction, and the law of excluded middle) and several forms of thinking (first of all, concept, judgement, and inference) which are studied in logic:

Logic is the science of the laws and forms of correct constructions of thoughts (Vinogradov & Kuz'min, 1954, p. 3).

The logical psychologism was preserved in Russian handbooks of logic till now:

The term 'logic' ... is used both to designate a set of rules that manage the process of thinking, reflecting reality, and to designate the science of the rules of reasoning and the forms in which it is carried out. <...> It [A.S.: logic] explores thinking as a means of knowing the objective world, those forms and laws in which the world is reflected in the process of thinking (Getmanova, 2011, p. 11).

Friedrich Ludwig Gottlob Frege (1848 – 1925) was one of the first to offer sharp criticism of logical psychologism and to convincingly show that logic, nevertheless, is completely autonomous from psychology and represents an intersubjective knowledge obtained through the study of methods of compositional combination of meaningful sentences with achieving a conclusion (Frege, 1934). Logic, in the end, is a more general science than mathematics.

According to Frege, logic studies our thinking as drawing conclusions, but it does not study the external world and its representations:

Die Gedanken sind weder Dinge der Außenwelt noch Vorstellungen. Ein drittes Reich muß anerkannt werden. Was zu diesem gehört, stimmt mit den Vorstellungen darin überein, daß es nicht mit den Sinnen wahrgenommen werden kann, mit den Dingen aber darin, daß es keines Trägers bedarf, zu dessen Bewußtseinsinhalte es gehört (Frege, 1918/1919, p. 69).

Thoughts are neither things of the external world nor representations. A third empire must be recognized. What belongs to this agrees with the idea that it cannot be perceived by the senses, but with things that there is no need for an actor to whose consciousness it belongs.

Frege's approach to the logical competence is now called *logicism*. According to this approach, mathematics is reducible to logic, and the logical competence itself is understood as an ability to draw correct conclusions based on meaningful sentences chosen as axioms or hypotheses (assumptions). This ability is associated exclusively with a knowledge of inference rules, which give consistent results. These inference rules are formulated within constructing a logical system and can be different. In particular, in the 20<sup>th</sup> century the so-called non-classical logics arose, which have axioms and inference rules a lot different from those used by the creators of mathematical logic (Gottlob Frege, Bertrand Russell and others), but they also provide consistent knowledge. For example, there are many-valued logics, where the law of excluded middle does not hold true, and paraconsistent logics, where the law of contradiction does not hold true, etc.

Thus, in logic there are no laws – it is just a study of inferences by their formalizations taking into account the fact that these formalizations can be different, because they are acceptable only due to their consistency:

<...> ein neues Feld der Forschung eröffnet, indem die Formalisierung des logischen Schließens zu einer systematischen Beweistheorie verwertet wurde, welche die Frage nach der Tragweite der logischen Schlußweisen, die von der traditionellen Logik nur in einer sehr speziellen Form gestellt und gelöst wurde, in systematischer Allgemeinheit behandelt und durch deren Untersuchungsmethode die Probleme der Grundlagen der Mathematik mit den logischen Problemen in unmittelbaren Zusammenhang treten (Hilbert & Bernays, 1968, p. 44).

<...> there was opened up a new field of research by utilizing the formalization of logical inference into a systematic theory of proof, which addresses the question of the scope of logical inference, which traditional logic posed and solved only in a very special way, dealt with in a systematic generality and whose investigation method brings the problems of the fundamentals of mathematics into direct connection with the logical problems.

Kurt Friedrich Gödel (1906 – 1978) showed, in turn, the boundaries of logicism and thus justified the autonomy of mathematics from logic (Gödel, 1931). As a consequence, the logical competence is now understood as a unique ability that is not directly related to a mathematical competence. This is the ability, first of all, to automatically draw conclusions, avoiding contradictions. Investigations of this ability began to be called formal logic:

Our subject is *logic* – or, as we may say more fully, in order to distinguish from certain topics and doctrines which have (unfortunately) been called by the same name, it is *formal logic*.

Traditionally, (formal) logic is concerned with the analysis of sentences or of propositions and of proof with attention to the *form* in abstraction from the *matter* (Church, 1996, p. 1).

Other names of the study of inferences: theoretical logic (*die theoretische Logik*), mathematical logic (*die mathematische Logik*), and symbolic logic (*die symbolische Logik*), e.g. see (Hilbert & Ackermann, 1949, p. 1). Now logic as such is understood as this investigation: “Logic is a science that studies principles and methods of inference” (Mortari, 2001, p. 2); “Logic is the study of the methods and standards of inference” (Kelley, 2014, p. 2). On the other hand, logic began to be studied in the context of cognitive science, also (Stenning & van Lambalgen 2011, Stenning & Lambalgen 2008).

### 3. Cognitive Biases and Informal Logic

From the new understanding of logical competence as an ability to mechanically draw consistent conclusions, some plans arose to create *artificial intelligence* – abstract machines (such as Alan Turing’s machines), which themselves are able, based on the data received at the input of the system, to give new data (like conclusions) at the output from the system so that these new data contain already a new knowledge in the form of processing the received data according to the required algorithms. By the beginning of the 21<sup>st</sup> century significant successes have been achieved in the creation of artificial intelligence systems, which have led to a series of information revolutions.

However, in the form of understanding the logical competence in computer science, it is no longer a natural ability of thinking – neither of animals nor of humans. It is presented exclusively in the implementation of electronic devices, where there is information at the input as well as algorithms that transform it to obtain information at the output. Moreover, these algorithms are taken from classical mathematics with the assumption that all actions in them are sequential and deterministic.

At the beginning of the 21<sup>st</sup> century, in computer science, a new direction has emerged, called *unconventional computing*, which studies the possibility of representing any complex natural process (mechanical, chemical, biological) as a specific system of logical (or, more generally, arithmetic) functions (Adamatzky, 2017). For instance, it was shown that *Physarum polycephalum*, the unicellular organism behaving as a giant amoeba, can approximate some simple algorithms of Kolmogorov-Uspensky machines

(Adamatzky, 2007) or algorithms of cellular automata (Shirakawa et al., 2015). This amoeba can be examined even as some biological logic gates (Adamatzky & Schubert, 2014; Jones & Adamatzky, 2010, Mayne & Adamatzky, 2015; Schumann, 2015; Whiting et al., 2014). Then these results were generalized to demonstrate that some proteins of amoeboid organisms are responsible for their intelligent behavior. Therefore these proteins can be considered a medium implementing different logical and arithmetic functions. So, some networks of these proteins such as actin filament networks can be presented as a logic or computation medium, too (Adamatzky & Mayne, 2015; Alonso-Sanz & Adamatzky, 2016; Schumann, 2016, 2017; Siccardi & Adamatzky, 2015, 2016; Siccardi et al., 2016). Hence, these studies of actin filament networks and behaviors of unicellular organisms are being carried out in order to find out what specific logical and mathematical problems can be encoded in the biochemical processes of living organisms, thus presenting the organisms themselves in the form of computers – in other words, “the logical competence” of unicellular beings is studied (Schumann, 2018). One of the important results of these studies is that noticeable cognitive distortions were shown in the process of receiving signals by the single-celled organisms with subsequent reactions of the organisms to them. It turns out that cognitive distortions and violations of the rules of logic are a natural and integral part of cognition already at the level of behavior of unicellular organisms. Their “logical competence” is not perfect – the way computers “think” is significantly different (Schumann, 2018, 2019; Shirakawa et al., 2020).

We can show, therefore, that logic is implemented biologically only in some biased forms. A pure logic does not exist in nature. As we remember, in the philosophy of the 19<sup>th</sup> century, the idea that logic is a natural way of thinking for all humans was dominant within the logical psychologism. Nevertheless, in the 20<sup>th</sup> century it was proved experimentally that humans do not think logically in its strict sense surely – their logical competence is imperfect. They make gross blunders even at the primary level of logic. For example, the logical conjunction  $A \& B$  is often understood not correctly – some referents do not know that its probability  $P(A \& B)$  must be always smaller than the individual probability  $P(A)$  or  $P(B)$  (Tversky & Kahneman, 2008). It is an insensitive blunder in understanding conjunction as one of the basic logical operations.

Let  $\vdash$  be a sign of logical deducibility. So,  $A \vdash B$  means that from  $A$  we can deduce  $B$ , i.e.  $A$  is a premise for inferring  $B$ , and  $\vdash B$  means that  $B$  is an axiom, i.e. it holds without premises. Then any inference of  $A$  from  $B$  in classical logic:  $A \vdash B$ , assumes that an appropriate implication  $A \Rightarrow B$  holds. This statement is called the deduction theorem: if  $A \vdash B$ , then  $\vdash A \Rightarrow B$ .

In classical logic there are the following two basic inference rules:

- *modus ponens*: If  $A \Rightarrow B$  and  $A$ , then  $B$  (it means that from the true implication  $A \Rightarrow B$  and the true statement  $A$  we can infer that  $B$  is true, also);
- *modus tollens*: If  $A \Rightarrow B$  and  $\neg B$ , then  $\neg A$  (it means that from the true implication  $A \Rightarrow B$  and the true negation of  $B$  we can infer the true negation of  $A$ , as well).

These inference rules are always valid. There are similar deductions which are invalid in fact:

- *affirmation of the consequent*: If  $A \Rightarrow B$  and  $B$ , then  $A$  (but it is invalid that from the true implication  $A \Rightarrow B$  and the confirmation of its consequent  $B$  we can infer that  $A$  is true);
- *denial of the antecedent*: If  $A \Rightarrow B$  and  $\neg A$ , then  $\neg B$  (but it is invalid that from the true implication  $A \Rightarrow B$  and the denial of its antecedent  $A$  we can infer the denial of its consequent  $B$ ).

Understanding that *modus ponens* and *modus tollens* are valid, while the affirmation of the consequent and the denial of the antecedent are invalid, constitute the nuts and bolts of elementary propositional logic. Without this understanding any reasoning will contain fallacies. But it turned out that sometimes even the scientists cannot draw elementary correct conclusions from conditionals presented by true implications. So, seventy-two scientists including psychologists, biologists, and physicists from a large US university completed a questionnaire (Kern et al., 1983) to check their abilities to rate the validity of four forms of propositional inference: *modus ponens*, *modus tollens*, affirmation of the consequent, and denial of the antecedent. As a result, some of them made serious fallacies, see Table 1. In the meanwhile, there was detected the failure of nearly half of the scientists to recognize the logical validity of *modus tollens*, one of the two fundamental inference rules of propositional logic. This experiment shows that even the scientists do not think purely logically with avoiding unintentional fallacies.

Table 1.

This table shows that the scientists make gross fallacies in understanding *modus ponens*, *modus tollens*, affirming the consequent, denying the antecedent. Even the logic training for them is not so helpful. The table is taken from the work of L.H. Kern, H.L. Mirels, V.G. Hinshaw (1983).

| REFERENTS                                                                          | PSYCHOLOGISTS | BIOLOGISTS | PHYSICISTS | OVERALL<br>PERCENTAGE<br>ACROSS<br>DISCIPLINES (NO<br>LOGIC TRAINING) | SOME LOGIC<br>TRAINING (N = 26) |
|------------------------------------------------------------------------------------|---------------|------------|------------|-----------------------------------------------------------------------|---------------------------------|
| 1. Recognizing<br>the logical validity<br>of <i>modus ponens</i> ,                 | 91.5%         | 96%        | 96%        | 94.5%                                                                 | 100%                            |
| <i>abstract scheme</i> :                                                           | 83%           | 92%        | 92%        | 89%                                                                   | 100%                            |
| <i>concrete example</i> :                                                          | 100%          | 100%       | 100%       | 100%                                                                  | 100%                            |
| 2. Recognizing<br>the logical<br>invalidity of denial<br>of the antecedent,        | 79%           | 66.5%      | 87.5%      | 77.5%                                                                 | 83.5%                           |
| <i>abstract scheme</i> :                                                           | 83%           | 58%        | 75%        | 72%                                                                   | 67%                             |
| <i>concrete example</i> :                                                          | 75%           | 75%        | 100%       | 83%                                                                   | 100%                            |
| 3. Recognizing<br>the logical invalidi-<br>ty of affirmation of<br>the consequent, | 79%           | 62.5%      | 83.5%      | 72.5%                                                                 | 84.5%                           |
| <i>abstract scheme</i> :                                                           | 83%           | 75%        | 75%        | 73%                                                                   | 87%                             |
| <i>concrete example</i> :                                                          | 75%           | 50%        | 92%        | 72%                                                                   | 82%                             |
| 4. Recognizing<br>the logical validity<br>of <i>modus tollens</i> ,                | 41.5%         | 58%        | 66.5%      | 55%                                                                   | 72%                             |
| <i>abstract scheme</i> :                                                           | 33%           | 33%        | 58%        | 41%                                                                   | 53%                             |
| <i>concrete example</i> :                                                          | 50%           | 83%        | 75%        | 69%                                                                   | 91%                             |

As it was shown (Sjöberg, 2017), the fallacy of affirming the consequent is very common for biologists. Sometimes they recognize the following reasoning as logically valid: “If bipolar patients have reduced cognitive processing speed, then we will observe slower reaction time on a test. We observe a slower reaction time, and therefore bipolar patients have reduced cognitive processing speed”. Furthermore, they often use the false analogy to extrapolate from sharing some characteristics between two groups of animals to sharing their other characteristics (Sjöberg, 2017).

Our reasoning with its biased forms is studied in two modern logical disciplines: critical thinking and informal logic (or argumentation theory). So, critical thinking is examined as an ability to detect cognitive biases to avoid them (Baker, 2020) and informal logic is defined as a science that studies both biased and unbiased inferences:

Logic ... is the branch of philosophy that studies the reasoning process and seeks to understand the differences between good and bad reasoning (Crews-Anderson, 2007, p. 8).

Logic is the study of the methods and principles used to distinguish correct from incorrect reasoning (Copi et al., 2014, p. 2).

Thus, logic is not natural even for humans and even for scientists and we study it to avoid some biased inferences. We can then wonder if we can expect to see a pure logic in nature – how far biological processes of *Physarum polycephalum* or actin filament networks can implement the pure logic.

#### 4. Explaining Optical Illusions due to the Lateral Inhibition and the Lateral Activation

We may observe the following fundamental duality in all active reactions of different organisms to external stimuli: each stimulus either attracts the organism or repels it. So, each stimulus which changes the behavior of individuals in fact, can be considered either an attractant (such as food pieces or sex pheromone) or a repellent (such as predators or poisons). This duality is topological: the attractant can be occupied **from** different directions, but from the repellent each individual moves **in** different directions (Shirakawa et al., 2012; Fackler & Grosse, 2008; Pollard & Borisy, 2003; Van Haastert & Devreotes, 2004). On the basis of this topological duality, it is possible to present (i) the attractants  $|p|$ ,  $|q|$ , ...,  $|r|$ , occupied by some individuals, as true propositional variables  $p$ ,  $q$ , ...,  $r$ ; (ii) and the repellents  $|\dot{p}|$ ,  $|\dot{q}|$ , ...,  $|\dot{r}|$ , avoided by some individuals, as true negations  $\neg p$ ,  $\neg q$ , ...,  $\neg r$ .

Evidently, different attractants (repellents) have a different intensity of their influence on individual behaviors: they can more or less attract (repel) organisms. This property can be

expressed by a degree of truth (falsehood) presented by the interval of real numbers  $[0, 1]$ , where it is possible to be closer to 0 (it means, do more to repel) or to 1 (do more to attract).

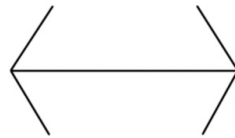
Now, let us consider several attractants  $|p|, |q|, \dots, |r|$  and several repellents  $|\dot{p}|, |\dot{q}|, \dots, |\dot{r}|$ . Surely that all the repellents  $|\dot{p}|, |\dot{q}|, \dots, |\dot{r}|$  must be avoided. Then this reaction of individuals can be treated as the following true conjunction:  $\neg p \& \neg q \& \dots \& \neg r$  (it means  $|\dot{p}|, |\dot{q}|, \dots, |\dot{r}|$  are avoided simultaneously). However, in respect to the attractants  $|p|, |q|, \dots, |r|$  reactions of individuals can be very different. First of all, there are the following two typical reactions: (i) fusion as an attempt of focusing only on one attractant (or on a small number of attractants) to occupy it (them) first and foremost; (ii) splitting as an attempt of occupying several attractants at once, the more the better.

Both possible reactions to the attractants are explained by the following two complementary mechanisms in perceiving signals on the intra- and inter-cellular levels: (i) a lateral inhibition corresponding to a fusion, it expresses some stress conditions which increase the contrast of one signal to react more directly and efficiently and to be better focused; (ii) a lateral activation corresponding to a splitting, it shows some comfortable conditions which decrease the contrast of signals to perform concurrent actions (Jones, 2015; Schumann & Kuznetsov, 2018). These mechanisms in perceiving stimuli are scale-invariant. First, they occur during assembling and disassembling of actin filament waves to cause a fusion of protoplasmic tubes of *Physarum polycephalum* or splitting them. Second, they occur between groups of neurons within the visual cortex to process visual cues more or less focused. Third, they occur between active zones of swarms to react to their environments to occupy attractants more or less focused, too.

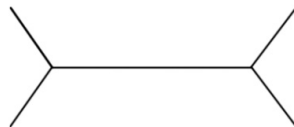
Repellents  $|\dot{p}|, |\dot{q}|, \dots, |\dot{r}|$  cause always a stress in individuals, therefore each individual tries to be concentrated on the repellents due to the lateral inhibition to avoid all of them in advance. But for attractants  $|p|, |q|, \dots, |r|$  each individual can be more laterally inhibited or more laterally activated. As a consequence, it starts to occupy all attractants or only some of them. The attractants in focus are connected by conjunction. The attractants which are not in focus are connected by disjunction. Hence, the lateral inhibition for signals  $|p|, |q|, \dots, |r|$  gives their increasing  $p \& q \& \dots \& r$ , while the lateral activation for signals  $|p|, |q|, \dots, |r|$  gives their decreasing  $p \vee q \vee \dots \vee r$ . The problem is that there is a smooth transformation between lateral inhibition and lateral activation. Therefore, there is the same smooth transformation between conjunction and disjunction which is difficult to be presented logically. In other words, conjunction as well as disjunction can be more or less biased (activated).

Many optical illusions occur, because our brain is trained to switch between lateral inhibition and lateral activation automatically under different conditions. For example, in

the Müller-Lyer illusion, see Fig.1, we perceive a length of a line depending on whether the line has the inward wings, see Fig.1(a) (when we follow the lateral inhibition effect), or the outward wings, see Fig.1(b) (the lateral activation effect). It is worth noting that T. Sakiyama and Y.-P. Gunji (2013) showed that the Müller-Lyer illusion holds for the group behavior of foraging ants. They explain this phenomenon by that each swarm of ants has the following two main logistic tasks: (i) to build a global route system connecting the nest with food sources to monopolize all reachable food sources (it corresponds to the lateral activation, that is, to the colony's ability to discover new food sources through exploration); (ii) to exploit effectively and efficiently each found food source (it corresponds to the lateral inhibition, i.e. to the colony's ability to concentrate on some food sources).



(a) Line with inward wings.



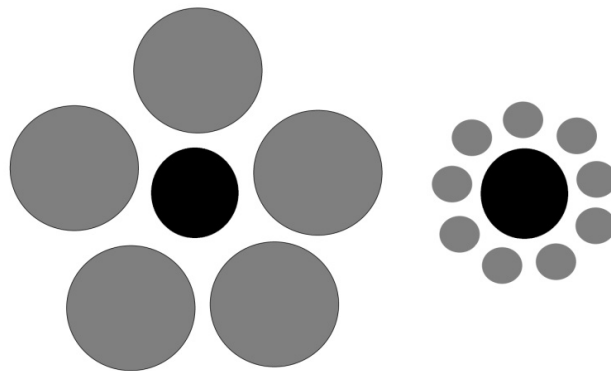
(b) Line with outward wings.



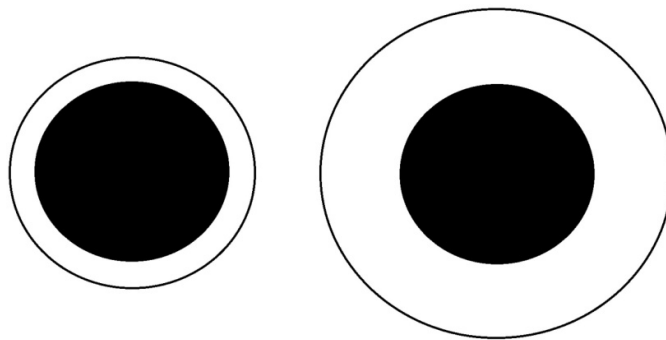
(c) Line without wings.

**Figure 1.** The Müller-Lyer illusion. The lines of (a), (b), (c) are of the same length. Nevertheless, it seems to us that the line with the inward wings (i.e. (a)) is shorter than line (b), and the line with the outward wings (i.e. (b)) is longer than line (a).

In the Ebbinghaus illusion, see Fig.2, as well as in the Delboeuf illusion, see Fig.3, perceiving the disc surrounded by bigger figures is laterally inhibited, therefore this disc seems to be smaller, while perceiving the disc surrounded by smaller figures is laterally activated, therefore this disc seems to be larger. It was shown (Howard et al., 2017) that a kind of Ebbinghaus or Delboeuf illusions can be observed in the behavioral patterns of honeybees. So, when the viewing distance was unrestricted, an appropriate illusion occurred. However, when the viewing distance was restricted, bees did not demonstrate perceiving the Ebbinghaus or Delboeuf illusions.



**Figure 2.** The Ebbinghaus illusion. The two discs in both centers are of the same size, but the right one seems to be larger than the left one because of the surrounded discs of different diameters.



**Figure 3.** The Delboeuf illusion. The two discs in both centers are of identical size, but the left one seems to be larger than the right one because of the surrounded rings of different diameters.

Note that at least one Grey parrot (*Psittacus erithacus*) reports the Müller-Lyer illusion under certain conditions (Pepperberg et al., 2008). The Ebbinghaus and Delboeuf illusions are well traced in perceptions of macaque monkeys and some other species of monkeys (Parrish et al., 2015; Huang et al., 2002). The point is that the perception of brightness in macaques and humans is quite similar because of common underlying neural computations in the two

species. The same illusions under some conditions are well traced in perceptions of red-footed tortoises (*Chelonoidis carbonaria*) and bearded dragons (*Pogona vitticeps*) (Santacà et al., 2019). Nevertheless, they are not so well reported for domestic dogs (Byosiére et al., 2020). Thus, different mechanisms of lateral inhibition and activation may alter perceiving or not perceiving different illusions in a particular species. This data could improve our understanding of animal perception. In any case this perception is always biased. So, a reaction of any organism to stimuli can be more conjunctive (laterally inhibited) or more disjunctive (laterally activated) than it is in fact.

Hence, the negation as a logical operation has its biological meaning presented as appropriate responses to repellent stimuli. The conjunction as a logical operation may be biologically treated as laterally inhibited responses to several stimuli at once. The disjunction as a logical operation may be understood as laterally activated responses to several stimuli simultaneously. This allows us to implement some logic gates within the biological reactions (Whiting et al., 2014, Schumann & Pancerz, 2019). But the problem is that these implementations can be realized only with strict limits and strong approximations. The matter is that the negation, conjunction, and disjunction are some pure ideal forms which are unchangeable, while repelling, lateral inhibition, and lateral activation are instable and exist only in some hybrid forms with being more inhibited or activated or being a mixture of those reactions. And that is why the lateral inhibition and lateral activation are more responsible for biases and cognitive distortions than for purely logical reactions.

## 5. Cognitive Biases of Humans and Animals

Behavioral scenarios of human and animals may be more or less rational – as an attempt to maximize own profits. If they are at least partly irrational (i.e. there are no forecasted maximal profits after carrying out these scenarios), they are called cognitive biases or distortions. There are the same two main reasons of many biases: stress as psychic reaction (analogous to the lateral inhibition with one neural network) and satisfaction (analogous to the lateral activation of neurons). For example, managers usually overestimate the probability of positive forecasting (Heaton, 2002) – they are not focused enough (i.e. not enough laterally inhibited – in the terms of neural networks) and pay more attention to information that supports their opinions. Among cognitive biases of managers there are the so-called framing (when they may frame the risky outcome differently, either considering the worst case or the best case) and confidence (their optimistic or pessimistic assessment of one's ability to perform, e.g. in a period of economic recession, people are less confident and more pessimistic about future economic conditions and, as a consequence, they avoid risky decisions) (Chen et al., 2001). There is also a 'herd behavior', when they copy actions of a

larger group due to the social pressure of conformity, e.g. some investors may follow others to buy the same stocks (Gilovich et al., 2002). The study of different cognitive biases and their influences on making financial decisions is presented in behavioral finance (Goldberg & von Nitzsch, 2001), a new approach in finance studying some real irrational patterns of decision-making. So, managers are not enough logical and their decisions are at least partly always biased – through a stress or satisfaction.

Now, cognitive biases and distortions are studied for animals, too, first of all for animals of zoo and aquaria to check their affective states to know whether their welfare is acceptable or it should be improved (Clegg 2018). It is stated (Spruijt et al., 2001) that a balance of positive and negative affective states is always preferable for managing zoo or aquaria. For instance, cognitive biases are analysed for bottlenose dolphins (*Tursiops truncatus*) (Clegg et al., 2017) and western lowland gorillas (*Gorilla gorilla gorilla*) (Daoudi et al., 2017). It is learnt that under stress conditions animals follow more pessimistic judgements about ambiguous stimuli, pay more attention to negative stimuli, and remember more often negative memories (Clegg 2018; Eysenck et al., 1991; Wright & Bower, 1992). As a consequence, it was detected (Robinson & Roiser, 2016) that, similar to humans, animals in their negative affective states exhibit negative affective biases on decision-making and memory tasks. Their emotional responses to a stimulus can be quick, but at the same time there are affective states which are longer (such as anxiety or depression) and they are not caused by a single stimulus but are the results of their experiences (Robinson & Roiser, 2016). There was developed a special system of judgment bias tasks (Roelofs et al., 2016) to assess emotional states of animals by a cognitive measure of their optimism and/or pessimism by recording behavioral responses to ambiguous stimuli.

Hence, cognitive biases are normal mechanisms of thinking not only of humans, but also of animals. People, like animals, do not think purely logically, which means that their logical competence, by definition, is far from perfect. At the same time, there are two main affects in decision-making (i.e. in building an inference): stress and pleasure. At the micro level, both affects are expressed by two mechanisms in the perception of signals: lateral inhibition (analogue of stress) and lateral activation (analogue of satisfaction).

Different forms of cognitive distortions in our reasoning are called *logical fallacies*. They are not directly related to stress and satisfaction as two essential moods of our emotions. Among them there are ‘motivational biases’ which reflect a desire, involved in inferences, to achieve a result in discussions. Usually, these biases are committed unintentionally (Correia, 2011). The matter is that “people are more likely to arrive at conclusions that they want to arrive at, but their ability to do so is constrained by their ability to construct seemingly reasonable justifications for these conclusions” (Kunda 1990, p. 480). In other words, people

tend to seem rational – based on some arguments, but they do not know logical norms well. For example, the alcoholic can justify his or her excessive drinking by affirming: “I am not an alcoholic. I drink because...” (Correia, 2011, p. 119). Nevertheless, his or her further reasoning thereby will be imperfect logically. Also, there are as the straw man fallacy presenting a strategy whereby the proponent misrepresents the some intentional fallacies such opponent’s position in a way that makes it easier to be refuted, and then refutes this made up position instead of the real one (Talisce & Aikin, 2006; Johnson & Blair, 1983; Walton, 1989). In any case each logical fallacy has an appropriate belief biasing the reasoning and an appropriate emotional distortion in accepting logically correct inferences (Cohen, 1992). Furthermore, logical fallacies are rooted in real discussion so that purely logical forms of inferring do not seem to be a real practice in the every-day argumentation (von Eemeren, 1986, Walton 1989).

## 6. Logical Competence as a kind of Social Practice

Within the framework of unconventional computing, the group or swarm behavior of animals is also being explored for their logical competence, namely, the logical competence is discovered in the behavior of so-called ‘social’ animals (such as bees, ants or horses), which coordinate their behavior among themselves, distributing specific roles from a leader or queen to workers. As a result of these studies, a special direction of unconventional computing arose, called *swarm computing* (or *swarm intelligence*) – in it the logical competence is subjected to a logical and mathematical analysis in the form as it is presented in a particular swarm (for example, in an anthill, a nest of bees, a herd of horses, a flock of birds, a school of fish, etc.), see (Bonabeau et al., 1999; Schumann, 2020). Swarms were found to be capable of solving complex logistic problems such as the traveling salesman problem (Dorigo, 1997) or the generalized assignment problem (Ozbakir et al., 2010).

But the swarm behavior is not social in the strict sense of the word – there are no pure conventions consciously shared among the members of the group. From the standpoint of behaviorism, any collective behavior is a complex system that is controlled by stimulating individuals. And in this regard, it would seem that the ‘social’ behavior of animals is just a simpler stimulation system than the behavior of humans. However, some inevitable cognitive distortions are also recorded in the group behavior of swarm animals, which shows the strict boundaries of behaviorism, its inapplicability in modeling complex forms of group behavior, in which noticeable cognitive distortions are simply inevitable, and, as a consequence, the same combination of stimuli can give different reactions, which contradicts the basic assumptions of behaviorism (Shirakawa et al., 2020; Schumann, 2019).

In the theory of symbolic interactionism, the creators of which are George Herbert Mead (1863 – 1931) and Herbert Blumer (1900 – 1987), the social behavior is analyzed initially in the light of possible cognitive biases and distortions that are responsible for attributing a special symbolic meaning to some objects when these items get an added value. The founders of this theory noticed that people act in relation to things based on symbolic, and not direct, meanings of these things. In other words, some cognitive biases involved in the cognition of people are fixed in symbolic meanings shared by others, which, thereby, provide social interactions of people, i.e. interactions beyond the biological behavior. The social behavior is built, therefore, on symbolic interactions, which are simply impossible in the swarm behavior due to the absence of the symbolic meanings as such. It turns out that the social behavior is impossible without a material culture – without some objects that have acquired an additional symbolic meaning, consciously shared by this group of people.

At the same time, in order to stabilize societies, there is a need for a long-term consolidation of symbolic meanings and other conditions of symbolic interactions. This is how social institutions are born with legal support as long-term forms of social contact. As a result, the logical competence acquires a new meaning – it is a form of intersubjective decision-making without cognitive distortions from some existing conventions (social norms, provisions, assumptions), which, in turn, are built on symbolic meanings (cognitive biases involved in the formatting of social space).

Thus, the logical competence of people can be understood in different ways: (1) as an individual mechanism of cognition in combination with cognitive distortions; (2) as a swarm cognition combined with cognitive distortions; (3) as a social cognitive mechanism that can be cleared of natural cognitive biases. This research analyzes the origins and early genesis of logical competence in the third sense.

So, the logical competence of a person as a certain practice of thinking, devoid of logical fallacies, is not an innate ability, since in its natural form it is not separable from cognitive distortions, but is a special social practice of making logically correct decisions, which once arose and was subsequently developed and spread through a direct learning.

The logical competence without distortions is implementable within *artificial intelligence* that is presented as some useful techniques of human cognitions which are free of cognitive biases in advance. Artificial intelligence now is the study of designing expert systems, which help us to analyze the environments and produce actions which maximize success according to some algorithms. The idea of some machines with drawing conclusions in an automatic way which support us in decision-making and finding consensus is quite old and has been popular since Raymundus Lullius (ca. 1235 – 1315) and Gottfried Wilhelm von Leibniz (1646 – 1716). For instance, Leibniz claimed that his analytic language of reasoning

based on arithmetic operations (*characteristica universalis*) offers the unbiased forms of conclusions and, therefore, it can be used for converting the Chinese empire to Christianity, because the *characteristica universalis* is the best way of finding any consensus including the religious one (Swetz, 2003).

Traditionally, Aristotle is considered the father of logic as a science of unbiased inferences. So, we can suppose that only after him we can trace back some samples of perfect logical competence. But it is not true. There is one tradition that is much older than the Aristotlian logic – it is a Sumerian-Akkadian legal hermeneutics established at least since Ur III (Culbertson, 2009; Schumann 2019, 2020), i.e. about four thousand years ago. Within this hermeneutics, laws were presented as law codes, where each article from the code was formulated as a conditional statement (implication): “If an act of  $p$  is conducted, then its actor is sentenced to  $q$ ”. As a consequence, each court should make a decision in accordance with *modus ponens*:

*Law article from a code:* “If  $p$ , then  $q$ ”;

*Fact-checking at the trial:* “There is a fact  $p$ ”;

-----

*Court decision:* “There is a sentence  $q$ ”.

There are excavated many legal documents written in Sumerian and different dialects of Akkadian: law codes, trial records, contracts, testaments, etc. And their authors possess such a high logical competence with demonstrating the good knowledge of various inference rules: *modus ponens*, *modus tollens*, introducing conjunction, introducing disjunction, etc. (Schumann, 2019, 2020). Furthermore, in these documents there are no logical fallacies and we can explicate ten or even more logical steps in inferring. This tradition is unique and has no analogies in other ancient cultures.

Hence, the Sumerian-Akkadian legal hermeneutics has its origin in Mesopotamia and has been expanded later through different empires: Neo-Assyrian, Neo-Babylonian, and Achaemenid, then through Hellenistic states, the Roman Empire, and some partly Hellenized states of Central and South Asia participating in trading networks of the Silk Road, such as the Kushan Empire. Laws from state to state can have been different, but the legal hermeneutics, at the same time, was very similar, i.e. the legal matters were changeable, but the logical forms in making legal decisions on the basis of laws were constant. First this hermeneutics was mainly in Akkadian, then in Aramaic and Greek, and only then in some other official languages of the Silk Road such as Bactrian or Sogdian.

Let us show that marriage contracts composed in Akkadian, Aramaic, Greek, and Bactrian have had the same structure since the Old Babylonian period and, accordingly,

they have belonged to the same joint legal hermeneutics indeed. Furthermore, it is known that Aramaic legal documents were a continuation of legal tradition in Akkadian. Then Greek became the *lingua franca* of diplomacy and international trade with many clichés borrowed from the Babylonian jurisprudence.

Marriage contracts were needed for obtaining a legal status of wife and husband in an appropriate society according to the Old Babylonian law codes. For example, in the Code of Eshnunna, §27-28, we read:

[§27] *šum-ma lú dumu. SAL l[ú b]a-lum ša-al a-bi-ša ù um-mi-ša i-ḫu-si-ma [ù]<sup>2</sup> [k] ir-ra-am ù ri-ik-⟨sa⟩-tim a-na a-bi-ša ù um-mi-ša la i[š-ku-u]n u<sub>4</sub>-mi mu.l.kam i-na é-šu li-ši-im-ma ú-ul dam*

[§28] *šum-ma u<sup>3</sup> ri-ik-sa-tim ù kir-ra-am a-na a-bi-ša ù um-mi-ša iš-ku-un-ma i-ḫu-us-si dam u<sub>4</sub>-um<sup>4</sup> i-na su-un lú [iṣ-ša-ab]-ba-tu i-ma-at ú-ul i-ba-al-lu-[u<sup>7</sup>]*

§27 If a man took another man's daughter without asking her father and mother and did not arrange for (lit. establish) a libation and marriage contract with her father and mother, though she live in his house for a year, she is not a wife.

§28 If ... he arranged for a marriage contract and libation with her father and mother and took her, she is a wife; the day she is caught with (another) man she shall die; she shall not live (Greengus, 1969, p. 505).

More formally:

§27 If  $r$  (a man and a woman live together) and not  $p$  (there is no marriage contract), then not  $q$  (she is not a wife).

§28 If  $p$  (there is a marriage contract), then  $q$  (she is a wife). If  $q$  (she is a wife) and  $s$  (she is caught with another man), then  $t$  (she gets death penalty).

The Code of Hammurabi, §128, has almost the same article as §27:

[§128] *šum-ma a-wi-lum aš-ša-tam i-ḫu-uz-ma ri-ik-sa-ti-ša la iš-ku-un SAL ši-i ú-ul aš-ša-at*

§128 If a man took a wife and did not arrange for her marriage contract, that woman is not a wife (Greengus, 1969, p. 505).

Let us assume that there are two facts: not  $r$  and  $p$ , checked at the trial, i.e. a man and a woman do not live together, but they have a marriage contract. Then, according to §28 and *modus ponens*, the court logically infers that  $q$ , i.e. she is a wife. If we have  $r$  and  $p$ , the court again infers  $q$  due to §28. Hence, the articles are formulated in the way that allows us

to apply inference rules of propositional logic to entail legal decisions automatically – for example, to prove that there is a lawful marriage in fact.

Something is treated as a fact in the Sumero-Akkadian hermeneutics only after checking at the trial. To be a fact any state of affairs should be either fixed in a well-composed legal document (past trial decision, contract, will, etc.) or testified by at least two witnesses. In legal documents one main thesis as one declarative or a series of declarations (e.g. that the woman is a wife) should be well formulated. Then the declarations are supported by several witnesses who sign (i.e. put their seals on) an appropriate document containing this main thesis. Each additional declarative of the legal document is formulated in the form of implication: “If a state of affair  $p$  takes place, then a legal decision  $q$  shall be made” so that the next court can use this conditional declarative for making the next decision  $q$  just by applying *modus ponens* after checking the fact of  $p$ .

Let us quote the marriage contract composed at the time of Erra-Imitti, the ninth king of the first dynasty of Isin who ruled in ca. 1868 – 1861 B.C.:

**Obv.**

1 <sup>m</sup>reš-tum dumu-munus ur-<sup>d</sup>utu  
<sup>md</sup>šu-i-lí-šu  
 dumu <sup>m</sup>ēr-ra-mu-ba-lí-iṭ  
 nam-dam-šè in-ni-tuku  
 5 mu-lugal-ka-ni in-pàd  
 túg-íḅ-lá nu-mu-un-zu-[a]-ni\in-du<sub>8</sub>  
 é a-ša ù níḡ-gur<sub>11</sub>-ra a-na [me-a-bi]  
 lú-lú-gin<sub>7</sub> al-dím  
 tukum-bi u<sub>4</sub>-kúr-šè  
 10 <sup>m</sup>reš-tum <sup>md</sup>šu-i-lí-šu dam-a-ni-éš  
 dam-ḡu<sub>10</sub> nu-me-en ba-na-an-du<sub>11</sub>  
 ½ ma-na kù-babbar ì-lá-e  
 ù tukum-bi <sup>md</sup>šu-i-lí-šu  
<sup>m</sup>reš-tum dam-a-ni-éš

**Rev.**

15 dam-ḡu<sub>10</sub> nu-me-en ba-⟨na⟩-an-du<sub>11</sub>  
 ½ ma-na kù-babbar ì-lá-e  
 mu-lugal-bi in-pàd-dè-eš  
 [igi] <sup>m</sup>na-bi-<sup>d</sup>EN.ZU dumu <sup>m</sup>na-bi-ì-lí-šu  
 igi [<sup>m</sup>a]-ta-na-aḥ dumu <sup>m</sup>na-bi-ì-lí-šu

20 *igi a-du-ba* *dumu èr<sup>d</sup>-en-líl-lá*  
*igi* <sup>md</sup>EN.ZU-*ki-ma-ì-lí-a*  
*dumu* <sup>md</sup>EN.ZU-*mu-ba-lí-ìt*  
*igi* <sup>m</sup>*a-ad-da-wa-qar* [*dumu xx*]-*lí-wa-qar*  
*igi* <sup>m</sup>*àm-ma-nu-um* [*lú*]-*nar*  
 25 *igi* <sup>md</sup>EN.ZU-*i-din-na-šu* DUMU *la-bi-šu*  
*iti* *giš-apin-du<sub>8</sub>-a* *ud-15-kam*  
*mu* <sup>d</sup>*èr-ra-i-mi-ti* *lugal*  
 4 *urudu ur-maḥ-gal-gal*  
*é-AN-nir-ra mu-na-dím*  
**Seal impression:** <sup>md</sup>*šu-i-lí-šu*

**Obv.**

1 *Rēštum* daughter of *Ur-Utu*  
*Šū-ilišu*  
 son of *Erra-muballit*.  
 took her for wife  
 5 he swore in the name of the king  
 he has loosened her ‘belt of virginity’  
 the house, the field and the possessions,  
 as much as [they may be]  
 they shared like equals.

**If** in future,

10 *Rēštum* to *Šū-ilišu*, her husband  
 says: you are not my husband  
 she will weigh out ½ mina of silver  
**and if** *Šū-ilišu*  
 to his wife

**Rev.**

15 says: you are not my wife  
 he will weigh out ½ mina of silver  
 They swore in the name of the king.  
 Witnesses:

**Nabi-Sîn** son of **Nabi-ilišu**

**Atanaḥ**son of **Nabi-ilišu**

**Aduba** son of **Warad-Enlila**

**Sîn-kîma-iliya son of Sîn-muballit**

**Adda-waqar son of ... -li-waqar**

**Ammanum [the] singer**

**Sîn-iddinaššu son of Labišu**

Fifteenth day of November

year king Erra-imitti

4 large copper lions

he fashioned on the Ziggurat.

Seal impression: Šū-ilīšu

(Al-Shamari & Al-Jalili, 2020).

This legal document is signed by several witnesses who thereby support the following two declarations: (1) Rēštum is a wife of Šū-ilīšu; (2) the property is owned by both spouses equally (50/50). Then there are two additional declarations formulated as conditionals: (1) **if** (tukum-bi) Rēštum declares that Šū-ilīšu is not her husband, **then** she pays ½ mina of silver; (2) **if** (tukum-bi) Šū-ilīšu declares that Rēštum is not his wife, **then** she pays ½ mina of silver.

