This dissertation offers an integrated, systematic account of the qualitative, symmetrical patterns that are shared by all thought experimental activities, which include the wide range of hypothetical questions employing subjunctive reasoning that are a routine part of science, as well as the wide range of appeals to imaginary experience made in the hypothetical case studies routinely used in philosophy (especially ethics), the social sciences (economics, sociology, history, politics, etc.), and legal studies — all of which are, obviously, domains in which “true experiments” cannot be implemented.

The development, construction, and the internal mechanism of the thought experiment process (the major constituent of which is the analogy) is examined, and a practical, useful and all-encompassing definition of thought experiments is presented, along with a unique, seven-part typology of thought experiments, based on the qualitative temporal differences between each class of thought experiment.

Submitted in Partial Fulfillment of the Requirements for the

*Degree of Graduate Diploma in Arts by Research*

in the School of History and Philosophy of Science,

*University of New South Wales.*

2004
By the end of 2001, I had completed the last requirements for my Masters degree in Cognitive Science, and I was preparing to undertake a research Ph.D. associated with my long-term interest in the nature, form, content, and application of efficacious hypnotic suggestion (and, to a lesser extent, ‘mental imagery’).

Yet, whilst I was, ipso facto, ‘academically qualified’ to be admitted to a Ph.D., I had no ‘research track record’; out of the fact that the Masters degree that I had just completed had been a course-work degree.

In order to address this issue, I undertook additional, preparatory studies: in the form of a Graduate Diploma in Arts by Research within the School of the History and Philosophy of Science.

Quite deliberately, and in order to prepare myself for (i.e., rather than ‘anticipate’) my later Ph.D. studies — and because they seemed, somehow, analogous sorts of activities — I decided to investigate the structured, imaginary activities that philosophers and scientists routinely conduct, called “thought experiments”.

The entire first year of this research and these studies were eventually abandoned due to devastating circumstances that were not of my making.

I re-commenced my research endeavours at the beginning of 2003 with a new, motivated, and enthusiastic supervisor.

From my own deep, long-term interest in both ‘hypnotic suggestion’ and ‘mental imagery’, my original intention had been to concentrate on the interesting work that was then emerging, in the field of cognitive psychology, on the influence of the prior imagining of specific ‘scenarios’ upon the decisions that individual subjects subsequently made (or, in the case of beliefs, etc., upon subsequent attitude changes).

It seemed to be a very fertile field.

Yet, and in particular, most of the work on the influence of imagining on attitudinal change had been conducted on ‘normal people’.

Yet, it was plain that the majority of those who routinely used thought experiments were those Cacioppo and Petty had identified as having a “high need for cognition”: 
viz., an intrinsic propensity to engage in and enjoy “effortful cognitive endeavors”, 
“analogous to individual differences in people's motivation to engage in effortful 
physical endeavors, which is related to but not the same thing as physical ability”.

They also had a strong preference for “central route processing” of information 
(i.e., thoughtful consideration of content, involving considerable cognitive effort), 
rather than “peripheral route processing” (i.e., quick, cursory judgment based on 
variables peripheral to content, and involving minimal cognitive effort).

I was initially drawn to investigate whether scientists and philosophers, having, 
on overall, a much “higher need for cognition”, and a greater propensity for “central route 
processing”, were more likely (without any conscious awareness of the fact) to be 
more deeply immersed — and, thus, more easily persuaded — by their own 
performance of a particular thought experiment (that had, of course, been constructed 
to deliver a particular outcome); all of which, in a sense, seems to be rather like 
opening the gates of the city, and trusting the Trojan Horse.

Despite the immense attractiveness of such matters, it was soon evident that my 
own need to produce strong evidence of an ability to conduct research would be far 
better served by investigating the mechanism of the thought experiments themselves 
(an investigation that was sufficiently ‘narrow’ to allow me to demonstrate a capacity 
to examine a question at an appropriate ‘depth’).

In the process, I eventually produced a practically useful definition of thought 
experiments, as well as a unique, seven-part typology of thought experiments (in 
2004), based on the qualitative temporal differences between each type of thought 
experiment, that seems to have become widely adopted since late 2006.

My research and the dissertation I produced was of such a standard that I was 
awarded a Postgraduate Award Scholarship by the University of New South Wales, 
and was admitted as a Ph.D. candidate, in the School of History and Philosophy of 
Science (I graduated Ph.D. in 2013).

Apart from some minor typographical corrections, the following is the complete 
text of the dissertation I presented for the Graduate Diploma in Arts by Research.

Lindsay B. Yeates, Kensington, NSW, July 2013
To the memory of my parents,

Leonard Alfred Yeates (1912-1991)

Norma Gwendoline Yeates, née Verran (1914-2002).

With love, respect, and gratitude.
Acknowledgements

First, I wish to thank Dr John Schuster, Head of the School of History and Philosophy of Science, and Dr. Helen Milfull, Manager, Faculty of Arts and Social Sciences, for their active, sympathetic, and practical support.

This project could not have been attempted, let alone completed, without the professional support, intelligent guidance, enthusiastic encouragement, good humour, and patient kindness of Dr. Anthony Corones, who so graciously agreed to take over the supervision of a project that would otherwise have been abandoned altogether.

It is substantially due to Tony’s efforts and influence that a comprehensively lost confidence has been restored; and I will always be grateful to him for teaching me so much by his personal example.

I am grateful to the staff of the Social Sciences and Humanities Library; especially to Kaz Kazim, who patiently taught me how to use the internet for research, and how to get the best from databases, and to Pamela O’Brien, who arranged for my borrowing privileges to be extended.

The expertise, diligence, and prompt efficiency of all of the staff of the Inter-Library Loans section, over an extended time, has made a very significant contribution to whatever success this project may have, and I am very grateful to them all.

I must also record my gratitude to Bill Sayers and the late Harold Pennefather, my English teachers in Year Eight and Year Eleven respectively, who, so enthusiastically, drew my attention to the powerful way that figures of speech communicate.

Finally, I want to thank my friends Bob Hall and Karen Kato-Hall for their friendship, social and emotional support and, also, Bob’s willingness to explain the intricacies of US English. I thank my colleague Margaret Tomko for her extended support, and for her time so generously given. I am grateful to Andrew Kapos for insightful comments he made about earlier drafts of this dissertation.

Lindsay Bertram Yeates

Kensington, NSW, July 2004
# Table of Contents

Acknowledgements ii  
Table of Contents iii  
List of Tables iv  
Table of Figures v  

*The Blind Men and the Elephant. A Hindoo Fable*, by John Godfrey Saxe (1816-1887) vi

Chapter One: Introduction 1  
Chapter Two: Approaches to Thought Experimentation 5  
Chapter Three: Comparative Cognitive Processes 40  
Chapter Four: Understanding: A Context for Thought Experimentation 77  
Chapter Five: Performance Aspects of Thought Experimentation 110  
Chapter Six: Constructing, Designing, and Polishing Thought Experiments 126  
Chapter Seven: Thought Experiments as Belief Attenuating Engines 138  
Chapter Eight: Conclusion 160

Appendix One: The Lady, or The Tiger 165  
Appendix Two: Emile Coué’s Thought Experiment 171  
Appendix Three: von Platen and the Demiurge 173  
Appendix Four: From an Episode of *Frasier* 178  
Appendix Five: A Country and Western Song 180  
Appendix Six: A Counterfactual Headline 182  
Appendix Seven: “Exchange of Bodies” Thought Experiments 183  
Appendix Eight: Foot’s 1967 “Tramcar Problem” and Thomson’s 1971 “Famous Violinist Problem” 185  
Appendix Nine: Consequences of the Free-Trade Agreement 189  
Appendix Ten: “The Monk Problem” 191

Bibliography 197
# List of Tables

| Table One: Four Basic Types of Scenario | 36 |
| Table Two: Science Fiction Pieces vs. Scenario Techniques | 38 |
| Table Three: Analogy in Discovery, Development and Evaluation of Scientific Theories | 51 |
| Table Four: A Taxonomy of the Four Different Types of Counter-Thought-Experiment which Reveal Deficiencies in the Original Thought Experiment. | 58 |
| Table Five: The Consequences of Attribution of Responsibility in Four Models of Helping and Coping | 84 |
| Table Six: The Typical Questions Examined by Each Type of Thought Experiment | 87 |
| Table Seven: Socratic & Scientific Five-Step Method for Analysing Concepts & Events | 88 |
| Table Eight: Five Distinct Types of Counterfactual Thought Experiment | 99 |
| Table Nine: Six Types of Narrator | 112 |
| Table Ten: The Distribution of Narratorial Functions in Actual Agents | 114 |
| Table Eleven: Temporal Symmetry Among Thought Experiments | 142 |
| Table Twelve: Enantiomorphic Symmetry Among Thought Experiments | 142 |
| Table Thirteen: Standard Format for Representing the Thought-Experimental Pathway | 159 |
List of Figures

Fig.1: The Internal and External Processes of the Rebuttal Analogy 56
Fig.2: The Analysis of a Rebuttal Analogy 57
Fig.3: The Metaphorical Process: Interaction between Principal and Subsidiary Subject 61
Fig.4: The Mechanism of the Dynamic Metaphor 64
Fig.5: The Catachrestic Metaphorical Process: Interaction between Un-Named Principal Subject and Named Subsidiary Subject 64
Fig.6: The Mechanism of the Catachrestic Substitution Metaphor 66
Fig.7: Models: Interaction between Primary and Secondary Domain 69
Fig.8: The Mechanism of the Scale Model 72
Fig.9: The Mechanism of the Analogue Model 73
Fig.10: The Mechanism of the Intellectual Model 75
Fig.11: The Thought-Experimental Narrative's Pathway: in the Context of Other Types of Narrative 117
Fig.12: The Iterative Nature of the Scientific Method 129
Fig.13: Thought Experiments; A Concept Map 140
Fig.14: The Prefactual Thought Experiment 143
Fig.15: The Counterfactual Thought Experiment 144
Fig.16: The Semifactual Thought Experiment 144
Fig.17: Prediction, Forecasting and Nowcasting 145
Fig.18: Hindcasting 145
Fig.19: Retrodiction or Postdiction 146
Fig.20: Backcasting 147
Fig.21: The Belief Attenuating Engine: Input and Output 152
Fig.22: The Belief Attenuating Engine: Anomaly Detection and Belief Attenuation 153
Fig.23: The Belief Attenuating Engine: A Challenge Identified 154
Fig.24: The Belief Attenuating Engine: Input and Output 154
Fig.25: The Belief Attenuating Engine: Challenge and Anomaly 155
Fig.26: The Belief Attenuating Engine: Inner Structure and Working Mechanism 156
The Blind Men and the Elephant.
A Hindoo Fable

It was six men of Indostan
To learning much inclined,
Who went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind

The First approached the Elephant,
And happening to fall
Against his broad and sturdy side,
At once began to bawl:
"God bless me! but the Elephant
Is very like a wall!"

The Second, feeling of the tusk,
Cried, "Ho! what have we here
So very round and smooth and sharp?
To me 'tis mighty clear
This wonder of an Elephant
Is very like a spear!"

The Third approached the animal,
And happening to take
The squirming trunk within his hands,
Thus boldly up and spake:
"I see," quoth he, "the Elephant
Is very like a snake!"

The Fourth reached out an eager hand,
And felt about the knee.
"What most this wondrous beast is like
Is mighty plain," quoth he;
"'Tis clear enough the Elephant
Is very like a tree!"

The Fifth, who chanced to touch the ear,
Said: "E'en the blindest man
Can tell what this resembles most;
Deny the fact who can
This marvel of an Elephant
Is very like a fan!"

The Sixth no sooner had begun
About the beast to grope,
Than, seizing on the swinging tail
That fell within his scope,
"I see," quoth he, "the Elephant
Is very like a rope!"

And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!

Moral:
So oft in theologic wars,
The disputants, I ween,
Rail on in utter ignorance
Of what each other mean,
And prate about an Elephant
Not one of them has seen!

John Godfrey Saxe (1816-1887)
Chapter One: Introduction

I keep six honest serving men
(They taught me all I knew);
Their names are What and Why and When
And How and Where and Who.

Rudyard Kipling (1865-1936)

This dissertation is an attempt to answer the question, “What is a thought experiment?”

From scholars such as Roy Sorensen we gain the generally accepted understanding that a thought experiment is “an experiment that purports to achieve its aim without the benefit of execution” (1992a, p.205); and we soon discover that thought experiments are quite different from the sorts of experiments with thoughts that cognitive scientists conduct to determine whether a subject can manipulate the mental image of a “b” into a “d”, then into a “p”, then into a “q” and, then, back into a “b” again.

A major complication for any coherent study of thought experimental activities is that the term “thought experiment” denotes a very wide range of different activities, of varying degrees of “thickness”,¹ that are conducted for many different reasons, by significantly different groups of people — groups with distinctly different disciplinary loyalties, driven by entirely different theoretical positions, pursuing many different goals, engaged in totally different enterprises, and seeking many different outcomes from the performance of the thought experiment. Finally, these groups have, as their target audience, a constellation of subjects who possess an extremely wide range of relevant theoretical knowledge.²

¹ See Folger and Turillo (1999) for an extensive discussion of “thickness”. Hubert (1999, p.374) describes the “thinness/thickness” feature:

Thought experiments are examples of what are often called “thin” cases. This is to say that they are intentionally schematic and devoid of detail. This thinness allows students an opportunity to test their judgments and clarify concepts without having to sort through various practical contingencies. It is important to keep in mind that there is no substitute for exposing students to real-life cases, which are by their nature “thick” with detail and moral complexity. But at this stage we are concerned only with helping students grasp certain central concepts and testing the coherence of alternative judgments.

² Skinner’s (1996) survey of the notion of “control” in the psychological literature made a similar discovery. She lists one hundred and eleven different “constructs of control and their definitions” (pp.566-570), finding that (p.549-550) the literature contained a “large number of terms, which, although different, nevertheless seem to be interrelated and partially overlapping”; and, confusingly, that whilst “some [of these] appear to be different labels for the same construct”, others are clearly “cases in which the same term is used to refer to very different constructs”, with the consequence that:

(a) “the large number of terms has produced some theoretical confusion about the boundaries of the topic of control, about the interrelationships among constructs and even about which constructs can be appropriately included in the study of control”,
My projected task was to examine the literature on thought experiments, isolate an issue of interest, and produce a dissertation that addressed that issue; however, the total lack of established order within the literature was immediately apparent, and this task was soon abandoned. By foraging into more extensive domains of the application of thought experimental activity, I soon discovered that the apparent disorder (and most of the heated methodological, terminological, ideological, and functional disputes between individuals) came from the individual participants’ commitment to single issues or to narrow, specialized fields of personal interest; each of which had certainly been able to provide valuable and detailed understandings of specific, highly technical issues — but in complete isolation from one another.

As my extensive study of the literature progressed, a number of significant patterns began to slowly emerge amongst the entirely different applications. These patterns gradually accumulated into a framework that provided an overarching understanding that could accommodate all of the different participants’ views. As a consequence, I am able, in this dissertation, to offer an integrated, systematic account of the qualitative, 

3 For example, there was no universally accepted definition of thought experiments, there were no reliable taxonomies available, there was no standard nomenclature (e.g., the term thought experiment itself was used by many to denote both category and sub-set), and the various types of thought experimental activity were universally considered in isolation from one another.

4 For example, some might only examine one type of experiment; others might only examine a single experiment.

To adapt Miller’s (2003, p.143) view of the evolution of Cognitive Science, and the independent attempts by “psychology, linguistics, neuroscience, computer science, anthropology and philosophy” to understand “cognition” (circa 1977), to the domain of thought experiments, “each [participant], by historical accident, [had] inherited a particular way of looking at [thought experiments]”; but, unlike Cognitive Science’s six aggregate disciplines, they had not yet “progressed far enough to recognize that the solution to some of [their] problems [depended] crucially on the solution of problems traditionally allocated to [those investigating] other [applications of the thought experimental process]”.

5 Given the overall disorder in the literature on thought experiments, it must be stressed that much time and effort has been spent (and exceptional care has been taken) to ensure that all of the charts, concept maps, tables, diagrams, representations, explanations and descriptions that appear in this dissertation are as symmetrical with one another as they can possibly be.

This is an essential part of this dissertation.

It recognizes that there are significant differences in terminology use over a wide range of entities; with the same terms being applied to different entities, or different terms being applied to the same entities. Each term used in this dissertation has been isolated and described — and the entity it denotes has been distinguished from others — in a productive way that ensures both a specific understanding of what this dissertation is addressing, and a seamless transfer of these notions to the wider literature.
symmetrical patterns that seem to be shared by all thought experimental activities. 6

Before reaching that point, however, we must examine the history and application of the term “thought experiment”. Chapter Two positions thought experimental activities within philosophy and science reasoning; whilst, at the same time, emphasizing that these activities are not unique to either philosophers or scientists. Various views on the function and purpose of thought experiments are discussed, along with the rôle of thought experiments in literature in general,7 in science fiction in particular, and their application in Kahn’s specific decision making tool, the “scenario”.

Chapter Three describes the major constituent of all thought experiments: the analogy. The concept of source-to-target mapping is discussed, and clear, symmetrical distinctions are made between the comparative cognitive processes of analogies, metaphors and models;8 and various important applications of each of these mechanisms are discussed.

Chapter Four offers a contextual framework for thought experimentation that is grounded in human need and activity, and discusses the psychological importance of explanations, and the extent to which a sense of control is important for general well-being. The practice of divination is offered in support of the notion that thought experiments simply formalize a number of the patterns of thinking and questioning that pervade all aspects of a normal, healthy mental life; and these questions are examined in some detail, and their various applications are described.

Chapter Five examines the directives for thought experiment performance: their constituent narratives. The notions of narrative, narrator, and reader are examined, along with the consequences of the narrative act. The importance of understanding that a reader actively performs (rather than just passively reads) a narrative, and the extent to which the reader’s degree of transportation by (and their level of immersion in) the text

---

6 That is, rather than an extended list of all of their quantitative differences considered in isolation.

7 From this view, all literature can be thought of as an act of thought experimentation; which, of course, explains how authors and their characters end up in places that they never imagined when they started writing.

8 These terms are, so often, used in a highly ambiguous and extremely confusing way; also, because each tends to be spoken of in isolation from the other two (or, by contrast, all three are sometimes spoken of as if they were interchangeable), there is considerable confusion about which specific mechanism is correctly denoted with which particular word.
narrative determines the extent to which they acquire a cognitive artifact is emphasized. The distinction between knowledge by description and knowledge by acquaintance is highlighted, and the extent to which the acquisition of knowledge by acquaintance generates a sense of psychological ownership is stressed.

Chapter Six examines the extensive process through which an explicit thought experiment is created, designed, developed, constructed and adapted until it reaches the final, polished form that is eventually published.

Chapter Seven introduces two distinctive qualitative patterns amongst thought experiments: their orientation in time, and their temporal direction and sense. These distinctions allow a concept map describing thought experiments, and a set of seven diagrams that isolate the unique features of different sorts of thought experimental activities to be created. Four additional characteristics of thought experiments are identified, and from this a productive, all-encompassing definition of thought experiments is produced which, in turn, permits the identification and description of the function of thought experiments to the extent that, by the end of the chapter, the reader will come to understand all thought experiments, collectively, as powerful devices for attenuating beliefs.

The Bibliography lists all of the works I consulted in the progress of this survey — a proportion of which deal with the consequences of imagining hypothetical scenarios upon subsequent decision making\(^9\) — and each of the ten Appendixes contain important material that readers may need to consult, from time to time.

\(^9\) Due to the way this project has evolved, there is no space for any further discussion on these interesting issues.
Chapter Two
Approaches to Thought Experimentation

While thought experimenting is a truly creative part of scientific practice, the basic ability to construct and execute a thought experiment is not exceptional. The practice is a highly refined extension of a common form of reasoning. It is rooted in our abilities to anticipate, imagine, visualize, and re-experience from memory. That is, it belongs to a species of thinking by means of which we grasp alternatives, make predictions, and draw conclusions about potential real-world situations we are not participating in at that time. (Nersessian, 1993, p.292)

In the quotation above, Nersessian suggests something important to understanding the nature of thought experiments: namely, that we think of them in terms of the cognitive processes involved in thought experimental practice. This cognitive orientation is central to my argument in this dissertation.

I begin my examination of approaches to thought experimentation in this chapter, therefore, with a cognitively oriented definition: thought experiments are formal, intentional, and structured acts of intellectual deliberation,\(^1\) driven by the link between an antecedent and a consequent,\(^2\) that are conducted exclusively in the imagination.\(^3,4\)

---

1. These are not the random, unregulated, passive, unsought imaginings of spontaneous reveries, daydreams, "brown studies" or the so-called flights of the imagination — "daydreams and fantasies are excluded... because they are not [concerned with] raising or answering questions" (Sorensen, 1992, p.207) — they are the purposive, regulated, active, and entirely voluntary imaginings that are often used by scientists and philosophers.

2. In the conditional statement "If you asked me, I would agree", "If you asked me" is the antecedent (or protasis), and "I would agree" is the consequent (or apodosis). Newstead, et al. (1997, p.56) identified seven sorts of conditional statement [(a)-(d), Evans (2002, p.984); (e)-(g), Newstead, et al.]:
   - (a) promise ([father to son] "If you pass the exam, I will buy you a new bicycle");
   - (b) tip ([friend to friend] "If you pass the exam, your father will buy you a new bicycle");
   - (c) threat ([boss to employee] "If you are late for work again, I will fire you");
   - (d) warning ([colleague to colleague] "If you are late for work again, the boss will fire you");
   - (e) temporal ("If the next southbound train is for Ealing then the one after it will be for Wimbledon");
   - (f) causal ("If the lorry is heavier than the legal limit then the alarm bell will ring"); and
   - (g) universal ("If the student is doing economics then he is a socialist")

3. Popper (1968) called them "Imaginary Experiments". White (1990, p.10) distinguishes between sensory imagination ("imagining that a wall has been painted blue"), and non-sensory imagination ("imagining that a promise has been broken"). Thought experiments routinely employ either sort.

4. Galton conducted research (1880) into the extent to which eminent scientists used "mental imagery". On the basis that he, himself, had a great personal ability to create, manipulate and employ vivid mental imagery, he was shocked to discover that most eminent scientists not only did not habitually employ mental imagery, but were also, generally, quite incapable of generating "mental images" at will (Galton, 1880). By contrast, he also cites the extraordinary case of eminent Egyptologist, Flinders Petrie, who could easily manipulate precise technical equipment in the spaces of his own imagination (Galton, 1982/1907, p.66):

   Mr. Flinders Petrie, a contributor of interesting experiments on kindred subjects to Nature, informs me that he habitually works out sums by aid of an imaginary sliding rule, which he sets in the desired way and
and have the function of sharpening the mind by narrowing it.\textsuperscript{5} The significance of this definition will become clearer as the argument of the dissertation develops; but for now, consider it a preface to my examination of the literature on thought experiments.

What relation do thought experiments have to real ones? Thought experiments are the imaginary, “proxy” experiments\textsuperscript{6} scientists conduct (a) prior to a real, “physical” experiment,\textsuperscript{7,8,9}or (b) in place of a physical experiment, when that physical experiment reads off mentally. He does not usually visualise the whole rule, but only that part of it with which he is at the moment concerned. I think this is one of the most striking cases of accurate visualising power it is possible to imagine.

\textsuperscript{5} And, whilst a narrowed mind may be able to productively think very incisively, it may also quite inadvertently ignore many otherwise significant things:

...a good scientific theory finds confirmation in meeting new facts successfully. But in a theoretical study such as philosophy the new facts which present themselves are determined more by one’s mental make-up than by an impartial sampling of reality; for it is association rather than experience which presents them. So even this "coherence test of truth" may indicate... not so much the objective validity of the theory as the groove in which the author’s mind runs. Instead of his theory being as wide as reality, his perception of reality may be as narrow as his theory. (Craik, 1943/1952, p.vii)

Despite Holmes attributing the view to Edmund Burke in his “Commencement Address” to Brown University in 1897 (“One heard Burke saying that law sharpens the mind by narrowing it” [Holmes 1920, p.164]), there is no trace of any such statement in Burke’s work. Perhaps Holmes was alluding to Burke's 1774 speech on American Taxation:

*He was bred to the law, which is, in my opinion, one of the first and noblest of human sciences; a science which does more to quicken and invigorate the understanding, than all the other kinds of learning put together; but it is not apt, except in persons very happily born, to open and to liberalize the mind exactly in the same proportion. (Gerhart, 1969, p.352)*

\textsuperscript{6} Cooper (1999), pp.266-267: Successful thought experiments have many advantages over successful real experiments. Thought experiments are always cheaper and often more ethical than real experiments. They have the advantage of being easily replicated – the reader of a thought experiment can repeat the experiment for himself rather than having to rely on reported results.

\textsuperscript{7} Ignoring the obvious fact that, in many senses, “mental experiments” are indeed physical.

\textsuperscript{8} Mach (1926/1976, p.136) strongly argued that these preliminary imaginary, proxy experiments (Gedankenexperiments) were “a necessary precondition for physical experiment”:

Our ideas are more readily to hand than physical facts: thought experiments cost less, as it were. It is thus small wonder that thought experiment often precedes and prepares physical experiments.

For Mach (p.145), the “method of physical and thought experiment” (which could significantly “enliven enquiry, tell us about new properties and promote insight into their connections”) was an invaluable interaction that promoted the “attunement of thoughts with facts and with each other”:

If [a physical] experiment does not produce the expected outcome, it may be a considerable drawback for the inventor or engineer. But the engineer will regard it as proof that his thoughts did not accurately correspond to the facts. It is precisely this sort of clearly expressed incongruity that can lead to new clarifications and discoveries.

Nersessian (1993), p.292:

Mach held that thought experiments are on a continuum with real-world experiments and saw them as providing empirical data with epistemological status comparable to that of rea-world experiments.

To Irvine (1991, pp.150-151), not only is the “parallel between physical experiments and thought experiments is a strong one”, but also “many thought experiments are meant to precede real experiments in which the original thought experiment’s premises are actually instantiated”.

\textsuperscript{9} Attempting to deliver a more accurate sense of Gedankenexperiment, Hempel (1965, p.164) labelled thought experiments “experiments-in-imagination”.

He spoke of two types, “intuitive experiments-in-imagination” and “theoretical experiments-in-imagination”, warning that the two were extreme archetypes “which are rarely realized in their pure form”, and labelled this kind of thought experiment an intuitive experiment-in-imagination:

An intuitive experiment-in-imagination is aimed at anticipating the outcome of an experimental procedure which is just imagined, but which may well be capable of being actually performed. Prediction is guided here by past experience concerning particular phenomena and their regularities, and occasionally by belief in certain general principles which are accepted as if they were a priori truths. Thus, in explaining the
Chapter Two: Approaches to Thought Experimentation — 7

is impossible to conduct.\textsuperscript{10}

And, moreover, the proxy experiment’s outcome can be so unequivocally clear there’s no need to conduct a physical experiment at all:

The contemplation of an imaginary scenario may lead us to new knowledge neither because it provides us with quasi-observational knowledge of abstracta, nor because it is actually an act of argumentative rehearsal. Rather… in the case of imaginary scenarios that evoke certain sorts of quasi-sensory intuitions, their contemplation may bring us new beliefs about contingent features of the natural world that are produced non-inferentially, but quasi-observationally; the presence of a mental image may play a crucial cognitive role in the formation of the belief in question. And this albeit fallible quasi-observational belief-forming mechanism may, in certain contexts, be sufficiently reliable to count as a source of justification. (Gendler, forthcoming, p.3)\textsuperscript{11}

As Damper (2001/2002, p.2) reminds us, the opposite may also be true: the virtual experiment may be so successful that it generates technological change:

We should note that Turing’s seminal (1936) paper ["On computable numbers, with an application to the Entscheidungsproblem"] is properly a thought experiment. He effectively invented a (virtual) digital computer — the Turing machine — so allowing him to solve a problem in mathematical logic. In this sense, his thought experiment was a trigger to subsequent revolutionary equidistribution of results obtained in rolling a regular die, or in anticipating similar results for a game with a regular homogeneous dodecahedron, certain rules of symmetry, such as the principle of insufficient reason, are often invoked; and similar principles are sometimes adduced in imaginary experiments involving levers and other physical systems with certain symmetry features. Imaginary experiments of this kind are intuitive in the sense that the assumptions and data underlying the prediction are not made explicit and indeed may not even enter into the conscious process of anticipation at all: past experience and the — possibly unconscious — belief in certain general principles function here as suggestive guides for imaginative anticipation rather than as a theoretical basis for systematic prediction. (p.164).

\textsuperscript{10} Hempel categorizes these as theoretical experiments-in-imagination, stating that:

The theoretical kind of imaginary experiment, on the other hand, presupposes a set of explicitly stated general principles — such as laws of nature — and it anticipates the outcome of the experiment by deductive or probabilistic inference from those principles in combination with suitable boundary conditions representing the relevant aspects of the imagined experimental situation. Sometimes, the latter is not actually realizable, as when the laws for an ideal mathematical pendulum or for perfectly elastic impact are deduced from more general principles of theoretical mechanics. The question what would happen if, say, the thread of a pendulum were infinitely thin and perfectly rigid and if the mass of the pendulum were concentrated in the free end point of the thread is answered here, not by "thinking away" those aspects of a physical pendulum that are at variance with this assumption and then trying to envisage the outcome, but by rigorous deduction from available theoretical principles. Imagination does not enter here; the experiment is imaginary only in the sense that the situation it refers to is not actually realized and may indeed be technically incapable of realization. (p.164-165)

It is also important to recognize that in a number of domains of “hard” science (e.g., astronomy) real, physical experiments are impossible to conduct in all cases. Also, "well constructed thought experiments can be used to supplement the findings of real experiments by allowing predictions to be made about the outcomes of experiments which have not yet been, and maybe never could be, performed” (Cooper, 1999, p.271).

\textsuperscript{11} Sorensen (1992), p.197:

The usual point of the experimental procedure is to answer a question. Yet sometimes merely thinking about the procedure answers the question; the thought sometimes renders the action superfluous.
advances in computer technology.

The term *thought experiment*, a direct translation of the German term *Gedankenexperiment*, first appeared in English translations of a paper Mach had published in 1897.\footnote{Or *denkexperiment*, as Fleck (1935/1979, p.21) would have it. Some authors (e.g., Hardy, 1992, p.298) anglicize the term as 'gedanken experiment' (i.e., a *gedanken* type of experiment).

If *gedankenexperiment* is used as an English word it is a "loan-word" (as with pizza, blitz, etc.); and if *thought experiment* is used, it is a "loan-translation" (or *calque*), because it has "a direct correspondence between the morphemes of the model and the new-coined equivalent" (Novotnà, 1967, p.614); in fact, the English expression 'loan-word' is also loan-translation (from the German *Lehnwort*).


Describing them as "surrogates for actual experiments", Mayo (1996, p.67) suggests that thought experiments "may be seen as experiments "done on paper"".\footnote{Witt-Hansen (1996) established that Mach was not first to use *Gedankenexperiment*. Dane, Hans Christian Oersted, first used the term circa 1812 (Witt, p.61). And, in 1820 (Über das Studium der allgemeinen Naturlehre at p.485) Oersted also used the term’s "Germanic" equivalent, *Gedankenversuch* (*Gedankenexperiment* is a German/Latin mix) — reminiscent of Orwell’s plea (1946/1968, pp.131-132) to choose Saxon-based English words in preference to Latin or Greek-based English word.}

Although positioning begins with an entity, it is not a process of invention; in fact, the entity may have been around for a long time. The term refers to the unique (perceptual) location of the entity in the subject’s mind (an analogy to the positioning of balls on a croquet lawn).
extended to the wide range of hypothetical questions that employ subjunctive reasoning which had long been a routine part of philosophy and science.