Marriage contracts in Aramaic have almost the same logical structure as those documents in Akkadian (Yaron, 1958), which is not surprising taking into account the fact that it is the same Babylonian legal tradition. Aramaic began to supplant Akkadian as a new legal language from the 6<sup>th</sup> century B.C.

Many legal documents in Aramaic were excavated in Elephantine (an island on the Nile). At the same place there were excavated legal documents in Greek also, but dated later than the Aramaic ones. It shows that Greek started to supplant Aramaic as a new legal language from the 4<sup>th</sup> century B.C.

Let us consider the marriage contract between Heraclides and Demetria, composed in 311/310 B.C. – at the time of Alexander, the satrape of Ptolemy:

1 Ἀλεξάνδρου τοῦ Ἀλεξάνδρου βασιλεύοντος ἔτει ἐβδόμῳ, Πτολεμαίου  
σατραπεύοντος ἔτει τεσσαρε-

σκαίδεκάτῳ(\*) μηνὸς Δίου. συγγραφὴ συνοικισίας Ἡρακλείδου καὶ Δημητρίας.  
λαμβάνει Ἡρακλείδης(\*)

Δημητρίαν Κώϊαν γυναῖκα γνησίαν παρὰ τοῦ πατρὸς Λεπτίνου Κώϊου καὶ τῆς μητρὸς  
Φιλωτίδος ἐλεύθερος

ἐλευθέραν προσφερομένην εἰματισμόν(\*) καὶ κόσμον (δραχμῶν) Α, παρεχέτω δὲ  
Ἡρακλείδης Δημητρίαι

5 ὅσα προσήκει γυναικὶ ἐλευθέραι πάντα, εἶναι δὲ ἡμᾶς κατὰ ταυτὸ(\*) ὅπου ἂν δοκῇ  
ἄριστον εἶναι βουλευομένοις κοινῇ

βουλῇ Λεπτίνῃ καὶ Ἡρακλείδῃ(\*). εἰάν(\*) δέ τι κακοτεχνούσα ἀλίσκηται [[ἀλίσκηται]]  
ἐπὶ αἰσχύνῃ τοῦ ἀνδρὸς Ἡρακλείδου Δημητρία,  
στερέσθω ὦμ(\*) προσηνέγκατο πάντων. ἐπιδειξάτω δὲ Ἡρακλείδης ὅτι ἂν ἐγκαλῇ  
Δημητρία ἔναντιον ἀνδρῶν τριῶν ,  
οὓς ἂν δοκιμάζωσιν ἀμφοτέροι. μὴ ἐξέστω δὲ Ἡρακλείδῃ γυναῖκα ἄλλην ἐπεισάγεσθαι  
ἐφ' ὅβρει Δημητρίας μηδὲ  
τεκνοποιεῖσθαι ἐξ ἄλλης γυναικὸς μηδὲ κακοτεχνεῖν μηδὲν παρ ευρέσει μηδεμιᾷ  
Ἡρακλείδῃ ἐν Δημητρίαν.

**10** εἰάν(\*) δέ τι ποῶν(\*) τούτων ἀλίσκηται Ἡρακλείδης καὶ ἐπιδείξῃ Δημητρία  
ἔναντιον ἀνδρῶν τριῶν , οὓς ἂν δοκιμάζωσιν(\*)  
ἀμφοτέροι, ἀποδότω Ἡρακλείδης Δημητρία τῇμ(\*) φερνὴν ἣν προσηνέγκατο  
(δραχμῶν) Α, καὶ προσαποτεισάτω ἀργυρί-  
ου Ἀλεξανδρείου (δραχμᾶς) Α. ἢ δὲ πράξις ἔστω καθάπερ ἐγ(\*) δίκης κατὰ νόμον  
τέλος ἐχούσης Δημητρία καὶ τοῖς μετὰ  
Δημητρίας πράσσουσιν ἐκ τε αὐτοῦ Ἡρακλείδου καὶ τῶν(\*) Ἡρακλείδου πάντων καὶ  
ἐγγαίων καὶ ναυτικῶν. ἢ δὲ συγγραφὴ  
ἦδε κυρία ἔστω πάντῃ πάντως ὡς ἐκεῖ τοῦ συναλλάγματος γεγεννημένου, ὅπου ἂν  
ἐπεγφέρῃ Ἡρακλείδης κατὰ

**15** Δημητρίας(\*) ἢ Δημητρία τε καὶ τοῖ(\*) μετὰ Δημητρίας πράσσοντες ἐπεγφέρωσιν(\*)  
κατὰ Ἡρακλείδου. κύριοι δὲ ἔστωσαν(\*) Ἡρακλεί-  
δης καὶ Δημητρία(\*) καὶ τὰς συγγραφὰς αὐτοὶ τὰς αὐτῶν φυλάσσοντες καὶ  
ἐπεγφέροντες κατ' ἄλληλως(\*). μαρτερες(\*)

Κλέων Γελῶιος, Ἀντικράτης Τημνίτης, Λῦσις Τημνίτης, Διονύσιος Τημνίτης,  
Ἀριστόμαχος Κυρηναῖος, Ἀριστόδικος  
Κῳιος.

(seal) (seal) (seal) (seal)

In the 7th year of the reign of Alexander son of Alexander, the 14th year of the satrapship of Ptolemy, in the month Dios. Marriage contract of Heraclides and Demetria. Heraclides takes as his lawful wife Demetria, Coan, both being freeborn, from her father Leptines, Coan, and her mother Philotis, bringing clothing and ornaments to the value of 1000 drachmae, and Heraclides shall supply to Demetria all that is proper for a freeborn wife, and we shall live together wherever it seems best to Leptines and Heraclides consulting in common.

If Demetria is discovered doing any evil to the shame of her husband Heraclides, she shall be deprived of all that she brought, but Heraclides shall prove whatever he alleges against Demetria before three men whom they both accept.

It **shall not be lawful** for Heraclides to bring home another wife in insult of Demetria nor to have children by another woman nor to do any evil against Demetria on any pretext. If Heraclides is discovered doing any of these things and Demetria proves it

before three men whom they both accept, Heraclides shall give back to Demetria the dowry of 1000 drachmae which she brought and shall moreover forfeit 1000 drachmae of the silver coinage of Alexander. Demetria and those aiding Demetria to exact payment shall have the right of execution, **as if** derived from a legally decided action, upon the person of Heraclides and upon all the property of Heraclides both on land and on water. This contract shall be valid in every respect, wherever Heraclides may produce it against Demetria, or Demetria and those aiding Demetria to exact payment may produce it against Heraclides, **as if** the agreement had been made in that place.

Heraclides and Demetria shall have the right to keep the contracts severally in their own custody and to produce them against each other.

Witnesses **Cleon, Gelan; Anticrates, Temnian; Lysis, Temnian; Dionysius, Temnian; Aristomachus, Cyrenaeon; Aristodicus, Coan.**

(Hunt & Edgar, 1932, p. 2–3).

We see that several witnesses signed their confirmation that (1) Demetria is a wife of Heraclides, (2) Demetria has a dowry with its description, (3) the property obtained after the marriage is owned now by both spouses equally (50/50). The two additional declarations formulated as implications: (1) **if** Demetria is caught with another man **and** this fact is testified by at least three witnesses, **then** she shall be deprived of the dowry; (2) **if** Heraclides brings home another woman with the testimony of three witnesses **or** he has children by another woman with the testimony of three witnesses **or** he is caught with another woman with the testimony of three witnesses, **then** he shall give back to Demetria her dowry and shall moreover pay 1000 drachmae. To infer from the second implication, we should apply the inference rule of deleting disjunction and only then *modus ponens*.

Legal documents in Bactrian (one of the Eastern Middle Iranian languages) were a continuation of legal tradition in Aramaic and Greek. Furthermore, Bactrian was being written in the Greek script and it supplanted Greek as a new official language in ca. 127 A.D. in today's Afghanistan and Pakistan, according to the edict issued by Kanishka I, the emperor of Kushans (Rabatak inscription).

Let us quote one of the Bactrian marriage contracts:

#### Document A

A1 χῆρονο ρ' ι Αὐρηζνο μαο σαχτο Αβαμοχοινο ρωσο καλδο νοβιχτο μο ολοβω-

A2 στογο μαλο αβο μο ρωβαγγο ἥαρο αβιοανδαγο σιδο στηβο ριζδο αζδηβδο  
ειμουανο αστ-

A3 ηβιγανο αζαβοργανο οιγαλφανο μολροδηβγανο κιδο ι ωνδο λιβο ζανινδο ταδο  
αβο μασκο

A4 πιδοναμγο νοβιχτιγινδο οδο κιδο ι ωνδο λιβο ναζανινδο ταδοιηνο αβο μο νιγο  
λιστοβαρο κι-

A5 ρδο πισο οινδοωρομοζδο κολαγανο ιανδαγοβιδο οδο πισο χοαδηολαδο  
ιαστονικανο οδο πι-

A6 σο οαραζωωρομοζδο χοασραογανο οδο πισο αβδαβοκο παβογανο οδο πισο  
ασπαλομιρο

A7 ιωλικανο οδο πισο βαγοβαο ιαβιτταδο οδο πισο μανδαρουανισο στηβιγανο  
αζαβοργανο

A8 οιγαλφανο μολροδηβγανο πιδοοασατο χοακαμο χοασινδο βαγοφαρνο ζαμωδο  
πορο

A9 οδο βαβο οδο πιδοκο βαγοφαρνο πορανανο ζαουοζαδγανο κιδο ωσο νοβανδο  
νινδοκο ι ω-

A10 χροβαδογανο πιδο αζαδο ασπασο ποροχηζαμο μισιδο ζιττο μανο βαγοφαρνο  
ασαφαγο

A11 φαροσηρο οδασαφαγο νωγοσανινδο ιαλωγδο ασωνωυο ειο ζινο κιδο ραλικο  
ναμο φαρι-

A12 μουανο πορανανο ολο φαρο βαβο οδο φαρο πιδοκο ιθαυατανο ληρηιο αζο  
βαγοφαρνο

A13 αβο μασκο νοβιχτισο ζινο κιδο ραλικο ναμο ασνωυο ασνωυογγγο αβο οισπο  
αστο σιδωμ-

A14 ηνο ωσο αστο οδο σιδανο ασο μαβρισο αλφανζαμηιο ιθασαγωνδο ασνωυο  
πορδο αβο

A15 λριγδο σαγωνδο αβο μο λαμο νοβανδανο οτο πιδοοασατο βαβο οδο πιδοκο  
αγγιτιδο αμαχ-

A16 ο μανο βαβο οδο πιδοκο αβο ραλικο ολο ολογωγγο αβο οισπο αστο σιδωμνηνο  
ωσο αστ-

A17 ο οδο σιδανο ασο μαβρισο αλφανζαμηιο φινζο φρομανζο ιθασαγωνδο αβο μο  
λαμο νοβα-

A18 νδανο οτανο μαπαδοαρχανηιο αζο βαβο οδαζο πιδοκο ανδαρο ολο αβο κιδο  
οδο μαζαδο πιδο-

A19 ρωφσο αβο λριγδο σιδανο ραλικο μαυοσινδηιο οδο καλδανο αζο βαβο αλαζο  
πιδοκο ανδαρο

A20 ολο κρινδηιο αλο αζαδο πιδορωφσο ληρινδηιο σιδανο ραλικο μαυοσινδηιο  
ταδανο

A21 λαυινδηιο ταοανο αβο ηραογανο γαζνο ζαροζιδγο κ· διναρο οδο δαφρηλο φαρο  
παδαρλο-

A22 γο [ο]το πιδοοασατο νινδοκο μοζδαοανινδο πορο οδο ιαμφοβανδαγο οδο παπο  
οδο ιατο

A23 νινδοκο πορανανο κιδο καδγο ωχροβαδογανο ριζδο κοοαδο αμαχο αβο ραλικο  
χοαδο ζ-

A24 ιττο οτο πιδιδο ιειρο υοσινδαμο οτανο μαπαδοαρχανινδηιο αζο νινδοκο οδαζο  
ιαμφο-

A25 βανδαγο οδαζο παπο οδαζο ιατο φαρo ραλικο ιαοαργο οδο ταραζο αβο λαδο  
οδο μασιδανο α-

A26 σo ραλικο αζιηιο οδο σιδανο ασο ραλικο πορο αζιηιο ταδανο πιδο  
χοβο χοαρηιο

A27 οτο πιδο υαργο κρηιο πιδο αζαδο ασπασο σαγωνδο νιαγο οδο πιδο  
κιρδο οδο σιδανο

A28 λογδο αζιηιο τακροαδανο πιδο οδο ιμαδο οδο γωυριγο υοσινδηιο  
αβηγο αβο λαδο ταδ[ο]

A29 ιανο λανινδηιο οδο καλδανο αζο νινδοκο αλαζο ιαμφοβανδαγο αλαζο  
παπο αλαζο ιατο οιζα-

A30 γγ[οδ]αρο κρηιο αλδανο φαρo ραλικο ιαοαργο αλο ταραζο λανινδηιο  
οδαλδο ιθαχοανινδ-

A31 ηιο κοοαδομηνο αβο ραλικο ζαδικανο βανζο αλο μαρηγο βουδο βηγο  
πιδο αζαδο ασπασο

A32 σαγωνδο νιαγο οδο πιδο ταδο λανινδηιο ταοανο αβο βρογανο γαζνο  
ζ[α]ροζιδγο κ' διναρο

A33 οδο δαφρηλο φαρo παδαρλογο οτομηνο ιθαησο χοανδο οδo  
οισαρo[ναοαο]ορο αβουδηιο

A34 ναμαγγo κωσoβο α' βαρζο α' οανοστoσιγο α' ραγαζο ' [...]δομιγγο α'  
λι-

A35 στoβανδιγο δ' βκομινo β' καφρινo ιογδιγο γ' ποσο β' [...]γανδο]μο μιδο  
γ'

## Document A

(It was) the year 110, the month Ahrezhn, the day Abamukhwini having elapsed, when this marriage contract was written <sup>2</sup>here in the borough of the city of Rob which is called Steb, with the cognizance of these <sup>3</sup>freemen of Asteb, the witnesses (and) those who have impressed (their) seals (upon it), who witness the present document and <sup>4</sup>(whose) signatures are written hereupon – and those who do not witness the present document have put (their signatures) on the other manuscript – <sup>5</sup>(namely) in the presence of **Wind-ormuzd Kulagan** the chief of the district, and in the presence of **Khwadew-lad Yastunikan**, and in the presence of **Waraz-ormuzd Khwas-rawgan**, and in the presence of **Abdabuk Pabugan**, and in the presence of **Aspal-mir** <sup>7</sup>**Yolikan**, and in the presence of **Bag-bandag** the master craftsman, and also in the presence of the other freemen of Steb, <sup>8</sup>the witnesses (and) those who have impressed (their) seals (upon the documents). (Then this) declaration was freely (and) willingly made (by me), Bag-farn, son of Zamod, <sup>9</sup>and (by us), Bab and Piduk the legitimate sons of Bag-farn, who now serve with Ninduk <sup>10</sup>Okhsh-badugan in free service.

Now: there has been requested by me, Bag-farn, from you, <sup>11</sup>Far-wesh, and from you, Nog-sanind, this woman whose name (is) Ralik (as) a fully privileged daughter-in-law, <sup>12</sup>(as) a wife for these (my) sons, Bab and Piduk, so that I, Bag-farn, **may treat** <sup>13</sup>the woman described herein, whose name (is) Ralik, (as) a daughter-in-law (to be treated)

like a daughter-in-law, in every home which we <sup>14</sup>now have and also which we may acquire in the future, as (one) ought to treat a daughter-in-law, <sup>15</sup>as (is) the established custom in the land.

And a declaration was made (by us), Bab and Piduk. We have received <sup>16</sup>Ralik – I, Bab, and (I), Piduk – (as) a wife (**to be treated**) like a wife in every home which we now have <sup>17</sup>and also which we may acquire in the future, (as) a lady possessing authority, as (is) the established custom in the land. <sup>18</sup>And I, Bab, and I, Piduk, **shall not have the right** to make another (woman our) wife, **nor** <sup>19</sup>to keep a free (woman as a) concubine, to whom Ralik should not agree; and **if** I, Bab, or I, Piduk,

<sup>20</sup>should make another (woman our) wife, **or** keep a free (woman as a) concubine, to whom Ralik should not agree, **then** <sup>21</sup>(we) shall give a fine to the royal treasury of twenty dinars of struck gold and the same to the opposite party.

<sup>22</sup>And a declaration was made (by us), Ninduk son of Muzda-wanind, and Yamsh-bandag and Pap and Yat, <sup>23</sup>the sons of Ninduk, whose house is called Okhshbadugan, that we ourselves have requested Ralik, <sup>24</sup>and in this matter we are in agreement, and (we) **shall not have the right** – I, Ninduk, and I, Yamsh-<sup>25</sup>bandag, and I, Pap, and I, Yat – to assign duties and tasks to Ralik, **nor** (to) whatever (child) <sup>26</sup>may be born from Ralik. And whatever son may be born from Ralik, then she may have (him) for her own, <sup>27</sup>and may put (him) for hire in free service as (his) grandfather and father did; and whatever <sup>28</sup>daughter may be born, then in as much as the father and the mother and the family may agree to give (her) away, then <sup>29</sup>(we) shall grant (their) wish. And **if** I, Ninduk, or I, Yamsh-bandag, or I, Pap, or I, Yat, <sup>30</sup>should do otherwise, **or** (if we) should assign duties or tasks to Ralik, **or** (if we) should claim thus, <sup>31</sup>with reference to Ralik's progeny, that (the child) is our slave-girl or slave rather than in free service <sup>32</sup>like (its) grandfather and father, **then** (we) shall give a fine to the royal treasury of twenty dinars of struck gold <sup>33</sup>and the same to the opposite party, and so too our claim and argument shall be [inva]lid.

<sup>34</sup>The dowry (consists of) one blanket, one pillow, one ..., ... cloak(s), ... one ..., <sup>35</sup>four bracelets, two ..., three pairs of shoes, two sheep, ... three measures of [wheat].

<sup>36</sup>...

(Sims-Williams, 2000, pp. 32–35).

In this document, several witnesses signed the declaration that (1) Ralik is a wife of two brothers, Bab and Piduk; (2) accordingly, she became a daughter-in-law of Bag-farn, the father of these brothers; (3) she has a dowry with an appropriate description. The two additional declarations formulated as conditional statements: (1) **if** (καλδανο) Bab **or** (αλαζο) Piduk make another woman their wife **or** (αλαζο) they keep a free woman as their concubine, to whom Ralik should not agree, **then** they shall give a penalty (of twenty dinars of struck gold) to Ralik **and** (οδο) to the royal treasury; (2) **if** (καλδανο) Ninduk **or** (αλαζο) Yamsh-

bandag **or** (αλαζο) Pap **or** (αλαζο) Yat assign duties or tasks to Ralik **or** (αλαζο) to the child who may be born from Ralik, **then** they shall give a penalty (of twenty dinars of struck gold) to Ralik **and** (οδο) to the royal treasury.

Thus, we see that marriage contracts in Akkadian, Aramaic, Greek, and Bactrian were composed within one and the same legal hermeneutics, existing for more than two thousand years at the time of the Bactrian contracts and well traced retrospectively. In marriage contracts there were one main declarative and a series of additional declarations formulated as implications and all these declaratives were confirmed by several seals of witnesses. In the trial records (transcripts) of Sumerian-Akkadian legal tradition continuing in Hellenistic and partly Hellenized states of Central and South Asia, we find already explicit and logically correct applications of different logical inference rules including *modus ponens*, *modus tollens*, and many others (Schumann, 2019, 2020). Meanwhile, these trial records are free from rhetoric tropes and logical fallacies and contain only a purely logical argumentation based on the usage of propositional logic.

## 7. Conclusions

So, the logical competence consists of the methods of implementing logical knowledge that are correct from the standpoint of logical norms and it is not our natural (innate) ability. These methods have different historical forms and can be found in the texts of some early states. Moreover, the earliest examples of perfect logical competence are found only in the documents of Mesopotamia, starting with legal documents and business correspondence of the Akkadian period (since the 24<sup>th</sup> century B.C.). By the time of the appearance of Aristotle's treatises on logic in the 4<sup>th</sup> century B.C. there was already a long history of the perfect logical competence within the Mesopotamian legal hermeneutics expanded later through Hellenistic states. The analysis of various forms of manifestation of this competence allows us to find out the nature of logical knowledge in its practical applications.

## References

- Adamatzky, A. (ed.), (2017). *Advances in Unconventional Computing*. Springer International Publishing.
- Adamatzky, A. (2007). Physarum machine: implementation of a Kolmogorov-Uspensky machine on a biological substrate. *Parallel Processing Letters*, 17(4), 455-467.
- Adamatzky, A., Mayne, R. (2015). Actin automata: Phenomenology and localizations. *International Journal of Bifurcation and Chaos*, 25(2).
- <https://doi.org/10.1142/S0218127415500303>
- Adamatzky, A., Schubert, T. (2014). Slime mold microfluidic logical gates. *Materials Today*, 17(2), 86-91.
- Al'churron, K.E., Bulygin, Ye.V. (2013). "Normativnyye sistemy" i drugiye raboty po filosofii prava i logike norm [«Нормативные системы» и другие работы по философии права и логике норм].
- Alonso-Sanz, R., Adamatzky, A. (2016). Actin automata with memory. *International Journal of Bifurcation and Chaos*, 26(1).
- Al-Shamari, M. Kh. J., Al-Jalili, M. (2020). Two Old Babylonian Marriage Contracts from Isin. *Iraq*, 82, 125-137.
- Arnould, A., Nicole, P. (1662/1992). *La logique ou l'Art de penser*. Gallimard.
- Baker, G.S. (2020). *Critical Thinking. A profound handbook to improve your memory and logical skills, level up your decision making and problem solving*. Sannainvest Ltd.
- Bonabeau, E., Dorigo, M., Theraulaz, G. (1999). *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press.
- Byosiére, S.-E., Chouinard, P. A., Howell, T. J., Bennett, P. C. (2020). Illusion susceptibility in domestic dogs. *Ethology*, 126(10), 949-965.
- Chelpanov, G. (1917). *Uchebnik logiki* [Учебник логики] (9<sup>th</sup> edition).
- Chen, J., Hong, H., Stein, J. C. (2001). Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices. *Journal of Financial Economics*, 61(3), 345-381.
- Church, A. (1996). *Introduction to Mathematical Logic*. Princeton University Press.
- Clegg, I. L. K. (2018). Cognitive bias in zoo animals: An optimistic outlook for welfare assessment. *Animals*, 8(7). <https://doi.org/10.3390/ani8070104>

- Clegg, I. L., Rödel, H. G., Delfour, F. (2017). Bottlenose dolphins engaging in more social affiliative behaviour judge ambiguous cues more optimistically. *Behavioural Brain Research*, 322, 115-122.
- Cohen. J. (1992). *An Essay on Belief and Acceptance*. Clarendon Press.
- Copi, I. M., Cohen, C., McMahon, K. (2014). *Introduction to Logic* (14th edition). Pearson Education Limited.
- Correia, V. (2011). Biases and fallacies: The role of motivated irrationality in fallacious reasoning. *Cogency: Journal of Reasoning and Argumentation*, 3(1), 107-126.
- Crews-Anderson, T. A. (2007). *Critical Thinking and Informal Logic*. Humanities-Ebooks.
- Daoudi, S., Badihi, G., Buchanan-Smith, H. M. (2017). Is mixed-species living cognitively enriching? Enclosure use and welfare in two captive groups of tufted capuchins (*Sapajus apella*) and squirrel monkeys (*Saimiri sciureus*). *Animal Behavior and Cognition*, 4(1), 72-90.
- Dorigo, M., Gambardella, L.M. (1997). Ant colonies for the travelling salesman problem. *Biosystems*, 43(2), 73-81.
- Eemeren, F. van, Grootendorst, R., Blair, A., Willard, C. A. (eds.), (1986). *Argumentation*. Foris Publications.
- Eysenck, M., Mogg, K., May, J., Richards, A., Mathews, A. (1991). Bias in interpretation of ambiguous sentences related to threat in anxiety. *Journal of abnormal psychology*, 100(2), 144-150.
- Fackler, O.T., Grosse, R. (2008). Cell motility through plasma membrane blebbing. *Journal of Cell Biology*, 181, 879-884.
- Frege, G. (1918/1919). Der Gedanke. Eine logische Untersuchung. In *Beiträge zur Philosophie des deutschen Idealismus*. Band 1 (pp. 58-77). Stenger.
- Frege, G. (1934). *Grundlagen der Arithmetik*. Marcus.
- Getmanova A.D. (2011). *Uchebnik logiki* [Учебник логики] (8<sup>th</sup> edition). KHOPUS.
- Gilovich, T., Griffin, D., Kahneman, D. (eds.), (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press.
- Gödel K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I. *Monatshefte für Mathematik und Physik*, 38(1), 173-198.
- Goldberg, J., von Nitzsch, R. (2001). *Behavioural Finance*. John Wiley & Sons.

- Greengus, S. (1969). The Old Babylonian Marriage Contract. *Journal of the American Oriental Society*, 89(3), 505-532.
- Heaton, J. B. (2002). Managerial optimism and corporate finance. *Financial Management*, 31(2), 33-45.
- Heymans, G. (1905). *Die Gesetze und Elemente des wissenschaftlichen Denkens. Ein Lehrbuch der Erkenntnistheorie in Grundzügen*. Barth.
- Hilbert, D., Ackermann, W. (1949). *Grundzüge der theoretischen Logik* (Dritte, verbesserte Auflage). Springer-Verlag.
- Hilbert, D., Bernays, P. (1968). *Grundlagen der Mathematik*. Band 1 (Zweite Auflage). Springer-Verlag.
- Howard, S. R., Avarguès-Weber, A., Garcia, J. E., Stuart-Fox, D., Dyer, A. G. (2017). Perception of contextual size illusions by honeybees in restricted and unrestricted viewing conditions. *Proceedings of the Royal Society B: Biological Sciences*, 284(1867), 20172278.
- Huang, X., MacEvoy, S. P., Paradiso, M. A. (2002). Perception of brightness and brightness illusions in the macaque monkey. *Journal of Neuroscience*, 22(21), 9618-9625.
- Hunt, A. S., Edgar, C. C. (1932). *Select Papyri*. Volume I: Private Documents. Loeb Classical Library 266. Harvard University Press.
- Johnson, R., Blair, J. (1983). *Logical Self-defense*. McGraw-Hill.
- Jones, J. D., Adamatzky, A. (2010). Towards Physarum binary adders. *Biosystems*, 101(1), 51-58.
- Jones, J.D. (2015). Towards lateral inhibition and collective perception in unorganised non-neural systems. In K. Pancerz, E. Zaitseva (eds.), *Computational Intelligence, Medicine and Biology: Selected Links* (pp. 103-122). Springer.
- Kelley, D. (2014). *The Art of Reasoning: An Introduction to Logic and Critical Thinking* (4<sup>th</sup> edition). W. W. Norton & Company, Inc.
- Kern, L. H., Mirels, H. L., Hinshaw, V. G. (1983). Scientists' understanding of propositional logic: An experimental investigation. *Social Studies of Science*, 13(1), 131-146.
- Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, 108(3), 480-498.
- Mayne, R., Adamatzky, A. (2015). Slime mould foraging behaviour as optically coupled logical operations. *International Journal of General Systems*, 44(3), 305-313.

- Mortari, C.A. (2001). *Introdução à Lógica*. Editora UNESP: Imprensa Oficial do Estado.
- Ozbakir, L., Baykasoglu, A., Tapkan, P. (2010). Bees algorithm for generalized assignment problem. *Applied Mathematics and Computation*, 215(11), 3782-3795.
- Parrish, A., Brosnan, S., Beran, M. (2015). Do you see what I see? A comparative investigation of the Delboeuf illusion in humans (*Homo sapiens*), rhesus monkeys (*Macaca mulatta*), and capuchin monkeys (*Cebus apella*). *Journal of experimental psychology. Animal learning and cognition*, 41. <https://doi.org/10.1037/xan0000078>
- Pepperberg, I. M., Vicinay, J., Cavanagh, P. (2008). Processing of the Müller-Lyer illusion by a grey parrot (*Psittacus erithacus*). *Perception*, 37(5), 765-781.
- Pollard, T.D., Borisy, G.G. (2003). Cellular motility driven by assembly and disassembly of actin filaments. *Cell*, 112, 453-465.
- Povarnin, S. I. (1916). *Logika* [Логика].
- Robinson, E., Roiser, J. (2016). *Affective Biases in Humans and Animals*. In T. W. Robbins, B. J. Sahakian (eds.), *Translational Neuropsychopharmacology* (pp. 263-286). Springer International Publishing.
- Roelofs, S., Boleij, H., Nordquist, R. E., van der Staay, F. J. (2016). Making decisions under ambiguity: Judgment bias tasks for assessing emotional state in animals. *Frontiers in Behavioral Neuroscience*, 10, 119. <https://doi.org/10.3389/fnbeh.2016.00119>
- Sakiyama, T., Gunji, Y.-P. (2013). The Müller-Lyer illusion in ant foraging. *PLOS ONE*, 12(8), 1-12.
- Santacà, M., Miletto Petrazzini, M. E., Agrillo, C., Wilkinson, A. (2019). Can reptiles perceive visual illusions? Delboeuf illusion in red-footed tortoise (*Chelonoidis carbonaria*) and bearded dragon (*Pogona vitticeps*). *Journal of Comparative Psychology*, 133. <https://doi.org/10.1037/com0000176>
- Schumann, A. (ed.), (2020). *Swarm Intelligence: From Social Bacteria to Humans*. CRC Press, Taylor & Francis Group.
- Schumann, A. (2019). *Behaviourism in studying swarms: logical models of sensing and motoring*. Springer International Publishing.
- Schumann, A. (2018). Decidable and undecidable arithmetic functions in actin filament networks. *Journal of Physics D – Applied Physics*, 51(3), 034005.
- Schumann, A. (2019). Did the Neo-Babylonians Construct a Symbolic Logic for Legal Proceedings? *Journal of Applied Logics – IfCoLog Journal of Logics and their Applications*, 6(1), 31-82.

- Schumann, A. (2020). Legal argumentation in Mesopotamia since Ur III. *Journal of Argumentation in Context*, 9(2), 126-154.
- Schumann, A. (2017). On arithmetic functions in actin filament networks. In *10th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, ACM. <https://doi.org/10.4108/eai.22-3-2017.152402>
- Schumann, A. (2015).  $p$ -adic valued logical calculi in simulations of the slime mould behaviour. *Journal of Applied Non-Classical Logics*, 25(2), 125-139.
- Schumann, A. (2016). Toward a computational model of actin filament networks. In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2016) - Volume 4: BIOSIGNALS, Rome, Italy, February 21-23, 2016* (pp. 290-297). <https://doi.org/10.5220/0005828902900297>.
- Schumann, A., Kuznetsov, A. V. (2018). Foundations of Mathematics under Neuroscience Conditions of Lateral Inhibition and Lateral Activation. *International Journal of Parallel, Emergent and Distributed Systems*, 33(3), 237-256.
- Schumann, A., Pancerz, K. (2019). *High-level models of unconventional computations. A case of plasmodium*. Springer International Publishing.
- Shirakawa, T., Gunji, Y.-P., Sato, H., Tsubakino, H. (2020). Diversity in the Chemotactic Behavior of Physarum Plasmodium Induced by Bi-modal Stimuli. *International Journal of Unconventional Computing*, 15(4), 275-285.
- Shirakawa, T., Sato, H., Ishiguro, S. (2015). Constrction of living cellular automata using the Physarum polycephalum. *International Journal of General Systems*, 44, 292-304.
- Shirakawa, T., Yokoyama, K., Yamachiyo, M., Gunji, Y.P., Miyake, Y. (2012). Multi-scaled adaptability in motility and pattern formation of the Physarum plasmodium. *International Journal of Bio-Inspired Computation*, 4, 131-138.
- Siccardi, S., and Adamatzky, A. (2015). Actin quantum automata: Communication and computation in molecular networks. *Nano Communication Networks*, 6(1), 15-27.
- Siccardi, S., and Adamatzky, A. (2016). Quantum actin automata and three-valued logics. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 6(1), 53-61.
- Siccardi, S., Tuszyński, J. A., Adamatzky, A. (2016). Boolean gates on actin filaments. *Physics Letters A*, 380 (1), 88-97.
- Sigwart, Ch. von (1895). *Logic*. Vol. I: the Judgment, Concept, and Inference (2nd edition, translated by Helen Bendy). Swan Sonnenschein & co; Macmillan & co.

- Sims-Williams, N. (2000). *Bactrian Documents from Northern Afghanistan*. Vol. I: Legal and Economic Documents. Oxford University Press.
- Sjoberg, E. (2017). Logical fallacies in animal model research. *Behavioral and brain functions*, 13(1), 3.
- Spruijt, B. M., van den Bos, R., Pijlman, F. T. (2001). A concept of welfare based on reward evaluating mechanisms in the brain: anticipatory behaviour as an indicator for the state of reward systems. *Applied Animal Behaviour Science*, 72(2), 145-171.
- Stenning, K., van Lambalgen, M. (2011). Reasoning, logic, and psychology. *Wiley Interdisciplinary Reviews. Cognitive Science*, 2(5), 555-567.
- Stenning, K., van Lambalgen, M. (2008). *Human reasoning and cognitive science*. MIT Press.
- Swetz, F. J. (2003). Leibniz, the *Yijing*, and the Religious Conversion of the Chinese. *Mathematics Magazine*, 76(4), 276-291.
- Talisie, R., Aikin, S. (2004). Two Forms of the Straw Man. *Argumentation*, 20(3), 345-352.
- Tversky A., Kahneman D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Tversky, A., Kahneman, D. (2008). *Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment*. In J. E. Adler, L. J. Rips (eds.), *Reasoning: Studies of Human Inference and its Foundations* (pp. 114-135). Cambridge University Press.
- Van Haastert, P.J., Devreotes, P.N. (2004). Chemotaxis: signalling the way forward. *Nature Reviews Molecular Cell Biology*, 5, 626-634.
- Vinogradov, S. N., Kuz'min, A. F. (1954). *Logika* [Логика] (8th edition). Государственное учебно-педагогическое издательство Министерства просвещения РСФСР.
- Walton, D. (1989). *Informal Logic*. Cambridge University Press.
- Whiting, J. G., de Lacy Costello, B. P., Adamatzky, A. (2014). Slime mould logic gates based on frequency changes of electrical potential oscillation. *Biosystems*, 124, 21-25.
- Wright, W. F., Bower, G. H. (1992). Mood effects on subjective probability assessment. *Organizational Behavior and Human Decision Processes*, 52 (2), 276-291.
- Yaron, R. (1958). Aramaic Marriage Contracts from Elephantine. *Journal of Semitic Studies*, 3(1), 1-39.
- Culbertson, L. E. (2009). *Dispute Resolution in the Provincial Courts of the Third Dynasty of Ur. A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Near Eastern Studies) in the University of Michigan*.

# SÀÁGUA AND BAUMTROG ON PRACTICAL RATIONALITY

\*

DAVID BOTTING

Flat 1, 38 Westfield Rd. Yeovil  
SOMERSET, BA21 3DA, U.K.

## Abstract

*In “Practical Rationality at Work – A New Argumentation Model” (2018, pp.193-230) Sàágua and Baumtrog offer what they call “a new ideal model of integrated practical reasoning and argumentation.” I will argue that their only really original contribution is a reconceptualization of argumentation schemes as producing pro tanto reasons, and such a reconceptualization is both wrong-headed in itself and makes their model less integrated than other models, not more. In fact, their work is itself an example of bad practical reasoning.*

**Keywords:** *practical reasoning; argumentation; means-end rationality*

## 1. Introduction

“Practical Rationality at Work – A New Argumentation Model” is a long paper and impossible to summarize in a few words. Its proclaimed aim is to give an “integrated ideal model of practical reasoning and argumentation.” In this introductory section I will give and criticize their general theoretical framework and conclude that not only does it fail to achieve that aim, but also the means Sàágua and Baumtrog choose to achieve that aim are actually counter-productive; their model is itself an example of bad practical reasoning, as it involves choosing means that cannot possibly, for broad theoretical reasons irrespective of the detail, achieve their ends. In the later sections I will enter into detailed criticism of the argumentation schemes they offer and show that they do not all conform to the same pattern, sometimes because they conclude with different kinds of reason, and sometimes because of a mismatch between what the scheme concludes and the question to which the scheme is meant to be providing an answer.

The first task then is to lay out the model and explain what is meant by their claim that it is an “integrated ideal model of practical reasoning”. The model is given on (2018, p.199):

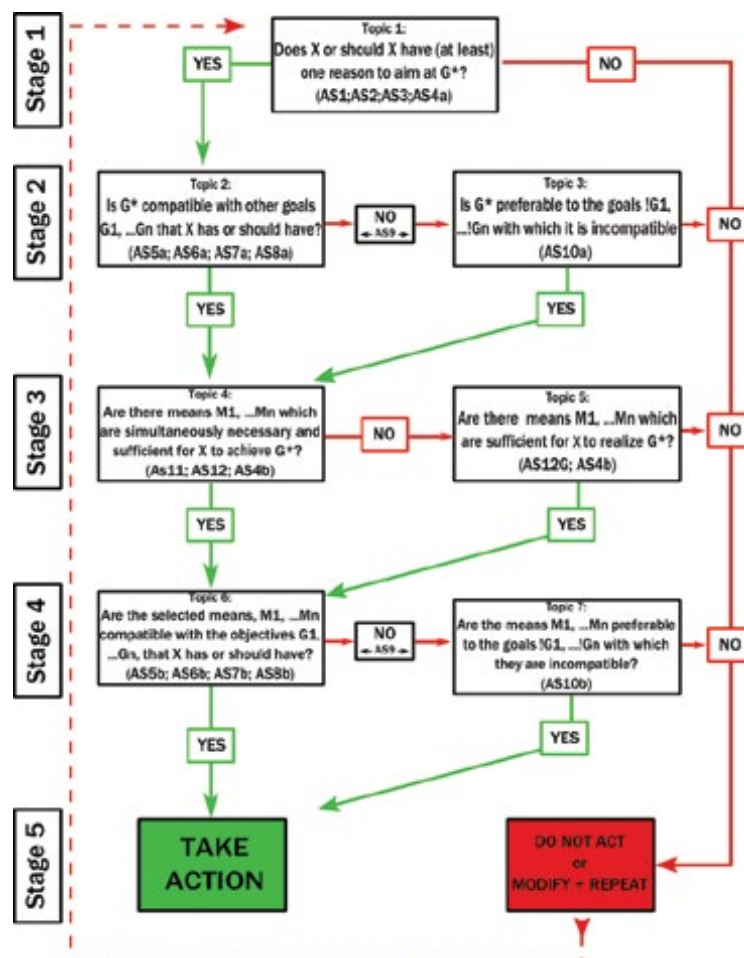


Figura 1

An agent navigates the model by answering each topic question in turn “Yes” or “No”, the answers to which determine at the final stage whether the agent acts or not. These answers depend in turn on the agent’s answers to the critical questions of the argumentation schemes appealed to with respect to each topic question: “In order to reasonably answer the Topic Questions, the agent must have reasons supporting their answer. Those reasons can be specified using an appropriate argument scheme. The model indicates what we consider to be the basic, necessary schemes to justify an answer to each one of the Topics” (Sàágua and Baumtrog, 2018, p.227). More than one argumentation scheme can be appealed to for any question, and the same argumentation scheme can be used in different stages. Most of the rest of Sàágua and Baumtrog’s paper is a comprehensive list of the argumentation schemes involved and their critical questions.

In what way or ways is this model an “integrated ideal model of practical reasoning and argumentation”, and how is it different, if it is different, from other models?

The first difference worth noticing is that whereas most models of practical reasoning (e.g., utility theory) are models of means-end reasoning and start at stage 3, Sàágua and Baumtrog’s model has an earlier stage in which the reasoner’s choice of goals is included as something for which reasons must be given, which reasons must stand the test of critical questioning. In Sàágua and Baumtrog’s words (2018, p.200) “We think that the assumption of the objectives themselves should be an object of reasoning and argumentation – and not only the choice of means”, and this is re-iterated in their conclusion: “With the aim of providing a model extending beyond mere instrumental reasoning, we have included consideration of the motivations for aiming at a goal as the first step in explicit practical reasoning and argumentation” (Sàágua and Baumtrog, 2018, p.227). This is one way in which they take their model to be integrated: moral considerations or in their word “ethics” are explicitly included in the model. However, as it turns out, making “the assumption of the objectives themselves . . . an object of reasoning and argumentation” only allows for a limited kind of rational criticism, as the critical questions are explicitly prohibited from questioning the choice of goals on *material* grounds: it does not allow the judgment “That choice of goal is wrong” but at best “That choice of goal is internally inconsistent with other goals you have or should have.” It will not rule out from being a reasonable choice of ends the very things that, as we will see in the next section, they want to rule out.

Another way in which Sàágua and Baumtrog take their model to be integrated, and indicated in the quote above where they say “argumentation or reasoning”, is that the stages can be interpreted as either stages of argumentation or stages of reasoning. As can be seen, there are five stages, and at each stage one or two topic questions, to which the answer is either “Yes” or “No”. Each question is to be answered by appeal to an argumentation scheme

or schemes. These questions and schemes can be taken as appealed to in an argumentative discourse or in monological reasoning. In Sàágua and Baumtrog's words (2018, p.194): "For those viewing the model with a mind to argumentation, . . . think in terms of commitments interlocutors have and can be held accountable for. Viewing the model as an instance of reasoning . . . view the model in terms of attitudes that connect to reasons in an inferential process." This claim to integration, as integrating argumentation and reasoning, is true as far as it goes, but I would deny that this is a difference between their model and other models.

Their last claim to integration is that it integrates elements from dialectical models with elements from BDI models. This might not be obvious from the diagram, but it occurs because the arguer or reasoner is required to have only a *pro tanto* reason to answer "Yes" to the Topic Question, and the argumentation schemes appealed to are taken to generate such *pro tanto* reasons.<sup>1</sup> However, I will show that this is actually inconsistent with the preceding claim to integration. By explicitly including an element from BDI models, or to be more specific, by building it into the very form of the schemes that their conclusions are *pro tanto* reasons, it is difficult to see how such schemes can be interpreted as or incorporated into any other model than a model of reasoning. If argumentation schemes are like this, they are not models of argumentation. But I will argue in section 3 that this innovation of theirs, which is their only real claim to originality, is misguided for all sorts of reasons not limited to the one just mentioned.

In three ways, then, their model includes more than other models would, or so Sàágua and Baumtrog claim, appealing to this novelty as justification for proposing their model as an improvement on those already available. If this extended scope of the model is a false ideal, or not really an extension but found also in other models, then by their own account their whole proposal is lacking in justification. I will argue that Sàágua and Baumtrog's innovations fail for one or other of these reasons. Let me be more specific: by an "integrated ideal model of practical reasoning and argumentation" the main (but not only) thing that they seem to mean by "integrated" is that the model can represent alternatively a piece of argumentation or a token of reasoning (by interpreting the model's components in different ways), but I will argue that this is to be found in models quite generally and not unique to their model, and by "practical reasoning" they mean more than just means-end reasoning but include also reasoning about the goals the agent aims to achieve, which reasoning the

---

1 There is a curiosity here connected to the fact that the Topic Question at Stage 1 simply is "is there a reason?" The argumentation schemes at Stage 1 justify answering "Yes" because their conclusions simply are those reasons, that is to say, those conclusions are truth-makers for an affirmative answer rather than providing further reasons for an affirmative answer. In short, they are not reasons for believing that we have a reason, but just are that reason. This is not so for the other topic questions, where the reasons generated by the argumentation schemes function straightforwardly as epistemic reasons rather than as truth-makers. This is a minor quibble, though.

model aims to bring into the scope of critical questioning, but I will argue that this is a false ideal. For these reasons, their proposal is lacking in justification.

I will largely set aside the extension of the model to include evaluation of one's choice of goals,<sup>2</sup> that is to say, the integration of practical reasoning and ethics, and focus instead on their claims to having integrated reasoning and argumentation—I will argue on completely general grounds that such a degree of integration is a false ideal and fails to take account of norms of reasoning that are not norms of argumentation and *vice versa*, but that not only does their model fail to achieve this false (and actually impossible) ideal, it actually fails to reach a level of integration that such a model ought to have and that other models do have. Nonetheless, the model might still have value if those parts, in particular the argumentation schemes they use, were correct. But I will argue that in so far as these argumentation schemes offer anything novel, they are wrong, both on general grounds and in points of detail. Therefore, the model is less than the sum of its parts and fails to meet its stated aim, and the parts are either unoriginal or, when they are original, wrong, and moreover their model combines those schemes in a way that often conflates instrumental reasons associated with means (by which I mean practical 'Pascalian' reasons favouring being in a particular psychological state, where being in that state is instrumental to bringing about the intended end) with moral reasons associated with ends (in which I include subordinate ends alongside the ultimate values they promote). The unavoidable conclusion is that their whole proposal is lacking in justification.

## 2. "Practical reasoning" and the choice of ends

Let us look at some examples of "practical reasoning" to see what verdict Sàágua and Baumtrog want their model to deliver. We will see that this motivates adoption of an objective conception of reasons. But, as I suggested above, the model can only deliver the verdict Sàágua and Baumtrog want if these objective reasons can be criticised on material grounds, and this is something that they do not want to allow out of fear of being seen to be saying "These ethical principles are the right ones!" When they talk about goals the reasoner should have, they do not mean those goals that are in conformity with any substantive list of ethical principles or values, but merely internal consistency with the reasoner's own value-system, and it should be noted that nowhere in the critical questions is the reasoner's value-system held up to scrutiny.

---

<sup>2</sup> For criticisms of this aspect in respect to an earlier paper of Sàágua and Baumtrog see Hitchcock (2016). I will make some further comments on this in the conclusion.

The Nazis engaged in practical reasoning when they reasoned that using Zyklon B was a more efficient means of achieving their end of eliminating the prisoners in concentration camps than other alternatives. As an example of means-end reasoning, this is a perfectly good piece of reasoning. Moreover, given their aims, the Nazis had instrumental reason to kill their prisoners with Zyklon B. But it seems that Sàágua and Baumtrog would not want to say that this is good practical reasoning, and this is their motivation for making evaluation of the original choice of goal an explicit aim of the model<sup>3</sup>—something has gone very wrong with our reasoning if this is our goal, and I think they would deny that the Nazis had good reason to kill their prisoners with Zyklon B (or at all, for that matter).

In saying this, though, they cannot believe that psychological states function as providing reasons, or at least it is not reasons of this kind that are being referred to in the verdict (which they want the model to give) that the Nazis had no reason to kill their prisoners with Zyklon B, for on such a subjective view of reasons all that is needed in order for an agent to have a reason is for the agent to desire the aim, no matter what the aim or how arbitrary the desire. The Nazis do have a reason to kill their prisoners on this subjective notion of a reason, as their desire to kill Jews gives them one. So, what issues from the model as the reason acted on is not this kind of subjective reason. By “reason”, then, Sàágua and Baumtrog appear to mean the kind of reasons that are generated by (or simply are) facts rather than by psychological states of the agent, that is to say, an objective notion of a reason. To illustrate: they say “‘X is thirsty’ is a reason (a consideration that counts in favour) for x to (intend to) drink water” (Sàágua and Baumtrog, 2018, p.197). Note that they do not say that X has a reason to drink *what he believes to be* water.

On this objective conception of reasons, a thirsty man does not have a reason to drink a glass of petrol if he mistakenly believes it to be water, but on the contrary has a reason not to drink it, his belief that it is water notwithstanding.<sup>4</sup> However, his belief that it is water will be a premise in his practical reasoning and when he concludes (mistakenly) that he has a reason to drink from the glass, we cannot fault his practical reasoning when he dies from poisoning. Nor can we fault the Nazis’ practical reasoning when they mistakenly think that they have reason to kill Jews with Zyklon B, but objectively do not. This illustrates the difference between a subjective and an objective conception of reasons. Objective reasons, unli-

---

3 Baumtrog’s preferred example is the mass murders of Anders Breivik (Baumtrog, 2015). Unfortunately, it will turn out that this evaluation of the original choice of goal does not extend to saying that the end is morally wrong, which is surely what we want to say, and which I think is what they want to say too. The kind of critical questioning and evaluation that is brought to bear on the choice of ends turns out to be very limited.

4 That this is Sàágua and Baumtrog’s conception of reason is further supported by their appeal to Scanlon (1998), as this is certainly Scanlon’s view.

ke subjective reasons, do not as such motivate the reasoner, as the reasoner might be quite unaware of having such a reason, and if he is unaware of such a reason he is entirely unable to act upon it, so objective reasons do not explain a reasoner's actions either. Objective reasons function normatively as a standard against which we can evaluate a reasoner's actions—when we say, as Sàágua and Baumtrog want to say, “The Nazis did not act (or value) reasonably” this means “The Nazis did not act in accordance with objective reasons.”

Insofar as an agent's practical reasoning accords to the model, he may think that he has reasons when in fact he does not, and insofar as the model is meant to be an ideal model of the agent's reasoning this does not matter much since the reasoning can be good either way, and the difference is only in the premises—if it were having the aim or desire that provided the reason then it is not clear how we could ever be in error about this and so the premise will be true. For example, it would be strange if we could somehow be mistaken about whether we *believe* that the glass contains water. We can, in contrast, be mistaken about whether the glass *actually does* contain water, so when the glass does not contain water the premise that it does will be false. But the reasoning itself is good and the agent is instrumentally rational either way.

So, this conception of reasons is not a problem as such, since believing that he has an objective reason and his actually having an objective reason is a distinction that the reasoner himself is unable to make. The reasoner takes himself to have objective reasons, and for him this explains why drinking what is in the glass is the right thing to do, though it is the subjective reasons (e.g., his normative belief that he has such and such an objective reason) that motivates and hence explains his actual behaviour. There are two kinds of explanation here, one normative and one causal. It is the subjective reasons that explain what we do when we act on a reason, because it is these that motivate, and play some kind of causal role in, our actual behaviour. But this is not what we are normally asking for when we ask someone why they act in a particular way. For instance, when we ask someone why they raised their arm, we do not expect them to answer “Well, there was a certain brain-event that activated the muscles in my arm and caused it to rise.” This is explanation of a kind, a causal explanation, but not what was sought, and we might quite justifiably think that the person answering this way has misunderstood the question. What we want and expect when we ask this question is the normative reason, and it is the objective reasons that determine whether this is or is not normatively the right thing to do and explains why we do it. If we asked the petrol-drinker why he drank what was in the glass he would not give the causal explanation in terms of subjective reasons, and if he did we are likely to answer “That is not what I meant!”—what he gives, and what we would accept in response, is a normative explanation and the normative reason, and it is this kind of reason that the petrol-drinker

means when he says, laying out his reasoning, that he has “reason” to drink what is in the glass. The petrol-drinker is sincere, and has argued entirely correctly, but on a false premise. If we convince him his premise is false he can still say “Well, I thought I had that objective reason” and that will still explain his behaviour and why he did what he did, but he will now be giving a causal explanation and not a normative explanation.

If we actually have the objective reasons that we think we do, we will always do the right thing, and quite correctly take ourselves to be doing the right thing. The problem is that however scrupulous an agent may be to try and believe only what is true there is always the possibility of being substantively wrong; since it is a fact that is providing the reason and not something like a desire or even a normative belief, it is always possible for either of the disagreeing parties to be mistaken about it. By using an objective conception of reasons Sàágua and Baumtrog seem to want to allow for just this possibility and would want to say that the Nazis’ choice of ends is substantively wrong, but they seem to want to say this without making any substantive assumptions of their own, and in the end they settle for insisting that the reasoner subject his choice of ends to close critical scrutiny and keep their fingers crossed that the reasoner thereby discovers some lacuna or defect in his assumptions. If Sàágua and Baumtrog really want their model to deliver the opposite verdict on the Nazis, the way to do so is to criticize this premise on the material grounds that it is false, which an objective conception of reasons allows them to do. Their reluctance to do this, presumably on the grounds that they do not want to be seen as favouring one set of substantive ethical principles over others simply because those principles are their own, and instead allowing only criticism of the choice of goals on other than material grounds and thereby laying the blame for the choice of bad goals on bad reasoning, leaves the adoption of an objective conception of reasons without any point as they could have done this with a subjective conception of reasons. Having practical premises that can be materially false is only useful if you are allowed to say that they are materially false.

But I have digressed slightly. The main point is that although they do not say so explicitly Sàágua and Baumtrog are adopting an objective conception of reasons, and must do so in order to be able to question the premises from which the Nazis, for example, argued; I am just not as optimistic as they seem to be about how much such questioning can actually achieve when those questions ask only for the agent’s reasons for accepting those premises and not the material truth of those premises. If an agent’s value-system is consistent, I do not see how any critical questioning that does not make value judgments opposed to those of the agent is going to show that that value-system and that choice of ends is wrong, and when we come to the critical questions it is noticeable that the reasoner only needs to show that he has reasons for thinking that his goal will promote *whatever* he positively values, even if

it is something like racial purity. The critical questions can pick up on when the reasoner is factually wrong about his goal promoting his value, or on when he is factually right but has poor reasons for thinking so, but do not pick up on morally egregious goals as such.

These objective reasons, including reasons for subordinate goals, are not instrumental reasons. The facts that make killing Jews a bad thing to do and give us objective reason not to kill Jews (whether we know this or not) do not say that it is bad on the grounds that it is a *means* to something else which we disvalue. Even if there is a “something else”, say the general value of human life, which we appeal to as our reason against killing Jews, this is because respecting Jewish life just is a case of respecting human life and so a subordinate goal, and not a means by which we respect human life. Now, it might be (and probably is) the case that we are more likely to respect human life if we assume not killing Jews as one of our goals than if we did not, and if so, there is an instrumental reason too not to assume this as a goal. But this is a ‘practical’ or ‘Pascalian’ reason insofar as it does not actually matter what the content of the goal is;<sup>5</sup> if Murphy’s Law applied and making this assumption was actually counter-productive and led to *more* killing of Jews there *could* be instrumental reason to make it our aim to kill Jews, but the moral reason would still obviously be to *not* kill Jews, and it is because of this reason that we would have the instrumental reason to assume killing Jews as our aim, though we would not have, or even believe that we have, reason to act on this aim and actually kill Jews.

So, there is a certain amount of ambiguity in asking whether we have a reason to assume something as a goal, as it might mean a moral reason or it might mean an instrumental reason of the practical kind, in analogy to the difference between having reasons for *p*’s being true (“epistemic” reasons) and reasons for being in the psychological state of believing *p* (“non-epistemic” reasons, e.g., positive consequences). When they talk about “assuming something as a goal” Sàagua and Baumtrog could be thinking of “assuming” as a voluntary act of adopting a particular *practical* attitude for which there can be non-epistemic reasons, or they could mean something more like a belief for which non-epistemic reasons – although they could exist – cannot really affect what beliefs we have.<sup>6</sup> In fact, I think that at some points their reasoning relies on assuming’s being a practical attitude, and at other points being a cognitive attitude. Of course, usually these things go together as luckily Murphy’s

---

<sup>5</sup> It is a state-given reason rather than an object-given reason.

<sup>6</sup> We might think, truly, that believing that God exists will allow us to live a more meaningful life, but we would not believe that God exists on these grounds for the simple reasons that it provides no grounds whatever for thinking this proposition to be true. We may, however, adopt a practical attitude towards the proposition that God exists to act as if it were true.

Law does not apply, and it might seem that this is a nit-picking difference. However, it is important to keep the differences clear for two reasons.

One is that sometimes the argumentation schemes are best interpreted as providing moral reasons (when the choice of ends is being considered) and sometimes as providing practical reasons, so the schemes ought to be revised to make clear what kind of reasons their conclusions are talking about. But even if this were done, having two types of reasons in the model would lead to problems later in the reasoning: when we are weighing up which of our multiple aims we ought to assume, it seems that it is the practical kind of reason that is at issue, as we want to know the conditional probability of the value being promoted having assumed something as a goal. But although we ought to assume this as a goal, it is not necessarily the case that we ought to act on this goal, the goals that we ought to act on being determined by the moral reasons – that is to say, the reasons supporting the truth of “Killing Jews is wrong” – rather than the practical reasons, and it is the moral reasons that we need to appeal to if we judge the Nazis’ practical reasoning as a case of bad reasoning. As mentioned above, we might be in the unlucky situation that the conditional probability of our actually killing Jews is higher if we assume as our goal not to kill Jews than if we assume as our goal to kill Jews. Then we ought to assume killing Jews as our goal, taking “assuming” here as a practical attitude, but we ought not to kill Jews, and the reason *why* we ought to assume killing Jews as our goal is *because* we ought *not* to kill Jews. This would be a bizarre and unfortunate condition to be in, but it would not be inconsistent.

Another is that we cannot get two reasons, or increase the support for choosing a particular end, by adding a practical reason for adopting that end (i.e., assuming it as a subjective goal) as an aim to the moral reason for endorsing that end. When the question is whether we should endorse the end, only the moral reason is relevant, and the fact (assuming it is a fact, which I have pointed out it might not be) that we may be more likely to act in conformity with our endorsed ends if we assume them as goals, does not give us an extra moral reason to endorse the end. Similarly, when the question is whether we should adopt the end, only the practical reason is relevant, as this is determined by how likely we are to satisfy that end once we have adopted it, and nothing else. Unfortunately, I will show that some of Sàágua and Baumtrog’s argumentation schemes conclude with practical reasons and treat “assuming as an end” as a practical attitude, and some do not, and this causes them to miscount the number of reasons in answering their topic questions.

For these reasons putting both kinds of evaluation in the same model is a dubious virtue and false ideal; although there is nothing wrong with questioning our original choice of ends, having both in the same model encourages us to mix up different kinds of reasons and different kinds of norms of reasoning that are better kept apart. As already said, there

was nothing wrong with the Nazis' practical reasoning, or the petrol-drinker's. Both were *instrumentally* rational, whether the reasons are taken to be objective or subjective, and I do not see why Sàágua and Baumtrog want their model to deliver the opposing verdict; whether reasoning is good or not does not depend on whether the premises are true or not, and we do not consider a model of deductive reasoning to be defective on the grounds that it does not tell us whether the premises of a valid deductive argument are true. Anyway, as we have seen, the premises' being false does not actually enter into the evaluation of the reasoning—only the agent's reasons for accepting the premises does.

So much for "practical reasoning". As I said, it is not the primary focus of this paper. The distinctions made here between objective, subjective, and practical reasons will only gain in importance when we come to analyse the argumentation schemes.

### 3. Integrating reasoning and argumentation

I will now show that being "integrated" in the sense of being a model of both argumentation and reasoning would also be of dubious merit if it were actually possible. It is not possible, however, because there are norms of good reasoning that are not norms of good argumentation, and *vice versa*. It is a false ideal to try to integrate both in the same model (even more so than the dubiousness of including evaluation of means and ends in the same model). *Some* integration is possible and even unavoidable. What I object to is their claiming such integration as a special virtue of their model that is lacked by others, and taking this in part to justify their work. This is simply not true. Insofar as their model might be thought to be integrated, *all* models of reasoning or of argumentation are integrated. But I will argue that their model is not even as integrated as this, that in fact by explicitly incorporating reasons in the schemes, the model is *less* integrated than other models, as such an element resists any alternative interpretation.

Let's go back to our Nazi example to illustrate what they mean when they claim that their model is integrated. The argument that the Nazis would use would be something like:

(Desire Premise) We want to kill prisoners.

(Means-End Belief) Using Zyklon B is a more efficient method of killing prisoners than the alternatives available to us.

(Conclusion) We intend to use (kill prisoners by using) Zyklon B.

The first premise is a desire, the second a belief, the third an intention. This is the basis of the B(elief) D(esire) I(ntention) model. We could imagine this easily as an argumentative discourse:

CONCENTRATION CAMP OFFICER: We have too many prisoners. We need to kill some of them.

SUBORDINATE: How? If we shoot them all it will waste ammunition. We need a more efficient way of killing them.

CONCENTRATION CAMP OFFICER: We'll gas them.

SUBORDINATE: That's not very efficient either.

CONCENTRATION CAMP OFFICER: I've heard of this stuff called Zyklon B. We can kill large numbers with only small amounts. It's cheap and easy to get.

SUBORDINATE (enthusiastically): Sounds ideal!

Equally we can imagine the officer arguing with himself "We need to kill more prisoners, and most of the ways are too inefficient. Zyklon B is the most efficient. So, I will kill prisoners by gassing them with Zyklon B." The same elements – the aim of killing prisoners, the efficiency of Zyklon B with respect to this aim, the decision to kill prisoners by this means – figure in all three.

It is, then, to some extent obvious that all three will have a similar structure and the same structural units, which merely have to be interpreted differently, as Sàágua and Baumtrog say. In this respect 'integratedness' is a natural aim for a model. But it is worth noting that it is because of the structural similarities between these things, and not because of a virtue of the particular model chosen, that there is this 'integratedness.' *All* models of reasoning will be models of argumentation for the reason that, and to the (limited) extent that, argumentation and reasoning are structurally the same.

Even so, not all models are the same, as there are norms of reasoning that are not norms of argument or argumentation: reasoning well is not identical to possessing a sound argument, since, for example, if one is not basing one's belief on the argument but on something else, or is not justified in believing that the argument is valid, even though it is, then one has not reasoned well, but there is nothing wrong with one's argument, which is perfectly sound, or with one's performance in argumentative discourse if one stated the argument's premises and then its conclusion (although there may be other reasons why, in this situation, one has not argued well). Also, one may have reasoned well even when those other norms are satisfied but the argument is not sound, that is to say, it has false premises. Basing one's beliefs on arguments that are valid and known to be valid is a norm of believing but not of argument, while having true premises is a norm of arguments but not of reasoning. Next, it is often claimed to be a norm of argumentative discourse that the arguers only use premises that the other party would agree with, or at least not use premises the other party would disagree with, but there is no such norm for reasoning or for argument. So, a model of good reasoning might have things that a model of argumentation would not have, and *vice versa*, some of which are relevant to practical rationality and some which are not (for example, if the subordinate failed to address his superior correctly or failed to 'take his turn'

then this is an error of performance in argumentative discourse, but of little or no relevance to practical rationality). Hence, Audi (2006) distinguishes between assessing the argument (by largely logical and material criteria), assessing the reasoning (by what he calls inferential criteria) and assessing the rationality of the doxastic states (by what he calls epistemic criteria). If being integrated is taken to imply that reasoning and argumentation can both be assessed by the same norms, even if interpreted differently, then it is making a substantive, and false, assumption; it would not be to a model's advantage to be 'integrated.'<sup>7</sup> As I will argue, Sàágua and Baumtrog's model does not in fact achieve this level of integration; the error is in claiming that it does when it does not and in claiming it to be a special virtue of their model not shared by others that it does so, when it would not be a virtue anyway even if it were achievable, and to the limited extent that it is achievable, it would not be a special virtue of their model anyway because *all* models are integrated to this extent as being so integrated follows directly from the structural similarities of what is being modelled. So again, even if Sàágua and Baumtrog trimmed back their claims about integration (or if I have been uncharitable in attributing to them a stronger claim about integration than they really meant), it would still be of no special virtue of their model that it is integrated to this extent, since all models are integrated to this extent.

In response, Sàágua and Baumtrog would probably point to their modification of argumentation schemes as illustrative of the way they intend to combine elements from commitment models and the BDI models used by theorists of reasoning. This is admittedly original and is what I will turn to now.

Of special interest to practical reasoning are considerations that count in favour of or against assuming some goal or performing some action. One thing that they are explicit on is the inferential character of these reasons. Reasons are *contributing* and *pro tanto* (Sàágua and Baumtrog, 2018, p.198). Moreover, they (or at least those that are proximate to assuming the goal/performing the action) are the outcomes of instances of *plausible* reasoning codified in argumentation schemes. Seeing argumentation schemes as producing *pro tanto* reasons is one of two innovations that they make to argumentation schemes, and

---

<sup>7</sup> Audi is interested in assessing practical reasoning and does not consider argumentation in discourse. My point (borrowed from Audi) is that although when someone reasons they are tokening the argument, the soundness of the argument tokened does not guarantee the goodness of the reasoning. Insofar as an integrated model fails to recognize this, it is a defect in the model. So, the advantages of 'integratedness' in a model should not be overstated, although it is certainly true that many of a model's components should easily lend themselves to re-interpretation in the ways described. It should be noted that Audi (2008) is cited by Sàágua and Baumtrog so one can only assume that they have read it.

Perhaps these differences would be attenuated if, as in most models (e.g., utility theory), it were only means-end rationality (which does not consider, for example, the truth of the reasoner's premises) that they wished to assess, but the model is specifically intended to go beyond this.

is an explicit inclusion of a BDI component in the model.<sup>8</sup> I will argue that far from making their model more integrated, this move actually makes their model *less* integrated. But the proposal has difficulties in itself independently of whether it serves their wider purpose.

First of all, by their own account, Sàágua and Baumtrog “consider argumentation to turn on the notion of conflict, and thus conceive of it as a dialectical situation” and that “argumentation begins when one or more parts of the process of reasoning come into conflict – it is the contestation of reasoning or its conclusion” (Sàágua and Baumtrog, 2018, p.195-6). Use of an argumentation scheme is supposed to resolve this conflict by transferring the acceptability of its premises to its conclusion. But if the conclusion is just a *pro tanto* reason, the conflict is *not* resolved. For there to be a conflict in the first place there must be (softening the “must be” to “could be” would not substantially change the objection) *pro tanto* reasons for and against the conclusion, so for an argumentation scheme to tell you that there is a *pro tanto* reason for the conclusion only unhelpfully re-iterates what you already knew.

---

8 The other is that instead of the normal set of critical question with “Yes/No” answers, they insert a pair of critical questions for each of the scheme’s premises with the aim of eliciting reasons for and against it. Here is their explanation (Sàágua and Baumtrog, 2018, p.202):

If you have a separate, critical Other asking the critical questions, then the wording of the questions can be less stringent in light of the opportunity for the Other to ‘press harder’ – so to speak – if the answerer does not provide a satisfactory answer. In the case where you are the only one responsible for asking and answering the critical questions, a more carefully worded question will make it harder to provide an unsatisfactory answer. As such, we find the wording of the questions is of great importance and have avoided using critical questions with only ‘Yes’ or ‘No’ answers.

...  
... [W]e have designed the questions as a pair of questions, the first of which asks for an explanation of the inferential term’s use and the second of which challenges it. Both parts are necessary, in our view, for the critical questioning of the inferential term to be adequate.

I find the notion of an “inferential term” obscure, but we do not need to dwell on it. I am not too concerned with this particular innovation. I will only make two remarks in passing.

The first is that the point of asking and answering questions whose answers are “Yes” and “No” is to transfer the burden of proof from one arguer to another. It is not clear to me whether the critical questions expressed this way does this, especially if the reason given by the answerer is *pro tanto*. (I think that they take these reasons to be *pro tanto* as well as the reason produced by the scheme, although it is not clear). The reservations I will express in a moment about the power of *pro tanto* reasons to conclude a stage of reasoning (in the model) also bear on their power to resolve the question of who has the burden of proof. Does a *pro tanto* reason resolve anything, apart from the question of whether one has a *pro tanto* reason (which, as it turns out, is the topic question at stage 1 of their model)?

The second is that what Sàágua and Baumtrog appear to have in mind in taking argumentation schemes to produce *pro tanto* reasons is that the conclusion of the argumentation scheme mentions a *pro tanto* reason, and not that the premises of the scheme produce *pro tanto* reasons for its conclusion (although they would not necessarily deny the latter). It is not clear whether they would hold all *pro tanto* reasons to be the outcomes of argumentation schemes (certainly they do not believe that all reasons whatsoever are contributing and *pro tanto*). If they did they would be faced with a regress, since the critical questions elicit reasons (I continue to assume these are *pro tanto* reasons) for the premises of the argumentation scheme, so if the only way we could get those reasons and answer critical questions were through using another argumentation scheme (which they explicitly allow for in particular instances) and answering that scheme’s critical questions, we would need an infinite number of instances of argumentation schemes, or an argumentation scheme with no critical questions to stop the regress. I am guessing, then, that they would agree that there are some *pro tanto* reasons that an arguer just has, and that the arguer does not have to appeal to an argumentation scheme for such cases.

Second of all, an argumentation scheme is meant to represent how someone would actually reason (being a kind of ‘naturalized logic’) and we do not in general reason that we have reasons. When we reason deductively, for example, we do not reason “I have a conclusive reason to believe my conclusion if I believe my premises, and I do believe my premises; therefore, I have a conclusive reason to believe my conclusion” and then proceed to act on this reason, but rather more simply “If my premises are true then my conclusion is true, and my premises are true; therefore, my conclusion is true.” Reasons-talk is generally used from an external standpoint to describe someone else’s reasoning or to formulate normative principles of reasoning (e.g., reason transmission principles, about which later), or when we want to abstract away from what we are reasoning about. We *might* reason in a second-order way that refers to reasons (and reason transmission principles) explicitly, but it certainly would not be very natural.

Now in Stage 1 of their model Sàágua and Baumtrog want to answer the question “Does X have (or should X have) at least one reason to assume  $G^*$ ?” which, quite obviously, requires X to reason about whether he has a reason. Here, any conflict over whether X has a *pro tanto* reason to assume  $G^*$  is settled by saying “Yes!” and exhibiting it to the disagreeing party. But this is part of the *content* of the reasoning task, whereas having the argumentation schemes conclude with having a reason (to assume  $G^*$  in these cases) is part of the *formal structure* of the scheme, and in fact of all schemes as Sàágua and Baumtrog have them, irrespective of content and what question they are answering.

Sàágua and Baumtrog have effectively taken the content of one of their topic questions as determining the structure of argumentation schemes generally, and this turns those schemes into second-order reasoning in cases where this is unnatural; or at least, it is only the content of their topic question at Stage 1 that makes having this kind of conclusion natural, and (going back to my first objection) it is only this question that a *pro tanto* reason can settle, precisely because it *is* the *pro tanto* reason that the topic question asks for. In short, the reason why these two objections do not seem to apply at Stage 1 is simply because of the topic question asked in Stage 1. Success at stage 1 may have misled Sàágua and Baumtrog into thinking that their modification works at all stages and in general, no matter what the content of the question. It does not.

Third of all, the argumentation schemes have detachable conclusions, which is to say that they resemble narrow-scope principles of rationality like:

(Reasoning Modus Ponens):  $B(p) \ \& \ B(\text{if } p \text{ then } q) \rightarrow R(B(q))$ <sup>9</sup>

---

<sup>9</sup> In words, if you believe that  $p$  and that  $p$  implies  $q$ , then you have a reason to believe  $q$ .

This leads to boot-strapping problems that Broome and others have pointed out.<sup>10</sup> Now, Sàágua and Baumtrog would counter-argue that what they mean is not analogous to this but instead to

(Reasoning Modus Ponens\*):  $R[B(p)] \ \& \ R[B(\text{if } p \text{ then } q)] \rightarrow R(B(q))$

Although the premises of the argumentation schemes do not themselves mention reasons, such reasons are elicited by the critical questions, so this is not boot-strapping. I would accept this.<sup>11</sup> But there are additional complications when it is a case of transferring reasons from ends to means:

(Practical Reasoning Modus Ponens):  $D(p) \ \& \ B(\text{if } p \text{ then } q) \rightarrow R(I(q))$ <sup>12</sup>

Suppose that although I desire to work, and in order to work I need to get out of bed, but that, as a matter of fact and despite my desire, I am *not* going to work. In those circumstances, do I have a reason to get out of bed, or even a reason to have the intention to get out of bed? The example is contentious. My complaint is that I think Sàágua and Baumtrog take it for granted that I do have such a reason.<sup>13</sup> Perhaps they would say that since they take the inference to be defeasible, it would not be inconsistent with the argumentation scheme for it to give either answer, that is to say that they would endorse

10 The classic counter-example is that believing something gives you a reason for believing just because you believe it, since this is  $B(p) \ \& \ B(\text{if } p \text{ then } p) \rightarrow R(B(p))$ .

11 I still have reservations about this. The reason indicated by the  $R$  in the consequent seems to be the objective reason, so even after critical questioning this would only imply that if I have reasons to believe that  $p$  and reasons to believe that  $p$  implies  $q$ , I therefore have reasons to believe that I have an objective reason to believe  $q$ . But it is not clear that I *do* have this second-order reason, as the same objections apply as before; if attaching the reason to the consequent is troubling then it is not clear that insisting that the reasoner has reasons for thinking that the antecedents hold is going to abate that problem, as it is still not clear that the consequent holds, that is to say, that the reasoner has reasons for believing that he has an (objective) reason. And even if I do have reasons for believing that I have an objective reason, it doesn't follow that I actually do have an objective reason. For example: I have a reason to believe that the glass contains water, and a reason to believe that if it is water, it is potable, but do I have a reason to believe that what is in the glass is potable on that conception of a reason in which, as I have pointed out earlier, I do *not* have a reason to drink it if in fact it contains petrol? But perhaps they would counter-argue that someone's reasons for believing  $p$ , for example, just are their reasons for believing that they have an objective reason to believe  $p$ . In that case, I think you do get Reasoning Modus Ponens\* and we just have to bear in mind that one can be mistaken about what objective reasons one has. In the scenario described, the petrol-drinker does not have an objective reason to think that the glass contains water (though he has reason to believe that he has this objective reason) any more than he has an objective reason to believe it contains something potable, or to drink it if thirsty (although, again, he has reasons to believe that he has these reasons). So, Reasoning Modus Ponens\* is still valid and I will set this reservation aside.

12 In words, if you desire that  $p$  and believe that  $p$  implies  $q$  (i.e.,  $q$ -ing is a necessary means to bringing about  $p$ 's being true), then you have a reason to intend to  $q$ .

13 The issue is perhaps easier to see when expressed with 'oughts'. I ought to work today, and it is necessary to get out of bed in order to work. But as a matter of fact I am not going to work when I get out of bed. Is it still the case that I ought to get up? The model of Fairclough and Fairclough [2012], which heavily influences Sàágua and Baumtrog's model, expresses its conclusion in terms of such 'oughts' and seems oblivious to the problem here. My point is that this is not uncontroversial, and although Sàágua and Baumtrog do not talk in terms of 'ought' I think they need to address this issue too.

(Practical Reasoning Modus Ponens\*):  $R[D(p)] \ \& \ R[B(\text{if } p \text{ then } q)] \rightsquigarrow R(I(q))$

Or perhaps they would say that, while I am lying in bed, it is correct practical reasoning to intend to get up, even if as a matter of fact I do not work after I have got up.

But there is a general problem in this area. By traversing the model one topic at a time, it is presumed that if all goes well an action will result. But all that the argumentation schemes can generate is a *pro tanto* reason, and this only defeasibly, and the best that can be said for such an undefeated *pro tanto* reason is that it hasn't been defeated so far, either by other *pro tanto* reasons or by the scheme's defeasibility conditions. A *pro tanto* reason is simply not strong enough either to lead to action or resolve a difference of opinion (or, more arguably, resolve the question of who has the burden of proof during the critical questioning). Also, what licenses the move to the last stage in the model (to be shown shortly) seems implicitly to be stronger than what licenses the moves between the previous stages. The implication is that once you have taken all the *pro tanto* considerations into account, the one left standing will be an all things considered reason (i.e., the net result after all the *pro tanto* considerations for and against have been weighed against each other), and an all things considered reason admittedly would be strong enough. But then we have to suppose that there are some argumentation schemes whose conclusions are all things considered reasons, or we have to suppose that there is some reasoning that is not represented in the model by which the *pro tanto* reason is made into an all things considered reason, and it is not clear what this reasoning could be, since a *pro tanto* reason can always be defeated in principle, and only follows defeasibly from the premises to boot—even if you have considered all the other reasons so far, the best you can say from this is that the reason hasn't *in fact* been defeated *yet*; one may still rationally prefer to wait than to act.

Or if we take the argumentation scheme to model commitments, what we get in the final stage is a “a discursive commitment that one intends to perform some act” (Sàágua and Baumtrog, 2018, p.194). Such a discursive commitment is only a commitment relative to that particular dialectical situation. We cannot detach any kind of all things considered commitment from this, as there may be other dialectical situations in which one would not have this commitment. In fact, we may be in this other dialectical situation at the same time as we are in the first; since we may be in more than one dialectical situation at a time, that is to say, be having a reason-giving exchange with more than one set of interlocutors at a time, one would simultaneously have a discursive commitment to act relative to one set of interlocutors and a discursive commitment not to act, or not have a discursive commitment to act, relative to another. Now, one may say that as long as one is in *one* of those situations, that situation provides at least a *pro tanto* reason. But then we have gone in a circle, as such a reason will not make reasonable a decision to act, but only make acting justifiable to one particular set of interlocutors (and possibly unjustified to another).

#### 4. A Critique of “Practical Rationality at Work”

I will go through each of the five stages and the argumentation schemes provided for them. I will argue that the schemes do not always produce the same kind of reason, some only providing instrumental or subjective reasons. I will also argue that many of the schemes are reducible to the same argumentation scheme, and that it is consequently wrong to think of these different schemes as providing numerically distinct reasons.

- a. Stage 1, Topic Question 1: Does X have (or should X have) at least one reason to assume  $G^*$ ?

Four argumentation schemes are given for this stage (Sàágua and Baumtrog, 2018, pp.204-208). More than one of these argumentation schemes can be used, which means, according to Sàágua and Baumtrog (2018, p.228), that there can be up to four reasons in Stage 1 for advancing to Stage 2 by answering the topic question affirmatively.

##### **Argumentation Scheme 1. Assumption of Objectives by Teleology (AS1)**

Premise 1: X has  $G^+$  as its finality.

Premise 2:  $G^*$  belongs to  $G^+$ .

Therefore, plausibly

Conclusion: There is a reason for X to assume  $G^*$ .

##### **Critical Questions for Argumentation Scheme 1**

CQ1: How does  $G^+$  really correspond to the finality of X?

How can  $G^+$  not correspond to the finality of X?

CQ2: How is  $G^*$  really a particular case of  $G^+$ ?

How could  $G^*$  not be a particular case of  $G^+$ ?

##### **Argumentation Scheme 2. Argument from Positive Values (AS2)**

Premise 1: value V is positive ( $= V^+$ ).

Premise 2:  $V^+$  positively values  $G^*$ .

Therefore, plausibly

Conclusion: There is a reason for X to assume  $G^*$ .

##### **Critical Questions for Argumentation Scheme 2**

CQ1: What reasons are there for attributing a positive value to V?

What reasons could count against attributing a negative value to V?

CQ2: What reasons are there for the positive evaluation of  $G^*$  by  $V^+$ ?

What reasons are there for  $G^*$  not to be positively evaluated by  $V^+$ ?

**Argumentation Scheme 3. Argument from Positive Consequences (AS3)**

Premise 1: If  $G^*$  is realized by X, then the consequences  $K_1, \dots, K_n$  will occur.

Premise 2:  $K_1, \dots, K_n$  are to be valued positively.

Therefore, plausibly

Conclusion: There is a reason for X to assume  $G^*$ .

**Critical Questions for Argumentation Scheme 3**

CQ1: What makes it plausible that  $G^*$  has  $K_1, \dots, K_n$  as consequences?

How could  $G^*$  not have  $K_1, \dots, K_n$  as consequences?

CQ2: Why should  $K_1, \dots, K_n$  be positively valued?

How could  $K_1, \dots, K_n$  not be positively valued?

**Argumentation Scheme 4. Argument from Ability (AS4a)**

Premise 1:  $G^*$  should be positively valued.

Premise 2: X has the ability to realize  $G^*$ .

Premise 3: X's ability to realize  $G^*$  is a necessary/enabling condition for the realization of  $G^*$ .

Therefore, plausibly,

Conclusion: There is a reason for X to assume  $G^*$ .

**Critical Questions for Argumentation Scheme 4a**

CQ1: How does X have the ability to realize  $G^*$ ?

What could prevent X from realizing  $G^*$ ?

CQ2: To what extent is the assumption of  $G^*$  by X a necessary/enabling condition for the realization of  $G^*$ ?

Which  $y$  exists (such that  $y \neq x$ ) whose ability to realize  $G^*$  is a necessary/enabling condition for the realization of  $G^*$ ?

Commentary: Each critical question is expressed by a pair of questions, the first member of each pair aimed at eliciting X's reasons for thinking that the corresponding premise<sup>14</sup> is true while the second member aims at eliciting X's reasons for thinking that it is false. Note, though, that reasons do not enter into the premises themselves; according to the schemes, X has a reason to assume  $G^*$  if the premises are true whether he believes those premises for good reasons or not, or even if he does not believe them at all (unless, perhaps, such cases are meant to be ruled out by the defeasibility conditions). As such, we get an appeal to a reason in the conclusion when there is no reason explicitly referred to in the premises, which state only

---

<sup>14</sup> In most cases, CQ1 corresponds to/questions Premise 1, CQ2 corresponds to/questions Premise 2, and so on, though this is not always the case.

facts. This is not necessarily a problem if facts can be reasons, so the form of the schemes is more evidence that Sáágua and Baumtrog are committed to this objective conception of reasons. Of course, when X appeals to the scheme in his reasoning he must believe the premises, and if he does not believe the premises for good reasons but draws the conclusion anyway, his reasoning will be rationally criticisable. X might have an objective reason but not be rational in believing that he does, or he might be rational in believing that he has an objective reason when he does not. In order to advance to the next stage, X must be reasonable in believing that he has an objective reason whether he does so in fact or not. As far as it goes, then, the critical questions express conditions of adequacy for reasoning according to the scheme.

Even so, I would deny that these are properly called argumentation schemes. Repeating a point made earlier, an argumentation scheme should be able to settle a disagreement, which is to say it should be able to determine, even if only temporarily and provisionally, which side of an argument an arguer should settle on when there are reasons on either side. To be told that there is a *pro tanto* reason on one side clearly does not do this, as this was granted from the outset; if there are reasons on either side, *a fortiori* there is reason on one side, and use of the argumentation scheme has not progressed the disagreement at all.<sup>15</sup>

Let's look at the schemes in detail, starting with AS1. First, we might ask what Premise 1 ("X has G+ as its finality") even means. In their example X is "NATO", G+ is "contribute to world peace and security", and G\* is "helping Ukraine increase its defensive power." Contributing to world peace and security is defined as NATO's mission and, to use their curious term, "finality". In some cases (though probably not here) it might be constitutive of something being X that it has G+ as an aim. As such it is difficult to see how any very illuminating answer can be given to the question "How does G+ really correspond to the finality of X?" beyond "That is how X is defined" or "It is stipulated that G+ is the finality of X." And the answer to "How can G+ not correspond to the finality of X?" is "It can't." Boiled down, CQ1 simply asks "Is Premise 1 true?" The same can be said of CQ2 in respect of Premise 2.