According to Sorensen (1992, p.186) “thought experiments evolved from ordinary experiments by a process of attenuation” and, today, the term “thought experiment” labels the wide range of appeals to imaginary experience made in the hypothetical case studies routinely used in philosophy (especially ethics), the social sciences (economics, etc.), and philosophy. 20

According to Trout, the essential feature of positioning is that it takes place inside the subject’s mind; with the implication that the first to be in that position is the easiest to remember (because the mind was heretofore “unoccupied” in that location). Positioning not only uniquely locates the entity in a subject’s mind, it also provides a generic name for something that, perhaps, hitherto could not be spoken of — a famous example of positioning is the concept of ‘cybernetics’; an infamous example is the concept of ‘recreational drugs’. 19

This catachrestic extension often causes confusion between those who use the term metaphorically and those who do not; leading to the question “Is this a real thought experiment or not?”

One of the things that sets contemporary analytic philosophy apart from its older variant is its ample use of thought experiments. For examples, see Appendix Seven and Appendix Eight.

Because there is a significant difference between the overall concerns of science and philosophy, the investigations philosophers and scientists conduct are quite easy to distinguish from one another: [Philosophical and scientific] investigations differ both in their methods (the former is a priori, and the latter a posteriori) and in the metaphysical status of their results (the former yields facts that are metaphysically necessary and the latter yields facts that are metaphysically contingent). Yet the two types of investigations resemble each other in that both, if successful, uncover new facts, and these facts, although expressed in language, are generally not about language (except for investigations in such specialized areas as philosophy of language and empirical linguistics). (Ackerman, 1993, pp342-343)

Gooding (1994) p.1038: Thought experiment makes observation easy, leaving us free to concentrate on assumptions and arguments rather than interpretation, instruments, and practical competence. That is why scientists find them so useful, and why philosophers prefer them...

[Thought experiments] are found in the deepest recesses of real experimentation, about matters rarely imagined in philosophers’ mental laboratories. In so far as philosophers can successfully conduct [thought experiments], this is because they share and participate in the same discursive practices as the authors of those [thought experiments]. Scientists are well aware that real experiment exposes matters that thought does not anticipate.

Davenport (1983, p.282) notes that “scientists have [always] been willing to decide about hypotheses on the basis of logical or thought experiments long before any physical evidence was available”. Beardsley (1950), p.220):

The hypothesis may consist of a single statement, or it may be a long and complicated string of statements (in which case it is often called a “theory”). The reconstruction of a crime at the end of a detective story, or in the files of the police, may consist of an elaborate hypothesis about what the criminal is supposed to have done: it may take dozens of pages and contain thousands of statements. But these statements are all parts of the same hypothesis, for together they make up a single explanation of all the known data about the crime.

And, as in the case of “real” physical experiments, there is often an extensive interaction between the theory-oriented “thinkers”, who create and articulate new concepts, and the action-oriented “doers”, who attempt to convert those new ideas into a physical reality:

In research practice... it is often the case that the categories which researchers had initially in mind come to be inappropriate in that they do not allow them to formulate stable regularities. And in the ensuing process of revising, newly forming and stabilizing categories and concepts, experimentation often plays a central role. Any new concepts have to prove themselves by enabling the formulation of stable and ever more general regularities of the experimental results. During the whole process, experimenting and forming of concepts are most closely intertwined. Acting and conceptualizing stabilize or destabilize each other on every step. (Steinle, 2002a, p.306)
sociology, history, politics, etc.), and legal studies; all of which are “fields of inquiry in which true experiments cannot be implemented” (Roese & Olsen, 1995, pp.12).

To Gomila (1991, p.88), thought experiments make an appeal to intuition, depicting situations which “surprise us, mobilize our intuitions, obliging us to clarify them, to bring to light our implicit knowledge of the concepts in virtue of which we think”.

Yet, instead of finding challenges that were based on the irrelevance (from its “logical or empirical impossibility”), or on the absurdity of the intuitions the experiment reports, Gomila found that "debate [routinely] centers round the conclusions, morals,

---

23 It is important to recognize that there is a significant difference between the hypothetical questions that employ subjunctive reasoning in law, and those in philosophy and science. The law must decide whether or not an existing law applies in a specific set of circumstances; whilst scientists and philosophers will leave a difficult question unresolved where no unequivocally clear answer can be reached.

24 Yet, these thought experiment are not generally “treated as practical cases requiring resolution”, but as a means for introducing "some important theoretical considerations" designed to “assist [students] in sorting through the complex real-life cases they will eventually examine” (Hubert, 1999, p.376).

25 Intuitions have always been highly significant in philosophical discourse because they "exhibit the shared character and inter-subjective nature of our experience” (Gomila, 1991, p.88); yet, "a thought experiment is only as reliable as the intuition it exploits” (Sorensen, 1991, p.253)

Research on naïve or intuitive physics has clearly shown that the majority of people have "striking misconceptions about the motion of objects in seemingly simple situations” (McCloskey, Washburn & Felch, 1983, p.636; see also McCloskey, 1983, and Norman, 1988, pp.36-39), and that their "intuitive explanations of motion differ greatly from Newtonian explanations” (Nersessian & Resnick, 1989, p.412).

Even more significant is the finding that “[even though some physics students] may have learned to manipulate the mathematical formalism of Newtonian physics [from their formal training, the majority] exit physics classes with their intuitive beliefs pretty much intact” (Nersessian & Resnick, 1989, p.412).

Accounts of the training of World War I pilots emphasize just how difficult it was to train the pilots to release their (hand-held) bombs some time before they were directly overhead their intended target. This is still an on-going issue in the training and the day-to-day performance of both parachutists and those responsible for dropping materiel and ordnance by parachute from aeroplanes.

In discussing the rôle experiments play in “providing grounds for belief in the truth of physical theories and in the existence of entities involved in those theories” Franklin (1999, pp.7-8) reports a colleague’s attempt to physically replicate Galileo’s thought experiment:

My colleague John Taylor has recently replicated the Galileo experiment. He dropped three steel balls weighing 16 pounds, 8 pounds, and 2 ounces, respectively, from the top of Gamow Tower on the University of Colorado at Boulder campus, and photographed them as they fell a distance of approximately 100 feet. The method of release used was to place the balls at the edge of a hinged platform. The platform was released and allowed to rotate freely. A simple calculation shows that the acceleration of the edge of the platform is 3/2g, where g is the acceleration due to gravity. Because the balls will accelerate at g, the platform falls out from under them, releasing them [simultaneously]...

I note, however, [in Taylor’s photograph] that the platform has the bottom of the three spheres at the same height. Because the spheres were of different sizes (5 inches, 4 inches, and 1 inch in diameter, respectively), the centers of mass of the spheres were at different distances from the center of the earth, and thus, the gravitational force on each sphere was slightly different. The largest difference is approximately one part in 10^3, which is negligible in comparison with the approximately 1% effect observed in the final result...

[Taylor’s second photograph on page 8 shows that] after a fall of approximately 100 feet the heaviest ball is about one foot ahead of the lightest one... and one inch ahead of the intermediate ball.

26 Canadian neurologist Don Weaver asked 100 of his epileptics the question: “what would be their dream epilepsy drug, if they could design one to do whatever they wanted” (Singer, 2004b, p.37). He "was expecting the majority of people to tell [him] about fewer side effects, and once-per-day dosing". Unexpectedly, “more than three-quarters of his patients said their dream drug would have stopped them developing epilepsy in the first place”. Impressed by their insight, Weaver totally changed his career path, and is now engaged in “the quest to create a drug that can stop epilepsy in its tracks”.

Singer reports that “a growing number of researchers” have joined Weaver’s quest.
consequences, etc., that the experiments are supposed to justify, trying to draw a better interpretation of the results”.  

Hempel warns (1965, p.165) that “intuitive experiments-in-imagination are no substitute for the collection of empirical data by actual experimental or observational procedures” especially because their “outcome is liable to be affected by preconceived ideas, stereotypes, and other disturbing factors”.  

---

27 Gomila (1991, p.88) feels that “this general pattern is evidence of the important role intuitions play in philosophy”, “because of their inter-subjective character”. Thought experiments are just one tool among many that philosophers use to evaluate their theories. But they are an important tool, for not only can they strengthen or weaken existing theories, they can also generate data that any future theory must take into account. Theories at the forefront of philosophical research are generally superior to their predecessors because the thought experiments of the past have broadened the evidence base upon which future theories must rest. (Schick and Vaughn, 2003, p.44)

28 According to Hempel (p.165), Lazarsfeld’s review of the Stouffer team’s 1949 publication, The American Soldier, provides an example of the extent to which “beliefs such as these could... affect the outcome and defeat the purpose of intuitive thought-experiments in sociology” (Hempel, p.165). In his review, Lazarsfeld (1949, p.379) speaks of how surveys seem to report things so commonplace that, “from time to time, the argument is advanced that surveys only put into cold storage observations that are already obvious to everyone”. He supplied a list of six statements, plus “brief interpretive comments”, about World War II American soldiers “that are typical of many survey findings” for his readers’ perusal (p.380):

1. Better educated men showed more psycho-neurotic symptoms than those with less education. (The mental instability of the intellectual as compared to the more impassive psychology of the-man-in-the-street has often been commented on.)

2. Men from rural backgrounds were usually in better spirits during their Army life than soldiers from city backgrounds. (After all, they are more accustomed to hardships.)

3. Southern soldiers were better able to stand the climate in the hot South Sea Islands than Northern soldiers (of course, Southerners are more accustomed to hot weather).

4. White privates were more eager to become non-coms than Negroes. (The lack of ambition among Negroes is almost proverbial.)

5. Southern Negroes preferred Southern to Northern white officers. (Isn’t it well known that Southern whites have a more fatherly attitude toward their “darkies”?)

6. As long as the fighting continued, men were more eager to be returned to the States than they were after the German surrender. (You cannot blame people for not wanting to be killed.)

Yet, the real significance of all these “obvious” statements is that not one of them are true: We have in these examples a sample list of the simplest type of inter-relationships which provide the “bricks” from which our empirical social science is being built. But why, since they are so obvious, is so much money and energy given to establish such findings? Would it not be wiser to take them for granted and proceed directly to a more sophisticated type of analysis? This might be so except for one interesting point about the list. Every one of these statements is the direct opposite of what actually was found. Poorly educated soldiers were more neurotic than those with high education; Southerners showed no greater ability than Northerners to adjust to a tropical climate; Negroes were more eager for promotion than whites; and so on.

If we had mentioned the actual results of the investigation first, the reader would have labelled these “obvious” also. Obviously something is wrong with the entire argument of “obviousness”. It should really be turned on its head. Since every kind of human reaction is conceivable, it is of great importance to know which reactions actually occur most frequently and under what conditions; only then will a more advanced social science develop. (p.380)

History also records that other, equally commonplace insights are never heeded; e.g., in his last speech as US President, General Dwight D. Eisenhower gave the (yet to be heeded) explicit and emphatic warning against ever trusting the “military-industrial complex” that seems so prescient today:

In the councils of government, we must guard against the acquisition of unwarranted influence, whether sought or unsought, by the military-industrial complex. The potential for the disastrous rise of misplaced power exists and will persist.

We must never let the weight of this combination endanger our liberties or democratic processes. We should take nothing for granted. Only an alert and knowledgeable citizenry can compel the proper meshing of the huge industrial and military machinery of defense with our peaceful methods and goals, so that security and liberty may prosper together.
As with physical experiments, the burden of proof lies upon the thought experiment’s designer. If the designer "makes an outlandish claim, then that will increase [his/her] burden of proof" (Rips, 1998, p.432); and, if there is insufficient support for the view proposed, the burden of proof will determine the experiment’s failure by default.

The literature typically describes thought experiments in terms of their function; e.g., Brown (1993, pp.33-48) categorizes them as:

1. **challenging** (or, even, **refuting**) a prevailing theory,
2. **confirming** a prevailing theory,
3. **establishing** a new theory, or
4. simultaneously **refuting** a prevailing theory and **establishing** a new theory through a process of mutual exclusion.

Others, like Norton (1991), do not accept “that thought experiments provide some

Although, as Rips (p.432) indicates, in law “burden of proof has technical meanings having to do with which party is obliged to produce evidence relevant to some decision”, he stresses that he is using the term in “a more informal sense”: “in which the person with the burden of proof is the one who has most to do to support his or her position”.

It is also important to recognize that any acceptance of the claims made by the author may produce “systematic shifts in the perceived burden of proof”:

If a claim becomes mutually accepted, then the participant who offered it need no longer defend it. Because there are fewer claims that the participant must now establish, his or her burden will be lighter. By contrast, a claim that is mutually rejected or open (i.e., not mutually determined) can increase a participant’s burden. The participant must find new ways to support claims that are currently under contention and new ways to shore up previous points when currently supporting claims fail. [Thus,] the smaller the number of mutually accepted claims and the larger the number of mutually rejected or open claims, the greater the burden of proof. (Rips, 1998, p.432)

See, for example, Anapolitanos (1991) and Brown (1993).

And, of course, as with all experiments, whilst a theory can never be conclusively confirmed by a positive result from a thought experiment, a theory can be conclusively refuted by a negative result.

Brown labels (a) as “destructive thought experiments” (p.34), (b) and (c) as “constructive thought experiments” (p.34), and (d) as “Platonic thought experiments”.

He further subdivides his “constructive” category into three further groups:

(a) **mediative** thought experiments: which “start with a given background theory” and “[act] like a midwife in getting out a new conclusion” (p.40); i.e., they “[facilitate] a conclusion drawn from a specific well-articulated theory”(p.36).

A mediative thought experiment might illustrate some otherwise highly counter-intuitive aspect of the theory thereby making it seem more palatable; or it may act like a diagram in a geometrical proof in that it helps to understand the formal derivation and may even have been essential in discovering the formal proof. (pp.36-37)

(b) **conjectural** thought experiments: which “[prod us] into conjecturing an explanation for the events experienced in the thought experiment”.

(c) **direct** thought experiments, which:

resemble mediative thought experiments in they start with unproblematic (thought-experimental) phenomena, rather than conjectured phenomena. On the other hand, direct thought experiments, like conjectural ones, do not start from a given well-articulated theory — they end with one. (p.41)
new and even mysterious route to knowledge of the physical world” (p.129), and believe
that thought experiments are a type of argument:34

A good thought experiment is a good argument; a bad thought experiment
is a bad argument. (p.131)35

Some believe thought experiments can be reduced to categorical syllogisms:36,37 the

---

34 To Nersessian (1993, pp.292), whilst thought experiments are a form of “reasoning”, the main
difference between a thought experiment and a “logical argument” (or any other form of “propositional
reasoning”) is that “a thought experiment involves constructing and making inferences from a mental
simulation” — which “makes a thought experiment both “thought” and “experimental””.
Nersessian (1993, pp.297):

While I agree with Norton that thought experiments can often be reconstructed as arguments, the
[mental] modeling function [of the thought experiment] cannot be supplanted by an argument. As
Norton acknowledges, the argument can be constructed only after the fact. That is, on my account, the
argument is not evident until after the thought experiment has been constructed and executed.
Exhibiting the soundness of a thought experiment by reconstructing it as an argument can perform an
important rhetorical function. However, real-world experimental outcomes can be recast in argument
form as well, but no one would argue that the experiment can be replaced by the argument. In similar
fashion, we need to differentiate between the reasoning that is done with the thought experiment and
that which is done with the reconstruction of it. On my view, thought experimenting is a complex form of
reasoning that integrates various forms of information — propositions, models, and equations — into
dynamic mental models. By linking the conceptual and the experiential dimensions of human cognitive
processing, thought experimenting demonstrates the undesirable real-world consequences of a
representation, thereby compelling representational change.

35 Norton (1991):

Thought experiments in physics provide or purport to provide us information about the physical world.
Since they are thought experiments rather than physical experiments, this information does not come from
the reporting of new empirical data. Thus there is only one non-controversial source from which this
information can come: it is elicited from information we already have by an identifiable argument, although
that argument might not be laid out in detail in the statement of the thought experiment. (p.129)

[The] necessary conditions for something to be a thought experiment [is that] it must be an argument, if
its conclusion is to be credible; [and] it must (i) posit hypothetical or counterfactual states of affairs to
warrant the “thought” label; ... [for] if they did not posit such states of affairs they would not be thought
experiments; they would be the description of a real experiment or state of affairs... and (ii) [it must] invoke
particulars irrelevant to the generality of the conclusion to warrant the “experiment” label. (p.130)

The presence of these particulars is what makes thought experiments experiment-like. Thus, in one
version of the thought experiment in which Einstein sought to demonstrate that the effects of acceleration
mimic those of gravitation, he asked us to imagine a physicist-observer who has been drugged and
reawakens closed up inside a box. That there is an observer, that the observer is a physicist, that the
physicist has been drugged, that he is enclosed within a box — all these are particulars which are irrelevant
to the generality of the conclusion which Einstein seeks to draw. Without particulars such as these, however,
thought experiments would not have their experimental appearance. (p.130)

Gendler (2000, p.54) imagines Norton asking, “If the thought experiment is not an argument, why
should we put faith in its conclusion?” She answers her own question in the following way:

Thought experiments rely on a certain sort of constructive participation on the part of the reader, and that
the justificatory force of the thought experiment actually comes from the fact that it calls upon the reader to
perform what I will call an experiment-in-thought.

An experiment-in-thought is an actual experiment; the person conducting the experiment asks herself:
“What would I say/judge/expect were I to encounter circumstances XYZ?” and then finds out the (apparent)
answer. This technique is common in linguistics, where the methodology is used to ascertain the
grammaticality of sentences, the meanings of phrases, the taxonomic categories of words, and so on. And it
is, on one view at least, a central element of moral reasoning: we think about particular imaginary cases,
observe the judgments that they evoke in us, and use these judgments as fixed points in developing our
moral theories.

36 The typical syllogism is comprised of three categorical statements: e.g., “All rectangles have four
sides; all squares are rectangles; therefore, all squares have four sides”.


The syllogism, in its various forms, is one of the most basic and elementary tools used in logic. Aristotle
sometimes uses the term syllogism loosely to mean a form of argument or discussion. [In his Prior Analytics,
Aristotle] gave the definition: “A syllogism is a discourse in which, certain things being stated, something
other than what is stated follows of necessity from their being so. More precisely, “syllogism” has a technical
meaning as a structured form of arguing from two premises to a conclusion.

Henle (1962, p.374) speaks of the ubiquity of “the practical syllogism” in everyday life:
mechanisms which manipulate data\textsuperscript{38} in a precise, structured, \textit{algorithmic} fashion and connect premisses to conclusions; and which, because of their structure, given precisely the same input, they will always generate precisely the same output in every case.\textsuperscript{39}

Others claim that thought experiments are \textit{enthymemes}\textsuperscript{40} rather than syllogisms:

\begin{quote}
A paragraph expressing a thought experiment is actually an enthymeme... an abbreviated formulation of an argument (in which the conclusion or premises are left unexpressed). (Sorensen, 1992, p.214)
\end{quote}

Crick (2004, p.25) understands the real situation to be:

\textbf{thought experiment : argument :: enthymeme : syllogism}\textsuperscript{41}

\begin{itemize}
\item If we include, as it seems we must, the practical syllogism in the "great business of life", that of drawing inferences, it becomes difficult indeed to accept the conclusion that human beings are unable to reason logically. A couple of contemporary examples will suggest how ubiquitous is the practical syllogism in everyday life. "It is too far to walk to the Public Library; I must take a subway or bus. The Fifth Avenue bus passes the Public Library. I must take the Fifth Avenue bus." "I do not want to wear the same dress two days in succession. I wore this dress yesterday; so I do not want to wear this dress today." It is difficult to see how the individual could cope with the ordinary tasks of life if the practical syllogism embodied techniques which... were not a natural mode of functioning of the conscious mind.

\item A syllogism manipulates given data in order to produce a \textit{valid} conclusion. The issue of the data’s objective truth is never considered. As Flew (1979, pp.111) indicates, not only is it possible to generate \textit{valid} (yet entirely \textit{untrue}) conclusions from arguments based on \textit{untrue} premisses, one can also state an entirely true fact as the apparent conclusion to a fallacious argument.

He provides the following example (factually true at the time of its publication) of the \textit{fallacy of the undistributed middle}:

\begin{quote}
Given that "All Communists claim to repudiate racial discrimination", and given that "Dr. Angela Davis claims to repudiate racial discrimination", then it must follow that "Dr. Angela Davis is a communist".
\end{quote}

\item Syllogisms are not easy to use and understand. Although they all share the pattern of three sequential categorical statements and display 256 possible permutations of the syllogistic form — some say 512 (Gilhooly, 1996, p.119) — only 24 of the forms are valid (Baum, 1981, p.225).

Syllogistic reasoning is such a complex activity, and the valid syllogistic forms are so far from being immediately obvious that early in the 13\textsuperscript{th} century CE a set of mnemonic names appeared for the different syllogistic moods and were incorporated into verses designed to make possible to remember the valid forms. The following example embodies 20 of the 24 valid forms (Rose, 1968, pp.102-103):

\begin{itemize}
\item Barbara, Celarent, Darii, Ferioque prioris;
\item Cesare, Camestres, Festino, Baroko secundae;
\item Tertia Darapti, Disamis, Datisi, Felapton, Bokardo, Ferison habet;
\item Quarta insuper addit Bramantip, Camenes, Dimaris, Fesapo, Fresison.
\end{itemize}

\item Here they use the term \textit{enthymeme} with its modern meaning of "a [syllogism] with a suppressed premise [sic]" (Beardsley, 1950, p.336).

According to Aristotle’s earlier usage in \textit{The Art of Rhetoric} [1991, 1.1 (1354a-1355b), pp.66-70; and 2.22 1 (1395ba-1397a), pp.194-197], \textit{enthymeme} denoted a \textit{rhetorical} argument, which argued from premisses which were only probably true (i.e., contrasted with a \textit{demonstrative} argument argument, which argued from premisses which were known to be true). The etymology of \textit{enthymeme}, implies that listeners are actively involved in the rhetorical discourse and construct the complete argument \textit{within their own mind}. Aristotle also notes (1356b, p.76) that although those who argue from examples are just as persuasive, those who rely on enthymemes get the most applause.


For more on demonstrative argument see Corcoran (1972).

\item This is a conventional representation of the sentence: "\textit{thought experiment is to argument in precisely the same way that enthymeme is to syllogism}".
Norton (p.131) categorizes thought experiments on the basis of the styles of argument they use to propose their conclusion (and present their reason for that proposition):42

(a) Type I experiments, which involve deductive reasoning — accepting certain pieces of evidence, and using that evidence as the basis for a subsequent conclusion — and typically use reductio ad absurdum arguments.44

(b) Type II experiments, which use inductive inferences, which accept a

42 People form conclusions from the information they possess; and the type of reasoning they use (i.e., deductive or inductive) is determined by the nature of the information they have. Both styles of reasoning are similar — a reason is offered and a conclusion is proposed — yet both are different in the way connexions are made between the reason offered and the conclusion drawn.

43 Deductive reasoning begins with general or universal assumptions that are known or believed to be true and uses those assumptions to arrive at particular conclusions:

Scientists use deductive thinking when testing theories. To describe this process (known more formally as the hypothetico-deductive method of scientific investigation) step-by-step: We begin with a proposed explanation for understanding observable phenomena (i.e., a theory), from which we derive specific predictions (i.e., hypotheses). We then collect and quantify the relevant observations (i.e., data). If the data match the predictions, we conclude that our theory has empirical support. (Levy, 1997, p.120)

Many totally misunderstand the term deduction due to the unfortunate, misapplication of the term by Doyle’s famous detective Sherlock Holmes. For example, the following interchange takes place between Watson and Holmes in Chapter I of The Sign of the Four (1888):

"But," Watson said, "you spoke just now of observation and deduction. Surely the one to some extent implies the other."

"Why, hardly," [Holmes] answered, leaning back luxuriously in his armchair, and sending up thick blue wreaths from his pipe.

"For example, observation shows me that you have been to the Wigmore Street Post Office this morning, but deduction lets me know that when there you dispatched a telegram."

"Right!" said [Watson]. "Right on both points! But I confess that I don’t see how you arrived at it. It was a sudden impulse upon my part, and I have mentioned it to no one."

"It is simplicity itself," [Holmes] remarked, chuckling at [Watson’s] surprise — "so absurdly simple that an explanation is superfluous; and yet it may serve to define the limits of observation and of deduction. Observation tells me that you have a little reddish mould adhering to your instep. Just opposite the Wigmore Street Post Office they have taken up the pavement and thrown up some earth, which lies in such a way that it is difficult to avoid treading in it in entering. The earth is of this peculiar reddish tint which is found, as far as I know, nowhere else in the neighbourhood. So much is observation. The rest is deduction."

"How, then, did you deduce the telegram?"

"Why, of course I knew that you had not written a letter, since I sat opposite to you all morning. I see also in your open desk there [Holmes and Watson shared the same sitting room at 221B Baker Street] that you have a sheet of stamps and a thick bundle of postcards. What could you go into the post office for, then, but to send a wire? Eliminate all other factors, and the one which remains must be the truth."

"In this case it certainly is so," [Watson] replied, after a little thought… (Baring-Gould, 1968, pp.612-613)

44 Reductio ad absurdum (lit., "reduction to the absurd"), also termed "indirect proof", "establishes the truth of a statement by showing that the contradiction of it is false and that, therefore, the statement must be true" (Gullberg, 1997, p.159). According to (Marcus, 1995) the reductio ad absurdum approach uses either of the following strategies:

(a) Proposition \( P \) is proved by taking \( \neg P \) as a premiss, and "demonstrating that, in conjunction with previously established premisses or axioms, a contradiction follows"; or

(b) Proposition \( \neg P \) is proved by taking \( P \) as a premiss, and "demonstrating that, in conjunction with previously established premisses or axioms, a contradiction follows".

45 Inductive reasoning begins with particular observations and, then, generalizes those observations into broader principles.

Scientists use inductive thinking when creating or building theories. Initially, the scientist collects enough data that a pattern begins to emerge. From this pattern, he or she induces the underlying principle that appears to account for the pattern. In this way, the scientist formulates an explanation (that is, a theory) on the basis of specific observations. For example, physicists, in their attempt to identify and understand a large number of physical phenomena in the universe, have inductively created theoretical constructs such as "force", "gravitation", and "energy". (Levy, p.121)

46 Norton’s categorization implicitly draws attention to another significant issue. According to Levy (1997, pp.120-121), anything involving:
Gendler (1998) holds that thought experiments provide “instructions for guided contemplation” (p.419): “by focusing on imaginary scenarios and making reference to particulars, [they] can provide a fulcrum for the reorganization of conceptual commitments” (p.415).  

(a) deductive reasoning is “theory-driven”, and employs “top-down” cognitive processing;  
(b) inductive reasoning is “data-driven”, and employs “bottom-up” cognitive processing.

According to Arthur (1994, p.406), we use deductive reasoning when solving well-structured problems; yet, because our “deductive rationality [breaks] down under complication” and “beyond a certain level of complexity human logical capacity ceases to cope” (i.e., “human rationality is bounded”), we must rely on inductive reasoning when addressing ill-structured problems. 

Top-down and bottom-up endeavours are not necessarily equal. Maass et al. (2001) who studied the extent to which “people spontaneously infer traits from behaviors, even in the absence of any explicit instruction to do so” (p.391), contrasted with how they “make deductive inferences about the specific behaviors of the individual, based on a global impression of that individual... which may be among the most important social inferences that are made” (p.392), confirmed the existence of an “induction-deduction asymmetry”.

Their experiments clearly indicated “that social inference processes are indeed asymmetrical, with behavior-to-trait inferences appearing more frequently and more spontaneously than trait-to-behavior inferences” (p.401); and, in fact, that “[inductive] inferences from behaviours to traits were almost four times as frequent as [deductive] inferences from traits to behaviors” (p.400).

Hohwy (2004) describes the current range of medical opinion on whether — ignoring physical brain damage or deficits — particular types of delusion are the consequence of a top-down (“from beliefs to perception”) or bottom-up (“from perception to belief”) mechanisms.

47 Norton’s unfortunate choice of Type I and Type II to label his thought experimental categories has nothing in common with the well-known statistician’s concept of “Type I” and “Type II” errors, identified nearly 70 years earlier — which will be discussed in Chapter Four.

48 Using Hanson’s (1962) terminology, we could say that with inductive reasoning, the evidence is “theory-laden”; whilst, by contrast, with deductive reasoning, the evidence is “theory-free”.

49 The two forms of reasoning are not interchangeable.

Beardsley (1950), p.220:
It is important to see that the inductive argument cannot be turned into a deductive argument. You cannot deduce “This is sugar” from “This is sweet”; you can only induce it. It is true that All sugar is sweet, but it is not true that All sweet things are sugar. If all sweet things were sugar, then the inference we have been discussing would be deductive, but since not all sweet things are sugar, the inference can only be inductive...

In an inductive argument, the facts are evidence for the hypothesis if, and only if the hypothesis explains the facts, and the more facts it explains, the more evidence there is for it.

Beardsley (1950, p.222) stresses the importance of two activities:

(a) clearly distinguishing between “what is assumed to be a fact, and what is proposed as an inference from the fact”; and

(b) (consequent upon this) asking three critical questions:

(i) “Whether the alleged facts are indeed facts”;

(ii) “Whether the facts are really evidence for the conclusion”; and, if the answer to both (i) and (ii) are YES, then

(iii) “How strong is the evidence?”.

50 It can also happen that these imaginary exercises impel users into a position where they irrevocably view things in another way. For example, once someone has read Kimon Nicolaides’ The Natural Way to Draw (1972) which, in simple terms, advocates drawing, not by drawing each substantial part of, say, a chair — which involves all sorts of problems, such as perspective — but by simply drawing lines that represent the border between the various components of the chair and the space the entire chair occupies, and they view

...
Chapter Two: Approaches to Thought Experimentation — 17

She claims this “guided contemplation” description encloses both Brown’s “platonic intuition” and Norton’s “argument” view (p.419); stressing that she “[prefers] to describe it as a reconfiguration of conceptual space” (p.420).\(^51\)

By introducing novel categories by which we make sense of the world, this reconfiguration allows us to recognize the significance of certain previously unsystematized beliefs... these beliefs are new; and... they are justified. Thus the thought experiment brings us to new knowledge of the world, and it does so by means of non-argumentative, non-platonic, guided contemplation of a particular scenario.\(^52\)

Based on Craik’s (1943/1952) assertion that most human thinking involves mental modelling, some (e.g., Nersessian, 1993, pp.291-292) believe that one performs a thought experiment through a process of “mental modeling” using analogue models.\(^53\)

objects from this perspective, their perspective of chairs will have been been considerably altered, regardless of any world-view they may have previously held in relation to the way to draw chairs. (I noticed a similar thing happen to my perception of rocks and trees, once I had learned precisely how the Chinese painted such things during my Asian Studies degree.)

\(^51\) Gendler (p.420):

By introducing novel categories by which we make sense of the world, this reconfiguration allows us to recognize the significance of certain previously unsystematized beliefs... these beliefs are new; and... they are justified. Thus the thought experiment brings us to new knowledge of the world, and it does so by means of non-argumentative, non-platonic, guided contemplation of a particular scenario...

Even if one were to provide an argumentative reconstruction that did almost perfectly capture the thought experiment’s demonstrative force, this would not show that the reason the thought experiment is successful is because, deep down, it is nothing more than an argument in disguise...

To the contrary, the success of the thought experiment may be a result of the way in which it invites the reader’s constructive participation, depicts particulars in ways that make manifest practical knowledge, and describes an imaginary scenario wherein relevant features can be separated from those that are inessential to the question at issue.

\(^52\) Gendler (p.420):

Even if one were to provide an argumentative reconstruction that did almost perfectly capture the thought experiment’s demonstrative force, this would not show that the reason the thought experiment is successful is because, deep down, it is nothing more than an argument in disguise...

To the contrary, the success of the thought experiment may be a result of the way in which it invites the reader’s constructive participation, depicts particulars in ways that make manifest practical knowledge, and describes an imaginary scenario wherein relevant features can be separated from those that are inessential to the question at issue.