It should be noted that there is no question here of whether G+ is a "finality" that X *should* have, nor, consequently, is there an objective reason for X to assume G+ let alone G\*; yet the

---

15 I would argue that the way they are presented these schemes are less like argumentation schemes than reason transmission principles. A simple example of such a principle that is relevant to practical reasoning is that if I have a reason for an end, and a certain means is sufficient for me to achieve that end, then I have a reason to carry out that means (as in Practical Reasoning Modus Ponens given above). This takes one premise in which "reason" is explicitly mentioned and one where no reason is appealed to in order to conclude with a reason. (Note, again, that on the objective conception of having a reason I have a reason to carry out the means even if I am ignorant of its being the means for the end in question.) Generally, transmission principles take one normative premise and one non-normative premise in order to draw a normative conclusion. Apart from AS1, the argumentation schemes in stage 1 all seem to have one normative premise [that something is valued or to be evaluated positively], another premise (or premises) that is non-normative, and conclude with the normative claim that X has a reason (to assume G\*).

mere fact that X has this finality is taken to be enough to provide a reason just as an agent's mere having of a desire, even an arbitrary one, is enough to provide that agent with a reason to satisfy that desire and ultimately act on it. The reason named in the conclusion of AS1 thus seems to be a merely subjective<sup>16</sup> reason, even though it is a reason for endorsing a goal  $G^*$  rather than a course of action. Probably the idea is that insofar as the topic question asks about goals X *actually has* as well as those X *should have*, AS1 provides a reason for taking X to actually have  $G^*$  as a goal irrespective of whether X *should* have it as a goal or not. My objection would then be that the topic question itself involves different kinds of reason but treats them as the same. Also, since the Nazis could just as easily have named the racial purity of the Aryan race as their “finality”, by extending the question to include goals X simply has, the model allows for exactly the kind of practical reasoning that Sàágua and Baumtrog want to say is bad. AS1 explicitly rules in the very thing that they want to rule out (unless the hope is that it is ruled out later because of conflicting with other goals, goals that X *should* have, about which more later).

Or if it is to be supposed that there is objective reason for X's having  $G^+$  as an aim then it seems that this can only be because  $G^+$  is positively evaluated, in which case AS1 collapses into AS2 or AS3. It might be thought that in answering the critical questions to AS1 it is quite permissible to appeal to other schemes, and thus that this is not necessarily a defect; we appeal, for example, to AS2 in answering CQ1, and the reason provided by AS2 is percolated upwards, so to speak. But the critical questions do not ask anything about whether having  $G^+$  as a finality is to be valued. Either AS1 concludes with a subjective reason that X simply has, or it is dispensable and should be replaced with an instance of AS2 or AS3. In neither case (though for different reasons) should the reason concluded by AS1 be added to the reason concluded by AS2 or by AS3, in one case because the reasons are different in kind, and in the other case because the reason concluded in AS1 is effectively the same as that concluded in AS2/3—they would effectively be the same reason, and if we added them together, we would effectively be double-counting the same reason.

In fact, I will argue later that including goals that X merely has is a mistake and leads to boot-strapping. Having a goal cannot provide a reason for having that goal or acting on it, any more than having a belief can provide a reason for having that belief. To be more precise, it is a mistake given their goal of wanting to say that certain reasonings like those of the Nazis, although instrumentally rational, are bad. If the goal is to give conditions of adequacy for instrumental rationality itself it is not a mistake, but simply reflects the fact that we can reason practically even about goals that we merely have even when we should not have those goals and perhaps even believe we should not have those goals.

---

<sup>16</sup> Note that being a subjective reason is not the same as being an instrumental reason.

Let's look at AS2 next. In analogy to AS1, the Critical Questions ask for reasons for the truth of the two premises. It is not clear what reasons X is supposed to have for attributing a positive value to V+, except to subordinate V+ to a yet further value; for 'ultimate' values I don't see that any very illuminating answer can be given to CQ1. In disanalogy to AS1, the reason named in AS2's conclusion is *not* a subjective reason; it is not enough for X simply to desire that G+ be the case or have G+ as a stipulated aim in order for there to be a reason for X to assume G\*, or even for X to think that he has such a reason.

As for AS3, it seems to reduce to AS2. If some consequences are to be evaluated positively (Premise 2 of AS3) then there must be some positive value or values V (Premise 1 of AS2) that evaluate those consequences positively, and if V evaluates G\*'s consequences positively, it seems like it must evaluate G\* itself positively (Premise 2 of AS2). Perhaps it might be argued that bad things sometimes have good consequences, but that we should not for that reason call those bad things good. But I think that AS2 should call G\* good (that is to say, V+ should evaluate G\* positively) in this context, as we would still in this case want to draw the conclusion that there is a reason for X to assume G\*. Thus, I do not see how the premises of AS3 can be true without the premises of AS2 being true. This redundancy is harmless to a large extent but may deceive us into thinking that there are more reasons than there really are, two when there is really only one. Unless the idea is that for each positively evaluated consequence there is a further reason to assume G\*, but AS3 does not have this result (though perhaps it should be modified so that it does).

AS4a is, by their own admission, a new argumentation scheme. The first thing to notice is that its Premise 1 ("G\* should be positively valued") actually follows from the premises of AS2, from which AS2's conclusion also follows. Hence, insofar as we take AS2's conclusion and AS4a's conclusion to be the same, that conclusion follows by virtue of Premise 1 alone, and Premises 2 and 3 are redundant; Sàágua and Baumtrog (2018, p.208) seem to say as much themselves when they note that they take Premise 1 to have already been established by AS2 or AS3. But perhaps we should not take them to be the same conclusion. Sàágua and Baumtrog say (2018, p.208) that AS4a concludes to there being a reason for X *rather than any other agent* to assume G\*, because X uniquely has the ability to bring G\* about, and this reason can be added to the reason provided by AS2 or AS3. But this seems like an instrumental reason. If G\* is likely to be brought about or occur without X's agency then X does not have this kind of reason to assume G\*, or so they seem to say, and this is true insofar as one is generally (though not always, as I have pointed out) more likely to bring G\* about once one has assumed it as one's goal, provided that one has the ability to bring it about. It is a Pascalian, practical, state-given kind of reason for the practical attitude of making something one's goal, rather than a moral, object-given reason for the cognitive attitude of valuing that thing.

The reasons named in the conclusion of AS1 and AS4a are not the kind of reasons by which one is allowed to move from Stage 1 to Stage 2, or if they are it is in virtue of being reasons for X to assume  $G^*$  on the basis of  $G^+$  being a goal X actually has, which is not going to rule out the very thing that they want to rule out and that the use of objective reasons is designed to rule out. Only AS2 and AS3 are relevant, as only these provide objective reasons, and they are effectively the same argumentation scheme. Or perhaps it is only the instrumental reason that ought to count, so it is AS4a that we should keep and AS2 and AS3 that are redundant, but then, as I have indicated, it is not necessarily the case that we would want to end up acting to achieve this goal; even though there might be a practical reason to assume killing Jews as an aim, this is not an aim that we *should* act on or even desire to act on. The aims that we want to end up acting on are those that we actually endorse, which is to say, those for which we think that there is an objective moral reason because they promote values. So, I think it is AS4a that must be rejected after all. Either way, there can only really be one reason, and not four, for answering “Yes” to the topic question.<sup>17</sup>

- a. Stage 2, Topic Question 2: Is  $G^*$  compatible with other goals  $G_1, \dots, G_n$  that X has or should have?

There are four argumentation schemes relevant to Stage 2 and this topic question.

#### **Argumentation Scheme 5. Argument from Negative Values (AS5)**

Premise 1: the value  $V$  is negative ( $= V^-$ ).

Premise 2:  $V^-$  negatively values  $G^*/M^*$ .

Therefore, plausibly,

Conclusion: There is a reason for X not to assume  $G^*/M$ .

#### **Critical Questions for Argument Scheme 5**

CQ1: What reasons are there for attributing a negative value to  $V$ ?

How could  $V$  not have a negative value?

CQ2: What reasons are there for the negative evaluation of  $G^*/M^*$  by  $V^-$ ?

How could  $G^*/M^*$  not be negatively valued by  $V^-$ ?

#### **Argumentation Scheme 6. Argument Contradicting Positive Values**

(AS6)

Premise 1: Value  $V$  is positive ( $V^+$ ).

---

<sup>17</sup> With the qualification that there can be a number of positive consequences. This has nothing, however, to do with the number of argumentation schemes that can be appealed at a particular stage, as Säägaa and Baumtrog would have it.

Premise 2:  $G^*/M^*$  contradicts (or inhibits)  $V+$ .

Therefore, plausibly

Conclusion: There is a reason for X not to assume  $G^*/M^*$ .

### **Critical Questions for Argument Scheme 6**

CQ1: What reasons are there for attributing a positive value to V?

How could V not be valued positively?

CQ2: What reasons are there to indicate  $G^*/M^*$  contradicting (inhibiting)  $V+$ ?

How could  $G^*/M^*$  be congruent with  $V+$ ?

### **Argumentation Scheme 7. Argument from Negative Consequences (AS7)**

Premise 1: If  $G^*/M^*$  is realized by X, the consequences  $K_1, \dots, K_n$  will occur.

Premise 2:  $K_1, \dots, K_n$  are to be negatively valued.

Therefore, plausibly

Conclusion: There is a reason for X not to assume  $G^*/M^*$ .

### **Critical Questions for Argument Scheme 7**

CQ1: What makes it plausible that  $G^*/M^*$  has  $K_1, \dots, K_n$  as consequences?

How could  $G^*/M^*$  not have  $K_1, \dots, K_n$  as consequences?

CQ2: Why should  $K_1, \dots, K_n$  be negatively valued?

How could  $K_1, \dots, K_n$  not be negatively valued?

### **Argumentation Scheme 8. Argument from Operational Incompatibility (AS8)**

Premise 1: G is an objective already assumed by X

Premise 2: G and  $G^*/M^*$  are operationally incompatible

Therefore, plausibly

Conclusion: There is a reason for X not to assume  $G^*/M^*$

### **Critical Questions for Argumentation Scheme 8**

CQ1: What reasons are there for taking G as an objective already assumed by X?

What reasons are there for doubting X already assumed G?

CQ2: What makes G and  $G^*/M^*$  operationally incompatible?

How could G and  $G^*/M^*$  not be operationally incompatible?

Commentary: AS5 and AS6 are obviously the flip-sides to AS2 and conclude with reasons against assuming  $G^*$ . Similarly with AS7 with respect to AS3. In analogy to what I said regarding these, I do not see how AS7's premises can be true without AS5's being true, in which case there is no real need for both of them and only one reason can result from their both being appealed to rather than two. The critical questions, once more, aim at eliciting reasons why the premises may be true firstly and false secondly.

Superficially, it might seem that CQ1 in AS5 and AS6 do question the choice of values on material grounds, which is the very thing I said does not happen. But as long as reasons can be given for the choice of values, I do not see that the questioner has any more to say, and when the value is an ultimate value, I do not see any very illuminating answer can be given to the question “What is your reason for valuing this?”

AS8 is the interesting argumentation scheme, and I think that it is unnecessary. First of all, you cannot get a reason to assume  $G$  just because you actually have assumed  $G$ , and equally, you cannot get a reason to not assume  $G^*$  because it conflicts with some  $G$  that you have assumed without reason (and Sàágua and Baumtrog are explicit that this is the kind of case that they are considering; otherwise, it would fall under one of the other three argumentation schemes). Insofar as there may be a reason here, it is only a subjective one.

What would happen if, because of operational incompatibility with such a  $G$  (and for no other reason), we answered “No” to the question “Is  $G^*$  compatible with other goals  $G_1, \dots, G_n$  that  $X$  has or should have?” Then we would be faced with the question “Is  $G^*$  preferable to  $G$ ?” as the next topic question, and it seems to me that the answer to that question would always be “Yes”, a goal that we have reason to have (as guaranteed by AS2/3) always being preferable to a goal that we have just assumed without reason.<sup>18</sup> We only need to worry about the goals that we *should* have, as these are the only goals which provide us with objective reasons to act on them. So, we may as well eliminate the “has” from the “has or should have” in the first and second topic questions and answer “Yes” to the second topic question without worrying about what goals  $X$  actually has, thus eliminating AS8 (and AS1).

Now, it might be argued that insisting on this objective conception of reasons has illegitimately limited what we might engage practical reasoning on. After all, we do reason practically about how to achieve goals that we merely have whether we should have them or not, and we compare goals that we merely have with each other. True. But it is the attempt to integrate ethics with practical reasoning that has led to this limitation. Now, if we used subjective reasons all the way through, the model would, I think, still apply: AS1 will give us subjective reasons for goals (namely, our actually having those goals), these goals can be tested against each other, and the strongest subjective reasons will be the ones we act on and have a subjective reason to act on. So, we could not eliminate AS1 or consideration of goals we actually have, as long as we maintain also that the reason we end up acting on is subjective, and thereby allowing in all the kinds of acting that Sàágua and Baumtrog want on ethical grounds to rule out. Remember that a mere desire provides a subjective reason.

---

<sup>18</sup> Note that CQ1 for AS8 does not establish that  $X$  assumed  $G$  for some reason, but only that there is reason to believe that  $X$  has assumed  $G$ .

- a. Stage 2, Topic Question 3: Is  $G^*$  preferable to the other goals  $!G_1, \dots, !G_n$  that X has or should have?<sup>19</sup>

This is where the actual weighing of reasons for and against assuming goal  $G^*$  takes place; it is “at this point where argumentation theory meets choice theory” (Sàágua and Baumtrog, 2018, p.218). This is done by appeal to one argumentation scheme with two variants, depending on which goal is preferable.

### **Argumentation Scheme 10. Argument Based on Rational Preference (AS10)**

#### **10.1 Variation in favour of $G^*/M$**

Premise 1:  $!G$  and  $G^*/M^*$  are contradictory.

Premise 2:  $(R^{\pm 1}, \dots, R^{\pm n})G^*/M^*$ .

Premise 3:  $(!R^{\pm 1}, \dots, !R^{\pm n})!G$ .

Premise 4:  $(R^{\pm 1}, \dots, R^{\pm n})G^*/M^*$  are preferable to  $(!R^{\pm 1}, \dots, !R^{\pm n})!G$ .

Therefore, plausibly

Conclusion: X should assume  $G^*/M^*$  (and abandon  $!G$ )

#### **10.2 Variation in favour of $!G$**

Premise 1:  $!G$  and  $G^*/M$  are contradictory

Premise 2:  $(R^{\pm 1}, \dots, R^{\pm n})G^*/M^*$

Premise 3:  $(!R^{\pm 1}, \dots, !R^{\pm n})!G$

Premise 4:  $(!R^{\pm 1}, \dots, !R^{\pm n})!G$  are preferable to  $(R^{\pm 1}, \dots, R^{\pm n})G^*/M^*$

Therefore, plausibly

Conclusion: X should assume  $!G$  (and abandon  $G^*/M^*$ )

### **Critical Questions for Argumentation Scheme 10**

CQ1: What makes the standard(s) used for the evaluation of the reasons associated with the goals/means the best for this situation?

Why might the standard(s) used for the evaluation of the reasons associated with the goals/means not be the best for this situation?

CQ2: What makes the standard(s) used to assess the probability or plausibility of the reasons used to justify the assumption of the goal/means and of the goal/means being assumed the best for this situation?

Why might the standard(s) used to assess the probability or plausibility of the reasons used to justify the assumption of the goal/means and of the goal/means being assumed not be the best for this situation?

---

<sup>19</sup> Again, I would make the topic question refer only to the goals X should have. This is to be taken for granted from now on.

Commentary: First of all, because of AS6 any positive reason for  $G^*$  will be a negative reason for  $!G$ , and *mutatis mutandis* for negative reasons, so Premises 2 and 3 contain effectively the same information.<sup>20</sup> Second of all, I am not sure what to make of the critical questions here. According to Sàágua and Baumtrog (2018, p.218) “these questions are challenging the goodness in the reasons and the accuracy of the probability of success, respectively.” Note that this differs from what we usually say, which is that we prefer the goal with the greatest expected utility, which is the value of the *goal*, and not the strength of the *reason*, multiplied by the likelihood of achieving the goal. Yet AS10 talks about which *reasons* are preferable in Premise 4, and about standards used for the evaluation of the reasons in CQ1. I fail to see what these standards are supposed to be. Certainly, we can talk about reasons and their strength *having done the utility calculation*, but this is on the grounds that we have most reason to attempt to satisfy those goals with the greatest expected utility. Hence, we can talk about reasons and utilities interchangeably to some extent, but I fail to see how we get to utilities, or the preferability of one goal to another, by evaluating reasons, or how one reason is preferable to another, unless it means a reason to achieve something whose utility is higher, other things being equal.

I think that tolerable sense can be made of what they say in this way. All objective reasons ultimately derive from a value. This is given by AS2, and I argued that AS3 could be reduced to AS2. The weight of the reason for assuming  $G^*$  could then be taken as the conditional probability of achieving the value  $V+$  if we assume the goal  $G^*$  multiplied by the value of  $V+$  itself. (Note, though, that this *does* seem to treat the reason for assuming  $G^*$  as a Pascalian kind of instrumental reason, as it is the weight of this kind of reason that corresponds to the conditional probability at issue, in which case it is AS2 and AS3 that are irrelevant to this calculation, and not, as I suggested earlier, AS4.) Wanting to remain neutral on how exactly  $V+$  is to be valued, and according to what kind of moral theory, Sàágua and Baumtrog ask only that the arguer be able to justify his choice—this is the point of CQ1.<sup>21</sup> Also wanting to

---

20 Whatever other goals may be incompatible with  $G^*$  will *eo ipso* be compatible with  $!G$ , and so on.

21 In short, they want to critically test the choice of moral standards without, however, dictating any substantive moral theory or principles. One wonders how much they can hope to achieve this way, provided that the reasoner avoids self-inconsistency. What Baumtrog wants to achieve is to make the practical reasoning of Anders Breivik unreasonable, or at least make it more difficult for extremists of all kinds to reason their way into performing acts of terrorism:

[W]hile there is no substantive moral assumption in the model, the model does assume that morals are a matter of reason. As such, it expects every perspective to be subject to the realm of reason. Thus, in the case where a moral is presented without reason, say on mere faith [as opposed to reasoned faith which is a different story], any reason supporting an alternate moral would win the day even if only in a “1-0” reason victory. Where a conflict of morals occurs, genuine argumentation is expected to bring out the side with the strongest reasons supporting it at the time. Where opposing reasons are presented, the opposition is expected to be resolvable by recognized argumentative means. [Baumtrog, 2015, p.123-4]

remain neutral on whether a probability measure or some kind of plausibility measure is better to use, they ask only that the arguer be able to justify this choice too—this is the point of CQ2.

- a. Stage 3, Topic Question 4: Are there means  $M_1, \dots, M_n$  that are simultaneously necessary and sufficient for  $X$  to achieve  $G^*$ ?

There are three argumentation schemes relevant to Stage 3 and this topic question.

#### **Argumentation Scheme 11. Necessary Condition Argument (AS11)**

Premise 1:  $X$  has the objective of realizing  $G^*$ .

Premise 2:  $\{NM_1, \dots, NM_n\}$  are necessary means for  $X$  realizing  $G^*$ .

Therefore, plausibly

Conclusion:  $X$  has a reason to carry out  $\{NM_1, \dots, NM_n\}$ .

#### **Critical Questions for Argumentation Scheme 11**

CQ1: What makes it plausible that  $\{NM_1, \dots, NM_n\}$  are necessary means for  $X$  realizing  $G^*$ ?

How could any of these means be suppressed while still allowing for the realization of  $G^*$ ?

#### **Argumentation Scheme 12. Sufficient Condition Argument (AS12)**

Premise 1:  $x$  has the objective of realizing  $G^*$ .

Premise 2: If  $x$  carries out  $\{NM_1, \dots, NM_n\}$ , then  $x$  realizes  $G^*$ .

Therefore, plausibly

Conclusion:  $x$  has a reason to carry out  $\{NM_1, \dots, NM_n\}$ .

#### **Critical Questions for Argumentation Scheme 12**

CQ1: How does carrying out all of the necessary means guarantee the realization of  $G^*$ ?

How might  $G^*$  remain unrealized despite carrying out the necessary means?

#### **Argumentation Scheme 4b. Argument from Ability (AS4b)**

Premise 1:  $M^*$  are necessary and sufficient (or at least sufficient) for  $G^*$ .

Premise 2:  $X$  has the ability to realize  $M^*$ .

Premise 3:  $X$ 's ability to realize  $M^*$  is a necessary/enabling condition for the realization of  $G^*$ .

---

I do not find it obvious what Baumtrog's argument is for such optimism. Perhaps someone who takes a moral view on faith is thereby unable to give reasons in response to CQ1, and this makes his reasoning bad. Or perhaps it is the reasoners' failure to ask themselves this question in the first place during their reasoning that is some kind of vice which makes their acting unreasonable. At best, this rules out a very limited number of cases. I am not going to be saying much about this. I will make the technical point that, insofar as their intention is to extend their model to include evaluation of our reasoning about values, that evaluation occurs here in CQ1 of AS10.

Therefore, plausibly,

Conclusion: There is a reason for X assuming  $M^*$ .

**Critical Questions for Argumentation Scheme 4b (AS4b)**

CQ1: How does X have the ability to realize  $M^*$ ?

What could prevent X from realizing  $M^*$ ?

CQ2: To what extent is the assumption of  $M^*$  by X a necessary/enabling condition for the realization of  $M^*$ ?

Is there any  $y$  (such that  $y \neq x$ ) whose ability to realize  $M^*$  is a necessary/enabling condition for the realization of  $M^*$ ?

Commentary: There is no Critical Question connected with Premise 1 in either AS11 or AS12 since this premise has been established in the earlier stages (assuming that “the objective of realizing  $G^*$ ” is the same as assuming  $G^*$  as a goal, which I have argued that in fact it might not be). Nor is there one in AS4b, which follows AS4a in this regard. As usual, the other Critical Questions elicit reasons for the other premises’ being true. It should be observed that now we are talking, and rightly so, about instrumental reasons: assuming  $M^*$  is a means to bringing about  $G^*$ .

While it is plausible that AS11 and AS12 both provide *pro tanto* reasons for X to carry out the means, neither individually provide reasons for answering “Yes” to the Topic Question. AS11 provides reasons for answering “Yes” to “Are there means  $M_1, \dots, M_n$  that are necessary for X to achieve  $G^*$ ?”, and AS12 provides reasons for answering “Yes” to “Are there means  $M_1, \dots, M_n$  that are sufficient for X to achieve  $G^*$ ?”, but neither answer the question “Are there means  $M_1, \dots, M_n$  that are *simultaneously* necessary and sufficient for X to achieve  $G^*$ ?” I think Sàágua and Baumtrog mean to say that *both* AS11 and AS12 must be appealed to,<sup>22</sup> but you do not thereby get two reasons for answering “Yes” or two reasons for X to carry out the means. In fact, it seems to me that Topic Question 4 is redundant and that Stage 4 should consist only of Topic Question 5; as long as there are sufficient conditions for X’s achieving  $G^*$ , then X has reason to realise those conditions and proceed to the next stage, and it should be noted that both topic questions, on receiving the answer “Yes”, go to the same Topic Question at Stage 4. Conditions that are simultaneously necessary and sufficient are just a special case of conditions that are sufficient.<sup>23</sup>

---

<sup>22</sup> I am assuming that that it is the same means that are referred to by  $NM_1 \dots NM_n$  in the instances of both AS11 and AS12. In other words, I am assuming that X has a set of means and asks firstly whether these means are necessary, and secondly whether they are sufficient. If the answer to both questions is “Yes” then clearly these means are simultaneously necessary and sufficient.

<sup>23</sup> The fact that certain conditions may be necessary, and a part of all sets of sufficient conditions, does not add anything (and is not actually taken account of in the model anyway—the model only considers necessary conditions

As for AS4b, like AS4a it concludes with an instrumental reason. Even on a more charitable construction, it should be clear that AS4b cannot license answering “Yes” to “Are there means  $M_1, \dots, M_n$  that are simultaneously necessary and sufficient for  $X$  to achieve  $G^*$ ?”, or insofar as it does, it does so precisely because its first, unquestioned, premise is “ $M^*$  are necessary and sufficient (or at least sufficient) for  $G^*$ ”, in which case the other premises are redundant as far as answering the topic question is concerned. Perhaps what Sàágua and Baumtrog have in mind is that having established this premise by use of AS11 and/or AS12 (which do question this premise)  $X$ ’s having the ability to carry out the means gives  $X$  an extra reason to carry out the means. AS4b can be appealed to in addition to AS11/AS12 but is useless otherwise. But if so then it seems to me that it belongs to a later stage, as it does not contribute in any way to answering the Topic Questions asked at Stage 3, which only concerns the existence of the means of achieving  $G^*$ , and not  $X$ ’s ability, or anybody else’s, to carry out those means.

a. Stage 3, Topic Question 5: Are there means  $M_1, \dots, M_n$  that are sufficient for  $X$  to realise  $G^*$ ?

There are two argumentation schemes here, one of which is once again AS4b. Here is the other:

### **Argumentation Scheme 12. Sufficient Condition Argument (AS12G)**

Premise 1:  $X$  has the objective of realizing  $G^*$ .

Premise 2: If  $X$  carries out  $\{NM_1, \dots, NM_n\}$  and  $\{SM_1 \text{ or } \dots \text{ or } SM_n\}$ , then  $X$  realizes  $G^*$ .

Therefore, plausibly

Conclusion:  $X$  has a reason to carry out  $\{SM_1 \text{ or } \dots \text{ or } SM_n\}$  (in addition to the reasons  $X$  may have to carry out  $\{NM_1, \dots, NM_n\}$ ).

### **Critical Questions for Argumentation Scheme 12**

CQ1: How does carrying out at least one of  $SM_i$ , where  $SM_i$  belongs to  $\{SM_1, \dots, SM_n\}$  (in addition to  $\{NM_1, \dots, NM_n\}$ , if there are such) guarantee the realization of  $G^*$ ?

How might  $G^*$  remain unrealized in spite of carrying out all of these means?

---

when they are at the same time collectively sufficient). Possibly Sàágua and Baumtrog are influenced here by Audi’s (2006, pp.110-11) necessary condition schema and sufficient condition schema. However, for Audi a stronger reason ensues upon a necessary condition than a merely sufficient condition. In the case of the latter “[s]ome kind of *prima facie* judgment does follow, since  $S$  has some reason to  $A$ ” which is even weaker than the (already weak) reason derived from the necessary condition (weak because achievement of an end may be made only marginally more likely on the satisfaction of its necessary condition). If the same kind of reason is to apply to both (i.e., to be the conclusion of both AS11 and AS12), it must be the weaker, and this seems to be the *pro tanto* reason they have insisted on throughout. In short, if they want the distinction between necessary and sufficient schemes to have any point, their schemes would need to have different conclusions and refer to different strengths of reason. But since the weakest kind of reason is all that is required to get to the next stage anyway, there is no real motivation for making the distinction, and so we may as well skip Topic Question 4.

Commentary: I would point out a certain ambiguity in AS12G's formulation. Is SM1 a sufficient condition in itself (in which case it already includes all the necessary conditions) or is it whatever further condition has to be added to the necessary condition in order for the conditions to be jointly sufficient? I think they intend the latter, for otherwise it would make no sense to say that X has a reason for SM1 (for example) that is additional to the reasons for NM1 . . . NMn (as one would end up double-counting reasons); also, it seems to be an implication of CQ1. The next ambiguity is in the conclusion and concerns the scope of the reason. Is the reason a reason for the disjunction (and not necessarily for each disjunct) or for each disjunct? If SM1 is the kind of condition that I have just mentioned, then X does have a *pro tanto* reason to carry it out. The same applies to SM2 and SM3, and so on. Which of the 'sufficient' conditions one ends by having a conclusive reason to bring about will depend on further factors (and there may not necessarily be a uniquely best condition to bring about), but that is the nature of *pro tanto* reasons.<sup>24</sup> I think that this is probably what Sàágua and Baumtrog intend. Even so, it seems to me that the whole thing is redundant, as the Topic Question is simply Premise 2 in AS12G, just as Premise 1 was in AS4b! Answering CQ1 amounts to answering Topic Question 5 without any help from the argumentation scheme.

- a. Stage 4, Topic Question 6: Are the selected means M1, . . . , Mn compatible with the objectives G1, . . . , Gn that X has or should have?

There are four argumentation schemes here: AS5b, AS6b, AS7b, AS8b. These are the same as AS5, AS6, AS7, and AS8 except instead of goals being incompatible with other goals, it is means (the set M1, . . . , Mn selected by the previous reasoning).<sup>25</sup> One wonders how exactly means can be incompatible with, or preferable to, goals, and whether this is really distinct from the kind of considerations at issue in Stage 2. However, consider that there may be three distinct ways of bringing about G\*, and while the first one M1 inhibits the achievement of goal Gn, the other two do not. Therefore, G\* itself is compatible with Gn but G\*-ing by M1-ing is not compatible with Gn, that is to say, G\* and Gn cannot both be achieved if the means chosen to achieve G\* is M1. It makes sense, then, to compare means with goals as well as to compare different goals. Sàágua and Baumtrog conceive of this as operating itera-

---

<sup>24</sup> Suppose that one has a reason for satisfying a disjunction, and that reason is *not pro tanto*. It seems to me that one then has a *pro tanto* reason (though not a *non-pro-tanto* reason) for satisfying any of the disjuncts—it is a consideration in any of the disjuncts' favour that by satisfying it one satisfies the disjunction. Distribution of reasons across the disjunction to the disjuncts is valid when the reasons so attached to the disjuncts are *pro tanto*, whether the reason for the disjunction itself was *pro tanto* or not.

<sup>25</sup> This was indicated in those schemes using the device G\*/M\*. G\* referred to the goal and was the thing in question in those earlier stages. The variants being discussed at this stage take the M\* side, M\* referring to the selected means.

tively, so that a means that is incompatible (like M1 in the example) is excluded from further consideration if there is another that is not incompatible. Only if all the possible means are incompatible do we need to decide which is preferable. But, if this is the case, then I do not see how  $G^*$  can be compatible with  $G_n$  when *all* the possible means of achieving  $G^*$  are incompatible with  $G_n$ . And if this is right, then there will *always* be a means of achieving  $G^*$  that is compatible with  $G_n$ . It is tempting then to think that an affirmative answer to Topic Question 6 is guaranteed by affirmative answers to Topic Questions 3 and 5. But I don't think so, for the reason that while there must be a means of achieving  $G^*$  that is compatible with  $G_n$ , there may not be a single means that is compatible with all the goals  $X$  should have. Suppose that M2, which is compatible with  $G_n$ , is incompatible with  $G_m$  which is not itself incompatible with  $G^*$  but would be preferable if both  $G^*$  and  $G_m$  could not both be brought about, which would indeed be the case if  $G^*$  is brought about by M2. Although for each goal  $G_i$  incompatible with a means to  $G^*$  there must be another means to  $G^*$  that is compatible with  $G_i$ , there may not be a means that is compatible with every such  $G_i$ .<sup>26</sup>

Commentary: Obviously, everything I said about the AS5, AS6, AS7, and AS8 apply equally to AS5b, AS6b, AS7b, and AS8b. The basic idea seems to be that if we can carry out the means without thereby doing something that is incompatible with a goal that we ought to have, then we can go ahead and do it. Thus, an affirmative answer to this Topic Question leads at stage 5 to the directive "Take Action."  $X$  has a *pro tanto* reason to assume  $G^*$  (established at Stage 1) and, because of this, a *pro tanto* reason to carry out a selected means  $M_1 \dots M_n$  (established in Stage 3) to achieving  $G^*$  and no *pro tanto* reasons not to carry out this means. If I have a criticism here it is that I think that  $X$ 's ability to carry out this means (mentioned in regards to AS4b) ought to come in here—plausibly, if  $X$  does not believe that he can carry out the means, his reasoning should not make it incumbent upon him to carry it out. Something has gone funny with your reasoning if it leads you to try to carry out actions that you do not believe yourself able to carry out (even if, in fact, you *are* able to carry them out). Ability has to figure into practical reasoning somewhere, but not where Sàágua and Baumtrog put it. Believing that you can do it does not create an extra reason for you to do it (except, possibly, instrumentally), but believing that you cannot do it does create a reason for you not to attempt it.

a. Stage 4, Topic Question 7: Are the selected means  $M_1, \dots, M_n$  preferable to the goals  $G_1, \dots, G_n$  with which they are incompatible?

---

26 The difference is between "For every goal  $G_i$  compatible with  $G^*$ , there exists a means  $M_i$  for achieving goal  $G^*$  that is compatible with  $G_i$ " and "There exists a means  $M_i$  for achieving goal  $G^*$  that, for every goal  $G_i$  that is compatible with  $G^*$ ,  $M_i$  is compatible with  $G_i$ ". Only the first is true, whereas the reasoning being rejected requires the second.

There is one argumentation scheme here: AS10b. This is the same as AS10 except where AS10 talks about the preferability of goals over other goals, AS10b talks about the preferability of means over other goals. Only if all the means  $M_1, \dots, M_n$  are incompatible with a goal  $X$  should have is this Topic Question asked (Sàágua and Baumtrog, 2018, p.226).

Commentary: The obscurity of how exactly the weighing of reasons is supposed to take place recurs here. It is simply stated in Premise 4 of the scheme which is preferable out of the assumed goal/the selected means and the goal that is incompatible with this, and the conclusion says little more than to prefer the preferable. Perhaps the idea is that as long as  $X$ 's choice of preferability relation is not unreasonable then, even if it is wrong and a different relation is objectively better suited to the task,  $X$ 's wrongness is not that of bad practical reasoning.

#### a. Stage 5

There is no topic question here. Traversing the model is to result in taking action or not taking action, for the reasons generated by the argumentation schemes appealed to in the earlier stages.

Commentary: I will repeat here my skepticism about how a reason that is only *pro tanto* can license acting in this stage. Again, I think that the idea is meant to be that when we have established at Stage 4 that there is a set of means that is compatible with the preferable set of goals, this makes the reasons for those means conclusive. This would mean in effect that the argumentation schemes used at Stage 4 do *not* generate *pro tanto* reasons at all but *all things considered* reasons, reasons that could then be detached and taken to license acting, or at least be a reason to form that intention. But the problem that remains is that by Sàágua and Baumtrog's account this is all contextual, and, taking the model as a model of argumentation, relative to a dialectical situation. It is quite possible then for the same arguer to be at the same time in a context or dialectical situation where such an action was not licensed or even prohibited, and even if that is not the case at the current moment, he knows that he could be in that situation. Is his acting then justified overall? He can give reasons for his acting that would be acceptable to a particular set of interlocutors, certainly, but it might still be wiser to wait than to act.

## 5. Conclusion

"If we are to be responsible academics, we must then justify the need for our current contribution; we must answer the question 'What is on offer here that cannot be found elsewhere?'" Sàágua and Baumtrog ask, and answer that "the new model is an *integrated*

model of practical reasoning and argumentation” which they then cash out by saying firstly that its components can be interpreted in term of reasons in an inferential process or as commitments in an argumentation, and then by saying that argumentation schemes should be interpreted in a way such that their conclusions are (*pro tanto*, contributing) reasons. These two claims are in conflict: if it is built into the formal structure of the argumentation scheme that its conclusion is a reason, then I do not see how it can be re-interpreted as a commitment. Certainly, we can reason as well as argue using argumentation schemes, but only because their formal structure does not dictate them being interpreted one way or another; it is only insofar as reasonings and argumentations have the same structural elements that re-interpretation is possible. All models of reasoning and argumentation are integrated to this extent, so it can hardly be claimed by Sàágua and Baumtrog to be a special virtue of their model that it does this too, and by building reasons into the formal structure of the scheme it is not clear that they can even claim this virtue. Also, when we do reason using argumentation schemes, we do not reason about what reasons we have; building reasons into the formal structure of the scheme gives an unnatural view of reasoning as second-order reasoning about our reasons when used as a model of reasoning. So, their model is not even as integrated as other models and is unnatural as a model of reasoning. Yet this is the only really original part of their proposal, as otherwise their model is no better than other models, which is to say that they would not be offering anything in their model that would not be found elsewhere and by their own account they would be irresponsible academics.

Nor are they even consistent in their use of *pro tanto* reasons. They fail to distinguish between the instrumental reasons that are properly the subject of discussion in Stages 3 and 4 from the objective moral or normative reasons that seem to be at issue in Stages 1 and 2, and this causes them to miscount the number of reasons in favour of assuming  $G^*$ . I have also explained, in connection with this, that it would be better to consider only the goals  $X$  *should* have and not those that  $X$  just has. Also, at stage 3 there is a mismatch between the conclusions of the Necessary and Sufficient Condition Arguments and the topic question they are intended to answer. The *pro tanto* reason concluded in the Necessary Condition Argument is that there is a reason for  $X$  to carry out what is necessary, and not to answer “Yes” to “Are there means  $M_1, \dots, M_n$  that are simultaneously necessary and sufficient for  $X$  to achieve  $G^*$ ?” Obviously, we cannot answer “Yes” using only this argumentation scheme because the scheme only concerns necessary conditions. Sàágua and Baumtrog would argue that we have to apply the schemes sequentially so that when we have applied the Sufficient Condition Argument as well, we have a set of necessary conditions that are jointly sufficient. Even so, the conclusion is once again to carry out these means and not to answer “Yes” to “Are there means  $M_1, \dots, M_n$  that are simultaneously necessary and sufficient for  $X$  to

achieve  $G^*$ ?” Also, there may not be *any* means that are necessary for  $X$  to achieve  $G^*$ , so really the topic question should be more simply “Are there means  $M_1, \dots, M_n$  that are sufficient for  $X$  to achieve  $G^*$ ?” as in Topic Question 5. It is true to say that one *can* answer “Yes” to this question, but only because Premise 2 of the Sufficient Condition Argument AS12G is “If  $X$  carries out  $\{NM_1, \dots, NM_n\}$  and  $\{SM_1 \text{ or } \dots \text{ or } SM_n\}$ , then  $X$  realizes  $G^*$ ”, and this is only another way of saying that these means are sufficient for  $X$  to achieve  $G^*$ ! In other words, Premise 2 simply says that you can answer “Yes” to the topic question. Hence, not only can we use the Sufficient Condition Argument and eliminate the Necessary Condition Argument, we do not even need all of the Sufficient Condition Argument but only its second premise and possibly the critical questions associated with testing this premise. Stage 3 more or less vanishes. Finally, I have argued that they implicitly treat the reasons produced at stage 4 to be all things considered reasons and not *pro tanto* reasons at all; if they really are *pro tanto* reasons, I don’t see why someone who has such a reason could not say “Well, I’ll wait and see. My reasons are only *pro tanto* after all, and only follow defeasibly on my premises and provisionally on my current context. I don’t want to act and then find myself in a context where it would have been better not to act.” Only if there is a necessity to act immediately would such an agent be rationally motivated to act.<sup>27</sup>

This is not to say that the argumentation schemes in Sàágua and Baumtrog’s model are complete rubbish. I would argue, though, that their plausibility relies entirely on their resemblance to reason transmission principles, to which Sàágua and Baumtrog have nothing original to add and even seem unaware of the controversies regarding the formulation of such principles; for example, the schemes all seem to be narrow-scope rather than wide-scope.<sup>28</sup> It is because of the resemblance of the argumentation schemes to such principles that makes expressing their conclusion as reasons superficially attractive. Nonetheless, it is an error of an extremely fundamental sort, and not only but not least because this makes it powerless to resolve at least a core kind of conflict (i.e, where there are *pro tanto* reasons on both sides). Also, such principles are descriptions from an external third-personal viewpoint of what

---

27 Would it make a difference if one were arguing only with oneself? At [2018, p.196] Sàágua and Baumtrog say: “Only when someone reasons or argues by himself, does the argumentation immediately result in an intention to perform the action (or not).” But I don’t think this avoids the objection if the inference is only defeasible. Suppose that you conclude that Tweety flies because it is a bird: although you know that there are penguins and other flightless birds, you are in a state of ignorance about whether Tweety is one of these kinds of birds, and you know that you are in such a state of ignorance. It is not obviously irrational to wait and try to answer this question before committing yourself to the conclusion. Moreover, it seems plausible that you *do* have a *pro tanto* reason to think that Tweety flies on the grounds that it is a bird and birds generally fly (and if so the inference is not actually defeated after all, at least if the conclusion is that there is such a *pro tanto* reason) but again it is not obviously irrational to hold fire on committing oneself on whether Tweety flies.

28 This is despite claiming the influence of Broome, who is one of the most notable arguers for the wide-scope view. I think this is possibly due to the influence of Fairclough and Fairclough (2012) as explained in footnote 14.

objective reasons someone has or what subjective reasons someone ought to have if they are rational, and does not describe reasoning itself, whether our own or even that person's. We do not reason using Reasoning Modus Ponens but by using *modus ponens*. A model that uses such principles, then, is not a model of practical reasoning as such, or at least it is not a model of first-order reasoning but treats practical reasoning as second-order reasoning.<sup>29</sup>

An integrated model of reasoning and argumentation is a false ideal, because reasoning and argumentation have different norms. A model should be integrated to the extent that a piece of reasoning is structurally identical to a piece of argumentation. But Sàágua and Baumtrog's model fails to be even as integrated as this because of their misjudged attempt to put reasons into the formal structure of the argumentation schemes, and this resists any re-interpretation in terms of commitments. Far from being able to justify its existence on the grounds of its integratedness, as they claim, their model is a disaster on this count. But if they recanted on this point, then the best that could be said for their model is that it was no worse than other models; their reconceptualization of argumentation schemes is their one real claim to novelty, and without it their own justification for the model collapses. Nor do they seem able to offer anything beyond traditional utility/choice theory as a way of weighing which goals to pursue, except to re-express the same problem in a confusing way as the weighing of reasons, whereas the weight of reasons can only be seen as a way of expressing the output of the utility calculation and not its input. In fact, it is a premise of their argumentation schemes at this stage that one set of goals is preferable to another, and not something that the schemes conclude.

Of the two ways in which their model is extended – the evaluation of the choice and/or preference-ordering of values, and its being applicable to both reasoning and argumentation – the second has turned out to be a disaster, and it is what I have focussed on in this paper. I have not said much about the first, and reasons of space forbid my doing so here. Baumtrog believes that we can argue about values, but ironically it is not clear whether this is anything more than an article of faith on his part. I would suggest that he pose the same critical question to himself: what reasons can I give for the claim that disagreements about values can be resolved rationally that are acceptable to someone who does not believe this?

---

29 Admittedly, one *could* have first-order reasoning by following such transmission principles, e.g., by saying in the second-order account that because one believes that there is objective reason for the goal, and in such a case there is objective reason for the means, one should also believe that there is objective reason for the means; in other words, the objective reasons become part of a first-order reasoning (as usual, the attitude drops out of the first-order account, making the objective reason analogous to *p* in an ordinary *modus ponens*—the reasoner takes there to be an objective reason *R* just as he takes *p* to be true). Even so, it would not be very natural. We do not in general reason about what reasons we have until that point where we wish to weigh reasons against each other, which is not what we are talking about here.

Without making substantive moral judgments about the Nazis' belief system, can we really conclude that they were unreasonable and/or guilty of bad practical reasoning? Setting aside those cases where we can trap such reasoners in self-inconsistencies, it is not obvious to me that we can.

But even if more belief systems are internally incoherent than I am supposing, it locates the error in the wrong place; it would be like objecting to torture not on the grounds that it is morally wrong but because it is not actually effective in getting people to tell the truth, that is to say, torture is objectionable on empirical grounds because torturers are factually mistaken about its effectiveness as a means; moreover, I do not see that the Nazis were wrong in any comparable way about the effectiveness of using Zyklon B to gas Jews. We want to say more about such cases that they made a mistake in inference or held beliefs on insufficient evidence! We do not want to allow an easy escape from a moral vice into an epistemic vice. We want to say that they are morally wrong and not just that they fell short of some ideal of rationality which we all fall short of very frequently! I suspect that this is the conclusion that the authors are aiming for and their rationale for adopting an objective conception of reasons, which is to say, ironically, that they have themselves chosen a means inappropriate to their ends: no criticism that does not make substantive moral assumptions has the means to evaluate an argument as having *morally* wrong premises or conclusions. At best critics can show that it results from bad reasoning, but to be guilty of bad reasoning somewhere is clearly a much less serious charge. Having an objective conception of reasons will only serve the end of evaluating the practical reasoning as bad if those evaluating the reasoning are allowed to criticise those reasons on material grounds. By their reluctance to make any such criticism, they may as well have stuck with a subjective conception of reasons, as indeed most models of practical reasoning do, since they take as their target instrumental reasoning and we reason instrumentally whenever we have a desire which we then reason about how to fulfil.

Sàagua and Baumtrog are guilty of bad practical reasoning: they choose means inappropriate to their ends. They choose a means to integration which I have argued is actually counter-productive for that purpose, namely re-conceptualizing schemes as having *pro tanto* reasons as conclusions. They rightly choose an objective conception of reasons in order to rule out certain choices of ends, but undermine this by including goals that the reasoner actually has (and in which objective reasons are not involved, although it seems that they fail to notice this) as well as those he should have (which do) as determining which actions to perform, allowing in the limit case where there are no conflicting goals the very choice of ends that they want to rule out as unreasonable. Finally, by choosing not to make any substantive moral assumptions of their own and thereby disallowing any challenge to

premises on the grounds of their material truth, the only criticism Sàágua and Baumtrog are allowed to make of a morally invidious choice of ends is that the agent is inconsistent and/or unreflective and not that they are immoral.

## References

- Audi, R. (2006). *Practical Reasoning and Ethical Decision*. New York: Routledge.
- Baumtrog, M. (2015). *Improving practical reasoning and argumentation*. Doctoral Thesis.
- Fairclough, I., & Fairclough, N. (2012). *Political Discourse Analysis: A Method for Advanced Students*. London: Routledge.
- Hitchcock, D. (2016). Commentary on Baumtrog's "Investigating the Impact of Moral Relativism and Objectivism on Practical Reasonableness". In D. Mohammed & M. Lewiński (eds.) *Argumentation and Reasoned Action: Proceedings of the 1st European Conference on Argumentation, Lisbon, 2015*. Vol. I, pp.47-51. London: College Publications.
- Sàágua, J. and M. Baumtrog. (2018). Practical rationality at work — a new argumentation model. In *Essays on Value and Practical Rationality*, eds. A. Marques and J. Sàágua, pp.193-230. Bern: Peter Lang International Academic Publishers.

# PRAGMÁTICA Y ARQUITECTURA MENTAL: UNA PROPUESTA DESDE LA MODULARIDAD MASIVA DE LA MENTE

\*

NICOLÁS ACUÑA LUONGO

Centro Interdisciplinario de Innovación Educativa, Facultad de Educación,  
Universidad Santo Tomás, Chile.

Dirección: Pasaje Los Peumos 1860. Talca

Este trabajo fue apoyado por la Agencia Nacional de Investigación y Desarrollo (ANID).

Proyecto FONDECYT N° 3210639

acuna.nicolas17@gmail.com

## Resumen

*Uno de los problemas fundamentales respecto de la naturaleza cognitiva de la pragmática se refiere al lugar que ocupa ella en la arquitectura de la mente. Al respecto existen varias propuestas. Una de ellas sostiene que forma parte de los sistemas centrales en el marco de la teoría de modular de Jerry Fodor. Sin embargo, Asa Kasher plantea una hipótesis basada en la arquitectura modular de Fodor, pero sí otorga un carácter modular al menos a algunos aspectos de la interpretación pragmática. En este trabajo sostenemos que ambas propuestas son incorrectas debido a inconsistencias con los hallazgos empíricos sobre procesamiento pragmático. Por el contrario, se aboga por una hipótesis basada en la modularidad masiva de la mente, en donde consideraciones evolutivas llevaron a que emerja un submódulo pragmático que es parte de un módulo más general de teoría de la mente.*

**Palabras clave:** pragmática, modularidad, modularidad masiva, arquitectura mental, ciencias cognitivas

### Abstract

*One of the fundamental problems with the cognitive nature of pragmatics concerns its place in the architecture of the mind. In this regard there are several proposals. One holds that part of the central systems within the framework of the theory of modular Jerry Fodor. However, Asa Kasher raises a hypothesis based on Fodor modulates architecture, but it gives a character to the modular least some aspects of pragmatic interpretation. In this work we argue that both proposals are incorrect due to inconsistencies with the empirical findings on pragmatic processing. Rather, we support a hypothesis based on the massive modularity of mind, where evolutionary considerations led to the emergence of a pragmatic sub-module that is part of a more general theory of mind module.*

**Keywords:** *pragmatics, modularity, massive modularity, mental architecture, cognitive sciences*

## 1. Introducción

En sus inicios la pragmática nació como una rama de la filosofía del lenguaje. Sin embargo, el desarrollo de este campo prontamente se movió hacia la interdisciplinariedad. Una de las teorías más importantes de la pragmática, la teoría de la relevancia (Sperber y Wilson, 2004; 1995), estudia cómo se construyen inferencias dentro de un procesamiento mental que es responsable de la interpretación de la comunicación humana. Con ello, la pragmática se movió hacia los estudios de la ciencia cognitiva. El resultado es un marco explicativo el cual está incluido en los estudios amplios de la cognición humana.

De acuerdo con este giro, la presente investigación considera a la pragmática de índole cognitiva como el estudio de los estados mentales de las personas que participan en un evento comunicativo (Bara, 2010). El análisis de las interacciones comunicativas y sus estados mentales subyacentes implican determinar las motivaciones individuales, creencias, objetivos, deseos e intenciones y, especialmente cómo estos estados mentales se relacionan con el componente lingüístico de la comunicación. A su vez, los estados mentales y las elecciones lingüísticas que se emplean en un evento comunicativo están determinadas en distinto grado por las condiciones contextuales en que dichos eventos ocurren. Esta concepción delimita una serie de fenómenos comunicativos que caen dentro del dominio de la pragmática. Entre estos fenómenos podemos encontrar los actos de habla, las presuposiciones, las implicaturas, las expresiones deícticas, las reglas conversacionales, el lenguaje no literal, reglas de cortesía, etc. (Airenti, 2017).

Bajo este marco, uno de los temas que corresponde determinar es la descripción de la arquitectura cognitiva (Thagard, 2012) y el rol funcional de la pragmática al interior de dicha arquitectura (Carston y Powell, 2008; Fodor, 2000; Segal, 1996; Sperber, 2002; Sperber y Wilson, 2002; Wilson, 2005). Esto implica evaluar qué tipo de sistema es responsable del procesamiento pragmático. ¿Es un sistema de resolución de problemas de carácter general o dedicado? ¿Con qué otro tipo de sistema interactúa? La presente investigación tiene como objetivo evaluar diversas teorías respecto del lugar del procesamiento pragmático en la arquitectura mental. En este sentido, proponemos que una teoría respecto del lugar del procesamiento pragmático en la arquitectura mental debe ser coherente con consideraciones propias de la psicología evolutiva concerniente al tipo de sistema cognitivo que ha sido naturalmente seleccionado para resolver un problema adaptativo particular (Sperber, 2002). Ello conlleva preguntarse ¿responde el procesamiento pragmático a un problema adaptativo que ha llevado a que emerja un sistema cognitivo que se encargue de dichos problemas?, o en su defecto ¿basta un mecanismo de orden general para explicar el procesamiento pragmático?

Probablemente la posición más influyente en torno a la arquitectura cognitiva se atribuye a Jerry Fodor (1983; 2000). Desde su perspectiva la mente posee una arquitectura híbrida. Ella se compone, por un lado, de sistemas perceptuales de input, incluyendo la percepción del lenguaje; sistemas de outputs y por otro, de un sistema central. Mientras que los sistemas de input y outputs son módulos mentales autónomos, el sistema central no es modular y es responsable de la formación de creencias y de la toma de decisiones, entre otras conductas. Para Fodor (2000), los procesos de inferencia pragmática son una función del sistema central: su objetivo es fijar una creencia respecto del significado del hablante (el contenido de su intención comunicativa) y dicho proceso es altamente sensible al contexto. Por lo tanto, el autor concluye que la fase de decodificación lingüística de un enunciado puede ser realizada por un sistema modular, mientras que la fase inferencial, la cual vincula la brecha entre el significado lingüístico y el significado pretendido por el hablante, es no modular. Sin embargo, también existen posiciones, dentro del marco de la teoría de la arquitectura mental de Fodor, que sostienen que la dimensión pragmática del lenguaje tiene un componente modular (Kasher, 1991). Bajo esta teoría, ciertos fenómenos pragmáticos se procesan al interior de un módulo pragmático lingüístico, mientras que otros ocurren en el sistema central. Desde nuestra perspectiva se considera que ambas propuestas poseen problemas que impiden explicar idóneamente, dentro del contexto de la psicología evolutiva, cómo ocurre el procesamiento pragmático. Por el contrario, nuestro objetivo es apoyar la idea de que la pragmática sí constituye un módulo especializado, pero dentro de un marco de una propuesta distinta al marco fodoriano. A saber, la teoría de la modularidad masiva de la mente (Carruthers, 2006; Sperber, 2002). En esta teoría, la mente se compone exclusivamente de módulos especializados dentro de los cuales uno de ellos estaría dedicado al componente pragmático del lenguaje (Sperber, 2000; Sperber y Wilson, 2002; Wilson, 2000; 2005).

En la primera sección se expondrán los principales postulados de la teoría modular de la mente de Jerry Fodor (1983; 2000). Se pondrá énfasis en las características que poseen los módulos de la mente dentro de esta teoría. Finalizaremos esta sección con el lugar que otorga dicho autor a la pragmática en el marco de su teoría modular. En el segundo apartado revisaremos críticamente la teoría modular pragmática de Asa Kasher (1991). Dicho autor asume la teoría fodoriana de la mente, pero se diferencia de Fodor en el sentido en que sí existe un módulo pragmático lingüístico encargado de al menos, algunos fenómenos pragmáticos (la fuerza de los actos de habla). A continuación, presentaremos una serie de objeciones a la propuesta de Kasher que nos llevarán a concluir que su propuesta es poco plausible.

En el tercer apartado presentamos las bases para la propuesta respecto del problema que guía la presente investigación. Nuestra propuesta afirma la existencia de un módulo

pragmático, pero no dentro del marco de una teoría modular fodoriana, sino dentro de la teoría de la modularidad masiva (Carruthers, 2006; Sperber 2002; Wilson, 2005). Para ello, en la tercera sección se presentarán los principales aspectos de dicha teoría, enfatizando el concepto amplio de módulo que utilizan diferenciándolo con la propuesta de Fodor (2000). En el cuarto apartado presentamos argumentos en favor de la existencia de un submódulo pragmático que forma parte de un módulo más amplio responsable de la atribución de estados mentales (habilidad frecuentemente denominada Teoría de la Mente). Dicho submódulo pragmático se apoya en un par de principios que regulan la comunicación (Sperber y Wilson, 1995; 2004) y de procedimientos de procesamiento exclusivos para la detección de los significados intencionales de índole lingüística y también ostensiva. Finalmente, concluimos con una serie de proyecciones y asuntos que debe asumir una teoría pragmática modular en el contexto de la teoría modular masiva.

## 2. La teoría modular de la mente y el procesamiento pragmático

Una condición general que toda teoría del funcionamiento de la mente debe explicar es cómo la mente construye representaciones de estados de cosas que son externas a ella (Stillings, y otros, 1995; Von Eckardt, 2012). Fodor (1980) sostiene que las representaciones consisten en estructuras semejantes a las estructuras lingüísticas (*language-like structures*) dentro de un lenguaje interno del pensamiento. Sin embargo, estas consideraciones generan otro tipo de preguntas que es necesario abordar. Una de ellas se refiere a cuál es la naturaleza de la relación entre esas representaciones. Según Fodor (1980), esas relaciones son expresadas en términos de propiedades formales y sintácticas que se separan de las propiedades semánticas. Esta distinción entre los atributos semánticos de las representaciones y las propiedades sintácticas internas es integral a las teorías de Fodor (1980; 1983; 2000) y a las de Chomsky (1980).

Fodor explica de qué forma esas representaciones son manipuladas para realizar las variadas operaciones cognitivas de nuestra mente. Para ejecutar esas operaciones, como por ejemplo las inferencias o el procesamiento lingüístico, la mente debe ser capaz de manipular las representaciones a las que accede. La manipulación procede por medio de propiedades sintácticas y formales de las representaciones de un modo semejante a las computaciones que ejecuta un computador (Fodor, 1980).

El contenido semántico de las representaciones no es esencial a las operaciones de manipulación de dichas representaciones. Consideremos el siguiente ejemplo. Supongamos que nuestra mente tiene acceso a dos representaciones P y Q, donde P corresponde a la proposición 'Héctor está feliz' y Q corresponde a la representación 'Si Héctor está Feliz entonces irá a bailar'. La conclusión basada en la regla *modus ponens* es R ('Héctor irá a bailar'). Para llegar a R es necesario realizar una computación que procede desde las propiedades

formales de P y Q y, específicamente, a partir de la identidad formal de P con el antecedente de Q. Que ‘Héctor’ refiere a un hombre particular y que “irá a bailar” refiere a la discoteca de la calle principal son representaciones semánticas de P y Q y, como tal, no son parte de la manipulación formal de las representaciones.

La teoría modular propone que la mente posee una compleja organización estructural. A nivel global, la mente posee distintas facultades innatas interrelacionadas. Por ejemplo, Chomsky (1966) sostiene que existe una facultad innata del lenguaje en nuestras mentes. Esta facultad involucra manejar los aspectos computacionales del lenguaje. Estos aspectos incluyen las reglas fonológicas, semánticas, pero principalmente las reglas sintácticas de carácter universal.

A nivel local, la mente se organiza en torno una serie de “módulos cognitivos”. Revisemos a continuación cada una de las características que poseen los módulos según Fodor (1983).

- *Los módulos son de dominio específico:* Cada módulo está restringido al tipo de contenido que puede tomar como input. Solo puede procesar aquellos contenidos que constituyen su dominio. Así, el sistema visual está restringido a los inputs visuales; el sistema auditivo a los inputs auditivos, etc.
- *Los módulos generan outputs simples o superficiales (no conceptuales):* el output de un módulo es superficial en el sentido en que no es conceptual. Por lo tanto, los módulos generan información de varios tipos, pero ellos no generan creencias o estados mentales. Por el contrario, según Fodor (2000), la fijación de creencias es un arquetipo de sistemas no modulares. Por ejemplo, el módulo visual podría entregar una representación de la superficie y los contornos de algún objeto del campo visual, pero no podría identificar dicho objeto con una silla, ni mucho menos generar la *creencia* de que la silla está presente.
- *Los módulos son innatos y localizados:* El desarrollo de los módulos y su crecimiento se encuentra bajo el control genético. Además, ellos se encuentran localizados en estructuras específicas del cerebro.
- *Los módulos son obligatorios y rápidos en su procesamiento:* Sus operaciones no están bajo el control voluntario, y ellos generan outputs más rápido que el sistema central no modular.
- *Los módulos están encapsulados y son inaccesibles:* Ellos se encuentran aislados de los demás módulos de la cognición y poseen operaciones internas que son inaccesibles a otros módulos. Estas dos propiedades están vincu-

ladas a la especificidad de dominio, pero son distintas de ella. Decir que un sistema de procesamiento está encapsulado es decir que sus operaciones internas no pueden generarse en virtud de información contenida fuera de dicho módulo. En contraste, sostener que un sistema es inaccesible significa que los demás módulos no poseen acceso a su procesamiento interno, sino que solo a sus outputs. Por otra parte, la especificidad de dominio se refiere a las restricciones sobre el input de un sistema. Afirmar que un sistema es específico de dominio significa que puede procesar solo inputs de un tipo particular, mientras que las propiedades de la encapsulación e inaccesibilidad corresponden a relaciones de acceso que se obtienen entre las operaciones internas de esos sistemas y los demás módulos. Por lo tanto, es posible que un módulo pueda carecer de especificidad de dominio y a la vez estar encapsulado y ser inaccesible (Carruthers, 2006; Sperber, 2002).

Los módulos cognitivos proveen los inputs a un sistema central el cual, a diferencia de los módulos, no son específicos de ningún dominio y tampoco están informacionalmente encapsulados. Este sistema central es el lugar donde ocurren las inferencias y, como tal, está asociado a procesos de formación de creencias y resolución de problemas.

¿Qué lugar tiene la pragmática en esta propuesta arquitectónica de la mente? De acuerdo con la perspectiva fodoriana clásica, la pragmática y las habilidades de lectura de intenciones (*mind-reading*) son sistemas cognitivos centrales los cuales son utilizados para atribuir estados mentales a los demás sobre la base de habilidades de razonamiento de propósito general. Fodor (1983; 2000) combinó la perspectiva chomskiana del lenguaje (Chomsky, 1966; 1980) con la perspectiva griceana de la pragmática (Grice, 1975; 1989). En su teoría, el lenguaje es un sistema modular autónomo y de dominio específico y el procesamiento pragmático es parte del sistema central de dominio general. La diferencia entre ambos sistemas no es el hecho de que la pragmática sea de carácter inferencial y el lenguaje no, sino que las inferencias pragmáticas son fuertemente dependientes del contexto y no existe en principio ningún límite respecto del tipo de información contextual que podría requerirse ni tampoco la fuente de dicha información. Para Fodor (2000), la interpretación pragmática es un proceso central debido a que su resultado depende de factores “globales” tales como el acceso libre a información de la memoria a largo plazo. Los procesos modulares, en contraste, son ‘locales’, ‘encapsulados’ y fuertemente restringidos al tipo de fuente de información en la cual ellos operan. De esta manera, el procesamiento pragmático en la arquitectura mental de Fodor no constituye un módulo específico de la mente, sino que es parte del aparato inferencial del sistema central multipropósito.

### 3. La pragmática modular de Kasher

Kasher (1991) asume que la modularidad es una descripción correcta de nuestras capacidades mentales. Su objetivo, sin embargo, es sostener que sí existe un componente modular de la pragmática. Para ello, Kasher propone la siguiente fórmula básica de la arquitectura mental de la pragmática:

$$\begin{aligned} \text{Arquitectura mental de la pragmática} &= \text{un módulo pragmático} \\ &+ \text{una parte pragmática del procesador central} \\ &+ \text{una interfaz pragmática.} \end{aligned}$$

Para llegar a esta descripción tripartita, Kasher examinó seis fenómenos pragmáticos: deixis (Cummings, 2005), presuposiciones léxico-pragmáticas (Cummings, 2005), fuerza de los actos de habla (Searle, 1975), actos realizativos (Austin, 1998), implicaturas conversacionales (Grice, 1975) y principio de cortesía (Leech, 1983). El autor argumenta que el procesamiento de los actos de habla requiere de un módulo propiamente pragmático que es lingüístico en su naturaleza; las implicaturas conversacionales y el principio de cortesía sugieren la presencia de un componente pragmático en el procesador central de la mente y los elementos deícticos junto a las presuposiciones pragmáticas proveen evidencia respecto de la existencia de una interfaz entre el módulo pragmático y el procesador central. A continuación, se presenta cada uno de los componentes de la arquitectura mental de la pragmática según Kasher (1991).

- Módulo pragmático

Kasher (1991) sostiene que el procesamiento de los actos de habla y, particularmente, los actos de habla indirectos requieren de la existencia de un módulo pragmático lingüístico. Este módulo es responsable del procesamiento sintáctico y semántico de los enunciados, desde los cuales se obtiene el componente o fuerza literal del acto de habla. En este sentido, la construcción de la dimensión locutiva involucra la atribución de un estado intencional basado en esa fuerza literal. Sin embargo, para explicar que la fuerza pretendida en diversos actos de habla difiere de la fuerza que es codificada en la expresión lingüística –actos de habla indirectos–, la fuerza literal del enunciado opera como presuposición al interior del procesador central. En este sistema, dicha presuposición –derivada de la decodificación lingüística– interactúa con las creencias sobre las intenciones comunicativas del hablante, con el resultado de que algunas presuposiciones se mantienen –en el caso de que la fuerza literal de una expresión sea tomada como la fuerza pretendida por el hablante–, otras presuposiciones son rechazadas, e incluso otras presuposiciones son utilizadas y luego rechazadas. En

este sentido, el componente modular pragmático propuesto por Kasher (1991) es el encargado de construir la presuposición de estado intencional que se deriva de la interpretación literal de un enunciado.

- Sistema central

Kasher apela a un sistema central para explicar el componente pragmático del lenguaje referente a las implicaturas conversacionales (Grice, 1975, 1989) y los principios de cortesía (Leech, 1983). Para ello, se basa en un argumento fundamental de la teoría de las implicaturas de Grice (1975). Nos referimos a que el principio de cooperación y las máximas están basados en un principio general de acción racional.

Aplicado al dominio de la acción, la racionalidad se basa en un principio de medios efectivos. De este modo, un agente racional elige aquellas acciones que son más efectivas de acuerdo con sus objetivos y con un mínimo costo. Es imposible circunscribir el rango de factores que juegan un papel en la elección de tales acciones. Por esta razón, Kasher sitúa el principio general de racionalidad al interior del sistema central.

Obviously, our principles of general rationality and of effective means are not domain specific in any reasonable sense and hence, if they constitute the principles of operation of some mental device, then the latter must be a central cognitive system which employs the former principle in forming beliefs and in planning action, not only when language plays an overt role. (Kasher 1991:578)

Por lo tanto, según Kasher (1991), si las implicaturas conversacionales están regidas por un principio que se sitúa al interior de los sistemas centrales, entonces el procesamiento de las implicaturas también ocurre en dichos sistemas.

Con respecto a la cortesía, Kasher sostiene que la aplicación apropiada del principio de cortesía demanda el conocimiento de una serie de factores al interior del contexto de un enunciado. Siguiendo a Leech (1983), el autor sostiene que estos principios varían según aspectos culturales y sociales. Tal dependencia contextual implica la localización de estas operaciones en un sistema central.

- Interface entre módulo pragmático y sistema central

De acuerdo con Kasher (1991), existen dos fenómenos pragmáticos que sugieren la existencia de interfaces entre un módulo específico y un procesador central. Las expresiones deícticas, tales como 'él' o 'aquí' revelan un complejo grupo de características de interface entre el lenguaje y módulos perceptivos por un lado, y entre el sistema central por el otro. En este

sentido, el referente de 'ella', en un enunciado, no está determinado solo por operaciones de un módulo lingüístico. Incluso si tal módulo existiese y contuviera el conocimiento de que 'ella' refiere a un ser humano femenino, la identificación del referente de 'ella', según Kasher, depende de la operación de un dispositivo central el cual posee acceso a información del módulo perceptual.

### *3.1 Problemas de la teoría de Kasher*

En la presente investigación, consideramos que la propuesta de Kasher genera una serie de problemas que la hacen insostenible como propuesta respecto de la arquitectura mental de la pragmática. Específicamente, para conceder la plausibilidad de la hipótesis respecto de que la fuerza de los actos de habla se procesa en un módulo pragmático lingüístico, primero se debe estar seguro de que el tipo de creencias que se identifica con la operación del sistema central no son esenciales al procesamiento que ocurre en el módulo pragmático-lingüístico. Sin embargo, si se da el caso de que existan procedimientos del sistema central que influyen en la decodificación de la fuerza literal de un acto de habla, entonces la hipótesis de un módulo pragmático pierde plausibilidad, pues viola las propiedades de especificidad de dominio, aislamiento y encapsulamiento. Supongamos el siguiente ejemplo:

‘¿Cómo se dice, hijo; cuando te dan una carta?’

Este enunciado tiene la fuerza literal de una pregunta. Sin embargo, también puede interpretarse como una petición (el hablante quiere que su hijo le agradezca por algo que recientemente ha recibido). De acuerdo con Kasher (1991), la fuerza literal de este enunciado entra al sistema central como una presuposición. En este sistema interactúa con una serie de creencias basadas en el conocimiento del contexto en el cual ocurre el enunciado. Mediante la interacción entre la presuposición y el conjunto de creencias del sistema central, se rechaza la fuerza literal y se la reemplaza por la fuerza de una petición. Sin embargo, contrario a lo que sostiene Kasher, existen casos en que las creencias del sistema central pueden determinar la asignación de la fuerza literal de un enunciado. Consideremos la palabra ‘carta’. Podemos asignar tres posibles significados de la palabra: objeto de juego de mesa, baraja, correspondencia escrita y menú de restaurante. Para llegar al contenido proposicional explícito del enunciado se debe escoger solo uno de los significados propuestos de la palabra ‘carta’. Claramente, la información visual que se obtiene del contexto físico de la expresión es por sí misma suficiente para lograr la desambiguación de la palabra. Sin embargo, de acuerdo con las suposiciones de una teoría modular, la información visual es procesada en primera instancia por un módulo perceptual visual y solo después de este procesamiento se vuelve accesible la información a una serie de procesos cognitivos centrales para formar la creencia de dicha percepción. La desambiguación sería uno de estos procesos. Se despren-

de de esto que el módulo pragmático sí es permeable a la información que es manipulada en el procesador central. Esto viola el principio de encapsulamiento informacional de cada módulo. Por lo tanto, no existe un módulo pragmático-lingüístico de carácter discreto.

Por otro lado, la presunción de que el principio de racionalidad ocurra solo al interior de un sistema central parece poco probable, puesto que el principio de racionalidad que subyace a la comunicación humana no es exclusivo de dicha facultad. Cummings (2005) sugiere que el principio de racionalidad de Grice (1989) no se circunscribe solamente a la acción intencional de los seres humanos, sino que es un principio que guía todas las operaciones biológicas. Es decir, las operaciones de un módulo, por ejemplo, visual perceptual, también están guiados por el principio de la adquisición de mayor beneficio informacional con menor costo de procesamiento. Por lo tanto, la idea de que las implicaturas se sitúan en el procesador central debido a que ellas se derivan de un principio general de racionalidad es imprecisa, pues incluso los módulos responden a este principio.

#### **4. Modularidad masiva**

Una de las críticas fundamentales a la arquitectura modular de la mente provienen desde el enfoque de la cognición corpórea (Barsalou, 2010). De acuerdo con diversas investigaciones (Desai, Binder, Conant y Seidenberg, 2010; Fischer y Zwaan, 2008), existe evidencia de que durante el procesamiento del lenguaje ocurre una coactivación de zonas cerebrales asociadas a la percepción y a aspectos motrices, incluso en etapas preconceptuales (Ostarek y Huettig, 2017). En este sentido, las propiedades de los módulos fodorianos han sido criticadas, especialmente aquellas referidas a la encapsulación. Si la construcción de significado durante el procesamiento del lenguaje depende de la coactivación de zonas perceptuales y motrices, entonces la idea de que los módulos son inaccesibles y están asilados de las operaciones de otros módulos no sería sostenible bajo un marco explicativo de la cognición corpórea.

Estos cuestionamientos llevaron a diversos científicos a proponer modelos alternativos en torno a la arquitectura cognitiva. Uno de ellos corresponde a la teoría de la modularidad masiva de la mente (Carruthers, 2006). En esta perspectiva se asume que la mente funciona exclusivamente en virtud de distintos módulos sin la necesidad de apelar a sistemas centrales. Sin embargo, la noción de módulo utilizada en esta teoría se basa en criterios diferentes a los aplicados por Fodor (1983, 2000). Bajo la teoría de la modularidad masiva, la arquitectura de la mente responde al resultado de consideraciones evolutivas (Carruthers, 2006; Sperber, 2002). El supuesto que se sigue es que la selección natural favorece soluciones específicas a problemas específicos. Un procedimiento cognitivo dedicado a tratar con un problema ambiental particular y recurrente es muy probable que sea más eficiente que un procedimiento más general

aplicado al mismo problema. De esta forma, la perspectiva evolucionista (Barkow, Cosmides y Tooby, 1992) sugiere el incremento y refinamiento de la modularidad mental en vez de cualquier tipo de fusión entre sistemas más generales para tratar con los problemas contingentes de un organismo. La propiedad crucial de un sistema modular, bajo esta perspectiva, corresponde a que, a la hora de enfrentar estos problemas ambientales, un mecanismo de propósitos especiales, sintonizado con las regularidades del dominio particular del problema, será más eficiente que un mecanismo de propósitos generales.

Sin embargo, la tesis de la modularidad masiva no puede utilizar el mismo concepto de módulo que utiliza Fodor (1983). Algunas propiedades de los módulos de Fodor necesitan abandonarse en la medida en que se asume que no existen sistemas centrales. En la teoría de la modularidad masiva existen módulos que pueden perfectamente generar outputs puramente conceptuales (por ejemplo creencias). Por lo tanto, la propiedad de superficialidad de los outputs no se aplica a los módulos en el marco de la teoría de la modularidad masiva.

La especificidad de dominio también debe abandonarse o reconceptualizarse en términos funcionales antes que en términos de dominios de contenidos o inputs específicos. Recordemos que, por especificidad de dominio, Fodor (1983) sugiere que los módulos pueden procesar solo un tipo de inputs, aquellos que pertenecen a su dominio. Sin embargo, los psicólogos evolucionistas que están a favor de la teoría de la modularidad masiva entienden la especificidad de dominio de una manera bastante diferente. En la perspectiva evolucionista el dominio de un módulo es lo que se supone que realiza, en vez del tipo de contenido que puede recibir como input (Carruthers, 2006; Sperber, 2002; Wilson, 2005). En la teoría de la modularidad masiva pueden existir módulos que reciban inputs de distinta naturaleza. Un ejemplo de ello es el módulo de la Teoría de la Mente (Wilson, 2005). Este módulo recibe inputs tanto de información visual, verbal o de otro tipo.

¿Cuál es el objetivo detrás de la reconfiguración de la arquitectura fodoriana? La teoría de la modularidad masiva está basada en consideraciones extraídas de la psicología evolutiva (Barkow, Cosmides, y Tooby, 1992). Una propuesta de arquitectura mental debe considerar el desarrollo evolutivo íntegramente. Es decir, la organización de la mente debe responder a las presiones de selección natural y adaptación. La propuesta de Fodor no considera dichos aspectos y si bien posee coherencia interna, carece de plausibilidad biológica en el sentido evolutivo (Carruthers, 2006).

## **5. La teoría pragmática modular de Sperber y Wilson**

Si se asume la perspectiva de la modularidad masiva, la pregunta concerniente a la pragmática es si ella forma parte de un módulo lingüístico, de un módulo general de Teoría de la Mente o de algún otro módulo específico. Una de las posiciones al respecto es que la

atribución del significado del hablante cae dentro del dominio de la capacidad de Teoría de la Mente, la cual consiste en la atribución de intenciones, creencias y otros estados mentales sobre la base de cualquier tipo de propósitos conductuales, sea este comunicativo o no comunicativo. Es decir, la interpretación pragmática ocurre en el módulo de la Teoría de la Mente (Bloom, 2002). Sin embargo, para Sperber y Wilson (Sperber, 2000; Sperber y Wilson, 2002; Wilson, 2000; Wilson, 2005), los procesos pragmáticos son realizados por un módulo de comprensión dedicado, de dominio específico y con sus propios principios y procedimientos. Ello porque existen regularidades manifiestas de las conductas comunicativas ostensivas (de las cuales la conducta verbal es la principal) que las diferencian de las conductas no comunicativas. De esta forma, estas regularidades generaron filogenéticamente el surgimiento de un módulo dedicado exclusivamente a la comprensión pragmática mediante la atribución de intenciones. Esto debido a que existen diferencias entre la atribución de intenciones en términos genéricos y la atribución de intenciones mediante la conducta lingüística.

Desde una perspectiva evolucionista (Wilson, 2005), lo que caracteriza a un módulo es la presencia de mecanismos dedicados (típicamente adaptaciones biológicas a regularidades en algún dominio) los cuales no pueden ser vistos como casos especiales de mecanismos más generales operando en un dominio amplio (Sperber, 2002). Para ejemplificar, consideremos las habilidades de Teoría de la Mente, las cuales permiten a los humanos predecir y explicar la conducta de otros en términos de sus estados mentales -creencias, deseos, intenciones, etc.- (Baron-Cohen, 1997; Carruthers y Smith, 1996). De acuerdo con la versión clásica de la tesis modularista (Fodor, 1983, 2000), la Teoría de la Mente debiese ser un proceso central antes que uno modular, porque su resultado depende de factores globales: no existe en principio una restricción en el tipo de información contextual que se requiere para predecir y explicar la acciones de los demás. Desde en una perspectiva evolucionista (Wilson, 2005), la cuestión no se refiere tanto a si los procesos involucrados son globales o locales, sino a si son realizados por mecanismos de propósito general o por mecanismos autónomos de propósitos especiales en sintonía con las regularidades existentes solo en el dominio de la conducta intencional. En la medida en que la Teoría de la Mente involucre tales mecanismos inferenciales de propósitos específicos, sería de carácter modular, pero en un sentido amplio.

De acuerdo con Grice (1975, 1989), la interpretación de un enunciado es un ejercicio de Teoría de la Mente. La interpretación del significado del hablante implica el reconocimiento de la intención comunicativa detrás de la conducta lingüística. Esto constituye un caso especial de un problema más general consistente en explicar la conducta de un individuo en términos de la atribución de estados mentales.

Tradicionalmente se ha descrito la teoría de la mente como un ejercicio de razonamiento reflexivo (un proceso central en términos de Fodor). Este ejercicio involucra la aplicación de habilidades de razonamiento general basadas en premisas explícitas e implícitas sobre las relaciones entre los estados mentales y la conducta. La propuesta de Grice (1975) es coherente con esta aproximación. Por ejemplo, su esquema para la derivación de las implicaturas conversacionales es un ejercicio estándar de teoría de la mente:

He said that P; he could not have done this unless he thought that Q; he knows (and knows that I know that he knows) that I will realize that it is necessary to suppose that Q; he has done nothing to stop me thinking that Q; so, he intends me to think, or is at least willing for me to think, that Q. (Grice, 1975:30-31)

Un oyente que siga la ruta de Grice tendría que incurrir en un complejo proceso de razonamiento discursivo para identificar una sola implicatura. Sin embargo, existe evidencia que sugiere una perspectiva más modular (Cummings, 2009). La mayoría de esta evidencia proviene de las pruebas que miden la Teoría de la Mente tales como las tareas de falsas creencias, en las cuales los participantes deben rastrear las creencias de los otros (Baron-Cohen, 1997). Tal habilidad puede variar en distintos grados. Los niños generalmente pasan las pruebas de creencias de primer orden a la edad de cuatro años, mientras que las personas con Trastorno del Espectro Autista (TEA) típicamente tienen problemas para reconocer intenciones (Happé, 1993; Baron-Cohen, 1997).

Las investigaciones en torno a la comunicación inferencial y la Teoría de la Mente sugieren una disociación entre ambos fenómenos. Estas disociaciones no son explicadas por la perspectiva griceana y fodoriana tradicional de Teoría de la Mente como un ejercicio de razonamiento de propósito general. Por ejemplo, el Síndrome de Williams (SW) es un extraño desorden genético que resulta en un bajo promedio de coeficiente intelectual (CI), combinado con buenas habilidades lingüísticas y sociales (Cummings, 2009). Sin embargo, dichos sujetos poseen pobres habilidades de razonamiento general. Ellos pasan las tareas de Teoría de la Mente con facilidad, mientras que sus habilidades para razonar en términos generales están seriamente dañadas (Segal, 1996). Esto sugiere que la teoría de la mente no puede ser un proceso consciente y reflexivo como el tipo de procesos ilustrado por Grice, sino que depende de un mecanismo inferencial dedicado el cual podría permanecer intacto al mismo tiempo que otros mecanismos de razonamiento de propósitos generales podrían estar dañados. Estas disociaciones también pueden ocurrir en la dirección opuesta. Por ejemplo, las personas con TEA grado 1 pueden tener buenas habilidades de razonamiento general combinadas con serias dificultades en habilidades de Teoría de la Mente (Cummings, 2009). Por lo tanto, la evidencia neuropsicológica y de estudios del desarrollo

parece favorecer una perspectiva de la teoría de la mente como un sistema modular de dominio específico antes que un proceso central reflexivo.

Todos estos hallazgos llevaron a Sperber y Wilson (2002) a sostener que la Teoría de la Mente no es un sistema central en el sentido fodoriano, sino que un módulo inferencial de dominio específico (o un grupo de módulos). Además, Wilson (2005) aplica una perspectiva modular masiva de la mente a la relación entre pragmática y Teoría de la Mente. Del mismo modo en que existen razones para pensar que la Teoría de la Mente no es meramente una aplicación de habilidades de razonamiento general a un dominio particular (dominio conductual), también existen buenas razones para pensar que la interpretación pragmática no es meramente la aplicación de habilidades de Teoría de la Mente a un dominio particular lingüístico o comunicativo ostensivo. La comunicación verbal presenta dificultades especiales y exhibe ciertas regularidades que no se encuentran en otros dominios. Este tipo de regularidades podría conllevar el desarrollo de un módulo de comprensión dedicado a la comunicación verbal y ostensiva, con sus propios principios y mecanismos (Sperber, 2000; Sperber y Wilson, 2002; Wilson, 2000; Wilson, 2005). Incluso, propuestas recientes han sugerido la existencia de un mecanismo adicional cuya función es determinar la confiabilidad de los hablantes y sus enunciados. Este mecanismo de vigilancia epistémica sería de carácter específico entendido en términos de la teoría de la modularidad masiva (Mazzarella, 2016; Mazzarella y Pouscoulous, 2021).

Según lo mencionado anteriormente, la perspectiva psicológica evolutiva trata a la cognición como una función biológica y a los mecanismos cognitivos como adaptaciones (Barkow, Cosmides y Tooby, 1992). La mente es un grupo de mecanismos cognitivos dedicados, los cuales evolucionaron mediante pequeños pasos incrementales, típicamente involucrando la selección de mecanismos como variantes que realizan más eficientemente (generando mayores beneficios con menores costos) que otras variantes de mecanismos (Sperber, 2002; Carruthers, 2006). Una forma en la que un mecanismo cognitivo dedicado puede contribuir a la eficiencia de una tarea es proveyendo procedimientos inferenciales de propósitos especiales en sintonía con las regularidades de algún dominio particular, el cual genera conclusiones confiables solo cuando se aplica al input de dicho dominio. Por ejemplo, la perspectiva modular de la Teoría de la Mente apela a procedimientos inferenciales de propósito específico los cuales trabajan confiablemente en el dominio de la conducta intencional, produciendo las mismas conclusiones derivables mediante mecanismos de razonamiento de propósito general, pero de una forma más eficiente (en relación con los parámetros de costo-beneficio). En estos mecanismos más rápidos y sencillos, las regularidades en las relaciones entre los estados mentales y la conducta no necesitan estar representadas y utilizadas como premisas explícitas en un proceso inferencial. Por lo tanto,

la propuesta de Grice (1975) respecto del reconocimiento de intenciones como una inferencia de propósito general se simplifica mediante la actividad de un módulo que recoge las regularidades propias del dominio de la conducta intencional. El resultado debiese ser una habilidad más automática e intuitiva respecto de cómo extraer conclusiones válidas en el dominio de la conducta intencional, pero sin ningún tipo de consciencia reflexiva respecto de cómo esas conclusiones son construidas.

Un ejemplo de un submódulo de teoría de la mente investigado corresponde a la Detección de la Dirección de la Mirada (*Eye Direction Detector*), la cual es vista como una habilidad presente tempranamente en los infantes para inferir lo que las personas están observando a partir de la dirección de su mirada (Baron-Cohen, 1997; Tomasello, 2003). Un infante que realiza este tipo de inferencias no está utilizando un módulo de Teoría de la Mente general e internamente no diferenciado para derivar conclusiones como las de (3) sobre la base de premisas explícitas de (1) y (2):

- (1) Si una persona P fija sus ojos en dirección al objeto O, entonces P está observando O.
- (2) Papá está fijando sus ojos en dirección al gato.
- (3) Papá está observando el gato.

Parece más razonable asumir que los seres humanos están equipados desde la infancia con un Detector de Dirección de la Mirada, el cual explota la fuerte correlación entre la dirección de la mirada y la percepción visual y directamente atribuye estados perceptuales y atencionales en virtud de la dirección de la mirada (moviendo directamente desde las premisas tales como (2) a conclusiones tales como (3)). Esta atribución puede en sí misma proveer un input para otros dispositivos dedicados, tales como los involucrados en la adquisición de palabras (Tomasello, 2003).