\(^53\) While Nersessian believes that thought experiments are heuristics, and not algorithms — precisely because, she says, “even if used correctly [they] may lead to the wrong solution or to no solution at all” (Nersessian, 1999, p.162) — it certainly seems, as will be discussed later, that all thought experiments are algorithms at least to the stage that the performer engages in the subjunctive reasoning invited by the experiment’s hypothetical question.

According to Dennett (1995, pp.50-51) an algorithm is “a foolproof recipe” (“a certain sort of formal process that can be counted on — logically — to yield a certain sort of result whenever it is “run” or instantiated”) in which “each constituent step, as well as the transition between steps, is utterly simple”, and provides “a tedious breakdown of the process into dead-simple steps, requiring no wise decisions or delicate judgments or intuitions on the part of the recipe-reader”, and is “simple enough for a dutiful idiot to perform” (emphasis added). At least to the extent that all thought experiments require a “dutiful idiot” for their performance, they are algorithmic (and the extent to which a thought experiment might be non-algorithmic is, very likely, be an index of its lack of refinement).

According to Hempel (1965, pp.165-166), intuitive experiments-in-imagination are “[used] by such writers as Max Weber and Howard Becker as a method of sociological inquiry”; and the “heuristic function [of their experiments] is to aid in the discovery of regular connections between various constituents of some social structure or process... [which] can then be incorporated into an ideal type and thus provide the basis for the explanatory use of the latter”.
Whilst some thought experiments are driven by a view that the real problem is one of consistently asking the wrong questions\textsuperscript{54} (i.e., rather than there being an inherent absence of solutions\textsuperscript{55}), all thought experiments seem to embody a patterned way of thinking that has, as its principal goal, the efficient “\textit{explanation, prediction, and control of events}”\textsuperscript{56} (Horton, 1982, p.204), and are designed to:

(a) challenge the prevailing status quo;\textsuperscript{57,58}

(b) extrapolate beyond (or interpolate within) the boundaries of already established fact;\textsuperscript{59}

\textsuperscript{54} In the sense that the \textit{Meno} is an example of the “right” questions being asked.

\textsuperscript{55} This is the point made by Chesterton’s detective, Father Brown (1955, p.382):

“Father Brown laid down his cigar and said carefully:

"It isn’t that they can’t see the solution. It is that they can’t see the problem."

\textsuperscript{56} La Caze (2002), p.174:

Thought experiments are powerful persuasive tools that are used to provide immediate evidence and support for a philosopher’s conclusions, rather than to merely clarify intuitions, as is purported. Philosophers often decide how people \textit{ought} to respond to a particular thought experiment and design the experiment to reinforce a certain conclusion. There is a tension between the claim that thought experiments are intended as genuine tests and the philosophers’ attempts to justify their own position. These thought experiments clearly work on the level of persuasiveness. They also work on the level of enframing in dictating the terms of the debate, and even come close to the level of expressibility, since it is difficult to see how certain questions could arise at all without the thought experiments.

In discussing the “three levels of embeddedness” in “images” — a term she uses to label “analogies, thought experiments, myths, metaphors and models” collectively (p.2) — La Caze (p.11) defines her usage of \textit{persuasiveness, enframing, and expressibility} as follows:

It may be thought that if it is accepted that ideas can sometimes be expressed without a particular image, then images have been reduced yet again to the status of mere rhetoric. This is not the case. I propose... that images are essential to philosophy... [and introduce] three levels of what I will call "embeddedness". One is the level of expressibility — one needs the image to express the thought. The second level involves the provision of a framework that gives substance to the thought expressed. The third is the level of persuasion and concerns the way in which the image makes the thought convincing. I call these levels expressibility, enframing, and persuasiveness. Images can work on any or all of these levels... (p.11)

\textsuperscript{57} This includes activities as varied as correcting misinformation (or misapprehension), identifying flaws in the argument(s) presented, and the long-term preservation of objectively established fact.

\textsuperscript{58} Sorensen (1992a, p.135) believes that most thought experiments are designed to \textit{disprove} specific assertions that some particular thing is permissible, or forbidden, or known, or believed, or possible, or necessary. Medawar (1969, p.52) states that “we carry out experiments more often to discriminate between possibilities than to enlarge the stockpile of factual information”. Gottfried (1991, p.131) maintains that whilst thought experiments “cannot tell us what IS right... they CAN tell us what CANNOT BE right”.

\textsuperscript{59} Here the thought experiment uses “our explicit and tacit knowledge of the ways in which entities of the type we are imagining behave in actual situations, which allows us to predict how they would behave in the hypothesised situation” (Cooper, 1999, p.268) in order to isolate and identify what Merton (1936) called the “unanticipated consequences” (now more commonly termed the “unintended consequences”, see Norton, 2004) of “purposive social action”. He emphasized that his term “purposive action... [is] concerned with "conduct" as distinct from "behavior". That is, with action which involves motives and consequently a choice between various alternatives” (p.895).

For examples see Reynolds & Kates (1995), and Henry (2003; at \textit{Appendix Nine}).

In 1990 the Victoria legislated to make safety helmets mandatory for all bicycle riders. Although there was the expected significant reduction the absolute number of head injuries, there was also a quite unexpected reduction in the number of juvenile cyclists; because wearing a bicycle helmet was not considered “cool” (Vulcan, Cameron & Heiman, 1992; Vulcan, Cameron & Watson 1992).
(c) predict and forecast the (otherwise) indefinite and unknowable future;
(d) explain the past;\textsuperscript{60}
(e) retrodict, postdict and postcast the (otherwise) indefinite and unknowable past;
(f) facilitate decision making, choice and strategy selection;\textsuperscript{61}
(g) solve problems\textsuperscript{62} and generate ideas;\textsuperscript{63}

\textsuperscript{60} For example, Ward (1998, p.218) describes the work of palaeontologists in terms of them performing thought experiments:

Physicists, the top carnivores of the trophic pyramid of scientists, have at their service the most expensive machines and instruments ever devised. To smash atoms, to find and identify the smallest particles in the universe, to see and measure the last whispers emanating from the Big Bang — such quests require brute instrumental strength. Nevertheless, physicists still routinely encounter phenomena and questions that are beyond the scope of even the most advanced technology. When this happens they must take an entirely different tack — an approach that may not solve a problem but that at least brings it to an intellectual point where it can be re-examined and perhaps can eventually be solved theoretically rather than experimentally or observationally. These exercises are "thought experiments". Einstein loved thought experiments — and needed them. Yet these devices are not the sole property of physicists. Scientists in other disciplines (such as paleontology) can use them too. Every time a paleontologist attempts to reconstruct conditions in some long-ago time, he or she is conducting a thought experiment.

\textsuperscript{61} For example, see Appendix Four, and the comet and Earth issue it contains; an example of the "sharpening the mind by narrowing it" function of thought experiments.

\textsuperscript{62} Goel (1995, p.4):

Problem solving — our ability to recognize dissatisfactory states of affairs and transform them into satisfactory states — is a fundamental human cognitive capacity.

On most definitions, it requires at least the following conditions: (i) that there be two distinct states of affairs, (ii) that the agent be in one state and want to be in the other state, (iii) that it not be apparent to the agent how the gap is to be bridged, and (iv) that bridging the gap be a rational, consciously guided (at least at the top executive level), multistep process.

Smith (1988, p.1491) describing a problem as a situation where:

(a) "a gap, difference, or disparity [exists] between the way things are and the way one wants them to be",
(b) "[that] problematic gap must be difficult to bridge or close", and
(c) "the gap must be important enough to inspire current or prospective solution activities",

defines a problem as:

an undesirable situation that is significant to and may be solvable by some agent, although probably with some difficulty. Since a problem is an "undesirable situation", it does not exist strictly as an objective state-of-the-world, nor as a subjective state of dissatisfaction. A problem is a relationship of disharmony between reality and one's preferences, and being a relationship, it has no physical existence. Rather, problems are conceptual entities or constructs. The term is an abstraction from the world of observables and is applied because it serves a useful function. Essentially the term is an attention-allocation device. Marking a situation as problematic is a means of including it in one's "stack" of concerns, placing it on an agenda for future attention and solution efforts. Thus, there is an element of arbitrariness in labeling a situation as problematic. One can apply the concept more or less liberally, depending on whether he prefers his attention to be loosely or tightly focused. In any complex, real world, situation, there are an unlimited number of concerns which could be identified as problems, but there are none which absolutely must be so identified.

Despite arguing (pp.1489-1490) they are "undifferentiable at a cognitive process level", Smith makes a significant distinction between decision making and problem solving at a "conceptual level"; on the basis that, with decision making there is a strong implication that "a choice will be made" which, in turn "[suggests] that alternatives exist or will be identified", whilst with problem solving the implication is that "problem solving is directed at the resolution of a problem":

\[\text{[and, for the problem solver] the focus is on achieving some goal, [with the implication of] it being difficult to determine how this might be done. If the decision-maker's question is, "Which alternative should I choose so as to maximize my preferences?", that facing the problem solver is, "How can I achieve my goal, move from an unsatisfactory state to one more desired?"}\]

\textsuperscript{63} In terms of assisting with solving problems and generating ideas, Do and her colleagues (2000, pp.484) speak of seven rôles for diagrammatic thinking as a "medium of thought" that seem equally true of thought experiments in general:

- generating concepts;
- externalizing and visualizing problems;
(h) move current (often insoluble) problems into another, more helpful problem space;\textsuperscript{64}

(i) attribute causation, preventability, blame and responsibility for specific outcomes;\textsuperscript{65}

(j) assess culpability and compensation in social and legal contexts;\textsuperscript{66}

• organizing cognitive activity;
• facilitating problem solving and creative effort;
• facilitating perception and translation of ideas;
• representing real world artifacts that can be manipulated and reasoned with; [and]
• revising and refining ideas.

Smith's study of structured, structurable and unstructured problems (1988, pp.1501-1502) produced a taxonomy of problems, comprised of four "high level problem categories", and nine "fundamental types of problem", that can assist our understanding of the purposes of various thought experiments:

General Problem Categories
State Change: The need to change some unsatisfactory state or to achieve some goal.
Performance: The need to improve the performance of some function or system.
Knowledge: The need to acquire certain knowledge, information, or understanding.
Implementation: The need to put some action or desired state-of-affairs into effect.

Problem Types
Goal-Setting: Determining what one wants and appropriate levels of relevant objectives.
Diagnosis: Providing causal explanations of why things are what they are.
Design: Determining how to achieve a desired state or end.
Description: Determining, through observation and thought, what happens to be the case.
Research: Acquiring knowledge through directed investigation.
Alternative Generation: Identifying possibilities or alternatives appropriate to some objective.
Prediction: Predicting future or unknown current states of affairs.
Evaluation: Assessing the worth of some entity against one's preferences or external standards.
Persuasion: Gaining the consent or conviction of others.

\textsuperscript{64} See, for example, "The Monk Problem" at Appendix Ten.

\textsuperscript{65} These issues also have significant applications in areas outside the law.
Frey and his colleagues (1985, pp.323-327) conducted field studies into the role of psychological factors in the recovery of accident victims (nearly all of whose injuries were expected to improve, and in almost all cases it was a realistic goal that the patient would "regain the health they enjoyed before the accident" [p.325]). They clearly demonstrated (pp.323-327) that, \textit{ceteris paribus}, those who felt that:

(a) their accident was unavoidable,
(b) their accident had been caused by chance or fate,
(c) they were, retrospectively, not at all responsible for their accident,
recovered significantly better and faster. They also found that those thinking they had present and future control over the future course of their recuperation recovered significantly better and faster (p.325).

\textsuperscript{66} French (1984) describes how a thought experiment was used to attribute responsibility in the case of the crash of the New Zealand flight TE-901 into Mt. Erebus in 1979.
The thought experiment was, in part, motivated by (a) Aristotle's assertion (\textit{Nicomachean Ethics}, II.9.1109b: 1980, p.47) that, generally, "we do not blame people for unintentionally "slightly deviating from the course of goodness", as long as they do not subsequently practice behaviour that makes such deviations a matter of character" (French, p.108) and, (b) by a principle enumerated by Mackie (1965) that the crash would not have occurred if certain conditions had not been present.
(Mackie, who spoke of causes which were "an insufficent but necessary part of a condition which is itself unnecessary but sufficient for the result", gave the example of a spark from a short circuit, in the presence of inflammable material, and in the absence of a sprinkler system, \textit{combining} to cause a house
(k) ensure the repeat of past success;
(l) examine the extent to which past events might have occurred differently;\textsuperscript{67} or
(m) ensure the (future) avoidance of past failures.\textsuperscript{68,69}

The literature often makes distinctions between well-structured problems,\textsuperscript{70} which have a specified goal, and ill-structured problems,\textsuperscript{71} which are (according to Star, 1983, p.224) “transformed into well-structured problems by the expedient of ignoring progressive complications or qualifications”.\textsuperscript{72}

The spark, he argued, is \textit{indispensable} — because the other conditions, in the \textit{absence} of the spark, would not have caused the fire — and, thus, the spark was the \textit{cause} of the fire \textit{in the presence of the other conditions} (1965, p.245.).

For another legal application of a thought experiment, see Brodsky (2003).

\textsuperscript{67} For example, Sorensen (1992b), p.29:
In 1961, a B-52 carrying two hydrogen bombs disintegrated in flight over North Carolina. Inspection of one of the bombs revealed that five out of its six safety devices had failed. The defense secretary, Robert McNamara, cited this near catastrophe as one of the things that had the greatest impact on his thinking about nuclear policy.

\textsuperscript{68} See, for example, Morris, Moore & Sim (1999), and Morris & Moore (2000) on industrial accidents.

\textsuperscript{69} In response to a serious miscarriage of justice involving a wrongful conviction for murder following false testimony provided by a gaol inmate — the innocence of the accused was demonstrated via DNA many years after he had been convicted and incarcerated — the Canadian Province of Ontario instituted an unprecedented examination of its entire legal system (see Kaufman, 1998; King, 1998; Findley, 2002). From a view that there must be many serious flaws in the Province’s entire legal system (rather than the case just being an isolated mistake), the enquiry took the form of an “innocence commission”.

Canada is the first country in the world to have “innocence commissions”. What they do is truly marvellous. They will perform a post-mortem on the case of a man who was wrongly convicted and find out what went wrong and what they can do to reduce the likelihood of it happening again in the future. (Nowack, 2003, p.48)

The Commission held 146 days of hearings, heard testimony from 120 witnesses, reviewed transcripts, exhibits, and other documents, and produced a two-volume, 1400-page report, complete with 119 specific recommendations for improving the criminal justice system. Those recommendations addressed problems with forensic science, the use of informant testimony, police investigation procedures, the performance and training of prosecutors and defense counsel, the instructions given the jury at trial, and the rules governing post-conviction and appellate review. (Findley, 2002, p.342).

\textsuperscript{70} These have a specified goal or target (e.g., the Tower of Hanoi).

\textsuperscript{71} These don’t have a specified goal or target (e.g., “What special present should I buy my wife for her fiftieth birthday?”); and, in fact, there may be multiple solutions possible. And, moreover, by contrast with a well-defined problem, it is often the case that you have no idea of what a solution might look like until it is reached; and, even then, although it is relatively easy to judge whether or not a solution is acceptable, it is impossible to judge whether or not the solutions that have been considered represent all possible solutions.

Simon (1973/1977, p.304), attributes the notion of “ill-structured problems” to Walter Reitman, who spoke of “ill defined problems”.

\textsuperscript{72} According to Simon (1999, p.676), cognitive scientists studying the way in which problems are solved by humans (cognitive psychology), by computers (artificial intelligence), and by computers interacting with their environment (robotics) have gradually developed a common terminology to facilitate the exchange of information of these somewhat similar domains:
To solve a problem, a \textit{representation} must be generated, or a preexisting representation accessed. A representation includes (1) a \textit{description} of the given situation, (2) \textit{operators} or actions for changing the situation, and (3) \textit{tests} to determine whether the goal has been achieved. Applying operators creates new situations, and potential applications of all permissible operators define a branching tree of achievable situations, the \textit{problem space}. Problem solving amounts to searching through the problem space for a situation that satisfies the tests for a solution...
Polished thought experiments are well-structured problems.

They often introduce interesting, important and valuable new perspectives on old mysteries and old questions; and, whilst making old questions irrelevant, they can also create new questions which, in themselves, may or may not be easily answered.\textsuperscript{73}

Popper claimed (1963/2002, p.69) that the principal benefit of thought experiments was their capacity to “[eliminate] an inadequate hypothesis” (through revealing conceptual errors\textsuperscript{74}), “when a more dogmatic attitude would eliminate it by eliminating us”; something which, he said, allows “our theories, our conjectures [to] suffer in our stead in the struggle for the survival of the fittest”.\textsuperscript{75,76}

In 1980 Daniel Dennett, criticized John Searle’s (1980/1982) collection of “Chinese Room” thought experiments on the grounds that they were a coercive aggregate of incorrectly applied intellectual exercises that “[had produced] favourable intuitions [only

---

\textsuperscript{73} Nersessian (1993, p.296):  
A thought experiment is usually so compelling that even in those cases where it is possible to carry it out, the reader feels no need to do so. The constructed situation, itself, is apprehended as pertinent to the real world in several ways. It can reveal something in our experience that we did not see the import of before, such as that the measurable current in a stationary and in a moving conductor cannot support the distinction made in the theoretical explanation of them as different phenomena. It can generate new data from the limiting case, for example, that in no medium lead and wood would fall at the same speed. It can make us see the empirical consequences of something in our existing conceptions, such as that the attributes called "gravitational mass" and "inertial mass" are the same property of bodies.

\textsuperscript{74} Gopnik (1998, p.113) speaks of science as being a form of "cognitive optometry": “a system that takes the devices we usually use to obtain a veridical picture of the world and corrects the flaws and distortions of those devices”. Yet, she also remarks that “the fact that most people over forty wear glasses is not, however, usually taken as an indictment of the visual system” (Gopnik, 2000, p.316)

\textsuperscript{75} Or, as Dennett (1995, p.375) would have it, “[they permit] our hypotheses to die in our stead”.

Recall that design typically occurs in situations where it is not possible or desirable to tamper with the world until the full extent and ramifications of the intervention are known in advance. After all, we only get one "run" on the world. Every action is irrevocable and may have substantive costs associated with it. Thus it is not surprising to find that designers produce and manipulate representations of the artifact rather than the artifact itself. (Goel, 1995, p.128).

\textsuperscript{76} According to Craik (1943/1952, p.59):

Human thought has a definite function; it provides a convenient small-scale model of a process so that we can, for instance, design a bridge in our minds and know that it will bear a train passing over it instead of having to conduct a number of full-scale experiments; and the thinking of animals represents on a more restricted scale the ability to represent, say, danger before it comes and leads to avoidance instead of repeated bitter experience.
because they had been] generated by misleadingly presented thought experiments”; therefore, argued Dennett, the collection was nothing but “an intuition pump, a device for provoking a family of intuitions... [which] is not, typically, an engine of discovery, but a persuader or pedagogical tool — a way of getting people to see things your way, once you’ve seen the truth” (1980, p.429).

Over time, for many, Dennett’s term intuition pump has become synonymous with thought experiments; and, in fact is often used in preference by various authors.

Recent correspondence (p.c., 2003) has emphasized that Dennett still believes that his intuition pumps are very different from the classic thought experiments of Galileo and Einstein:

[Intuition pumps] are not supposed to clothe strict arguments that prove conclusions from premises. Rather, their point is to entrain a family of imaginative reflections in the reader that ultimately yields not a formal conclusion but a dictate of "intuition". Intuition pumps are cunningly designed to focus the reader's attention on "the important" features, and to deflect the reader from bogging down in hard-to-follow details. There is nothing wrong with this in principle. Indeed one of philosophy's highest callings is finding ways of helping people see the forest and not just the trees. But intuition pumps are often abused, though seldom deliberately. (Dennett, 1984, p.12)

77 Dennett (1995) clearly states that the term intuition pump was intentionally derogatory:

If you look at the history of philosophy, you see that all the great and influential stuff has been technically full of holes but utterly memorable and vivid. They are what I call "intuition pumps" — lovely thought experiments. Like Plato's cave, and Descartes's evil demon, and Hobbes' vision of the state of nature and the social contract, and even Kant's idea of the categorical imperative. I don't know of any philosopher who thinks any one of those is a logically sound argument for anything. But they're wonderful imagination grabbers, jungle gyms for the imagination. They structure the way you think about a problem. These are the real legacy of the history of philosophy. A lot of philosophers have forgotten that, but I like to make intuition pumps. I like to think I'm drifting back to what philosophy used to be, which has been forgotten in many quarters in philosophy during the last thirty or forty years, when philosophy has become a sometimes ridiculously technical and dry, logic-chopping subject for a lot of people — applied logic, applied mathematics. There's always a place for that, but it's nowhere near as big a place as a lot of people think.

I coined the term "intuition pump," and its first use was derogatory. I applied it to John Searle's "Chinese room," which I said was not a proper argument but just an intuition pump. I went on to say that intuition pumps are fine if they're used correctly, but they can also be misused. They're not arguments, they're stories. Instead of having a conclusion, they pump an intuition. They get you to say "Aha! Oh, I get it!" (p.182)

78 In some cases it seems that the authors concerned have not been aware of the term's generally derogatory connotations.

79 Dennett continues (loc. cit.):

Perhaps the most frequent abuse is deriving a result — a heartfelt intuitive judgment — from the very simplicity of the imagined case, rather than from the actual content of the example portrayed so simply and clearly. Might it not be that what makes the wasp's fate so dreadful is not that her actions and "decisions" are caused but precisely that they are so simply caused? If so, then the acknowledged difference between the object of our intuition pump and ourselves — our complexity — may block our inheritance of the awfulness we see in the simple case. Perhaps we should laugh, not shudder; perhaps this intuition pump is like that nightmare snake who swallows his tail and keeps on going until he's completely eaten himself up.

80 In answer to my question on the apparent equivocation of the word pump in the term "intuition pump"; specifically, was it denoting an activity of (a) "revealing and making available a hitherto hidden resource of intuition, whose magnitude remains otherwise unaltered -- and, therefore, a "pump" in the sense of bringing water to the surface from a flooded mine shaft 1,200 feet below the surface", or an
The imaginary cases presented in thought experiments can offer significant levels of challenge to even the most active of intellects; both in terms of the manner in which “such imaginary cases are to be constructed, and how they are to be taken when constructed in various ways” (Williams, 1970, p.161).

Whilst some (e.g., Schick and Vaughn, 2003, p.43) simply define a conceivable situation as being “coherently imaginable”, others differ: in his Sixth Meditation, Descartes made a very clear distinction between the things he could conceive and the things he could easily imagine, arguing that many of the things he could conceive (e.g., a 1,000 sided figure), he could not possibly imagine.

activity of (b) “increasing the magnitude of an already identified and otherwise available resource of intuition by a factor of 10 -- in the sense that a steroid-filled body-builder is "pumped-up" just before going on stage to pose in a competition (and, therefore, a "pump" in the sense of the foot-pump used to inflate a rubber raft)”, Dennett replied:
I meant to stress that these are artifacts, to be reverse engineered as if they were machines. Since they are designed to evince intuitions that one otherwise might not have, your first meaning [viz., (a) above] is somewhat closer to what I have intended by the term. (personal correspondence 2003)

All thought experiments assume that “thought experimenters are... perfectly competent in their ability to follow procedures and as perceivers of events and outcomes” and “the veridicality of [their] perceptual and experimental competencies is placed beyond question” (Gooding, 1994, p.1031).

According to Laymon (1991, p.172) the success of a simulation of a real experiment depends on the extent to which users are already familiar with the domain in question and have the capacity “to imagine for themselves the sorts of experimental refinements” needed to perform the thought experiment; and this capacity is usually based on their experience with “analogous experiments that [has demonstrated] that such refinements can be developed”.

Descartes was emphatic on the “traditional distinction made between (sensory) imagining on the one hand, and (non-sensory) non-imaginistic conceiving on the other” (Gendler & Hawthorne, 2002, p.9); particularly to support the view expressed in his Second Meditation, that the imagination is totally unsuitable for examining the nature of the soul.

Hume expresses a far more flexible view in his Treatise of Human Nature:
‘Tis an establish’d maxim in metaphysics, that whatever the mind clearly conceives includes the idea of possible existence, or in other words, that nothing we imagine is absolutely impossible. We can form the idea of a golden mountain, and from thence conclude that such a mountain may actually exist. We can form no idea of a mountain without a valley, and therefore regard it as impossible. (1739-1740/2001, 1.2.2, p.26, emphasis in original)

Notwithstanding this, in the specific context of thought experiments, it is clear that the ease with which a proffered scenario can be imagined is critical to the scenario’s utility.

Scarry (1995), p.1:
When we speak in everyday conversation about the imagination, we often attribute to it powers that are greater than ordinary sensation. But when we are asked to perform the concrete experiment of comparing an imagined object with a perceptual one — that is, of actually stopping, closing our eyes, concentrating on the imagined face or the imagined room, then opening our eyes and comparing its attributes to whatever greets us when we return to the sensory world — we at once reach the opposite conclusion: the imagined object lacks the vitality and vivacity of the perceived; it is in fact these very attributes of vitality and vivacity that enable us to differentiate the actual world present to our senses from the one that we introduce through the exercise of the imagination.

Descartes (1641/1968, pp.151-152) asserts that imagination is "nothing other than a certain application of the faculty of knowing to the body which is immediately presented to it" — or, in his Second Meditation, “imagining is nothing other than contemplating the figure or image of a corporeal object” (op.cit., p.106) — and produces the following in support of this assertion:
We must never assume that two apparently alternative solutions offered for the issues described in a particular narrative were aimed at solving the same problem.\footnote{Gendler and Hawthorne (2002, pp.1-2) argue that performing thought experiments is a "pervasive feature of our mental life — both in day-to-day decision making and in philosophical reasoning":

We have, it seems, a capacity that enables us to represent scenarios to ourselves using words or concepts or sensory images, scenarios that purport to involve actual or non-actual things in actual or non-actual configurations. There is a natural way of using the term ‘conceive’ that refers to this activity in its broadest sense.

When we engage in such conceivings, the things we depict to ourselves frequently present themselves as possible, and we have an associated tendency to judge that they are possible. Indeed, when invited to consider whether something is possible, we often engage in a deliberate effort to conceive of it; upon finding ourselves able to do so, we conclude that it is. We may even decide that something is impossible on the basis of our apparent inability to conceive of it.}

For example, when I imagine a triangle, I not only conceive it as a figure composed of three lines, but moreover consider these three lines as being present by the power and internal application of my mind, and that is properly what I call imagining. Now if I wish to think of a chiliagon \(\{a\,1,000\text{-sided figure}\}\), I indeed rightly conceive that it is a figure composed of a thousand sides, as easily as I conceive that a triangle is a figure composed of only three sides; but I cannot imagine the thousand sides of a chiliagon, as I do the three of a triangle, neither, so to speak, can I look upon them as present with the eyes of my mind. And although, in accordance with my custom of always making use of my imagination when I think of corporeal things, it may come about that, in conceiving a chiliagon, I picture confusedly to myself some figure, yet it is very evident that this figure is not a chiliagon, since it differs in no way from the figure which I would picture if I thought of a myriagon \(\{a\,10,000\text{-sided figure}\}\) or of any other many-sided figure, and since it in no way serves for the discovery of the properties which constitute the difference between a chiliagon and other polygons.

But if it is a question of considering a pentagon, it is indeed true that I can conceive its figure, as well as that of a chiliagon, without the help of imagination; but I can also imagine it by applying the attention of my mind to each of its five sides, and at the same time to the space which they enclose. Thus I know dearly that I need a particular effort of the mind in order to imagine, which I do not need in order to conceive; and this particular effort of mind shows dearly the difference between imagination and pure intellecction or conception.

Locke (1975, pp.368-369; II,XXIX, 13) held a similar view:

Our complex Ideas being made up of Collections, and so variety of simple ones, may accordingly be very clear and distinct in one part, and very obscure and confused in another. In a man who speaks of a Chilaëdron, or a Body of a thousand sides, the Idea of the Figure may be very confused, though that of the Number be very distinct; so that he being able to discourse, and demonstrate concerning that part of his complex Idea, which depends upon the Number of a Thousand, he is apt to think, he has a distinct Idea of a Chilaëdron; though it be plain, he has no precise Idea of its Figure, so as to distinguish it, by that, from one that has but 999 sides. The not observing whereof, causes no small Error in Men's Thoughts, and Confusion in their Discourses.

White (1990, p.174) complains that most contemporary writers "do not seem to bother about any differences between the imaginable and the conceivable", and some "even equate them".

He speaks of Wittgenstein using "sich denken" and "sich vorstellung" interchangeably (sometimes in the same passage), and how "sich denken" is intermittently translated as both "imagine" and "conceive". He also describes how some "argue that certain perfectly conceivable things, such as an unperceived physical object, are nevertheless unimaginable", how others "argue that one can imagine things, such as being Napoleon, which one cannot conceive" — or, even, Grayling’s view that "one can imagine, but not conceive, that a Gorgon turns men into stone" (p.221).

\footnote{Berg and Calderone (1994, p.125) draw our attention to the fact that different individuals interpret problems differently: "problem solvers interpret the same everyday problem in a variety of ways" and "individuals of different ages and genders may focus on different interpretations" of the same issue, and that "different domains (e.g., school domain vs. family domain), may pull for different interpretations", and "these different interpretations [have] implications for how individuals [think] about potential strategies for dealing with the problem".}

They may operate in a completely different problem space; e.g., given the issue of underage pregnancy, one solution might address the problem of making contraception and abortion freely available to girls over 10, whilst another might address that of girls remaining sexually abstinent until marriage. Duncker’s "Monk Problem" (Appendix Ten) is an example of a solution existing in one problem space, but not in another.

Fleck (1935/1979, pp.111-115) also speaks of a significant difference between the “esoteric knowledge” of a “specialized expert” and the “exoteric knowledge” of a “general expert” in all fields of natural science.

Often, there’s an extremely long exoteric-esoteric continuum. Fleck distinguished between a “vivid, simplified and apodictic” report confirming the presence of diphtheria written for the “esoteric general practitioner” and the complex report written for the “esoteric expert team” in the diagnostic laboratory.
It is also highly likely that cultural differences in the way that both designers and users process various sensory inputs will produce additional difficulties.\textsuperscript{87}

**Literature as Thought Experiment**

In this section, I move on to examine approaches to thought experimentation that view it in terms of literature. Davenport (1983, p.282) defines thought experimentation as “testing hypotheses in the mind — logically rather than physically”, and believes that:

Thought [experimentation] is even more visible in social science than in natural science. Perhaps this is because it is notoriously more difficult in the social sciences to set up and repeat physical experiments. Like natural science, however, social science depends on a combination of thought experiment and physical experiment, for hypotheses must be considered in terms of their logical implications and connections as well as in terms of their producing satisfactory physical results. (p.283)

He argues (pp.283-284) that literature is a form of thought experiment:\textsuperscript{88}

\textsuperscript{86} People approach thought experiments with different intellectual resources and capacities; despite their academic refinement, “scientists bring ordinary human cognitive resources and limitations to bear on their scientific representational, reasoning and decision-making practices” (Nersessian, 1999, p.160).

Gendler (forthcoming, p.4) asks the essential question: “Is the thought experiment that I perform when I read Galileo’s text the same as the thought experiment Galileo performed when he wrote it?”

\textsuperscript{87} Maruyama (1980, pp.596) refers to the work of a Japanese researcher, Tsunoda, who found that: natural sounds such as wind, waves, animal cries, bird songs, and insect songs are processed primarily in the dominant brain hemisphere in Japanese individuals but in the non-dominant hemisphere in Europeans. Similarly, nonverbal human vocalizations such as laughing, snoring, sighing, and yawning were processed primarily in the dominant hemisphere in Japanese individuals and in the non-dominant hemisphere in Europeans. Furthermore, Tsunoda found that Japanese brought up in the Americas showed the same pattern as Europeans, while some Europeans brought up in Japan showed the same pattern as Japanese. Therefore these differences are cultural rather than genetic.