A pesar del hecho de que las habilidades de la Teoría de la Mente pueden ser descompuestas en una serie de módulos subespecializados, la perspectiva estándar ha sostenido que la atribución de estados mentales requerida para la comprensión pragmática será automáticamente generada por mecanismos de Teoría de la Mente generales los cuales aplican a todos los dominios. Sin embargo, Sperber y Wilson (Sperber, 2000; Sperber y Wilson, 2002; Wilson, 2000; Wilson, 2005) sostienen que existen serios problemas en esta perspectiva. En primer lugar, el rango de intenciones que puede ser razonablemente atribuido a un agente sobre la base de alguna conducta no comunicativa en una situación particular es generalmente bastante limitada (por ejemplo, si extendiendo mi brazo hacia la alacena será para sacar una tasa o un vaso). Sin embargo, en la comunicación lingüística el entorno físico coloca

pocas restricciones al contenido de un enunciado (comparado con el caso anterior). Esto genera que exista una brecha entre el significado lingüístico y el significado pretendido por el hablante. Es decir, existe un vasto rango de significados posibles que en un momento dado un hablante pudo o quiso comunicar. No queda claro cómo los procedimientos estándar para atribuir las intenciones sobre la base de la conducta no comunicativa puedan generar la atribución adecuada del significado del hablante, excepto en casos triviales (Sperber, 2002; Sperber y Wilson, 2002).

En segundo lugar, mientras ambos tipos de atribución de estados mentales involucran claramente capacidades metarepresentacionales (la capacidad de representar las representaciones mentales de los otros), existe una importante diferencia en la complejidad de la aplicación de ambos casos. La comunicación ostensiva involucra tanto una intención informativa como una intención comunicativa (Sperber y Wilson, 2004; 1995), de tal forma que el oyente debe reconocer que el hablante *intenta hacerle creer* que él pretende que su oyente crea un cierto grupo de proposiciones (Sperber, 2000). En este caso existen 3 niveles de metarepresentación, mientras que en la comprensión de acciones ordinarias no comunicativas solo es necesario un único nivel de atribución de intención. Por lo tanto, existen dos procedimientos distintos de atribución de intenciones. Uno es activado por la comunicación lingüística (o cualquier comunicación ostensiva) y otra por la observación de la conducta intencional no comunicativa. En este sentido, para la teoría de la relevancia, la pragmática es un submódulo de un módulo de la Teoría de la Mente que utiliza sus propios principios y procesos. A continuación, describiremos cuáles son dichos principios y regularidades.

### *5.1 Pragmática como un submódulo del módulo de teoría de la mente*

La teoría de la relevancia (Sperber y Wilson, 2004; 1995) es una teoría sobre la cognición y la comunicación humana basada en la definición de la relevancia y dos principios generales: un primer principio cognitivo de relevancia y un segundo principio comunicativo de relevancia. La relevancia es caracterizada como una propiedad de los inputs de un proceso cognitivo y analizado en término de las nociones de efecto cognitivo y esfuerzo de procesamiento. Bajo esta teoría, la relevancia es una noción de costo-beneficio (el costo corresponde al esfuerzo mental requerido y los beneficios a los efectos cognitivos conseguidos), y la tendencia universal hacia el incremento de la eficiencia de un sistema cognitivo mediante la maximización de la relevancia encaja perfectamente con la perspectiva evolutiva expuesta en la sección anterior. Esta tendencia es descrita en el primer principio cognitivo de la relevancia

- (4) La cognición humana se orienta hacia la maximización de la relevancia.

Se sigue de este principio que la atención humana y los recursos de procesamiento tienden automáticamente a asignarse a la información que parece ser lo suficientemente relevante para valer la pena su procesamiento. Esto no se debe a que conscientemente elijamos esta opción, sino que a la forma en que nuestros sistemas cognitivos han evolucionado.

As a result of constant selection pressure towards increasing efficiency, the human cognitive system has developed in such a way that our perceptual mechanisms tend automatically to pick out potentially relevant stimuli, our memory retrieval mechanisms tend automatically to activate potentially contextual assumptions, and our inferential mechanisms tend spontaneously to process them in the most productive way. It is against this cognitive background that communication takes place (Wilson, 2005:1145).

Sabiendo que es probable que los demás pongan atención a los estímulos más relevantes de su ambiente y a procesar aquellos de tal forma de maximizar su relevancia, las personas son capaces de producir estímulos los cuales probablemente atraerán la atención de los demás. Según la teoría de la relevancia (Sperber y Wilson, 1995, 2004), la comunicación ostensiva involucra dos tipos de intenciones:

- (5) Intención informativa: La intención de informar a la audiencia respecto de algo.
- (6) Intención comunicativa: La intención de informar a la audiencia de la propia intención informativa.

¿Cómo el hablante indica a su audiencia que él está tratando de comunicarse con ellos en esta forma manifiesta e intencional? La teoría de la relevancia sostiene que la comunicación manifiesta involucra el uso de un estímulo ostensivo, designado para atraer la atención de la audiencia sobre el significado del hablante. De esta forma, la comunicación se sirve del principio general de relevancia de la cognición, conformándose así el segundo principio de la relevancia, en este caso de índole comunicativa. De acuerdo con este segundo principio, el uso de un estímulo ostensivo activa una expectativa de relevancia a nuestros oyentes que, de otro modo, bajo las acciones o conductas ordinarias no comunicativas, no conllevaría.

- (7) Principio comunicativo de relevancia: Cada enunciado (o cualquier acto de comunicación manifiesta) comunica la presunción de su propia relevancia óptima.

Se sigue de este principio que el hablante, con el mero acto de emitir un enunciado a otro, comunica que su enunciado es el más relevante o compatible con las preferencias del oyente y que al menos es lo suficientemente relevante como para que valga la pena procesarlo.

La regularidad de la comunicación ostensiva descrita en el principio comunicativo de la relevancia motiva un proceso de comprensión (mediante la construcción de inferencias) de propósitos especiales el cual es automáticamente aplicado al procesamiento en-línea de los inputs verbales. El oyente toma el significado lingüístico decodificado; siguiendo el camino de menor esfuerzo de procesamiento, utiliza la información contextual disponible para enriquecer el nivel explícito del enunciado y para complementar el nivel implícito hasta que la interpretación satisfaga su expectativa de relevancia (Sperber y Wilson, 2004; 1995).

El procedimiento de comprensión guiado por la teoría de la relevancia es de carácter heurístico, rápido y parsimonioso. Es decir, es un proceso inferencial de propósitos especiales justificado por la regularidad descrita en el principio comunicativo de la relevancia, el cual permite generar conclusiones confiables solo cuando es aplicado a los inputs del dominio de la comunicación ostensiva (Sperber y Wilson, 2002; Wilson, 2005). En general, a un observador no le corresponde esperar que las acciones no comunicativas de los otros sean lo suficientemente relevantes como para merecer su atención. Sin embargo, cuando interpreta enunciados, al oyente le corresponde tal expectativa (debido al principio comunicativo de la relevancia). El procedimiento de comprensión de la teoría de la relevancia corresponde a la construcción automática de una hipótesis del significado del hablante sobre la base de la descripción del enunciado más la información contextual disponible, de la misma manera en que el Detector de Dirección de la Mirada automáticamente construye una hipótesis respecto de lo que un individuo está observando sobre la base de la dirección de su mirada (Bara, 2010; Tomasello, 2003). Bajo esta perspectiva, las regularidades descritas en los principios cognitivos y comunicativos de la relevancia proveen una base adecuada para un mecanismo de comprensión dedicado a la comunicación ostensiva, un submódulo del módulo de la teoría de la mente.

## 6. Conclusiones

El aporte de la presente investigación consiste en aplicar criterios de evaluación de teorías pragmáticas basadas en las consideraciones biológicas dentro de un marco evolucionista (Barkow, Cosmides y Tooby, 1992). En la presente investigación se defiende y argumenta en favor de una perspectiva cognitiva de la pragmática y se propone su rol funcional en el marco de una arquitectura modular masiva de la mente. En este sentido, proponemos que una teoría respecto del rol funcional de la pragmática en la arquitectura mental debe ser consistente con toda la evidencia empírica en torno a cómo las habilidades pragmáticas operan en

términos cognitivos. A su vez, una buena teoría pragmática debe ser coherente con el marco de la psicología evolutiva. De acuerdo con esta última idea, las explicaciones de nuestras funciones cognitivas, como las habilidades pragmáticas, requiere que dichas funciones sean explicadas en virtud de cómo ellas están organizadas para satisfacer las demandas que las presiones ambientales ejercen sobre los seres vivos. Ello porque no solamente se debe exigir a una teoría que posea coherencia interna y manifestación empírica, sino que también ella debe poseer una consistencia con los demás campos científicos adyacentes. Bajo la revisión de la evidencia empírica (Airenti, 2017; Bara, 2010; Cummings, 2009; Happé, 1993; Papafragou, 2002; Tomasello, 2003) y un análisis de consistencia argumental, la teoría de la pragmática modular y de la modularidad masiva cumplen con tales exigencias. Bajo el análisis de las investigaciones empíricas sobre las habilidades pragmáticas y las consideraciones evolutivas se propone que el modelo que mejor encaja y satisface estos requerimientos corresponde a la perspectiva de la modularidad masiva de la mente, entendiendo la pragmática como un submódulo de habilidades de Teoría de la Mente. Dicha organización es consistente con los resultados observados en personas con déficit pragmático y además es coherente con respecto a consideraciones evolutivas.

La propuesta compartida en esta investigación es de carácter intuitiva y si bien existen antecedentes empíricos que la apoyan (sobre todo respecto de estudios de trastornos neurológicos y del desarrollo), todavía requiere ser desarrollada con más exhaustividad. Uno de los problemas que una perspectiva de la pragmática modular debe resolver, en el contexto de la modularidad masiva, se refiere a cómo se accede a las representaciones contextuales que se requieren para realizar una interpretación pragmática. Dado que no existe una restricción en principio en el tipo de fuente de información contextual que es utilizada en la construcción de una hipótesis respecto del significado pretendido por el hablante, el proceso sería global en el sentido fodoriano. De cualquier forma, no se sigue de todo esto que toda la información contextual tenga que estar activamente procesada para que la interpretación se lleve a cabo. El procedimiento de comprensión de la teoría de la relevancia sigue el camino de menor esfuerzo, utilizando cualquier tipo de información contextual que esté altamente activada mediante el trabajo automático de un sistema cognitivo en un momento dado. Sin embargo, para dar cuenta de la docilidad computacional que se requiere, es indispensable que se estipule de una manera sencilla cómo el módulo pragmático accede a información contextual sin que ocurra una explosión computacional. Una respuesta tentativa es que la presentación del estímulo comunicativo activa en-línea (*online*) una serie de representaciones en los mecanismos de memoria, las cuales son accesibles por parte del módulo pragmático. De todas formas, esta última hipótesis debe ser desarrollada con más profundidad.

Por otra parte, es necesario también enfatizar que, en el caso de que se asuma que la pragmática es un sistema que funciona de manera rápida, automática y con sus propios principios y procedimiento, la pregunta que se sigue es ¿cuál es el dominio de este módulo? La respuesta diferencia aún más la propuesta de Sperber y Wilson con la de Asa Kasher. El dominio del módulo pragmático, según Sperber y Wilson, no lo constituyen únicamente los estímulos lingüísticos, sino que cualquier hecho comunicativo ostensivo. La teoría de la relevancia sostiene que la pragmática corresponde a un dominio más amplio que al de la conducta lingüística. Su dominio se vincula a los estímulos ostensivos de carácter comunicativo, sean estos basados en un código lingüístico o no.

## Bibliografía

- Austin, J. (1998). *Cómo hacer cosas con las palabras*. Barcelona: Paidós.
- Bara, B. (2010). *Cognitive pragmatics: the mental processes of communication*. Massachusetts: MIT Press.
- Barker-Plummer, D. (22 de Junio de 2003). *The Stanford Encyclopedia of Philosophy*. Obtenido de The Stanford Encyclopedia of Philosophy: <http://plato.stanford.edu/archives/sum2013/entries/turing-machine/>
- [Consulta 02/05/2021]
- Barkow, J., Cosmides, L., y Tooby, J. (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford: Oxford University Press.
- Baron-Cohen, S. (1997). *Mindblindness*. Nueva York: A Bradford Book.
- Barsalou, L. (2010). Grounded cognition: Past, present, and future. *Topics in cognitive science*, 2(4): 716-724.
- Bloom, P. (2002). Mindreading, Communication and the Learning of Name for Things. *Mind and Language*, 17(1): 37-54.
- Carruthers, P. (2006). The case for massively modular models of mind. En R. Stainton (eds.), *Contemporary debates in cognitive science* pp. 3-21. Oxford: Blackwell Publishing.
- Carruthers, P., y Smith, P. (1996). *Theories of Theories of Mind*. Cambridge: Cambridge University Press.
- Carston, R., y Powell, G. (2008). Relevance theory—new directions and developments. En E. Lepore, y B. Smith, *The Oxford Handbook of Philosophy of Language*, pp. 341-360. Oxford: Oxford University Press.
- Chomsky, N. (1966). *Cartesian Linguistics: A Chapter in the History of Rationalist*. Nueva York: Harper y Row.
- Chomsky, N. (1980). *Rules and Representations*. Nueva York: Columbia University Press.
- Cummings, L. (2005). *Pragmatics a multidisciplinary perspective*. Nueva Jersey: Lawrence Erlbaum Associates.
- Cummings, L. (2009). *Clinical Pragmatics*. Cambridge: Cambridge University Press.
- Desai, R., Binder, J., Conant, L., y Seidenberg, M. (2010). Activation of sensory–motor areas in sentence comprehension. *Cerebral Cortex*, 20(2): 468-478.

- Fischer, M., y Zwaan, R. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly journal of experimental psychology*, 61(6): 825-850.
- Fodor, J. (1980). *The Language and Thought*. Harvard University Press: Cambridge.
- Fodor, J. (1983). *The modularity of the mind*. Cambridge: MIT Press.
- Fodor, J. (2000). *The mind doesn't work that way*. Cambridge: MIT Press.
- Grice, P. (1975). Logic and Conversation. En P. Cole, y J. Morgan (eds.), *Syntax and Semantics Vol. 3: Speech Acts*, pp. 41-58. Nueva York: Academic Press.
- Grice, P. (1989). *Studies in the Way of Words*. Cambridge: Harvard University Press.
- Happé, F. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, 48: 101-119.
- Happé, F., y Loth, E. (2002). Theory of Mind and Tracking Speakers Intentions. *Mind and Language*, 17: 24-36.
- Kasher, A. (1991). Pragmatics and the modularity of the mind. En S. Davis (ed.), *Pragmatics a reader*, pp. 567-582. Oxford: Oxford University Press.
- Leech, G. (1983). *Principles of Pragmatics*. Londres: Logman.
- Mazzarella, D. (2016). Pragmatics, modularity and epistemic vigilance. *Argumenta*, 1(2): 181-193.
- Mazzarella, D., y Pouscoulous, N. (2021). Pragmatics and epistemic vigilance: A developmental perspective. *Mind and Language*, 36(3): 355-376.
- Ostarek, M., y Huettig, F. (2017). Spoken words can make the invisible visible—Testing the involvement of low-level visual representations in spoken word processing. *Journal of Experimental Psychology: Human Perception and Performance*, 43(3): 499.
- Papafragou, A. (2002). Mindreading in verbal communication. *Mind y Language*, 17: 55-67.
- Searle, J. (1975). Una taxonomía de los actos ilocucionarios. En L. Valdés Villanueva (ed.), *La búsqueda del significado*, pp. 328-339. Madrid: Paidós.
- Segal, G. (1996). The modularity of theory of mind. En P. Carruthers, y P. Smith (eds.), *Theories of Theories of Mind*, pp. 141-157. Cambridge: Cambridge University Press.
- Simon, H. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106: 467-482.

- Sperber, D. (2000). Metarepresentations in a evolutionary perspective. En D. Sperber (ed.), *Metarepresentations: A Multidisciplinary Perspective*, pp. 117-137. Oxford: Oxford University Press.
- Sperber, D. (2002). In Defense of Massive Modularity. En E. Dupoux (ed.), *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler*, pp. 47-58. Oxford: A Bradford Book.
- Sperber, D., y Wilson, D. (1995). *Relevance*. Cambridge: Blackwell.
- Sperber, D., y Wilson, D. (2002). Pragmatics, Modularity and Mind-Reading. *Mind y Language*, 17: 3-23.
- Sperber, D., y Wilson, D. (2004). Relevance Theory. En L. Horn, y G. Ward (eds.), *The Handbook of Pragmatics* pp. 607-632. Oxford: Blackwell Publishing.
- Stillings, N., Weisler, S., Chase, C., Feinstein, M., Garfield, J., y Rissland, E. (1995). *Cognitive Science. An Introduction*. Cambridge: Bradford.
- Tager-Flusberg, H. (1999). *Neurodevelopmental Disorders*. Cambridge: MIT Press.
- Thagard, P. (2012). Cognitive architectures. En K. Frankish, y W. Ramsey (eds.), *The Cambridge Handbook of Cognitive Science*, pp. 50-71. Cambridge: Cambridge University Press.
- Tomasello, M. (2003). *Constructing a Language: a Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press.
- Varley, R. (2002). Science without grammar: scientific reasoning in severe agrammatic aphasia. En P. Carruther, S. Stich, y M. Siegal (eds.), *The cognitive Basis of Science*, pp. 99-116. Cambridge: Cambridge University Press.
- Von Eckardt, B. (2012). The Representational Theory of Mind. En K. Frankish, y W. Ramsey (eds.), *The Cambridge Handbook of Cognitive Science*, pp. 29-49. Cambridge: Cambridge University Press.
- Wilson, D. (2000). Metarepresentation in linguistic communication. En D. Sperber (ed.), *Metarepresentations: An interdisciplinary Perspective*, pp 411-450. Nueva York: Oxford University Press.
- Wilson, D. (2005). New directions for research on pragmatics and modularity. *Lingua*, 115: 1129-1146.

# EL ROL DEL RAZONAMIENTO ABDUCTIVO EN EL RAZONAMIENTO PROBATORIO DEL DERECHO: UNA PERSPECTIVA ORIENTADA HACIA LA PRÁCTICA<sup>1</sup>

\*

JOSÉ ENRIQUE  
SOTOMAYOR TRELLES

Investigador del Centro de  
Investigación Interdisciplinar  
Ciencia y Sociedad (CIICS) de  
la Universidad de Ciencias y  
Humanidades (UCH), Lima, Perú.  
jesotomayor@uch.edu.pe

GABRIEL EDUARDO  
USCAMAYTA

Abogado por la Universidad Andina  
del Cusco. Posgrado en Defensa  
de la Competencia y Propiedad  
Intelectual por la Universidad  
Católica San Pablo, Perú.  
guscamayta@prestonlegal.org

NIELS J. APAZA JALLO

Abogado por la Universidad  
Nacional del Altiplano, con  
segunda especialidad en  
Argumentación Jurídica por la  
Pontificia Universidad Católica del  
Perú y predocente en la Pontificia  
Universidad Católica del Perú.  
niels.apaza@pucp.edu.pe

## Resumen

*El presente artículo parte de una revisión y relativización de la tajante distinción entre contexto de descubrimiento y contexto de justificación, originada en la filosofía de la ciencia del empirismo lógico y presente como un presupuesto de varios estudios sobre la argumentación jurídica. A continuación, analiza la propuesta de desarrollar una lógica del descubrimiento para el contexto de descubrimiento, una de cuyas principales herramientas sería el razonamiento abductivo. Finalmente, aplica los alcances de tal discusión al razonamiento probatorio jurídico, a partir del análisis de algunos casos.*

**Palabras clave:** *contexto de descubrimiento; contexto de justificación; lógica del descubrimiento; razonamiento abductivo; razonamiento probatorio.*

---

<sup>1</sup> Una versión previa de este trabajo fue revisada por el profesor Miguel León Untiveros, quien formuló valiosas observaciones y sugerencias. Asimismo, agradecemos los comentarios y críticas planteadas por los evaluadores anónimos, pues nos permitieron identificar errores e imprecisiones a lo largo del texto.

### Abstract

*The article departs from an analysis and relativization of the sharp distinction between the context of discovery and the context of justification, originated in the philosophy of science of logical empiricism, but present as a presupposition of several studies on legal argumentation. Next, it analyzes the proposal to develop a logic of discovery for the context discovery, one of whose main tools would be abductive reasoning. Finally, it applies the scope of such discussion to legal evidential reasoning, developing the analysis of some cases.*

**Keywords:** *context of discovery; context of justification; logic of discovery; abductive reasoning; evidential reasoning.*

## **1. Introducción**

El presente trabajo tiene tres objetivos: (i) debilitar la distinción tajante entre contexto de descubrimiento y contexto de justificación, (ii) analizar la estructura y función de la abducción en el razonamiento probatorio y (iii) aplicar todo ello al análisis de casos que involucren razonamiento probatorio, extraídos de la jurisprudencia penal peruana.

Para el logro de tales objetivos, como punto de partida, explicaremos nuestra posición sobre por qué la abducción no ha recibido un estudio detenido en el derecho —en comparación con el razonamiento deductivo, inductivo, analógico e, incluso, conductivo— a lo largo del tiempo y cómo es que ahora aparece como una herramienta potente para explicar parte de nuestro razonamiento sobre los hechos, principalmente en el contexto de descubrimiento.

A continuación, explicaremos qué se puede entender por razonamiento abductivo y cuál es su estructura mediante algunos ejemplos.

Finalmente, defenderemos la tesis de que la incorporación de un análisis sobre la abducción nos lleva, desde una teoría de la argumentación jurídica centrada en el contexto de justificación, hacia una compleja teoría sobre la racionalidad en el análisis y toma de decisiones, con lo que se incorpora la racionalidad del proceso de descubrimiento. Esta tesis será ejemplificada a partir del análisis de dos casos.

Como apunte previo, es importante precisar que el artículo no desarrolla de forma detallada todas las críticas que se han formulado a la distinción entre contexto de descubrimiento y contexto de justificación. Esta es una tarea que, aunque necesaria y prometedora, no puede ser abordada aquí. En su lugar, se muestra que dicha distinción puede ser cuanto menos debilitada o relativizada, y que al hacerlo se puede obtener una perspectiva más amplia y fructífera sobre el razonamiento relacionado con cuestiones de hecho en un caso.

## **2. Sobre la dicotomía contexto de descubrimiento - contexto de justificación**

Algunos de los autores asociados con lo que Manuel Atienza (2020) llama “teoría estándar de la argumentación jurídica” (TAJ) influyente hasta la actualidad, partieron de la distinción entre el contexto de justificación (CJ) y el contexto de descubrimiento (CD), con la finalidad de limitar su objeto de estudio (Atienza, 2013, pp. 114-116). Tal orientación no es excluyente de la TAJ producida en el mundo hispanoamericano, sino que también parece ser una característica presente en algunas aproximaciones de teoría del derecho en el mundo anglosajón (Anderson, 1996, cap. 1). De hecho, la investigación de Anderson (1996) le lleva a identificar a Richard Wasserstrom y Neil MacCormick como dos de los más conocidos defensores de la distinción aplicada al derecho (pp. 11 y ss.). En efecto, MacCormick es claro al señalar que el proceso que le interesa estudiar es el de la argumentación entendida

como justificación, esto es, como el ofrecimiento público de un conjunto de razones que fundamentan una decisión de acuerdo a derecho (MacCormick, 2003/1978, pp. 15 y ss.).

Quienes aceptaban esta distinción, y concentraban sus esfuerzos en el CJ, podían centrarse en las razones justificativas de una decisión, dejando de lado las razones que explicaban el proceso psicológico-cognitivo de formación de la misma. De hecho, tal como parece afirmar García Figueroa, este enfoque tiene una connotación estratégica que permitió a los teóricos del derecho reducir su labor cognitiva, eligiendo una de las dimensiones (contextos) y dejando de lado la otra a ramas más especializadas como la psicología o la sociología jurídica (Gascón Abellán y García Figueroa, 2015, pp. 140-141). El problema es que, como apunta Anderson, la psicología del proceso de decisión judicial es una de las ramas menos estudiadas de las teorías sobre la decisión judicial (Anderson, 1996, p. 1).

Ahora bien, ¿a qué nos referimos, con exactitud, cuando hablamos de CJ y CD? La distinción se origina en el contexto de la filosofía de la ciencia del empirismo lógico. Autoras como Atocha Aliseda (2006, p. 6) o Jutta Schickore (2018) coinciden en sostener que fue Hans Reichenbach quien propuso la dicotomía en la década de 1930. En su formulación temprana, el CJ hace referencia a la formulación sistemática de una teoría científica, entendida como un conjunto de proposiciones sustentadas en razonamientos, mientras que el CD se refiere a la concepción de una idea, entendida como un proceso psicológico por parte del investigador (Reichenbach, 1938, pp. 3-16). No obstante, Aliseda matiza la distinción al proponer dos sentidos en que se aborda el CD: un sentido amplio y uno restringido. En sentido restringido el CD solo hace referencia a la concepción inicial de una idea, mientras que el sentido amplio incluye todo el proceso que va desde la concepción de una idea hasta el momento previo a su formulación como una teoría sistemática.

En la teoría del razonamiento jurídico se emplea esta distinción para identificar, indistintamente, dos cuestiones: (i) los criterios de corrección de una decisión y (ii) los factores que condicionan tal decisión. Dicha distinción ha sido utilizada como un instrumento de la argumentación jurídica para evaluar las decisiones judiciales, teniéndose preferencia por los criterios de corrección de la decisión (CJ) frente a los criterios condicionantes de la misma (CD). Ello se justifica en tanto los criterios de corrección pueden ser controlables interpersonalmente, en la medida que se concretizan en algún documento público (una sentencia, por ejemplo), mientras que los factores que condicionan la decisión necesitan de estudios empírico-psicológicos para ser analizados (Pino, 2017, p. 39).

Ahora bien, llegados a este punto, resulta pertinente recordar que la distinción entre CJ y CD no se origina de la teoría del Derecho. Esta es, más bien, resultado de una importación desde la parcela de la filosofía de la ciencia, donde se empleó desde inicios hasta

mediados del siglo XX, en el periodo de auge y progresivo declive del empirismo lógico (Schikore & Steine, 2006, p. vii). Es precisamente la filosofía de la ciencia un lugar en el que, desde hace varios años, la distinción se ha ido atenuando, debido a una importante variedad de críticas y nuevas investigaciones (para un repaso sobre diversas estrategias de distinción, asunciones compartidas y el desarrollo de algunas críticas y refinamientos de la distinción en la filosofía de la ciencia, véase Hoyningen-Huene, 2006, pp. 119-131). De hecho, consideramos que la misma suerte de progresivo debilitamiento de la distinción se viene desplegando en la teoría del Derecho (en esa línea véase, por ejemplo, Accatino, 2002). Veamos, entonces, de forma esquemática, algunos sucesos en la filosofía de la ciencia que explican la progresiva atenuación a la que hemos hecho referencia.

Las primeras concepciones sobre la distinción entre CD y CJ indicaban que el primero de estos conceptos daba cuenta de los procesos de generación de hipótesis y teorías, mientras que el segundo remitía a la comprobación y validación de estas. De hecho, se creía que no existían reglas lógicas que permitan construir una máquina del descubrimiento (*discovery machine*) que imite la actitud creativa de los genios que planteaban hipótesis, cuestión que era distinta cuando se hablaba del contexto de justificación, que sí era un área que podía ser rigurosamente estudiada (Arabatzis, 2006, p. 215).

Theodore Arabatzis afirma que esta distinción estaba ligada a las siguientes premisas: primero, que la filosofía de la ciencia era una empresa normativa que buscaba generar criterios capaces de establecer cuándo algo era ciencia y cuándo no; segundo, que se entendía que el CD estudiaba la “creatividad científica”; y, tercero, y en línea de lo anterior, que no existían reglas para mejorar la “creatividad científica” (Arabatzis, 2006, p. 215).

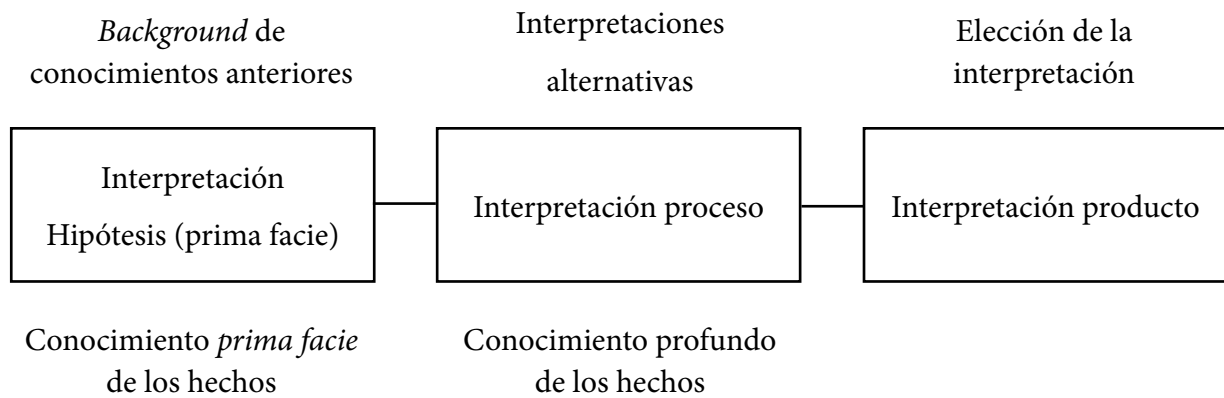
Uno de los argumentos contemporáneos empleados para atenuar la concepción clásica de la separación entre CD y CJ consiste en mostrar que la ciencia, al entender el *descubrimiento* como un logro epistémico que permite revelar la existencia de algo, presupone que tal descubrimiento solo tiene validez después de haber convencido (en el plano de la justificación) a la comunidad científica relevante, de acuerdo con criterios establecidos endógenamente por dicha comunidad (Arabatzis, 2006, p. 217). Es decir, que el descubrimiento de *algo* depende de la justificación de ese *algo*, ya que es a partir de la justificación que se utilizarán las reglas y se cumplirá con los procedimientos impuestos por la comunidad para darle consistencia, y como tal, si supera el umbral de racionalidad impuesto, se tendrá a este *algo* como *descubierto*. Arabatzis (2006) justifica lo anterior señalando que “(...) una mera hipótesis, en el sentido de que una nueva entidad existe, no calificaría como un descubrimiento de tal entidad. La justificación de una hipótesis tal sería una característica constitutiva del descubrimiento” (p. 217, traducción propia). Desde esta perspectiva, parece que la justificación es ya parte del descubrimiento.

En el terreno de la teoría del Derecho, la dicotomía CJ/CD también empezó progresivamente a atenuarse y a perder fuerza. Por ejemplo, Atienza (2017) señala que la TAJ —al centrarse exclusivamente en el CJ— es probable que “(...) haya limitado excesivamente el alcance de la teoría, e incluso que haya impedido una comprensión cabal del razonamiento justificativo si es que la distinción en cuestión no puede trazarse de manera estricta: o sea, si hubiera elementos de carácter explicativo que jugaran también un papel en la justificación. Y esto último es lo que parece suceder” (p. 16). Y, en efecto, tal como sucedió en la filosofía de la ciencia, en el derecho parece que también se pueden extraer algunas razones explicativas del discurso justificativo.

Por su parte, Anderson recuerda que en la tradición del realismo jurídico estadounidense existía una preocupación por el proceso mental que derivaba en la decisión judicial. Esta preocupación llevó a autores como Jerome Frank a interesarse en las actividades mentales desplegadas por el juez para tomar una decisión, de forma tal que se pueda identificar sesgos o prejuicios en el razonamiento. La identificación era, desde luego, instrumental, pues lo más importante era desarrollar mecanismos de control para evitar el ingreso de factores irracionales en el proceso de razonamiento (Anderson, 1996, p. 5). De esta forma, se puede encontrar una nueva crítica a la distinción tajante entre CD y CJ, en la medida que la misma no nos permite desarrollar una más integral teoría sobre la racionalidad en el proceso de análisis y toma de decisiones jurídicas. Al adoptar una distinción tajante entre CD y CJ corremos el riesgo de dejar el proceso de descubrimiento en una suerte de *caja negra* que no es sometida a los estándares de racionalidad que aplicamos al CJ, y ello puede tener como resultado una perspectiva parcial sobre la racionalidad del razonamiento jurídico.

También Giorgio Pino (2017) ha desarrollado tres argumentos que permiten sostener la necesidad de atenuar de la dicotomía entre CJ y CD en el razonamiento jurídico: (i) el curso dialéctico de la interpretación; (ii) la fuerza civilizadora del contexto de justificación; y, (iii) las ideologías jurídicas. Veamos con más detalle cada uno de estos argumentos.

Con referencia al curso dialéctico de la interpretación, Pino entiende que interpretar consiste en asignar sentido a disposiciones normativas (en el mismo sentido Guastini, 2018; Canale, 2019, entre muchos otros). Dicha asignación corresponde a un ejercicio volitivo que implica elegir entre alguno de los sentidos de entre todos los disponibles y que son adscribibles a una disposición. Podemos entender que el proceso de interpretación implica, en líneas generales, tres momentos hasta la elección final del sentido de la disposición normativa: (i) la interpretación hipótesis; (ii) la dilucidación de las interpretaciones disponibles; y, (iii) la interpretación producto. Estos momentos se recogen en el siguiente gráfico:



**Gráfico 1:** Proceso interpretativo (Giorgio Pino)  
Fuente: elaboración propia

Puede entenderse que el fenómeno interpretativo se inicia con una hipótesis inicial sobre la interpretación de una disposición normativa, de acuerdo con un primer acercamiento a los hechos del caso concreto. Isabel Lifante, siguiendo a Gianformaggio, ha llamado a este primer momento “interpretación noética”. En este “(...) se produce una captación del significado como un pensamiento intuitivo, es decir, una captación intelectual inmediata de una realidad inteligible (...)” (Lifante, 2019, p. 2). Ahora bien, parece que se *descubre* la mejor interpretación en el transcurso del proceso judicial, cuando el juez conoce mejor los hechos y analiza otras posibles interpretaciones de la disposición (o disposiciones) que las partes eventualmente aportan en sus argumentos. Después de todo este proceso, el juez justifica la elección de su interpretación, ya sea confirmando su inicial hipótesis interpretativa o eligiendo otra que fue develada en el transcurso del proceso de análisis. Esto nos sirve para entender que, en realidad, *sí* se puede *descubrir* cierta actividad cognitiva cuando el juez justifica su decisión en el documento público. De hecho, al respecto, Pino señala lo siguiente:

En efecto, en la justificación de la premisa normativa mayor del silogismo decisonal, el juez muchas veces no se limita a enunciar la interpretación preseleccionada y los argumentos a su favor, sino que explícitamente repasa posibles objeciones y posibles alternativas (lo que, en parte, se debe a la propia estructura triádica del proceso, que hace que el juez deba también tomar en cuenta los argumentos y los contraargumentos presentados por las partes). En estos casos, bastante frecuentes, parece que el límite entre el contexto de justificación y el contexto de descubrimiento se atenúa; que el primero “imita” al segundo (Pino, 2017, pp. 312-313).

En segundo término, Pino hace referencia a la “fuerza civilizadora del contexto de justificación”. Clásicamente se concibe que la justificación de la decisión judicial funciona como una garantía de las partes involucradas en el conflicto, así como un mecanismo de rendición de cuentas para la sociedad, en tanto implica el uso de poder que solo será legítimo si se encuentra justificado. Esta manera de aproximarse a la cuestión recuerda a las funciones endo y extraprocesales del deber de motivación propuestas por autores como Taruffo (Taruffo, 2006, pp. 309ss.; comentarios en Sotomayor, 2021a, pp. 23-31). El que los jueces justifiquen o motiven sus decisiones presupone que estos saben cómo hacerlo. Saber motivar puede entenderse aquí como saber justificar una decisión con razones obtenidas del derecho. Por una cuestión de formación, entonces, los jueces debieron aprender antes a razonar con reglas y justificar decisiones con las mismas. Es decir, el conocimiento de los rudimentos del razonamiento jurídico es una precondition básica de su capacidad para motivar.

Siguiendo esta línea de razonamiento, al analizar la decisión justificada de un juez puede *descubrirse* no solo las razones de la decisión, sino también que el juez sabe motivar y que, por tanto, obtuvo una adecuada formación (Pino, 2017, p. 314). Tal como recomienda Pino, este argumento debe manejarse con pinzas debido a su mera plausibilidad. De hecho, el argumento de Pino es en cierta medida problemático pues propone un cambio de perspectiva desde quien juzga hacia quien observa el acto de juzgar. En todo caso, mostraría que no solo para quien juzga es del todo clara la distinción entre CJ y CD, sino que este mismo problema de vaguedad afecta al observador del acto de juzgar.

Finalmente, Pino hace referencia a un tercer argumento, sobre “ideologías jurídicas”. Con tal expresión Pino se refiere al esquema de trabajo que tiene todo juez. Asimismo, por esquema de trabajo puede entenderse todas aquellas técnicas interpretativas y de análisis, así como sus preferencias sobre estas, esquema con el cual enfoca los casos que conoce.

Al igual que en el segundo argumento, al concretizar la justificación en un documento público, este puede ayudar a *descubrir* el esquema de trabajo del juez, así como sus preferencias interpretativas para el caso solucionado. Pino advierte, nuevamente, que este argumento es meramente plausible, pero que puede ayudar a aclarar la razón por la que la distinción entre CD y CJ es menos tajante de lo que se imagina.

\*\*\*\*

De lo dicho hasta este punto, puede entenderse que en el discurso jurídico se emplea la idea de contexto de descubrimiento en el sentido de *descubrir*, *prima facie*, las razones genuinas de la toma de decisiones por parte del juez. Consideramos, además, que ulteriormente se puede distinguir entre *descubrir* las razones genuinas de una decisión en un sentido fuerte o en sentido débil.

El sentido fuerte, supondría el *descubrimiento* de las razones genuinas y evidenciadas de la decisión, acudiendo a los dos momentos del descubrimiento en la filosofía de la ciencia: la generación de una eventual hipótesis sobre las razones genuinas de la decisión en concreto, una investigación empírica y, finalmente, la justificación final de dicha hipótesis según las teorías existentes y aprobadas por la comunidad científica para darla por *descubierta*. Sería algo así como generar una teoría sobre las razones genuinas de la decisión  $x$  aprobada por la comunidad científica. Se parte aquí de la premisa de que *descubrir* es un logro epistémico que permite identificar la existencia de algo, y que para que ese algo tenga existencia debe estar validado por las teorías existentes y la comunidad científica.

A diferencia del sentido fuerte, el sentido débil supondría la intuición de las razones genuinas de la decisión con el planteamiento de la hipótesis, una investigación lógica de la misma y su posterior justificación. Este proceso puede caracterizarse como el *descubrimiento* de *algo* sin necesidad de acudir a las reglas ni la aceptación de la comunidad científica. Bastaría, entonces, con el uso de herramientas lógicas interpersonales que permitan explicar las hipótesis. Su finalidad radicaría en explicar *prima facie* una hipótesis al margen de la aceptación científica.

Creemos, entonces, que para el caso del razonamiento jurídico, el sentido débil de *descubrimiento* resulta más relevante, debido al tiempo limitado y a las finalidades para las que es utilizado: justificar una decisión en el marco de las reglas de la lógica y las normas jurídicas.

#### 4. La lógica en el contexto de descubrimiento

Como hemos desarrollado en la sección precedente, en la filosofía de la ciencia de la primera mitad del siglo XX existió una actitud reticente frente a la posibilidad de postular algo así como una “lógica del descubrimiento”. Karl Popper es un buen representante de esta actitud<sup>2</sup>, pues en una de las primeras secciones de *La lógica de la investigación científica* — titulada “Eliminación del psicologismo” — señala que “La etapa inicial, el acto de concebir o inventar una teoría, no me parece que exija un análisis lógico ni sea susceptible de él. La cuestión acerca de cómo se le ocurre una idea nueva a una persona —ya sea un tema musical, un conflicto dramático o una teoría científica— puede ser de gran interés para la psicología empírica, pero carece de importancia para el análisis lógico del conocimiento científico” (Popper, 2017, pp. 37-38). Pensando en la actitud de Popper y otros filósofos de la ciencia, ante el proceso de descubrimiento de teorías, ideas y métodos, Fann (1970) concluye que “(...) estos filósofos ven al descubrimiento de nuevas ideas como una mera conjetura, azar, intuición, corazonada o algún salto mental por parte del científico, que solo

---

2 Citas equivalentes en la obra de otros filósofos de la ciencia se encuentran en Fann (1970, pp. 1-4).

queda abierto para una investigación de tipo histórico, psicológico o sociológico. Cuando ellos hablan de la lógica del descubrimiento, solo se preocupan por la prueba de hipótesis” (p. 2, traducción propia). En suma, se niega la posibilidad de que la lógica del descubrimiento tenga una estructura racional equivalente a la que, por oposición, podríamos llamar lógica de la justificación de una teoría.

La sección precedente ha analizado de qué forma este tipo de formulaciones se relacionan con la distinción entre contexto de descubrimiento (CD) y contexto de justificación (CJ), desarrollada en el seno de la filosofía de la ciencia del empirismo lógico. Por ello, en esta sección nos concentraremos en responder una pregunta más concreta: ¿en qué sentido podemos hablar de una lógica del descubrimiento? Es decir, ¿en qué sentido empleamos aquí la palabra “lógica”? Por su parte, en la siguiente sección trataremos de responder a la pregunta sobre cómo opera dicha lógica, es decir, bajo qué esquema inferencial o argumentativo. Ambas preguntas serán abordadas priorizando los problemas jurídicos relacionados con el razonamiento probatorio.