Poor children perceive the sizes of particular coins as being much larger ("accentuated") than rich children (Bruner and Goodman, 1947; Munroe, et al., 1969; Kirkland and Flanagan, 1979). Other research (Kirkland and Flanagan 1979) supports the hypothesis that, because (a) “persons in poor countries have greater subjective need than persons in wealthier countries”, and (b) “a country’s coinage allows institutional expression of the level of need”, “poor countries are more likely to circulate relatively large coins, and wealthy countries are more likely to circulate relatively small coins” (p.307).

It also seems axiomatic that, as well as real world objects being perceived very differently by different people, the significance of various entities will vary greatly from person to person.

\textsuperscript{88} In Mach’s view:

The thought experiment (Gedankenexperiment) [is] the common origin, or inner prototype, both of scientific experiment (primarily in physics) and of fiction (Romanschreiber, Dichter). It exists in the higher intellectual level (auf höherer intellektueller Stufe) and usually precedes its concrete realization either in science or in literature. The differentiation takes place only as a derivation from the Gedankenexperiment. In order to become a scientific experiment, it has to respect the given structure of the factual world and correspond to it [gute Abbilder der Tatsachen], whereas translated into fiction it is free to combine all elements and levels of reality, and to draw conclusions from them in a way that does not correspond to its given structure. (Moser, 1989, p.62)
The idea that literature can be viewed as thought experiment... is commonsense because it is easy to see that many kinds of stories are like thought experiments... these stories dramatize certain hypotheses about society and enable us to see the logical conceptual implications of these hypotheses, and so it seems commonsense to say that such stories are thought experiments.

And:

...if the concept of "the Muse" stands for the non-rational element of literature, then "thought experiment" stands for the partial rational control which both author and reader have over the powers of the muse. (p.306)

Moreover, using the thought experiment as "a model for composition" is extremely attractive because "it explains how writers can create things which, when they started, they did not conceive of" (p.297).

Yet, regardless of the extent to which general literature may be thought-experiment-

In particular, Mach (1926/1976, p.136) specifically had this to say on the extent to which both the scientific experiment and the work of fiction were the external, physical realizations of their inner, intellectual counterpart, the Gedankenexperiment:

Besides physical experiments there are others that are extensively used at a higher intellectual level, namely thought experiments. The planner, the builder of castles in the air, the novelist, the author of social and technological utopias is experimenting with thoughts; so, too, is the hardheaded merchant, the serious inventor and the enquirer. All of them imagine conditions, and connect with them their expectations and surmise of certain consequences: they gain a thought experience. However, while the former combine in phantasy certain conditions that never occur together in reality, or imagine these conditions accompanied by consequences that are not connected with them, the latter, whose ideas are good representations of the facts, will keep fairly close to reality in their thinking. Indeed, it is the more or less non-arbitrary representation of facts in our ideas that makes thought experiments possible. For we can find in memory details that we failed to notice when directly observing the facts. Just as in memory we may discover a trait that suddenly reveals a man's character hitherto misread, so memory offers new and so far unnoticed features of physical facts and helps us to new discoveries. (1926/1976, p.136)

He cites Aesop's Fables, Jesus' parables, Plato's dialogues, the speeches of Thucydides, Wells' The Sleeper Awakes, Orwell's 1984, and Asimov's Foundation Trilogy as examples.

He continues:

However, despite this commonsense observation, it is paradoxical to view literature as thought experiment because this clashes with the traditional art/science polarity. So paradoxical does it seem to view literature as thought experiment from which we can gain cognitive knowledge, that literary critics and theorists have frequently classed such books as I mention above as "didactic" and therefore "outside of" or "on the margins of" literature. This is done in order to indicate that the more a work of literature is like a thought experiment, the less it is like literature (because the less non-cognitive and non-rational it is). That such a crude distinction between didactic and literary works cannot stand is easily demonstrated. For social hypotheses are also tested in such clearly literary works as Aeschylus' Oresteia, Shakespeare's history plays, Dickens' Bleak House, Tolstoy's Anna Karenina, etc.

Davenport, pp.294-295:

Looking at all writing as one, fiction and nonfiction, it is evident that if writers are not going to write wholly at the dictation of the muse, but must write partly by the trial and error methods used by science, that writers must learn from their own works as they write them. They must learn, as scientists learn, by evaluating what they have tried, rejecting errors, and trying again. One method writers can use to do this learning is thought experiment...

[Literature as thought experiment] is not necessarily intended as a description of what writers think they are doing, or what they consciously do, but of what they actually do, whether consciously or unconsciously, in the whole process of composing. Thought experiment is an abstract model, in other words, of the sort of learning which, if it takes place, goes part way to explaining how composition is possible.

On this view, composing works by a bootstrap process: the writer [begins] writing, gets ideas from what he has written, writes more, sees conflicts in what he has written, either strikes out, or resolves (or ignores) the conflicts, and so on, getting a better idea of how his finished composition will look as he goes along.
like, all works of science fiction\footnote{In the context of the term science fiction, ‘science’ means ‘craft’ or ‘knowledge’ (i.e., as opposed to ‘belief’ or ‘opinion’); see, for example, Idier (2000, p.261).} are thought experiments.\footnote{Pratchett, Cohen & Stewart (1999, p.47):

Many diehard science purists look askance at science fiction. But such sniffiness is unfounded. After all, the central tenet of science fiction is nothing as crude as predicting the future, but dealing with the "worlds of if". At its best, a science-fiction story is a thought experiment. What would happen if this or that amazing ingredient turned out to be real? The good thing about "if" is that it does not imply belief. Mathematicians make perfectly logical deductions from false hypotheses: what, for example, if there were to be a counter-example to Fermat’s Last Theorem? It’s useful for them to think this way even if the result of such thinking turns out to be that there isn’t one. Similarly, science-fiction fans don’t believe the events in stories they read about aliens on distant planets any more than they believe that UFOs have landed on this one. What they do relish is putting their brains in gear to check the story is true to known science except for an added ingredient or two, that humanoid robot or this antigravity.}

Even in their earliest days, contemporaries Verne and Wells, who “fully deserve the [joint] credit they are given for moulding and popularizing science fiction” (Lambourne, Shallis and Shortland, 1990, p.14), displayed significant differences in their orientation: Wells was far “more deeply driven by science” and, by contrast with those of Verne, his works were far more concerned with science than with technology.

At first sight the careers of Verne and Wells seem strangely paradoxical. Verne, without any formal scientific training, wrote in a manner that emphasized factual accuracy and scientific credibility. Wells, who had received such a training, produced a kind of fiction that avoided scientific detail; he was more concerned with exploring ideas rather than setting them to work. (p.13)\footnote{Lambourne, Shallis and Shortland (p.14):

The crucial point perhaps is Wells’ clearer understanding of the distinction between science and technology. It is Wells who shows the greater concern with overall order and general system. It is Wells who always strives to integrate his ideas into large-scale schemes based on some underlying structure. It is surely Wells who has been more imbued with the scientific spirit of his age, the spirit of evolution, of thermodynamics and of all those other grand scientific theories that were being developed alongside his writings. Despite their differences, or perhaps because of them, Verne and Wells stand together as the Galileo and Newton of science fiction. They may not have been the first of their kind but viewed from the present they are certainly the foremost. They were both prolific generators of ideas; inventing themes that other writers would develop and extend. Together, they helped to determine the form and nature of science fiction and they inspired innumerable people, particularly young people, with the desire to read more and to create more of the same kind of literature.}

According to Lambourne, Shallis and Shortland, three streams of science fiction emerged in the 20\textsuperscript{th} Century:

1. "Imaginary Science" science fiction: involves a body of imagined science that is clearly “the natural outcome of normal scientific progress” (p.36):

Imaginary science is a term used to describe the network of scientific ideas, principles, laws, and hypotheses that authors imply as a background to their stories. Broadly speaking imaginary science is any kind of imagined ‘science’ that might plausibly be conceived as arising out of the continued progress of known science. (p.35)
(2) "Pseudoscience” science fiction: involves:
...activities such as ufology (the study of flying saucers), astrology and a whole variety of “psychic” sciences (p.37).95

(3) "Hard Science” science fiction: involves “factual science” and is:
[the sort of] science fiction in which the major impetus for the exploration which takes place is one of the so called hard, or physical, sciences, including chemistry, physics, biology, astronomy, geology and possibly mathematics, as well as the technology associated with, or growing out of, one of those sciences (pp.37-38).96

Idier (2000) identified five general classes:

(1) Anticipation Stories: which “usually takes place in the near future and is rather plausible, almost like an extrapolation of our present society” (p.259).

(a) Technological Foresight: which “takes place in some fathomable future and makes plausible guesses about technology and society” (p.260).97

(b) Social Foresight or Environmental Foresight: which “seeks to analyse the consequences of technological progress for mankind and its environment” (p.260).98


(3) Moral Stories: where “technology is merely used as a social or psychological factor” (p.260).99

95 Lambourne, Shallis and Shortland (p.37):
Those [fictional characters] who indulge in these activities may take them very seriously, they may even be scientists themselves and approach their subject in a rigorously scientific manner, but nonetheless, a sceptical scientific community will not accept them, even though it may not dismiss them out of hand...

Some good examples of the use of pseudoscience are to be found amongst the small group of novels in which magic is used as an extension to science...

In the context of fiction the use of magic with its hierarchy of witches warlocks and wizards, its spells and potions and its master practitioners such a Merlin and Gandalf, is just as valid as the use of quantum theory with its postgraduate students, research fellows and professors, its wave equations and uncertainty relations and its founding fathers Schrodinger, Heisenberg and Dirac.

96 They cite “Bainbridge, William Sims, Dimensions of Science Fiction (Harvard, 1986)”, as the source of this quotation, noting that Sims attributed it to L David Allen.

97 Idier (2000, p.260) comments:
Although plausible projected technology plays a central role [in these stories], philosophical questions always arise, as should be the case as technology develops. It should be noted that these stories generally assume continuous technological progress and many of them present technology as a positive development for mankind.

98 Idier (2000, p.261) comments:
Many social foresight books are quite pessimistic about technological evolution and the future of mankind in general. Problems like pollution, overpopulation, the dehumanisation of mankind, and regression due to a man-provoked catastrophe are often found in this kind of future exploration.

99 Idier (2000), p.260:
(4) **Horror Stories**: “although not usually regarded as hard-core science fiction, horror stories expose extreme deformations and transformations of the human body and mind as a direct or indirect result of uncontrolled technology” (p.260).

(5) **(Pseudo-)Scientific Visions**: “A marginal type of science fiction story [which gathers] different ideas or visions about future technologies and concurrent social changes” (p.260).

Lambourne, Shallis and Shortland, also describe how science\(^\text{100}\) may be involved in a science fiction story (pp.39-48):

1. Using science to provide a description of a real but relatively unfamiliar environment, the description being based on scientific information available at the time of writing.
2. Using science to provide a description of an imaginary environment that is as consistent as possible with established facts and principles.
3. Using a piece of scientific information as the basis of a puzzle.
4. Using science to justify the existence of devices or processes.
5. Using the scientific process itself or using a credible scientific setting for a story.
6. Using science peripherally, to justify a device or process, or to provide a generally ‘scientific’ background.

Dismissing ancient works such as Plato’s account of Atlantis,\(^\text{101}\) it’s widely accepted that the first modern example of a work of alternate history\(^\text{102,103}\) is Geoffroy-Chateau’s

---

\(^{100}\) In this context, ‘science’ has the far more technology-oriented “hard science” meaning of the “systematic field of study or body of knowledge that aims, through experiment, observation, and deduction, to produce reliable explanation of phenomena, with reference to the material and physical world” (Lafferty & Rowe, 1994, p.523); i.e., as distinct from the simple, far more basic “knowledge” meaning of the “science” in the term science fiction.

\(^{101}\) On the grounds that they were not generally thought of as fictional “speculations on alternative possibilities of history” Greenberg (1996, p.ix).

\(^{102}\) Alternate history (N.B., not alternative history) established itself as the preferred label among science fiction writers; though some refer to it as “allohistory” (allo-, ‘different’), in the sense of Lewis’ “possible worlds”. Historians generally speak of counterfactual history; though some refer to virtual history. The French call it “l’uchronie”.

\(^{103}\) It is a work of alternate history, rather than being a work of alternate history speculation; such as the counterfactual question examined by Edward Gibbon in his *Decline and Fall of the Roman Empire* (Chapter 52), where he speculates on the extent to which Islam might have spread throughout Western Europe if the French had not defeated the Muslim invaders at the Battle of Poitiers (732 CE):

A victorious line of march had been prolonged above a thousand miles from the rock of Gibraltar to the banks of the Loire; the repetition of an equal space would have carried the Saracens to the confines of Poland
Chapter Two: Approaches to Thought Experimentation

Napoleon and the Conquest of the World 1812-1823\textsuperscript{104} (the second, Renouvier’s _Uchronie\textsuperscript{105}_); an earlier (1833) novel by Disraeli\textsuperscript{106} (The Wondrous Tale of Alroy also known as _The Prince of the Captivity, a Wondrous Tale_) is sometimes mentioned as the first English alternate history;\textsuperscript{107} Hawthorne’s “P.’s Correspondence” (1845) is often cited as the first English alternate history short-story;\textsuperscript{108} the first English-language alternate history novel was Holford’s _Aristopia: A Romance-History of the New World_, published in 1895;\textsuperscript{109} the first English-language science fiction alternate history short-story was Hale’s (1881) “Hands Off”\textsuperscript{110} and the first “scholarly” English work of

and the Highlands of Scotland; the Rhine is not more impassable than the Nile or Euphrates, and the Arabian fleet might have sailed without a naval combat into the mouth of the Thames. Perhaps the interpretation of the Koran would now be taught in the schools of Oxford, and her pulpits might demonstrate to a circumcised people the sanctity and truth of the revelation of Mahomet.

\textsuperscript{104} Published, in French in 1836 (Napoléon et la conquête du monde, 1812-1823); it has Napoleon turning away from Moscow before the winter of 1812; and, without the losses he suffered in the winter of 1812, Napoleon went on to conquer the world (Greenberg, 1996, p.ix; Dozois & Schmidt.1998, p.xii).

\textsuperscript{105} Its title derived from _Utopia_ (not + place), _Uchronia_ (not + time) was published in French in 1857. Written from a view that small alterations in history could have significant consequences, it has the Roman Empire survive by keeping Christianity in check (Greenberg, 1996, p.ix).

\textsuperscript{106} Subsequently British Prime Minister.

\textsuperscript{107} It is not; it is a “secret” or “hidden history”; a narrative in which it is revealed that:

(a) hitherto secret events are now revealed (e.g., the plethora of “Now It Can Be Told” accounts that emerged after World War II);

(b) something we know about the past is revealed to be wrong (e.g., the famous BBC recording of Winston Churchill’s pre-Dunkirk speech of 4 June 1940 [“we shall fight on the beaches”, etc.] was not Churchill speaking in the House of Commons, but was an actor, Norman Shelley, recorded imitating Churchill’s voice in the BBC studios the following morning); or

(c) history, as understood, is the result of an intentional manipulation of historical fact by vested interests (e.g., the total suppression in Japanese school textbooks of any mention of the infamous 1937 “Rape of Nanking” by Japanese military personnel; see Chang, 1998).

The critical distinction between an alternate history and a secret history is that, in the case of a secret history, the front page of this morning’s Sydney Morning Herald would appear unchanged, whilst the front page would be very different in the case of an alternate history. (see http://www.uchronia.net).

Disraeli’s work speaks of a historical figure, a 12th Century Jewish messiah—who in fact, was destroyed after only a few minor successes—and has him founding a global empire based in Baghdad.