Comencemos, entonces, con la pregunta de esta sección. Aunque no habría sido el primero —Gascón Abellán (2014, p. 152) menciona el tipo de inferencia aristotélica denominada “reducción” o “retroducción” como un antecedente—, se atribuye a Charles S. Peirce haber estudiado con cierto detalle la estructura de una forma de inferencia distinta a la deducción y a la inducción —aunque también de tipo ampliativo o sintético como la inducción— que llamó abducción (Walton, 2005, p. 3). A la vez, se le atribuye haber sostenido que la abducción es un tipo de inferencia que nos permite dar cuenta de la lógica de descubrimiento. No obstante, como reconocen especialistas en la obra del filósofo estadounidense, es en realidad bastante difícil reconstruir una suerte de teoría unitaria sobre la lógica del descubrimiento y sobre la abducción en la obra de Peirce. Ello, en buena medida, se debe al carácter fragmentario, caótico e incluso a veces contradictorio de la obra del autor pragmatista (Kraus, 2003, p. 237). En la medida que el presente artículo no propone una exégesis de la obra de Peirce, podemos evadir y dejar de lado esta compleja cuestión.

Sin embargo, un punto importante en la propuesta de Peirce es que en la postulación de un tipo de inferencia distinto a la deductiva e inductiva, encontramos una primera interpretación posible para la idea de una lógica del descubrimiento. En efecto, si, desde una perspectiva amplia, por “lógica” entendemos una disciplina dedicada a la reconstrucción y análisis de diversos tipos de inferencia, entonces la tarea a dilucidar en una lógica del descubrimiento consiste en (i) caracterizar el tipo de inferencia más común en lo que genéricamente podemos llamar *actividades de descubrimiento*, y (ii) analizar la estructura de tal tipo de inferencia, es decir, la forma de sus premisas y el tipo de conclusión a la que se arriba.

Podríamos responder, entonces, señalando lo siguiente: la lógica de la justificación remite al tipo de inferencia que se emplea para justificar la conclusión de un razonamiento en el marco de una teoría científica, una sentencia judicial, entre otras formas de discurso. Estas inferencias pueden ser de tipo deductivo, en el que la conclusión se sigue necesariamente a partir de las premisas, siguiendo algún esquema válido de razonamiento; e, inductivo, en el que la conclusión se sigue —al menos en la llamada “inducción probabilística— con algún grado de probabilidad a partir de las premisas<sup>3</sup>. Por su parte, la lógica del descubrimiento remite al tipo de inferencia que se emplea para formular explicaciones o hipótesis que luego serán corroboradas. La forma básica de este tipo de inferencia puede seguir un esquema similar al inicialmente propuesto por Peirce y que será analizado con mayor detalle en la siguiente sección:

Se observa un hecho sorprendente E.

Si H fuese verdadera explicaría que se haya producido E, por lo que E pasaría a ser algo ya no sorprendente.

Por lo tanto, hay razones para pensar que H es verdadera (formulación de Gascón Abellán, 2014, p. 152).

Un ejemplo, tomado del análisis que Giovanni Tuzet (2014, pp. 129-132) realiza de *Los crímenes de la calle Morgue* de E. A. Poe, puede ser de utilidad llegados a este punto.

Se observa el hecho sorprendente de que se ha encontrado, en un departamento de París, una escena impresionante: dos mujeres han sido asesinadas con ensañamiento, una de ellas estrangulada y luego su cuerpo clavado en la chimenea, mientras que la otra mujer ha sido decapitada. Asimismo, el hecho es sorprendente porque aparentemente no hay vías de escape del lugar y los testigos declaran haber escuchado la voz de alguien que parece ser un extranjero, aunque ninguno reconoce el idioma que esta hablaría.

Si la hipótesis de que una criatura no humana ha cometido el crimen (por ejemplo, un animal muy fuerte como un orangután) fuese verdadera, esta explicaría que se haya producido el hecho sorprendente, por lo que este ya no lo sería tanto.

Por lo tanto, hay razones para pensar que la hipótesis de la autoría de una criatura como un orangután es verdadera.

---

3 Recientemente, Tuzet menciona también al razonamiento analógico (2021, p. 129). Es importante apuntar que dicho tipo de razonamiento recibe un tratamiento amplio en la obra de autores de la tradición hermenéutica como Arthur Kaufmann (1999, cap. 5).

En la formulación de Gascón Abellán, la lógica del descubrimiento atiende a la pregunta siguiente: ¿Por qué sucede este fenómeno? Mientras que corresponde a una fase hipotético-deductiva la comprobación de las hipótesis postuladas.

Ahora bien, hasta este punto hemos seguido los pasos de la intuición de Peirce de que una lógica del descubrimiento podría consistir en el análisis de un tipo de inferencia, la abductiva, que difiere en sus rasgos distintivos, de las inferencias deductiva e inductiva. Pero esta no es la única alternativa para dotar de contenido a la idea de una “lógica del descubrimiento”. Gascón Abellán postula que un segundo sentido en el que podemos hablar de una lógica del descubrimiento es entendiendo “lógica” como una suerte de metodología o racionalidad en la realización de un proceso o procedimiento. Este procedimiento se desarrolla en un arco temporal que nos lleva desde las hipótesis que se pueden sugerir a partir de pruebas primeras hasta la corroboración de tales hipótesis: “Podría decirse, por lo tanto, que el proceso de ‘descubrimiento’ de una hipótesis no es estrictamente independiente del proceso para su ‘validación’, pues ambos forman parte de una misma metodología inductiva” (Gascón Abellán, 2014, p. 157). Evidentemente no se trata de un procedimiento de tipo algorítmico, en el que el desarrollo de cada una de las fases cumple una suerte de protocolo estricto de instrucciones, sino de un procedimiento flexible en el que los elementos creativos se combinan con una fase de corroboración más parametrada.

Si entendemos “lógica” en este segundo y más flexible sentido, tenemos buenas razones para concluir que es posible postular una lógica del descubrimiento, entendida como un procedimiento ordenado que nos permite postular hipótesis que serán luego evaluadas. Más aún, si en el sentido más restringido de “lógica” buscamos un tipo de inferencia con una estructura típica, debemos concluir que la inferencia por antonomasia del proceso de descubrimiento es de tipo abductivo (aunque ello no impide que inducciones y deducciones estén presentes en el mismo procedimiento). Por tales razones, podemos concluir que (i) sí es posible hablar de una lógica del descubrimiento, y (ii) en qué consiste tal lógica dependerá de si adoptamos un sentido amplio de “lógica”, o más bien uno más acotado: así, la lógica del descubrimiento podrá consistir en un procedimiento de formulación de hipótesis a partir de datos u observaciones iniciales (sentido amplio) o en un tipo de inferencia empleada típicamente para la postulación de hipótesis, tal como ocurre con la inferencia abductiva (sentido restringido). De esta forma, pues, damos respuesta a la pregunta sobre en qué sentido se puede hablar de una lógica del descubrimiento. En la siguiente sección, tomando el camino del sentido restringido, nos preguntamos por la estructura de la inferencia abductiva.

## 5. Modelos de razonamiento abductivo

De acuerdo con Douven (2017a), a quien seguiremos de cerca en lo sucesivo, el término “abducción” se emplea, en la literatura filosófica, en dos sentidos distintos:

1. En un primer sentido de corte histórico, hace referencia al razonamiento explicativo en la generación de hipótesis.
2. En un sentido más contemporáneo, hace referencia al lugar del razonamiento explicativo en la justificación de hipótesis. En este segundo sentido, también se puede hablar de la “inferencia a la mejor explicación”.

Sobre el primer sentido, Douven (2017b), sostiene que la teoría de la abducción de Peirce se ubica principalmente en el contexto de descubrimiento, es decir, en la fase de investigación en la cual buscamos generar teorías explicativas que serán posteriormente evaluadas. Peirce parece tener esto en mente cuando sostiene, por ejemplo, que la abducción remite al proceso de formación de hipótesis explicativas. La pregunta, entonces, es en qué sentido podemos decir que la abducción es una operación lógica. En Peirce ello se resuelve a través de un esquema de razonamiento sobre la abducción, que es el siguiente:

Un hecho sorprendente C es observado.  
Pero si A fuese verdad, C sería algo natural.  
Entonces, hay una razón para sospechar que A es verdad.

Douven (2017b) nos recuerda que Harry Frankfurt (1958) desarrolla una crítica contra esta propuesta, y sostiene que aquí no hay una inferencia que conduzca a una nueva idea. Ello es así en la medida que la hipótesis explicativa A debe haber ocurrido antes de que uno infiera que esta es verdadera, pues ya aparece en la segunda premisa del esquema de Peirce. Por tal razón, Frankfurt se decanta por una segunda lectura de la abducción en Peirce, mediante la cual esta es un procedimiento no tanto para *inventar* una hipótesis para un hecho sorprendente, sino más bien para adoptar una hipótesis que lo explique. Aquí la abducción funcionaría como un filtro para determinar cuáles de las hipótesis sobre un hecho sorprendente pueden pasar a ser evaluadas en una siguiente fase.

El segundo sentido de “abducción” de los analizados por Douven (2017a) hace referencia al razonamiento explicativo desde el punto de vista de su estructura inferencial. Aquí la abducción se caracteriza como una de las tres más importantes formas de inferencia, siendo las otras la deducción y la inducción: mientras que la deducción es un tipo de inferencia en que la conclusión se sigue necesariamente de las premisas, esto no ocurre en la inducción y en la abducción. De esta forma, mientras que la deducción es analítica, la inducción y

la abducción son sintéticas (*vid.* Niiniluoto, 2018, p. 8). Si bien ello permite diferenciar a la deducción de la inducción y la abducción, es todavía insuficiente para captar el rasgo distintivo de esta última frente a la inducción.

Una posible forma de diferenciar entre inducción y abducción consiste en sostener que en la abducción no solo ni principalmente se formula una conclusión basada en frecuencias estadísticas y probabilidades, sino que el centro se encuentra en una apelación a consideraciones explicativas<sup>4</sup>. Así, la hipótesis formulada explicaría coherentemente la colección de datos observados, y debido a ello se propone aceptarla.

Ahondemos en lo señalado en el párrafo precedente. A diferencia de la deducción, la abducción es un tipo de razonamiento no monotónico, en la medida que la adición o sustracción de premisas de un conjunto puede alterar la calidad de la explicación propuesta<sup>5</sup>. Pero es, además, un razonamiento que apunta hacia una explicación falible sobre un conjunto de datos. Por ejemplo, a partir de (a) la declaración de un testigo que señala haber visto a un hombre parecido a Manuel rondar la casa de Gonzalo minutos antes de la hora aproximada de la muerte (producida por puñaladas) de este, (b) de la lectura de varios correos electrónicos y mensajes de Whatsapp en los que Manuel amenaza a Gonzalo con hacerle daño y hasta matarlo, y (c) un video de seguridad que muestra que Manuel salió temprano de su oficina el día de la muerte de Gonzalo, el fiscal conjetura que la mejor hipótesis explicativa sobre quién pudo asesinar a Gonzalo es que lo hizo Manuel. De esta forma, piensa el fiscal, se contaría con un móvil —el odio de Manuel hacia Gonzalo—, así como con la posibilidad física de hacerlo —Manuel salió temprano del trabajo y pudo llegar a las inmediaciones de la casa de Gonzalo— y con una declaración que coloca a una persona que físicamente aparenta ser Manuel en las inmediaciones del lugar del crimen. Sin embargo, si a este conjunto de datos se agregara (d) una coartada por parte de Manuel —este, por ejemplo, se dirigió a casa de sus abuelos, donde permaneció hasta horas después de la hora de muerte de Gonzalo—, y esta fuese confirmada por declaraciones de testigos, o incluso por imágenes fotográficas o de video, la hipótesis original perdería gran parte de su poder explicativo. Con ello, podemos concluir que la incorporación de nuevas premisas en un razonamiento abductivo puede alterar la inferencia, ya no garantizando, en el caso de nuestro ejemplo, que Manuel es el probable autor del crimen.

---

4 En la misma línea, sostiene Tuzet (2021) que “En sentido estricto una inducción es una inferencia que generaliza datos; sus conclusiones no versan sobre casos particulares, sino sobre clases en su totalidad, mientras que las conclusiones del razonamiento probatorio [que para Tuzet es principalmente una inferencia a la mejor explicación de corte abductivo] en un contexto jurídico conciernen al caso particular de que se discute” (p. 133).

5 Desde luego, tanto la inducción como la abducción son tipos de razonamiento no monotónico. No obstante, lo que aquí resulta relevante es enfatizar la orientación explicativa de la abducción. En ese sentido, este trabajo no desarrolla una distinción refinada entre abducción e inducción, labor que se reserva para algún trabajo posterior.

Douven (2017a) analiza algunas formulaciones esquemáticas de la abducción tomando en consideración algunas propuestas existentes en la literatura especializada. Una de estas versiones es la que llama “versión de libros de texto de la abducción” (*textbook version of abduction*), que tiene varias posibles formulaciones. Analicemos estas con mayor detalle:

Primera formulación: Dada la evidencia  $E$  y candidatas explicaciones  $H_1, \dots, H_n$  para  $E$ , inferimos la verdad de aquella  $H_i$  que mejor explique  $E$ .

Esta primera versión exhibe, desde el punto de vista de Douven, varios problemas: en primer lugar, presupone los conceptos de “explicaciones candidatas” y de “mejor explicación”. Para la noción de “mejor explicación”, se suele recurrir a virtudes de las teorías, tales como simplicidad, generalidad, coherencia y establecimiento de la teoría. Ahora bien, ninguna de estas virtudes posee una caracterización unánime en la literatura especializada, e incluso los críticos más extremos llegan a sostener que estas virtudes teóricas no tienen más que un rol retórico en la ciencia.

Un segundo problema de tal caracterización es que fuera de la vaguedad de los conceptos involucrados, esta podría ser demasiado fuerte: recordemos que la formulación habla de “la verdad” de una de las hipótesis. Hay quienes consideran que tal exigencia es muy elevada, y que puede ser reemplazada por expresiones como “probable verdad”, “verdad aproximada”, o una combinación de ambas.

Un tercer problema que considera Douven (2017a) es que en esta caracterización de la abducción, podríamos trabajar con un mal conjunto de explicaciones, siendo una de estas menos mala que las otras hipótesis propuestas. Ello es así pues solo se evalúa la relación entre la evidencia  $E$  y un conjunto de candidatas explicaciones. Para superar tal problema, y otorgar confiabilidad a la inferencia, deberíamos suponer que la hipótesis explicativa tiene el potencial de ser la mejor respecto de cualquier otro conjunto de hipótesis que podríamos haber concebido. No obstante, esta suposición parece ser, de acuerdo con los críticos, implausible, aunque hay intentos en el sentido de que se podría formular un universo finito de hipótesis que agoten el espacio lógico.

Finalmente, otra estrategia de crítica consiste en mostrar que la construcción de la primera formulación es asimétrica, en el sentido de que inicia con una premisa comparativa (sobre diversas hipótesis en pugna) y termina con una conclusión absoluta, que infiere la verdad de una de las hipótesis. Si ello es así, debemos evaluar otras alternativas.

Segunda formulación: Dada la evidencia  $E$  y candidatas explicaciones  $H_1, \dots, H_n$  de  $E$ , inferimos la verdad de aquella  $H_i$  que mejor explique  $E$ , siempre que  $H_i$  sea una explicación, en tanto explicación, suficientemente satisfactoria/buena.

Aquí, nuevamente, tenemos un problema de vaguedad, en la medida que se debe determinar qué cuenta como explicación satisfactoria/buena. No obstante, se logra salvar algunas de las críticas formuladas contra la primera versión, en la medida que un mal conjunto de explicaciones no contarán como explicaciones “suficientemente buenas o satisfactorias”. No se logra salvar, sin embargo, la última de las críticas a la primera formulación, en la medida que se mantiene la falta de simetría entre premisas comparativas y una inferencia general de verdad. Ahora bien, esa crítica se ve matizada en la medida que la hipótesis explicativa en esta segunda formulación no solo es comparativamente la mejor del conjunto analizado, sino que es en sí misma buena/satisfactoria como explicación. De esta forma, la doble vertiente de exigencias (comparativas y sobre la calidad de la hipótesis) parece otorgar un mejor fundamento a la pretensión de verdad de la conclusión.

Finalmente, una tercera formulación se articula del siguiente modo:

Tercera formulación: Dada la evidencia  $E$  y candidatas explicaciones  $H_1, \dots, H_n$  de  $E$ , si  $H_i$  explica  $E$  mejor que cualquiera de las otras hipótesis, inferimos que  $H_i$  está más cerca de la verdad que cualquiera de las otras hipótesis.

Douven (2017a) considera que si bien aquí se mantiene el problema de definir que cuenta como “cercano a la verdad”, la formulación posee la ventaja de que preserva la simetría entre premisa y conclusión de la inferencia. No obstante, un problema con esta tercera formulación es que la exigencia de que la hipótesis sea suficientemente buena o satisfactoria ha sido eliminada. Con ello, si bien se salva el problema de la falta de simetría entre premisa y conclusión, reingresa el problema de contar con un conjunto de hipótesis malas que son evaluadas.

Dados los inconvenientes que exhiben las tres opciones evaluadas por Douven (2017a), aquí proponemos una cuarta formulación, que, además, tiene una especial relevancia en el derecho y, en particular, en el razonamiento probatorio:

Propuesta de formulación: Dada la evidencia  $E$ , que en el razonamiento probatorio del derecho consiste en un conjunto de datos obtenidos a partir del análisis de medios probatorios; y dadas las candidatas explicaciones  $H_1, \dots, H_n$  para  $E$ , si (i)  $H_i$  explica  $E$  mejor que cualquiera de las otras hipótesis, y si (ii)  $H_i$  es una explicación, en tanto que explicación, suficientemente satisfactoria/buena para  $E$ , inferimos que  $H_i$  está más cerca de la verdad que cualquiera de las otras hipótesis.

Evaluemos si esta propuesta sobrepasa las críticas formuladas por Douven contra la primera formulación, varias de las cuales se extendían a las otras alternativas analizadas por el autor mencionado. En primer término, nuestra propuesta mantiene el problema de cierta

indeterminación en conceptos como “mejor explicación”, “cercanía a la verdad”, entre otros. Por ello, será menester tratar de clarificar tales conceptos para el caso del razonamiento probatorio, que es el centro del análisis aquí. No obstante, en nuestra propuesta se salva el segundo problema, pues no se infiere la verdad de la hipótesis sino la “mayor cercanía la verdad”. Por ello, nuestra propuesta no es tan fuerte como la primera. Asimismo, se salva la crítica del posible conjunto de malas explicaciones, en la medida que nuestra propuesta recurre a dos criterios: comparativo (entre  $H_i$  y las otras candidatas explicaciones) y sobre la calidad de la hipótesis, por lo cual la  $H_i$  debe ser, en sí misma, una explicación suficientemente satisfactoria para E. Una estrategia análoga se encuentra, por ejemplo, en las dos condiciones que componen la propuesta de un estándar de prueba para el proceso penal (el estándar de más allá de toda duda razonable) propuestas por Ferrer Beltrán (2007, p. 147).

Entonces, el concepto de suficiencia puede corresponderse, para el razonamiento probatorio, con la noción de estándar de prueba, pero esto es algo que queda por desarrollar en nuevas investigaciones.

Finalmente, nuestra propuesta salva el problema de la falta de simetría entre premisa y conclusión, en la medida que la premisa es comparativa y la conclusión no es absoluta sino altamente plausible (por ello está definida en términos de cercanía a la verdad). Si el intento emprendido hasta este punto es exitoso, hemos logrado ofrecer una caracterización de la estructura de la inferencia abductiva que la diferencia de otras inferencias como la deductiva o la inductiva. Queda, entonces, por analizar cómo la abducción se materializa en el análisis de casos relacionados con el razonamiento probatorio.

## 6. La práctica de la abducción en el razonamiento probatorio jurídico

De acuerdo a Bonorino (2014) “(...) uno de los problemas al que nos enfrentamos quienes tratamos el tema del razonamiento abductivo es el poder ofrecer un ejemplo claro y reconocible de abducción” (p. 163). Sin embargo, ello no es impedimento para asumir el reto de dicha empresa y ofrecer dos ejemplos reales y recientes<sup>6</sup> de cómo la abducción, primero, nos permite dotar de racionalidad al contexto de descubrimiento y, segundo, cómo el análisis de casos en concreto, en el marco de un análisis abductivo, nos permite elaborar un diseño institucional para acercarnos a la verdad, según expusimos en las secciones precedentes.

Esta empresa, además, resulta necesaria para disipar y superar la brecha que suele existir entre la teoría, los diseños intelectuales y la práctica jurisprudencial de los juzgados en el

---

6 Ambos casos fueron resueltos por la Sala Penal Permanente de la Corte Suprema de Justicia de la República del Perú, y están disponibles en la página web del Poder Judicial ([www.pj.gob.pe](http://www.pj.gob.pe)), en los ítems de “Consulta de expedientes judiciales - Supremo” y “Jurisprudencia”. También están disponibles en otras páginas web, como [www.gacetajuridica.com.pe](http://www.gacetajuridica.com.pe) o [www.lpderecho.pe](http://www.lpderecho.pe).

derecho, y así evitar que existan dos mundos paralelos de teoría y práctica. El objetivo, entonces, es que ambos ámbitos puedan colaborar de forma fructífera.

### *1. Un primer caso: Recurso de Nulidad N.º 2098-2019 Lima<sup>7</sup>*

El 4 de julio de 2018, la policía intervino y detuvo a Nilso Lerma, de nacionalidad colombiana, cerca del local de la empresa de transportes Flores Hermanos, ubicada en el distrito de La Victoria, Lima. Días antes, Nilso Lerma había enviado una encomienda desde la ciudad de Lima hacia Puno, y luego había viajado hacia esta última ciudad. Sin embargo, no logró tal cometido y solo llegó hasta Arequipa, ciudad que está en camino hacia Puno. Luego de este viaje trunco regresó a Lima a recoger la encomienda que había enviado antes. La detención de Lerma se debió a que actuaba de forma sospechosa y pretendía recoger la encomienda previamente enviada. Al realizarse el registro del bien enviado, se encontraron dos cilindros de pintura que tenían compartimientos acondicionados, y al interior de estos se hallaron 29 paquetes donde existían restos de vegetales, tallos y semillas, similares a la marihuana.

Ante estos primeros hallazgos, se realizaron las siguientes diligencias: (i) se redactó un acta de registro personal e incautación, donde se dejó constancia de la encomienda y los bienes hallados en poder del intervenido: tres celulares, la boleta de envío de la encomienda y “nada de dinero” (sic); (ii) la sustancia encontrada fue sometida a los exámenes químicos y de pesaje correspondientes, y se determinó que era marihuana, con un peso bruto de 27,945 kilogramos; y, (iii) se recibió la declaración del intervenido, quien señaló que el 19 de junio de 2018 llegó a la ciudad de Piura, proveniente desde Colombia, y fue allí que casualmente conoció a Ítalo Madrid, quien durante varios días lo ayudó a conocer esta ciudad, le hizo diversos favores, lo atendió amablemente y le pidió el favor de que lo ayude a enviar una encomienda de pintura de Lima a Puno. De acuerdo con la narración de Lerma, Ítalo Madrid no podía realizar el envío porque extravió su documento de identidad. Nilso Lerma accedió al pedido, de acuerdo a su versión de los hechos, a pedido de Madrid, y debido a la amabilidad y la gran hospitalidad con la que fue atendido. Lerma también indicó que no conocía el contenido ilícito de la encomienda que envió ni participó en su acondicionamiento, y que vino al Perú a conocer los restos arqueológicos de Machu Picchu. Asimismo, precisó

---

7 En el ordenamiento jurídico peruano coexisten dos normas procesales bajo cuyas reglas se siguen los procesos penales: el Código de Procedimientos Penales (derogado, pero que aún se aplica en los casos penales ocurridos con anterioridad al 2004) y el Código Procesal Penal de 2004 (de vigencia progresiva en todo el país). El recurso extraordinario que permite llegar a la Corte Suprema de Justicia, para el caso del Código de Procedimientos Penales, es el Recurso de Nulidad (los casos evaluados poseen tal condición). Por su parte, en el caso del Código Procesal Penal, es el Recurso de Casación. Más allá de los matices propios de cada norma procesal, en ambos casos la Corte Suprema conoce todos los casos elevados a dicha instancia, las cuales son numerados según el orden de ingreso a dicha instancia, seguido del nombre de la Corte Superior de Justicia de donde proviene el caso.

que el personal de la empresa de transportes, días después de haber enviado la encomienda y cuando se encontraba en Arequipa, se comunicó con él y le dijo que del paquete estaba goteando un líquido, por lo que tenía que recogerlo, y fue al realizar esta acción que fue detenido por la policía. Finalmente, añadió que consignó su propio nombre como destinatario de la encomienda enviada.

A partir de lo descrito hasta aquí, tenemos los siguientes “datos probatorios disponibles” (D):

- **D1:** Nilso Lerma e Ítalo Madrid Riofrío aparentemente se conocieron en Piura; además, este le hizo muchos favores y lo atendió amablemente en dicha ciudad.
- **D2:** Nilso Lerma envió una encomienda de Lima a Puno.
- **D3:** Dicha encomienda tenía camuflados 27,945 kilogramos de marihuana.
- **D4:** Nilso Lerma indicó que desconocía del contenido ilícito de la encomienda y precisó que estaba viajando a Cusco, a conocer Machu Picchu.
- **D5:** *Días después de haber realizado el envío de la encomienda*, Lerma regresó a la empresa de transportes, a recoger el bien enviado, debido a que le informaron que el mismo estaba goteando. No se encontró en su poder suma dineraria alguna.
- **D6:** Solo se intervino a Nilso Lerma cuando este se acercó a recoger la encomienda, y se advirtió que el envío tenía como destinatario al mismo remitente.

A partir de los datos probatorios descritos, se advierte que estamos ante un caso misterioso: de un lado, el intervenido envió una encomienda que tenía camuflada una cantidad considerable de marihuana y, de otro, indicó que hizo ello como agradecimiento por la hospitalidad y los muchos favores que un ciudadano peruano —que casualmente conoció en Piura— le realizó. Pero también señaló que desconocía del contenido ilícito del envío (creía que era pintura) y añadió que tampoco participó en el acondicionamiento de la droga.

¿Por qué creer que una persona que viajaba a Machu Picchu, de ser cierto ello, haya enviado una encomienda con marihuana a Puno?, ¿será cierto que hizo ello como agradecimiento por la hospitalidad que recibió en Piura?, ¿es razonable actuar de esa forma?, ¿había alguna forma de conocer del contenido ilícito del envío?, ¿habrá participado en su acondicionamiento?, ¿qué habría hecho usted o una persona razonable en similares circunstancias? Estas interrogantes, entre otras muchas que podrían surgir, hacen que el caso sea uno misterioso (además de ser bastante común en los tribunales peruanos).

*a. Las posibles hipótesis sobre lo ocurrido*

A partir de los “datos probatorios disponibles” podemos formular las siguientes “hipótesis explicativas”<sup>8</sup> (H) sobre lo acontecido:

- **H1:** Nilso Lerma (NL), con la asistencia de Ítalo Madrid, envió, desde Lima a Puno, una encomienda que tenía camuflada marihuana y conocía del contenido ilícito del envío. Respaldan esta afirmación el hecho de que envió una encomienda (**D2**) que tenía camuflados 27,945 kilogramos de marihuana (**D3**) y, días después, regresó a recogerla a Lima (**D5**).
- **H2:** NL envió a Puno una encomienda (**D2**) que tenía camuflada marihuana (**D3**), pero no conocía de su contenido ilícito (**D4**), pues realizó el envío en gratitud por la hospitalidad y los favores que le hizo Ítalo Madrid (**D1**).
- **H3:** NL envió la encomienda (**D2**) que tenía camuflada una cantidad considerable de marihuana (**D3**), sin que haya intervenido en ello Ítalo Madrid u otra persona, pues la policía solo intervino al primero (NL); además, al momento de realizar el envío, consignó como destinatario de la encomienda su propio nombre (**D6**).

De todas estas hipótesis podemos excluir a la tercera (**H3**), pues consideramos que el investigado debió requerir la asistencia de terceras personas para adquirir la sustancia ilícita en el Perú, camuflarla en los cilindros de pintura, obtener dinero para realizar todos los actos preparatorios del traslado y envío de la encomienda y, después, ejecutado propiamente dicha actuación (envío). Estas razones debilitan la tercera hipótesis. Además, tampoco existen mayores “datos probatorios disponibles” que doten de mayor credibilidad a tal hipótesis.

Asimismo, podemos afirmar que existen suficientes razones y “datos probatorios disponibles” que dotan de credibilidad a las dos primeras hipótesis explicativas (**H1** y **H2**), debido a que, de un lado, Nilso Lerma realizó el envío de una encomienda donde se camufló una cantidad considerable de marihuana y después fue a recogerla; y de otro, se presume su inocencia, por tanto, también debe considerarse su versión de los hechos. De modo que estas dos hipótesis resultan sostenibles.

*b. Actuaciones a realizar para resolver el caso*

---

<sup>8</sup> Seguramente podrían formularse más hipótesis; sin embargo, únicamente con el fin de presentar un caso de la forma más clara posible, no haremos mayores proposiciones hipotéticas que podrían complicar o hacer más tediosa la lectura y el entendimiento del caso y, por tanto, del ejemplo propuesto.

A partir de los “datos probatorios disponibles” y las “hipótesis sostenibles advertidas”, resulta claro que se necesita la producción de nuevos elementos que permitan determinar la verdad de lo ocurrido. Entre dichas actuaciones tenemos las siguientes:

- **A1:** Ver los mensajes de texto que envió y recibió NL *días antes de* la remisión de los bienes ilícitos y su intervención. De la lectura de estos mensajes se advirtió que el intervenido se comunicó con Ítalo Madrid y allí coordinaron sobre cuestiones tales como (i) el traslado de la encomienda donde se escondió la marihuana, desde Piura hacia Lima; (ii) su envío desde Lima hacia Puno y (iii) los 500 dólares *que le pagarían por realizar dicha labor ilícita*. De estas conversaciones también se advirtió que (vi) Lerma viajó a Arequipa y allí intentó recoger la encomienda, pero que ello no se concretó porque la encomienda se quedó en Lima, y, (v) ante la llamada de los representantes de la empresa de transportes, Lerma regresó a Lima, manifestando a Ítalo Madrid que tenía miedo de que hayan descubierto la droga o le estén tendiendo una trampa para detenerlo. Finalmente, se advirtió que (vi) *Ítalo Madrid le dijo que por seguridad borre las conversaciones que mantuvieron*.
- **A2:** Recibir las declaraciones de los policías que intervinieron a NL. Los efectivos policiales Cali Lima y Anderson Carrasco indicaron que por información de inteligencia tomaron conocimiento del envío de una encomienda con marihuana camuflada, por lo que intervinieron a Nilso Lerma y, en el envoltorio que envió a Puno, se encontraron 29 paquetes de marihuana.
- **A3:** Recibir las declaraciones de los representantes de la empresa de transportes Flores Hermanos. Yony Ordoño indicó que participó en la intervención de Nilso Lerma y detalló que en la encomienda se encontró marihuana. Por su parte, Julisa Huamán señaló que recibió la encomienda, que NL no se quiso identificar y solo ante su insistencia entregó su carnet de identidad colombiana; también dijo que el investigado le indicó que el paquete enviado contenía herramientas.
- **A4:** Recabar las grabaciones de seguridad del momento de envío de la encomienda donde se camufló la marihuana. De la revisión de estas imágenes se advierte que los relatos brindados por Yony Ordoño y Julisa Huamán son veraces.

- **A5:** Identificar a Ítalo Madrid Riofrío y recibir su declaración. Esta persona aceptó su responsabilidad, pero no otorgó información adicional sobre lo ocurrido.

*c. Evaluación definitiva de las hipótesis*

Del análisis de los nuevos elementos actuados no resulta posible proponer nuevas hipótesis explicativas del caso. Por ello, ahora debe identificarse cuál es la “mejor explicación del caso”, partiendo de las alternativas existentes. Para ello se analizará qué elementos confirman o rechazan alguna de las dos “hipótesis sostenibles”, es decir, se evaluará los puntos fuertes y débiles de ambas hipótesis.

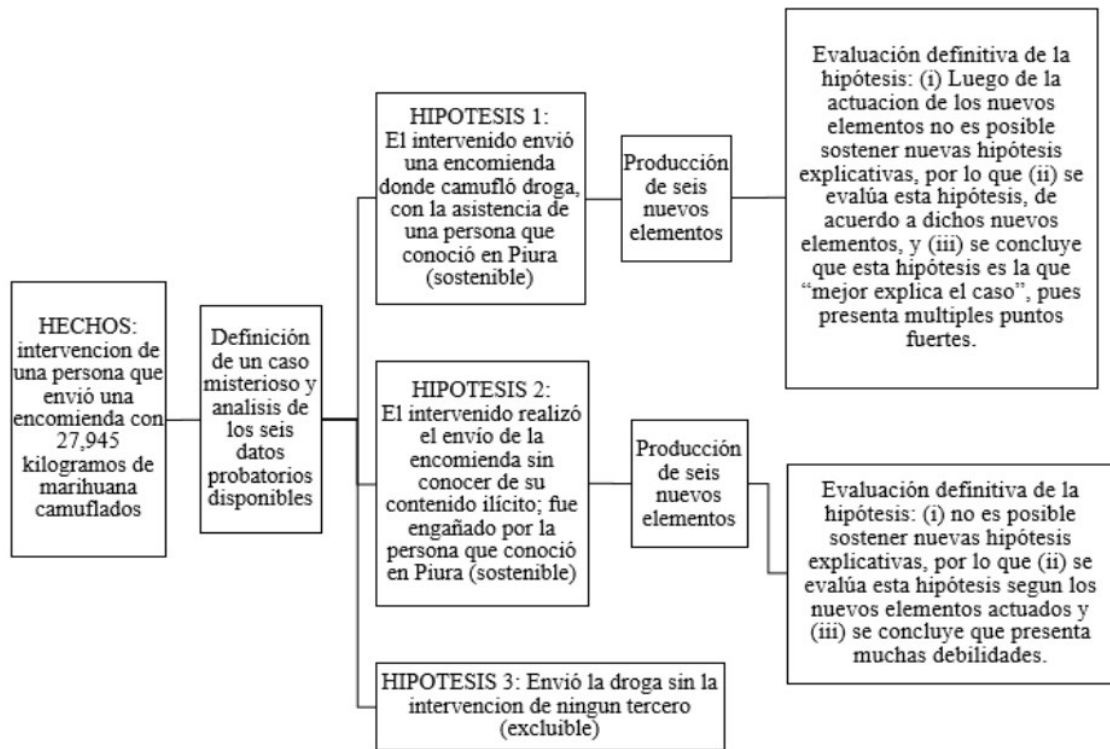
Con relación a la primera hipótesis (**H1**) advertimos que esta cuenta con puntos sólidos, pues NL envió una encomienda de Lima a Puno (**D2**), que contenía 27,945 kilogramos de marihuana (**D3**), que estaban camuflados en el interior de dos cilindros de pintura. Además, días después de haber realizado el envío de la encomienda regresó a la empresa de transportes a recoger el bien enviado (**D5**). Esto, a su vez, es respaldado con los mensajes de texto y las conversaciones por mensajes de texto que mantuvo NL (**A1**) con el otro responsable del delito, las declaraciones de los policías Cali Lima y Anderson Carrasco (**A2**), los relatos de Yony Ordoño y Julisa Huamán (**A3**) y la visualización de los videos de seguridad de la empresa de transportes (**A4**). No encontramos puntos débiles en esta hipótesis.

Respecto a la segunda hipótesis (**H2**) advertimos puntos fuertes y puntos débiles. Sobre los primeros, NL e Ítalo Madrid aparentemente se conocieron en Piura (**D1**), aun cuando ello no fue corroborado, y solo se intervino a NL cuando pretendía recoger la encomienda (**D6**). Sin embargo, estos datos iniciales adolecen de muchos puntos débiles, pues Ítalo Madrid no otorgó información alguna sobre lo ocurrido (**A5**) y se limitó a aceptar su responsabilidad; tampoco se encontró en posesión del intervenido dinero alguno o bienes suficientes para que puedan hacer presumir su viaje a Machu Picchu (**D5**). Este análisis nos permite concluir que la segunda hipótesis (**H2**) presenta importantes puntos débiles que hacen posible su exclusión.

De modo que, después de haber pasado *de las pruebas a las hipótesis* (narración de hechos, análisis de datos, formulación de todas las hipótesis y selección de las que resulten posibles, donde se hizo una indagación necesaria, la reconstrucción de los hechos con hipótesis decisivas y el establecimiento de las direcciones de la investigación) y, luego, *de las hipótesis a las pruebas* (análisis decisivo de las hipótesis planteadas, a partir de los nuevos elementos producidos, y la verificación de las mismas —confirmación o rechazo de las hipótesis con las pruebas decisivas, a partir de su análisis de puntos fuertes y débiles—)

podemos concluir que la mejor explicación del caso o la hipótesis que mejor explica el caso (definitiva) es **H1**, esto es, que NL envió, desde Lima hacia Puno, una encomienda que tenía camuflada marihuana y conocía del contenido ilícito del envío<sup>9</sup>.

Todo el análisis precedente puede graficarse de la siguiente manera:



**Gráfico 2:** Síntesis del análisis del Recurso de Nulidad N.º 2098-2019 Lima  
Fuente: elaboración propia.

El análisis del caso descrito, que fue recientemente resuelto por la Corte Suprema de Justicia, nos permite dotar de relevancia práctica a los argumentos expuestos en las secciones precedentes, y así defender la tesis de que la incorporación de un análisis sobre la abducción nos lleva hacia una teoría de la argumentación jurídica centrada no solo en el contexto de justificación, sino también en el contexto de descubrimiento, esto es, hacia una compleja teoría sobre la racionalidad en el análisis y la toma de decisiones, capaz de incorporar la racionalidad del proceso de descubrimiento (en la misma línea véase Sotomayor, 2021b, pp. 107-128).

<sup>9</sup> Esta también fue la conclusión de la Corte Suprema de Justicia, que condenó a Nilso Lerma como autor del delito de tráfico ilícito de drogas, en perjuicio del Estado peruano.

Ahora bien, nuestro análisis no concluye aquí, pues a continuación pasaremos revista a un segundo caso, siguiendo el mismo esquema descrito, pero esta vez avanzaremos un poco más: a partir de un análisis abductivo, primero introduciremos una nueva hipótesis (después de la actuación de nuevos elementos) y, luego, intentaremos elaborar un diseño institucional que nos permita llegar a la verdad del caso (a futuro).

## 2. *Un segundo caso: Recurso de Nulidad N.º 1562-2019 Selva Central*

La menor de iniciales G. M. G. quedó embarazada a los 12 años de edad. Ello motivó que Iván Villaizan, que era su pareja sentimental, sea detenido, pues la menor jurídicamente no podía consentir ningún acto sexual<sup>10</sup>. Esta persona, al ser detenida, indicó que no mantuvo relaciones sexuales con la menor embarazada y que esta, cuando tenía 11 años, convivió y practicó actos sexuales con Fray Huacho, en la vivienda donde residía –la menor– con su hermana y abuela, aprovechando que aquel –Fray Huacho– laboraba en una obra de construcción cercana.

Ante estas primeras declaraciones, se realizaron las siguientes diligencias: (i) se recibió la declaración de la menor, quien señaló que antes de mantener relaciones sexuales con Iván Villaizan, que era su conviviente y el padre del niño que esperaba, convivió y mantuvo relaciones sexuales con Fray Huacho, y añadió que el abogado Percy Coronado le pidió que niegue haber mantenido relaciones sexuales con Iván Villaizan, para que esta persona sea liberada; (ii) se recibió la declaración de Avelina García, madre de la menor, quien genéricamente indicó que su hija convivió con Fray Huacho; (iii) se practicó un examen médico legal a la menor, el cual concluyó que tenía 18 semanas de embarazo y presentaba signos de “desfloración antigua”; y, (iv) se practicó un informe social donde se indicó que dos de las tías de la menor señalaron que su sobrina convivió y mantuvo relaciones sexuales con Fray Huacho.

A partir de lo descrito, tenemos los siguientes “datos probatorios disponibles” (D):

- **D1:** La menor vivía con su hermana y su abuela (no con su madre), y a los 12 años quedó embarazada.
- **D2:** Ella indicó que convivió, respectivamente, con Fray Huacho e Iván Villaizan, y que este último era el padre del niño que esperaba.
- **D3:** G. M. G. *añadió que el abogado Percy Coronado le pidió que mienta, para que Iván Villaizan (patrocinado del abogado y, a la vez, pareja de la menor) sea liberado.*

---

<sup>10</sup> Es importante precisar que aquí no realizaremos un análisis y evaluación desde una perspectiva o enfoque de género, con el único fin de que podamos reconstruir el problema de la forma más fidedigna posible respecto de lo resuelto por la Sala Penal Permanente de la Corte Suprema de Justicia de la República.

- **D4:** Avelina García e Iván Villaizan indicaron que la menor convivió con Fray Huacho, en la casa donde aquella –a menor– vivía con su hermana y su abuela, antes de quedar embarazada. A similar conclusión arribó el informe social actuado.

A partir de los datos probatorios descritos, se advierte que estamos ante un caso complejo: de un lado, una menor de 12 años quedó embarazada y, de otro, no se sabe con quién mantuvo relaciones sexuales, a pesar de que jurídicamente no tenía libertad para consentir ello (ni quién es el padre del niño que espera). Su madre y su pareja sentimental indican que convivió con Fray Huacho, niegan la paternidad de Iván Villaizan y el abogado Percy Coronado aparentemente indujo a la menor a que mienta para que su patrocinado sea liberado. Asimismo, no se recibió la declaración de Fray Huacho, debido a que no se le pudo ubicar en la localidad donde ocurrieron los hechos. ¿Qué habrá ocurrido en verdad?

*a. Las posibles hipótesis sobre lo ocurrido*

A partir de los “datos probatorios disponibles”, podemos formular las siguientes “hipótesis explicativas” (H) de lo acontecido:

- **H1:** La menor convivió (**D4**) y fue agredida sexualmente por Fray Huacho (**D2**).
- **H2:** La menor convivió (**D4**), fue agredida sexualmente por Fray Huacho (**D2**) y quedó embarazada de esta persona (**D1**).
- **H3:** La menor convivió (**D4**), fue agredida sexualmente por Iván Villaizan (**D2**) y quedó embarazada de esta persona (**D1**), pero su abogado le dijo a la menor que mienta, para que su pareja sea liberada.
- **H4:** La menor convivió (**D4**) y fue agredida sexualmente por Fray Huacho (**D4**) e Iván Villaizan (**D2**), es decir, por ambas personas.

De estas hipótesis explicativas, podemos excluir a **H2** (la menor no dijo que el padre del niño que esperaba fuese Fray Huacho, por lo que esta hipótesis no cuenta con datos probatorios que la amparen<sup>11</sup>) y **H4** (las hipótesis **H1** y **H3** engloban mejor a esta hipótesis), debido a que presentan mínimos datos probatorios a su favor y varios datos en su contra<sup>12</sup>,

---

11 Es importante no perder de vista que en esta etapa del análisis solamente contamos con datos probatorios obtenidos de las declaraciones de la menor y su madre, así como un examen médico legal a la menor, y un informe social.

12 Quizás alguna de estas hipótesis también podrían resultar sostenibles, en mayor o menor medida, según los datos

a diferencia de **H1** y **H3**, sobre las cuales únicamente existen datos probatorios que las respaldan y, a su vez, las dotan de mayor credibilidad. De modo que estas dos últimas hipótesis resultan sostenibles.

*b. Actuaciones a realizar para resolver el caso*

A partir de los “datos probatorios disponibles” y de las “hipótesis sostenibles” advertidas, es necesaria la “producción de nuevos elementos” que permitan determinar la verdad de lo posiblemente ocurrido. Entre dichas actuaciones tenemos:

- **A1:** Recibir la declaración de la menor, a efectos de que ofrezca mayores detalles sobre lo que, de forma genérica, había señalado antes. Esta declaración busca establecer con quién mantuvo relaciones sexuales y convivió, quién es el padre de su hijo y cuál fue la participación del abogado Percy Coronado en su inicial declaración. Así, se recibieron dos declaraciones de la menor, donde reiteradamente señaló lo siguiente:
  - Vivía con su abuela y su hermana, no con su madre.
  - No brindó detalle alguno sobre su presunta convivencia o realización de actos sexuales con Fray Huacho, porque ello nunca ocurrió.
  - Sindicó a Fray Huacho para que liberen a Iván Villaizan, que era su conviviente y el padre del niño que esperaba, por recomendación del abogado Percy Coronado, quien le dijo que no pasaría nada si mentía, pues no podrían identificar o detener a Fray Huacho, ya que no sabían dónde estaba. Así, con la declaración falsa, lograría que se libere a su pareja.
  - Tanto Iván Villaizan como los miembros de su familia presionaron a G. M. G. para que mienta, pues le decían que por su culpa se detuvo a su conviviente y que su hijo se quedaría sin padre.
  - Conoció de vista a Fray Huacho debido a que su abuela le vendía comida en una obra, más nunca tuvo relación sentimental o amical alguna con esta persona. Más aún, G. M. G. señaló que, debido a su

---

probatorios disponibles que consideremos pertinentes; incluso podría existir una quinta hipótesis (otra persona, no identificada, agredió sexualmente a la menor y es el padre del niño que espera); sin embargo, nuevamente con el fin de presentar un caso de la forma más clara y concreta posible, no haremos mayores formulaciones que podrían complicar o hacer más tediosa la lectura y el entendimiento del caso y, por tanto, del ejemplo propuesto.

inmadurez, trataba despectivamente a Fray Huacho, quien le parecía una persona de la sierra que “no se arreglaba físicamente”.

- Cuando quiso decir la verdad, Iván Villaizan se la llevó a otro lugar.
  - Cuando quedó embarazada convivía con Iván Villaizan y la familia de esta persona, quienes la presionaban y coaccionan para que no diga la verdad y así se evite perjudicar a su conviviente y real agresor, con quien tuvo relaciones sexuales a los 11 años.
  - No tuvo ninguna comunicación, contacto o cercanía con Fray Huacho, antes o después de quedar embarazada; tampoco tenía vínculo de dependencia económica, afectiva o familiar con esta persona.
- **A2:** Recibir la declaración de la hermana y la abuela de la menor, debido a que al momento de ocurridos los hechos G. M. G. vivía con ellas.
    - Ernestina Chumbes (la abuela) señaló que vendía comida a Fray Huacho en una obra donde esta persona trabajaba, pero que no tenía mayor cercanía con dicha persona. También indicó que la menor es su nieta, vivía con ella y que esta nunca convivió con Fray Huacho. Asimismo, precisó que el enamorado de la menor era Iván Villaizan y que el abogado Percy Coronado fue quien indujo a mentir a la menor, de forma tal que se libere a Iván Villaizan.
    - Yesenia Miranda (la hermana) indicó que vivía con la menor G. M. G. y su abuela. También señaló que Fray Huacho jamás vivió con su hermana y que ella dijo eso para que liberen a Iván Villaizan y todo fue organizado por el abogado Percy Coronado. Además, precisó que cuando la menor quiso decir la verdad, Iván Villaizan se la llevó de la casa.
    - Ambas personas también señalaron que no tuvieron comunicación, contacto o cercanía alguna con Fray Huacho y precisaron que esta persona estuvo en el pueblo por un periodo corto, en que laboró en una obra de construcción, y después se fue, sin que lo hayan vuelto a ver.
  - **A3:** Recibir la declaración de Fray Huacho. Esta persona negó cualquier amistad, cercanía, convivencia o realización de algún acto sexual con la menor. Asimismo, indicó que vivía en otro lugar y que nunca tuvo contacto con la menor, a quien solo conocía porque su abuela le vendía comida en

la obra donde él trabajaba, pero una vez que concluyó esta actividad se regresó a su ciudad de origen, sin saber nada de lo supuestamente ocurrido.

*c. Evaluación definitiva de las hipótesis*

Del análisis de los nuevos elementos actuados resulta posible sostener una nueva hipótesis explicativa del caso (nótese que esto no ocurría en el primer ejemplo que analizamos). La nueva hipótesis es que la menor se retractó para eximir de responsabilidad a Fray Huacho (H5). Por ello, debe identificarse cuál de las tres hipótesis sostenibles (las dos iniciales o esta nueva) es la que “mejor explica el caso”. Para ello se analizará qué elementos confirman o rechazan las “hipótesis sostenibles”. Además, con un fin estrictamente metodológico, primero se analizará la nueva hipótesis y, después, se evaluará las otras dos hipótesis sostenibles.

Con relación a la nueva hipótesis explicativa (H5) advertimos que esta presenta múltiples puntos débiles, por lo siguiente:

- Avelína García (madre de la menor) indicó que su hija convivió con Fray Huacho. Sin embargo, la menor no vivía con su madre, pues estaba al cuidado de su hermana y su abuela. Ello resta veracidad al relato de Avelína García. Además, las dos tías de la menor no ratificaron sus presuntos relatos (los que sustentaron el informe social). La menor, en su primer relato, tampoco brindó detalles de su presunta convivencia con Fray Huacho o sobre los actos sexuales que mantuvo con este (su relato fue genérico). Asimismo, el nombre de Fray Huacho apareció solo después de que Iván Villaizan (conviviente de la menor) fue detenido y cuando la menor estaba embarazada. Ello nos permite concluir, luego de analizar la solidez o debilidad del inicial relato de la menor, que su incriminación inicial es débil (conclusión parcial 1).
- La menor reiteradamente indicó que no convivió ni mantuvo relaciones sexuales con Fray Huacho y que dijo eso para que liberen a Iván Villaizan, que era su conviviente y el padre del niño que esperaba, todo por recomendación del abogado Percy Coronado, quien le dijo que no pasaría nada si narraba dicha versión, y que con ello lograría que liberen a su pareja. Añadió que Iván Villaizan y toda la familia de esta persona la presionaron para que mienta, pues le decían que por su culpa se detuvo a su conviviente. También precisó que conoció de vista a Fray Huacho porque su abuela le vendía comida en una obra de construcción, pero ella nunca tuvo alguna

relación sentimental o amical con él, e incluso añadió que por su inmadurez trataba despectivamente a esta persona. Este relato es corroborado por la declaración de Ernestina Chumbes, abuela de la menor, quien indicó que su nieta vivía con ella y nunca convivió con Fray Huacho; además, precisó que el enamorado de la menor era Iván Villaizan y que el abogado Percy Coronado fue quien pidió a G. M. G. que mienta para que así suelten al padre de su nieto. Por su parte, Yesenia Miranda, hermana de la menor, indicó que vivía con la agraviada y que Fray Huacho jamás vivió con su hermana. También indicó que la menor dijo eso para que liberen a Iván Villaizan y que todo fue organizado por el abogado Percy Coronado. Además, precisó que cuando G. M. G. quiso decir la verdad, Iván Villaizan se la llevó a otro lugar y no supieron nada de ella. Esto nos permite concluir que existe coherencia interna y exhaustividad en el nuevo relato (retractación), el que también cuenta con capacidad corroborativa (conclusión parcial 2).

- La menor reiteradamente señaló que cuando quedó embarazada convivía con Iván Villaizan y la familia de esta persona. Esto hace que sea razonable pensar que en dichas condiciones haya mentido para que se libere a su pareja sentimental. Sería con tal finalidad que la menor habría sindicado falsamente a Fray Huacho, que era una persona que casualmente trabajó en el lugar. Además, es razonable pensar que cuando haya querido decir la verdad, la familia de su pareja la haya presionado o coaccionado para que no perjudique a su conviviente y real agresor, pues vivía con estas personas, y para todo ello contó con la asistencia del abogado Percy Coronado, que defendía a su pareja. Evidentemente dichas conjeturas se sostienen en máximas de la experiencia que podrían ser analizadas críticamente con mayor detalle, pero que para los fines de nuestro argumento basta con sostener que son plausibles. Todo lo hasta aquí mencionado hace razonable la tesis de que la menor brindó una inicial información falsa, que resulta proporcional con el fin buscado y la acción de denunciar falsamente (conclusión parcial 3).
- También podría ocurrir que la menor, al tener contacto con Fray Huacho o la familia de este, haya podido ser manipulada o influenciada para que cambie su versión sobre los hechos. Sin embargo, desde que se formuló la denuncia indicó que no tuvo ninguna comunicación, contacto o cercanía con Fray Huacho (la agraviada, su hermana y su abuela indican que esta

persona estuvo en el pueblo por un periodo corto, en que laboró en una obra de construcción, y después se fue, sin que sepan nada de él), lo que es corroborado por esta persona, quien negó haber vivido en el lugar donde ocurrieron los hechos por un tiempo mayor al que trabajó en el lugar. Distinto sería el caso si Fray Huacho fuese un familiar o allegado de la menor, pues allí podría haber presionado a esta o su familia, o haberles ofrecido alguna dádiva para que se retracten. Esto nos permite concluir que tampoco existieron probados contactos entre la menor y su presunto agresor, que permitan inferir que fue manipulada o influenciada para que cambie su versión sobre los hechos o se retracte (conclusión parcial 4).

- Un último elemento a considerar es que con la detención o liberación de Huacho no se afecta a la menor, pues ella no tenía ningún vínculo de dependencia económica, afectiva o familiar con esta persona, lo que si ocurría con Iván Villaizan, pues este era su conviviente y padre del hijo que la menor esperaba. Además, vivía con la familia de esta persona y señaló que el abogado Percy Coronado la conminó a sindicarse a Fray Huacho, porque no identificarían ni ubicarían a esta persona, en la medida que no era del lugar, nadie lo conocía y se encontraba lejos. De este modo, la intensidad de las consecuencias negativas generadas con la denuncia en el plano económico, afectivo o familiar de la menor son pequeñas o insignificantes (conclusión parcial 5).

De este modo, la nueva hipótesis (**H5**) puede ser descartada, en la medida que presenta numerosos puntos débiles. Asimismo, el análisis descrito también permite descartar **H1**<sup>13</sup> y corroborar **H3**, pues —de un lado— ninguno de los nuevos elementos actuados permite sostener que la menor convivió (**D4**) y fue agredida sexualmente por Fray Huacho (**D2**) y, de otro lado, estos nuevos elementos hacen posible concluir que la menor convivió (**D4**), fue agredida sexualmente por Iván Villaizan (**D2**) y quedó embarazada de este (**D1**), y que el abogado Percy Coronado y la familia de Iván Villaizan la conminaron para que mienta y así su pareja sentimental pueda quedar libre.

De modo que, nuevamente, después de haber pasado *de las pruebas a las hipótesis* y, luego, *de las hipótesis a las pruebas*, podemos concluir que la mejor explicación del caso o la hipótesis definitiva es **H3**.

---

13 Esta también fue la conclusión de la Corte Suprema de Justicia, que absolvió a Fray Huacho de la acusación formulada en su contra por la presunta comisión del delito de violación sexual de menor de edad.

Ahora bien, este análisis también podría graficarse de similar forma a la realizada en el primer caso. Sin embargo, consideramos más importante resaltar otro aspecto: el análisis abductivo del caso también nos permitió elaborar un diseño institucional que nos permite aproximarnos a lo que probablemente sucedió (o tal vez, incluso, a la verdad del caso), donde se consideran todas las conclusiones parciales a las que se arribó en el análisis de confirmación o rechazo que se hizo de **H5**. De este modo, para estimar como válida la retractación de una presunta víctima de un delito sexual —a futuro— pueden evaluarse los siguientes aspectos:

1. La solidez o debilidad de la declaración incriminatoria, a la luz de los elementos corroborativos actuados.
2. La coherencia interna y exhaustividad del nuevo relato y su capacidad corroborativa.
3. La razonabilidad de la justificación de haber brindado una versión falsa o errónea, verificando la proporcionalidad entre el fin buscado y la acción de denunciar falsamente.
4. Los probados contactos que haya tenido o podido tener el investigado con la víctima, que permitan inferir que esta haya sido manipulada o influenciada para cambiar su versión.
5. La intensidad de las consecuencias negativas generadas con la denuncia en el plano económico, afectivo y familiar de la presunta agraviada<sup>14</sup>.

El razonamiento sobre todos estos aspectos se materializa en un discurso justificativo, pero se va elaborando a través de un complejo proceso de descubrimiento, en el que la formulación de hipótesis y su proceso de corroboración se retroalimentan constantemente. Ello, nuevamente, muestra cómo no solo a nivel teórico-conceptual, sino también a nivel práctico, corresponde debilitar la distinción tajante entre CD y CJ a la que hemos hecho referencia a lo largo de este artículo.

Finalmente, es importante mantener cautela sobre las conclusiones que se pueden extraer de nuestro análisis. Esta cautela se fundamenta, principalmente, en dos motivos: en el hecho de que se trata de una reconstrucción racional que omite algunos aspectos procesales y en el hecho de que el análisis que hemos realizado se refiere a la estructura de la inferen-

---

<sup>14</sup> Estos elementos fueron recogidos por las Salas Penales Permanente y Transitorias de la Corte Suprema de Justicia de la República del Perú en el Acuerdo Plenario N.º 1-2011/CJ-116. En el ejemplo comentado se volvió a insistir en estos elementos y analizó el caso a la luz del mencionado Acuerdo Plenario.

cia abductiva del caso, más no a la satisfacción (o no satisfacción) del estándar de prueba relevante. Este último aspecto es fundamental para la determinación de la calidad de la explicación ofrecida por la que incluso podría ser la mejor hipótesis dentro de las disponibles.

## 7. Conclusiones

La principal conclusión a la que arribamos es que la dicotomía CJ/CD ha ido perdiendo fuerza progresivamente. Ello se ha producido tanto por avances en la discusión al seno de la filosofía de la ciencia, como por los avances en la teoría del derecho y de la argumentación jurídica. En este último ámbito se entiende, ahora, que en la justificación hay componentes propios del descubrimiento. Los ejemplos puestos sobre el proceso de interpretación de enunciados normativos, así como los referidos a la generación y justificación de hipótesis probatorias, lo confirman.

Asimismo, a grandes rasgos, el proceso de descubrimiento se puede desarrollar a través de una “lógica del descubrimiento”, expresión que puede hacer alusión tanto a una metodología para el análisis, como a un tipo de inferencia que corresponde especialmente al proceso de descubrimiento. En el segundo de estos sentidos, la inferencia buscada es de tipo abductivo. Precisamente el presente artículo avanzó una propuesta de caracterización del razonamiento abductivo que, además de salvar críticas formuladas contra otras versiones de dicho razonamiento, es especialmente fructífera para la comprensión del razonamiento probatorio.

Finalmente, los ejemplos propuestos, recientemente resueltos por la Corte Suprema de Justicia peruana, muestran cómo la abducción, primero, nos permitió dotar de racionalidad al contexto de descubrimiento y, segundo, cómo el análisis de casos en concreto, en el marco de un análisis abductivo, nos permitió elaborar un diseño institucional para acercarnos a la verdad. Además, nos ayudó a disipar y superar el divorcio que suele existir entre la teoría, los diseños intelectuales y la práctica jurisprudencial de los juzgados, de modo tal que fue posible que ambos ámbitos puedan colaborar de forma fructífera.

## Referencias

- Accatino, D. (2002). Notas sobre la aplicación de la distinción entre contextos de descubrimiento y de justificación al razonamiento judicial. *Revista de Derecho*, (13): 9-25.
- Aliseda, A. (2006). *Abductive Reasoning. Logical Investigations into Discovery and Explanation*. Springer.
- Anderson, B. (1996). *"Discovery" in Legal Decision-Making*. Springer.
- Arabatzis, T. (2006). On the inextricability of the context of discovery and the context of justification. En: J. Schikore, & F. Steinle (eds.), *Revisiting Discovery and Justification: Historical and philosophical perspectives* (pp. 215-230). Springer.
- Atienza, M. (2020). What is the Theory Legal Argumentation for? *International Journal for the Semiotics of Law*, (33): 147-153.
- Atienza, M. (2017). Algunas tesis sobre el razonamiento judicial. En Aguiló, J. & Grández Castro, P. (eds.), *Sobre el razonamiento judicial. Una discusión con Manuel Atienza* (pp. 11-42). Palestra.
- Atienza, M. (2013). *Curso de argumentación jurídica*. Trotta.
- Bonorino Ramírez, P. R. (2014). Ni deducción ni inducción: abducción. En: J. García Amado & P. Bonorino Ramírez (Coords). *Prueba y razonamiento probatorio en el Derecho. Debates sobre la abducción* (pp. 163-187). Comares.
- Canale, D. (2019). *En busca de lo implícito. Ensayos sobre razonamiento e interpretación en el Derecho*. Universidad Externado de Colombia.
- Douven, I. (2017a). Abduction. En: Zalta, E. (ed.), *The Stanford Encyclopedia of Philosophy*. Recuperado de: <https://plato.stanford.edu/archives/sum2017/entries/abduction/>
- Douven, I. (2017b). Peirce on Abduction (Supplement to Abduction). En: Zalta, E. (ed.), *The Stanford Encyclopedia of Philosophy*. Recuperado de: <https://plato.stanford.edu/entries/abduction/peirce.html>
- Fann, K. (1970). *Peirce's Theory of Abduction*. Martinus Nijhoff.
- Ferrer Beltrán, J. (2007). *La valoración racional de la prueba*. Marcial Pons.
- Frankfurt, H. (1958). Peirce's Notion of Abduction. *Journal of Philosophy*, (55): 593-596.
- Gascón Abellán, M. (2014). ¿Lógica del descubrimiento para la prueba? En: García Amado, J. y Bonorino, P. (eds.), *Prueba y razonamiento probatorio en Derecho. Debates sobre abducción* (pp. 149-162). Comares.

- Gascón Abellán, M., & García Figueroa, A. J. (2015). *La argumentación en el Derecho. Algunas cuestiones fundamentales* (2da. ed.). Palestra.
- Guastini, R. (2018). *Interpretar y argumentar* (2da. ed.). Centro de Estudios Políticos y Constitucionales.
- Hoyningen-Huene, P. (2006). Context of Discovery versus Context of Justification and Thomas Kuhn. En: *Revisiting Discovery and Justification: Historical and Philosophical perspectives on the context distinction* (pp. 119-131). Springer.
- Kaufmann, A. (1999). *Filosofía del Derecho*. Universidad Externado de Colombia.
- Kraus, M. (2003). Charles S. Peirce's Theory of Abduction and the Aristotelian Enthymeme from Signs. En: Van Eemeren, F.; Blair, J. A.; Willard, C.; Snoeck, F. (eds.), *Anyone who has a view. Theoretical contributions to the study of argumentation* (pp. 237-254). Springer.
- Lifante, I. (2019). En defensa de una concepción constructivista de la interpretación jurídica. *Revus. Journal for Constitutional Theory and Philosophy of Law*, 39, pp. 1-21. DOI: 10.4000/revus.5423
- MacCormick, N. (2003[1978]). *Legal Reasoning and Legal Theory*. Oxford Clarendon Press.
- Niiniluoto, I. (2018). *Truth-Seeking by Abduction*. Springer.
- Popper, K. (2017). *La lógica de la investigación científica* (2da. ed.). Tecnos.
- Pino, G. (2017). La teoría del razonamiento judicial de Manuel Atienza. Notas al margen. En Aguiló, J. & Grández Castro, P. (eds.), *Sobre el razonamiento judicial. Una discusión con Manuel Atienza* (pp. 305-327). Palestra.
- Reichenbach, H. (1938). *Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge*. The University of Chicago Press.
- Schickore, J. (2018). Scientific Discovery. En: Zalta, E. (ed.), *The Stanford Encyclopedia of Philosophy*. Recuperado de: <https://plato.stanford.edu/archives/sum2018/entries/scientific-discovery/>
- Schickore, J. & Steinle, F. (2006). Introduction: revisiting the context distinction. En: *Revisiting Discovery and Justification: Historical and Philosophical perspectives on the context distinction* (pp. vii-xix). Springer.
- Sotomayor, J. E. (2021a). Apuntes históricos, conceptuales y jurisprudenciales sobre el deber de motivación de las resoluciones judiciales. En: Alvites, E. (coord.), *La Constitución frente a la sociedad contemporánea. Treinta años de la Maestría en Derecho Constitucional de la Pontificia Universidad Católica del Perú* (pp. 23-52). PUCP.

- Sotomayor, J. E. (2021b). *Argumentación jurídica: una introducción*. Zela.
- Taruffo, M. (2006). *La motivación de la sentencia civil*. Tribunal Electoral del Poder Judicial de la Federación.
- Tuzet, G. (2021). *Filosofía de la prueba*. Marcial Pons.
- Tuzet, G. (2014). Usos jurídicos de la abducción. En: García Amado, J. y Bonorino, P. (eds.), *Prueba y razonamiento probatorio en Derecho. Debates sobre abducción* (pp. 121-147). Comares.
- Walton, D. (2005). *Abductive Reasoning*. The University of Alabama Press.



# HOW UNIVERSAL GENERALIZATION WORKS ACCORDING TO NATURAL REASON

\*

KYLE S. HODGE

Department of Philosophy, University of South Florida.  
kshodge@usf.edu

## Abstract

*Universal Generalization, if it is not the most poorly understood inference rule in natural deduction, then it is the least well explained or justified. The inference rule is, prima facie, quite ambitious: on the basis of a fact established of one thing, I may infer that the fact holds of every thing in the class to which the one belongs—a class which may contain indefinitely many things. How can such an inference be made with any confidence as to its validity or ability to preserve truth from premise to conclusion? My goal in this paper is to explain how Universal Generalization works in a way that makes sense of its ability to preserve truth. In doing so, I shall review common accounts of Universal Generalization and explain why they are inadequate or are explanatorily unsatisfying. Happily, my account makes no ontological or epistemological presumptions and therefore should be compatible with whichever ontological or epistemological schemes the reader prefers.*

**Keywords:** *universal generalization, reasoning, informal reasoning, informal logic, pedagogy*

## 1. Introductory Remarks\*

Universal Generalization is a natural, deductive rule of inference in virtue of which a universal proposition may be validly inferred from a singular proposition which involves a generalized or arbitrary particular. Almost everything turns on what it means for the particular at issue to be “generalized” or “arbitrary.” This is because, *prima facie*, it is a non sequitur to say that because something holds for one case in some class, it must thereby hold for all cases in that class. Common sense seems to dictate that a universal claim may only be validly inferred on the basis of either the conjunction of the proofs conducted on every particular of a class one by one, or else that it follows trivially from some definitions or stipulations which are themselves expressed as universal propositions. And yet, Universal Generalization appears to be a valid inference from a particular to a universal which does neither. How can this be?

The simple fact is that this inference rule is almost universally poorly understood. Even worse, it is almost always poorly and inadequately explained. Although philosophers working today can successfully employ Universal Generalization, their understanding of it seems to remain sufficiently inchoate as to make communication to others ineffectual. A successful and pedagogically useful account of Universal Generalization will be one in which the connection between a singular proposition involving a generalized particular and the corresponding universal proposition is rendered clear and explicit in a finite number of evident parts or steps. Indeed, every canonical rule of natural reasoning ought to be understood in this manner. My intention is to provide an informal remedy to this issue and to make Universal Generalization better and more widely understood by providing a novel and adequate account and justification of it. Moreover, I shall do so in a way which fills the explanatory gaps in extant accounts, and also in a way which does not have epistemological or ontological costs.

## 2. Extant Accounts

It is unfortunate that it is difficult to find a compelling answer to a question that one ought to expect from undergraduate logic students:<sup>1</sup> how exactly does Universal Generalization work? The extant accounts of Universal Generalization which aim to render the rule intuitive to students of logic are all correct, but they do not do the explanatory work that they should do and which they purport to do. Each account mentioned here is made complete by the solution I shall offer in section 3.

---

\* I want to acknowledge my good friend Jack Short for the stimulating conversation we had which lead to the crucial insights in section 3.3.

1 Cf. Kit Fine’s recollection of the genesis of his book-long effort at explaining Universal Generalization in terms of real arbitrary objects [1985b, p. vii].

## 2.1—“It Could Have Been Any $x$ ” and the Appeal to Arbitrariness

Perhaps the most common way of explaining why Universal Generalization is valid is because it is evident that the inference from “any  $x$ ” to “every  $x$ ” is valid (since they are logically equivalent). If *any* particular  $x$  in class  $S$  has property  $F$ , then *every* particular in  $S$  has  $F$ . For example, if *any* marble you pick from the jar will be red, then *every* marble in the jar is red. It is easy to see why we may infer that every particular in  $S$  has  $F$  when any (i.e. each) particular in  $S$  has  $F$ . Thus, one reasons as follows:

1. If it could have been any particular  $x$  in  $S$  which was proven to have  $F$ , then every particular in  $S$  must have  $F$ . (by logical equivalence of “any” and “every”)
2. But it could have been any particular  $x$  in  $S$  which was proven to have  $F$ . (intuitive)
3. Therefore every particular in  $S$  must have  $F$ .

This explanation is lacking for want of a way to clearly demonstrate (2). Suppose we try and convince our students that an instance of (2) is true—that any particular triangle may be proven to have internal angles equal to two right angles. We proceed to work the proof on several examples which were seemingly chosen at random. We might even tell the more incredulous of the students: “draw as many triangles as you like, no matter how bizarre they are, and I shall conduct the proof on all of them.”<sup>2</sup> While this dialectical strategy indicates that the unique properties of particular triangles are not important to the proof, this suffices only to communicate the inchoate idea underlying Universal Generalization to another person. This is especially so since this sort of seemingly random selection does not necessarily establish the universality of anything, and may even mislead us to prematurely generalize our results.<sup>3</sup> Now, the students will likely surmise that they will not be able to find a counterexample to the conclusion that every triangle has internal angles equal to two right angles—but how you *deduced* that conclusion from a few particular cases remains quite inchoate at best. By failing to explain why the inference from one or many particulars to *any* or *every* particular is valid, this approach does little to clarify the procedure of the inference. Thus, an adequate understanding of Universal Generalization cannot be gained by the appeal to “it could have been any  $x$ .”

---

2 Cf. E. M. Barth (1974, pp. 52–57); you are to imagine yourself simultaneously facing “every possible opponent” armed with a potential counterexample to the proposition as you are conducting the proof (p. 56).

3 For some examples, see: Baez (2018) and Guy (1988).

A closely related attempt at explanation is the appeal to arbitrariness (also “arbitrary selection” or “selection of an arbitrary particular,” etc.). This approach is common in logic textbooks when it comes time to explain Universal Generalization independently of formal rules.<sup>4</sup> There are usually two ways in which a particular is said to be arbitrary:

1. it was selected at random or with no care as to which particular of the class it is
2. its properties are general and indefinite (whether in fact or in representation)

Each way has its own problems. The issue with (1) is that it is just the warmed over version of the “could have been any” or “could have picked any” approach discussed above. The issue with (2) is that, despite the epistemological and ontological questions it can raise,<sup>5</sup> it explains Universal Generalization no better than (1) does. Suppose that I have just now concluded Euclid’s proof at *Elements* Book 1, Proposition 32 on an “arbitrary particular triangle” (however you like to conceive of it). Insofar as its properties are general and indefinite, the arbitrary particular triangle either has or represents all and only those properties that every triangle has. As such, the appeal to arbitrariness is simply an appeal to the properties common to all triangles. Yet, trivially, all non-arbitrary particular triangles also have these properties. So the appeal to the arbitrariness of an arbitrary particular triangle is superfluous; as far as the validity of the proof is concerned, whether the particular is “arbitrary” does not matter. Thus, the appeal to arbitrariness does not explain, or even help to explain, the validity of Universal Generalization.

## 2.2. Repeatability

Carlo Cellucci argues for an account of Universal Generalization as being explicable in terms of the *repeatability of a proof sequence*.<sup>6</sup> A repeatable proof sequence is a proof sequence that produces the same conclusion no matter which particular of the class the proof sequence is applied to. On Cellucci’s view, proof sequences are repeatable in virtue of their schematicity: “they are argument-schemata that, given any object in the domain, will yield a proof which is specific to that object.”<sup>7</sup> As such, Cellucci concludes that “in the universal generalization problem, what is primary is the repeatability of the proof rather than the generalizability of the result. The latter is simply a corollary of repeatability.”<sup>8</sup>

---

<sup>4</sup> For example, Theodore Sider in *Logic for Philosophy* (2010, pp. 99, 102) and Irving Copi in *Symbolic Logic* (1979, p. 72); cf. also the remarks in Cellucci (2009, pp. 4, 8–11).

<sup>5</sup> I am here putting aside the myriad issues with this approach which are sufficient to hinder an easy explanation. See Cellucci (2009, p. 9) for an example.

<sup>6</sup> *Ibid.*, pp. 13–15.

<sup>7</sup> *Ibid.*, p. 11.

<sup>8</sup> *Ibid.*, p. 13.

By inferring universality from repeatability, rather than from generalizability, Cellucci simply relocates the site at which the explanation of Universal Generalization must occur. Originally, the question was how we establish that the conclusion of a given particular proof is generalizable; now, the question is how we establish that a given proof sequence is repeatable with respect to a given class of particular objects. The answer to the latter question is forthcoming: universal propositions form the schematic character of a proof (i.e. they *are* the argument schemata), and they are derived either from other universal propositions or from definitions. The consequence of this, however, is that there simply is no Universal Generalization. Cellucci promises an explanation of that inference rule, but instead he eliminates it. This is because particulars, which were originally the site of generalization, play no role in the deduction of universal propositions.

As such, we are still left with no explanation of Universal Generalization.

### 3. How Universal Generalization Works

The procedure of universalizing a particular via generalization occurs in three stages:

1. Identification of a particular.
2. Generalization of the particular.
3. Universalization on the basis of the generalized particular.

Let us consider each in turn.

#### 3.1. Identifying a Particular

The identification of a particular is the simplest step and usually consists in merely asserting such things as “let  $n$  be an integer” or “let  $ABC$  be a triangle,” or by indicating the same by writing the letter ‘ $n$ ’ or by drawing a triangle whose vertices are labeled ‘A,’ ‘B,’ and ‘C’ respectively. One may also ostend a particular or give a sufficient description of one.

There may be other ways of identifying particulars, e.g. direct experience, modal quantification, etc., but this seems to me to be the most common sort of case; I leave it to the reader to consider others.

#### 3.2. Generalizing a Particular

In a particular demonstration that some particular  $c$  in some definite class<sup>9</sup>  $S$  has some property  $F$ ,  $c$  is a generalized particular on the condition that *every property by virtue of which the particulars in  $S$  could possibly differ is omitted from the premises of the demonstration*.<sup>10</sup>

---

<sup>9</sup> A definite class is a class which is defined by a finite set of necessary and sufficient conditions.

<sup>10</sup> I was influenced on this matter by George Berkeley’s account of Universal Generalization in his *A Treatise Concerning*

The particular is said to be *generalized* inasmuch as every property of  $c$  involved in the demonstration is common to all particulars in  $S$ .

I note that the conclusion of this demonstration on the particular—the singular proposition that  $c$  in  $S$  has  $F$ —is itself the premise from which the universal proposition will be validly inferred on the basis of the fact that  $c$  is a generalized particular.

### 3.3. Universalizing a Generalized Particular: Validly Inferring a Universal Proposition

Once we have demonstrated of  $c$  that it has  $F$  in the manner described above,  $c$  is such that it cannot differ from any other particular in  $S$  with respect to the properties involved in the demonstration because all differentiating properties were omitted. Thus, it follows that it cannot differ from any other particular in  $S$  with respect to  $F$ , for  $F$  was itself proven of  $c$  on the basis of properties with respect to which the particulars in  $S$  cannot differ. So, since  $c$  cannot differ with respect to  $F$  from any other particular in  $S$ , it follows that *every* particular in  $S$  must have  $F$ , for if one thing cannot differ from anything else in the same class with respect to a certain property, it follows that they must all have that property. So we have validly inferred a universal proposition from a singular proposition involving a particular.<sup>11</sup>

This is the rule of inference we call Universal Generalization.

## 4. Euclid's Demonstration as Exemplary of Universal Generalization

It will be instructive to see the procedure that I have laid out above exhibited in an example. For this purpose, we look quite naturally to Euclid. When Euclid constructs his demonstration of Proposition 32 in Book 1 of the *Elements*, he follows the procedure I have set out.

First, he identifies a particular triangle,  $ABC$ , by stipulation.

Second, he generalizes  $ABC$  by omitting from the premises of the particular demonstration anything concerning the lengths or proportions of the sides or the measures or proportions of the angles. Since these are the only possible ways in which triangles may differ from one another,  $ABC$ , as it appears in the premises and conclusion of the particular demonstration, is fully generalized.

---

the *Principles of Human Knowledge*, Intr. section 16. I extend this Berkeleian insight by identifying what must be omitted from the premises of a particular demonstration in order to generalize a particular. This addition is vital to clearly articulating, and so explaining, how Universal Generalization works.

11 This may be considered from another point of view. For any universal proposition  $\phi$  which states that all particulars in some class  $S$  have some property  $F$ , a counterexample consists in showing that there is a *reason* in virtue of which some particular  $c$  in  $S$  does not have  $F$ . This reason must be that there is some important *difference* between  $c$  and other particulars in  $S$ . Since the valid use of Universal Generalization cannot be subject to a counterexample, there must be something in the inference procedure which removes or precludes the sorts of differences that would figure in true counterexamples. So the relevant notion is that  $F$  is universal relative to  $S$  on the condition that every particular in  $S$  cannot differ with respect to  $F$ . From which comes the procedural method of *omitting* from the premises of the particular demonstration all properties by virtue of which the particulars in  $S$  could differ from one another. Thus, the impossibility of a counterexample coincides with the procedure of generalizing a particular. Hence, the omission procedure grounds and explains the validity of the inference.

Since Euclid proves of the generalized particular triangle *ABC* that the sum of its internal angles are equal to the sum of two right angles, he may validly conclude the universal proposition that this fact holds of all triangles. This conclusion is inferred validly because, since all differentiating properties were omitted, *ABC* must have its properties in common with every triangle. In other words, *ABC* cannot differ from any other triangle with respect to any property involved in the demonstration. So *ABC* cannot differ from any other triangle with respect to the sum of its internal angles, for that was the conclusion of the particular demonstration. Thus, because *ABC* cannot differ from any other triangle with respect to the sum of its internal angles, the sum of the internal angles of *every* triangle must be equal to two right angles.

## 5. Comments on the Foregoing Account

### 5.1. *Filling the Explanatory Gaps in Extant Accounts*

My account is able to explain why “it could have been any *x*,” as well as why such proof sequences are repeatable.

In a particular demonstration that *c* in *S* has *F*, *c* may be replaced with *d* (or any other particular in *S*) because the omission procedure is such that any differences between *c* and *d* cannot figure in the particular demonstration. So, with respect to the particular demonstration, *c* and *d* are the same. Thus, we may rightly say that it could have been *d*, or *e*, or any other particular in *S*.

Furthermore, my account shows why such demonstrations are repeatable. With respect to the particular demonstration, *c* and any other particular in *S* are the same. Since each particular in *S* is the same, the proof conducted on each particular may also be the same. Hence, the same proof sequence may be repeated across numerically distinct particulars because they are the same with respect to the proof.

### 5.2. *Metaphysical and Epistemological Consequences*

The account I have presented is free of the ontological and epistemological consequences that are typically associated with inferences by Universal Generalization. By means of the omission procedure, one avoids the question of whether there are, or whether we are able to cognize, abstract general ideas (à la Locke) on which we reason to deduce universal conclusions. We likewise avoid the question of whether there “really” exist, in addition to ordinary particulars, such things as arbitrary or indefinite particulars which scholars such as Kit Fine defend.<sup>12</sup>

---

12 Fine says that “in addition to individual objects, there are arbitrary objects: in addition to individual numbers, arbitrary numbers; in addition to individual men, arbitrary men” (1983, p. 55) and that “our talk of arbitrary triangles or indefinite triangles is to be taken at its face value as evincing reference to arbitrary objects” (1985a, p. 57); cf. the exposition in Macnamara (1988, p. 305) and Santambrogio (1988, pp. 630-631).

This account requires only that we deal with ordinary particulars while omitting from our premises the ways in which they can differ from other particulars in the same class; that is all that makes a particular “generalized.” It does not require, nor does it preclude, according the particulars any special epistemological or ontological status and it does not require or preclude any extraordinary cognitive powers on our part. Hence, it need not have any epistemological or ontological consequences.

### 5.3. *Some Important Pedagogical Implications of the Account*

One of the consequences that I am most excited about is the way the account can make this notorious rule of inference clear and explicit to students of logic, mathematics, and critical reasoning. For example, a simple and sound explanation might be:

Let’s leave out of the proof every way  $ABC$  can differ from other triangles. Then, we can agree that whatever is shown to be true of  $ABC$  in the proof must be true of all triangles because things that don’t differ must be the same. So, when we do this and show that  $ABC$  has internal angles equal to two right angles, we get to infer that every triangle has internal angles equal to two right angles because it is the same as all the other triangles in this respect.

In addition to the inadequate explanations of Universal Generalization discussed in section 2, many students of logic rely, by rote, on the restrictions imposed on the use of the rule in the predicate calculus. My account helps us to see why those restrictions are as they are. The restriction that  $Fa$  must be deduced from premises outside of the particular sub-proof in which there are no occurrences of  $a$  in order for  $\forall xFx$  to be validly deduced from  $Fa$  is simply the formal correlate of the omission procedure. As such, an understanding of the inference rule and the formalism which models it can be made to coincide.

## 6. Concluding Remarks

I understand the significance and novelty of this account to consist in the clarity with which it communicates the procedure and validity of Universal Generalization. Extant explanations fail at two specific places. The first is explaining the generalization of a particular or what makes for a generalized particular. The second is explaining the inferential connection between a singular proposition involving a generalized particular and the universal proposition which is inferred on its basis.

Although my account need not have any implications for formal languages employing Universal Generalization, it does have implications for logic or reasoning in the broader sense of thinking or reasoning well. This is because it helps make clear to natural reason

the manner in which this inference procedure works. When the conditions for successfully applying the inference procedure are clear and evident, it is easier to correctly employ the procedure, detect errors in the attempted use of the procedure by others, and to imaginatively extend it into domains in which others have not recognized its applicability. That is what we should wish for concerning every sort of inference for the sake of extending our stock of knowledge.

## Bibliography

- Baez, J. (2018). Patterns that Eventually Fail. *Azimuth*. <https://johncarlosbaez.wordpress.com/2018/09/20/patterns-that-eventually-fail/>
- Barth, E. M. (1974). *The Logic of the Articles in Traditional Philosophy: A Contribution to the Study of Conceptual Structures*. D. Reidel Publishing Company.
- Cellucci, C. (2009). The Universal Generalization Problem. *Logique et Analyse*, 52(205), 3–20. <https://www.jstor.org/stable/44084910>
- Copi, I. M. (1979). *Symbolic Logic* (5th ed.). Macmillan Publishing Co., Inc.
- Fine, K. (1985a). Natural Deduction and Arbitrary Objects. *Journal of Philosophical Logic*, 14(1), 57–107. <https://doi.org/10.1007/BF00542649>
- Fine, K. (1985b). *Reasoning With Arbitrary Objects*. Basil Blackwell.
- Fine, K., & Tennant, N. (1983). A Defense of Arbitrary Objects. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 57(1), 55–90. <https://doi.org/10.1093/aristoteliansupp/57.1.55>
- Guy, R. K. (1988). The Strong Law of Small Numbers. *The American Mathematical Monthly*, 95(8), 697–712. <https://doi.org/10.2307/2322249>
- Macnamara, J. (1988). [Review of the book *Reasoning With Arbitrary Objects*, by K. Fine]. *The Journal of Symbolic Logic*, 53(1), 305–306. <https://doi.org/10.2307/2274449>
- Santambrogio, M. (1988). [Review of the book *Reasoning With Arbitrary Objects*, by K. Fine]. *Noûs*, 22(4), 630–635. <https://doi.org/10.2307/2215463>
- Sider, T. (2010). *Logic for Philosophy*. Oxford University Press.

\* \* \*