\textsuperscript{108} However, as it concerns a letter written by a deranged individual, there are some problems with this claim. In the story, the deranged protagonist, P., writes (in 1845) of a world where luminaries such as Robert Burns (died 1796), John Keats (died 1821), Napoleon Bonaparte (died 1821), Percy Bysshe Shelley (died 1822), Lord Byron (died 1824), Sir Walter Scott (died 1832) were all still alive, and where Charles Dickens (died 1870) and Henry Wadsworth Longfellow (died 1882) were already dead.

\textsuperscript{109} It tells of how the early settlers in Virginia discovered a pure gold reef around 1607CE, and set up a utopian colony (Aristopia). The novel relates the history of Aristopia, and how it used its wealth to strengthen its society and, eventually, take over the whole of North America some 300 years later. (Dozois & Schmidt.1998, p.xii; http://www.uchronia.net).

\textsuperscript{110} In his story, Hale travels back through time to see Joseph escaping from the slave-traders (and, thus, no longer becoming the adviser to Pharaoh).

According to Hale’s account, without Joseph to advise Pharaoh, the Egyptians were over-run by the Caananites, Judaism died out, Greek culture had no chance to flourish, Rome was destroyed, great chaos descended on the entire human race and, eventually, civilization was totally destroyed.
alternate history was Squire’s collection, *If It Had Happened Otherwise: Lapses Into Imaginary History*, first published in 1931.\(^{111}\)

The narratives of the alternative histories of science fiction (e.g., Dozois & Schmidt, 1998, and Greenberg, 1996) are identical in form and content with those of historians’ (e.g., Ferguson, 1999a; Cowley, 2000; and Cowley, 2001).

However, it is important to recognize the significant difference between the two: a science fiction account begins and ends with the narrative, whilst the account of an historian tends to engage in a lot of post-narrative discussion.

**The Scenario**

Of all the different sorts of hypothetical question that require subjunctive reasoning, one unique type demands special, isolated attention: the *scenario*.\(^{112}\)

Despite the wide range of meanings “scenario” has recently attracted,\(^{113}\) scenario is a term that is still used in a precise fashion by strategic planners; thus, it is important to clarify the similarities and differences between scenarios and thought experiments.

The scenario is a specific thinking tool (Kahn, 1962/1971, p.150), strongly focussed on decision-centred analyses, designed to aid the imagination,\(^{114}\) stimulate creativity, and

---

\(^{111}\) This collection, also published as *If: or, History Rewritten*, included papers such as “*If Don John of Austria had Married Mary Queen of Scots*” (G.K. Chesterton), (“*If Lee had not Won the Battle of Gettysburg*” (Winston Churchill), “*If Napoleon had Escaped to America*” (H.A.L. Fisher), “*If Byron had Become King of Greece*” (Harold Nicolson), “*If Booth had Missed Lincoln*” (Milton Waldman), and “*If the Emperor Frederick had not had Cancer*” (Emil Ludwig).

\(^{112}\) Rather than being a technical term (like syllogism), scenario is a jargon term (like the philosophical “incontinence”). It has a far more widespread, better known, and far more universally understood meaning than the narrow meaning a small group of technical people hope to denote with their specific use of the term. Kahn (Kahn, et al., 1978, p.600) explains the reason for the choice:

> As near as I can tell the term “scenario” was first used in this context in a group I worked with at the Rand Corporation. We deliberately chose the word to deglamorize the concept — that is in writing the scenarios for various situations we kept saying, “Remember, it’s only a scenario”, the kind of thing that is produced by Hollywood writers — whether hacks or geniuses. There is no a priori concept that a scenario should be taken seriously or that it is intended to reflect aspects of the real world. Some scenarios do; others do not. Scenarios are simply a more or less imaginative sequence of events that are put together so that each event forms a context for the other events and so that there is some continuity over time in the “narrative”.

Stout (1998, p.29):

> If we take the word scenario back to its original theatrical context, what this implies is a three-stage composition: first, a stage setting before the characters have appeared; next, the action of the play; third, the possible resolutions — scenarios redrawn and more constrained, with the benefit of knowledge of the intended action. Once the process has begun, it should continue in a building wave, future scenarios being redrawn as a consequence of actions influenced by the scenarios of earlier rounds.

\(^{113}\) For example, the popular expression *worst-case scenario* (“the worst possible outcome that could ever happen given a certain set of initial circumstances”); a notion no doubt suggested by the technical concept of a *worst-case analysis*.

\(^{114}\) Kahn’s account of the structure and application of scenarios appears in “*Some Strange Aids to Thought*” in his *Thinking About the Unthinkable* (1962/1971, pp.133-185):
reveal “novel possibilities” which would otherwise go unnoticed. 115 Scenarios were first used by Kahn’s USAF defence-analyst research team at the RAND Corporation in the late 1940s/early 1950s. Later on, at the Hudson Institute, they adapted their narrower “war-planning scenarios” (which were “alternative paths resulting in alternative outcomes”) to the wider “business applications” of identifying “alternative outcomes of trends and events by a target year regardless of the precise sequence of events”.

Driven by the view that “the business environment was uncertain and could evolve in totally different ways”, they provided their users with “descriptions of future conditions rather than accounts of how events might unfold” (Millett, 2003, p.17). 116

To Vazsonyi (1982), the true scenario is comprised of a number of thought experiments amalgamated into a single activity, and involves a hypothetical sequence of events that specifically focus the user’s attention on:

(a) revealing the step-by-step causal processes involved in the evolution of the hypothetical situation, and
(b) identifying potential decision points (for preventing, diverting or facilitating the process) for each of the scenario’s participants, at each step. 117

Godet (2000, p.8) identifies two scenario approaches, prospective 118 and strategic, 119 which imply five questions for the individual (or organization) concerned: “What can and

---

115 Kahn (1962/1971, pp.150-151) stressed that scenarios:
(a) counteract the natural “tendency to indulge in "carry-over" thinking even when it is clear that 1965 [from the perspective of 1962] cannot be the same as 1945 or even 1960”.
(b) “dramatize and illustrate possibilities that might otherwise be overlooked”, and
(c) are valuable because they are “one way to force oneself and others to plunge into the unfamiliar and rapidly changing world of the present and the future”.

116 Stout (1998), p.28:
Their purpose is to help us imagine possible futures, so that we can plan and act in ways that will have very long-term consequences, in spite of ignorance about the future state of the world — and even to turn universal uncertainty to individual advantage.

117 Vazsonyi (1982) p.36:
The scenario is particularly suited to dealing with events taken together, integrating several aspects of a situation. The analyst can get a feeling for events and the branching points dependent on critical choices. Situations and decisions can be explored systematically, and the scenario itself can be used as a context for discussion of various assumed possibilities.

118 Which are designed to anticipate “possible and desirable changes” (p.8).

119 Which are designed to work out and assess “possible strategic choices so as to be prepared for expected changes (preactivity) or provoke desirable changes (proactivity)” (p.8).
might happen?” “What can I do?” “What am I going to do?” “How am I going to do it?” and, the most important question of all, “Who am I?”.

Whilst scenarios are not involved with the prediction of the future outcome of a single set of events, they specifically examine multiple alternative futures in an explicit, imaginative and structured way.

---

**120** Although “the purpose of scenario planning is not to predict the future; but rather, to show how different forces can manipulate the future in different directions” (IDEA, 2003), and “scenarios are not forecasts [because] if we could make a confident forecast, we should certainly not need scenarios” (Stout, 1998, p.29), they are driven by a theoretical position that the future is predictable; for example:

If a prediction is a definitive statement of what the future will be, then scenarios are heuristic statements that explore the plausibilities of what might be. (Staley, 2002, p.78).

A scenario is not a future reality but a way of foreseeing the future, thereby throwing light on the present in terms of all possible and desirable futures. (Godet, 2000, p.18).

In all cases, a scenario is far more useful for furnishing insights into prospective patterns and relationships than it is for exact predictions. (Kahn, et al., 1978, p.76).

...a scenario is not an end in itself; it only becomes meaningful when its results and implications are embodied in real action. (Godet, 2000, pp.19-20).

---

van der Heijden (1996, p.103), emphasizing that “forecasting requires that we first decide what we want to forecast” makes a distinction between forecasting and scenarios (which provide a far more “conceptual description of a future, based on cause and effect”):

Forecasting is done by experts, away from where the decisions are being made. That means that the decision maker who receives the result of the forecasting activity does not know the underlying thinking process and the uncertainties that have been taken into account to produce the prediction. The danger to the decision maker derives from this disconnection. If (s)he decides to use the forecast he does not know what risk assumptions enter his decision process. He is no longer in a position to see the different possibilities as they could unfold. He has shifted his responsibility on to the expert, who is not accountable. Uncertainty falls between two stools, the linear process runs out to an algorithmically inevitable outcome and thinking tends to stop. The decision process lacks basic information and therefore is essentially a chance event.

---

**121** According to Wollenberg, Edmunds and Buck (2000, pp.68-69) scenarios can be employed equally well to (a) identify decision options and (b) assess decision options. The table below, based on that appearing at Wollenberg, et al. p.69, has been amended on the basis of Wollenberg (28 August 2002):

<table>
<thead>
<tr>
<th>DIMENSIONS</th>
<th>PURPOSE OF SCENARIO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IDENTIFY OPTION</strong></td>
<td><strong>ASSESS OPTION</strong></td>
</tr>
<tr>
<td>Aim of scenario analysis</td>
<td>Identification of options based on different possible futures</td>
</tr>
<tr>
<td>Scenario application</td>
<td>Source of information for indicating options</td>
</tr>
<tr>
<td>Decision options</td>
<td>Are the result of scenario analysis and are variable</td>
</tr>
</tbody>
</table>

---

van Notten et al. (2003) offer a complex typology of 14 different characteristics aggregated in three themes: (a) project goal, (b) process design, and (c) scenario content.


---

**122** Kahn (1962/1971, p.152) elaborates further:

The use of scenarios has been criticized, both as being paranoid and schizophrenic. In the first case, the criticism is sometimes that only the paranoid personality, unjustifiably distrustful and suspicious, could conceive of such plots. This criticism hardly seems relevant, or, if relevant, justified. The analyst is, of course, interested in what devilish means others might contrive to destroy him; he is also interested in what they
In general, scenarios are one of four types:

<table>
<thead>
<tr>
<th>EXPLORATORY</th>
<th>ANTICIPATORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>DESCRIPTIVE</td>
<td>Given the causes, what are the effects?</td>
</tr>
<tr>
<td></td>
<td>Given the effects, what are the causes?</td>
</tr>
<tr>
<td>NORMATIVE</td>
<td>Given the means, what goals can be reached?</td>
</tr>
<tr>
<td></td>
<td>Given the goals, what means can be used?</td>
</tr>
</tbody>
</table>

Godet and Roubelat (1996, p.167) make a further distinction between: (a) possible scenarios (“everything that can be imagined”), (b) desirable scenarios, a sub-set of possible scenarios, and (c) realizable scenarios (“all that is possible, taking account of constraints”), a sub-set of desirable scenarios.

Kahn (op.cit., p.151-152) claims six additional benefits for scenario use:

1. they call the user’s attention to unseen possibilities;

might not do. To the extent that such criticism is justified, it pertains to the plausibility of the particular scenario and not to the methodology.

The second “diagnosis” may be more to the point, the criticism in this case being that fictional plots may be so divorced from reality as to be not only useless but misleading and therefore dangerous. However, one must remember that the scenario is not used as a predictive device. The analyst is often dealing with the unknown and unknowable future. It is hard to see how there can be a sure divorce from a reality which does not yet exist. Imagination has always been one of the principal means for dealing in various ways with the future, and the scenario is simply one of many devices useful in stimulating and disciplining the imagination. To the extent that scenarios may be divorced from reality, this again seems a criticism more apt for particular scenarios than for the methodology. If a scenario is to be plausible, it must, of course, relate at the outset to some reasonable version of the present, and must throughout relate rationally to the way people could behave though it is important not to limit oneself to the most plausible, conventional or probable situations and behavior.


They serve to call attention, sometimes dramatically and persuasively, to the large range of possibilities that must be considered in strategic analysis, some of which may escape notice if an analysis is done using conventional analytic techniques.

According to Schoemaker (1993, p.194):

In terms of inquiring systems, scenarios are Hegelian in their underlying philosophical premise. The scenario method courts contradiction and paradox. In contrast, the traditional approaches of decision analysis and forecasting tend to be Leibnizian. They seek a single truth and representation of reality.


The method of trial and error is not, of course, simply identical with the scientific or critical approach — with the method of conjecture and refutation. The method of trial and error is applied not only by Einstein but, in a more dogmatic fashion, by the amoeba also. The difference lies not so much in the trials as in a critical and constructive attitude towards errors; errors which the scientist consciously and cautiously tries to uncover in order to refute his theories with searching arguments, including appeals to the most severe experimental tests which his theories and his ingenuity permit him to design.

The critical attitude might be described as the result of a conscious attempt to make our theories, our conjectures, suffer in our stead in the struggle for the survival of the fittest. It gives us a chance to survive the elimination of an inadequate hypothesis — when a more dogmatic attitude would eliminate it by eliminating us. (There is a touching story of an Indian community which disappeared because of its belief in the holiness of life, including that of tigers.) We thus obtain the fittest theory within our reach by the elimination of those which are less fit. (By 'fitness' I do not mean merely 'usefulness' but truth.) I do not think that this procedure is irrational or in need of any further rational justification.

Davis (2002, p.4), describes one instance of the value of scenario use to Shell Petroleum:

The first real use of scenarios [by Shell] was in 1972. Of the six developed, one scenario suggested that disruptions to oil supply could result in a sharp rise in prices. In a period of continuing expansion, with a long history of stable prices, this was contrary to prevailing expectations.
(2) a user is forced “to deal with details and dynamics which he might more easily avoid treating if he restricted himself to abstract considerations”;\textsuperscript{124}

(3) they illuminate the \textit{interaction} of factors (psychological, social, political, military, etc.) which may fall outside the scope of traditional analysis;

(4) they examine these factors in a way that “permits the comprehension of many interacting elements at once”;

(5) With their often over-simplified “thinness”, they can illustrate questions, issues or principles “which would be ignored or lost if one insisted on taking examples only from the complex and controversial real world”; and

(6) They encourage users to consider alternative possible outcomes of real past and present events.\textsuperscript{125}

A year later it was more than a scenario — the oil shocks were the most significant discontinuity our industry had faced. But Shell was prepared: \textit{by building scenarios addressing global economic and oil issues, we were already thinking about what most others in the industry had deemed unthinkable}. (emphasis added)

\textsuperscript{124} Millett (1988, p.66) comments that “as rigorous as scenarios can be, they do not stifle creativity. On the contrary, scenarios allow consideration of low probability, even wildly imaginative, outcomes and disruptive events”.

\textsuperscript{125} Kahn (Kahn, et al., 1978, p.489) describes seven \textit{motivations} for creating medium- and long-term scenarios and “synthesizing them with current and short-term planning and decision making”, all of which, in addition to the benefits already listed above, are equally applicable to thought experiments:

(a) They “supply understanding, context, early warning, and increased sensitivity to current and near-term issues and events”.

(b) “Normative projections describe and elaborate, often dramatically, where the nation wants to go and how it hopes to get there”.

(c) “Disaster scenarios do the exact opposite, i.e., project where the nation does not want to go, perhaps in an alarming fashion”; which means that they are “effective in provoking discussion, exhortation, and polemics by making certain fears vivid and persuasive”.

(d) “All of the above can provide inspiration, argumentation, and ammunition for bureaucratic and political action and also furnish useful language, examples, metaphors, analogues, and other references and illustrations for use in evaluation and discussion”.

(e) “They can also be used for reflecting on and reinforcing a country’s basic philosophy of modernization and of its future”.

(f) “Scenarios often provide an extraordinarily useful context for serious discussion of issues that would otherwise be too politically or emotionally disruptive to be discussed seriously and openly”.

(g) “Fun, scholarship, or entertainment”.
Idier (2000) has identified valuable distinctions between the far more “possible” class of scenario techniques and the far less “possible” works of science fiction, when considered as pieces of “foresight literature”:

---

126 Which he describes as being “tools for helping one to think about the consequences of present decisions and to prepare for a changing environment” (p.259) which do not involve just “one global story but several minor stories, thus allowing the broadest possible spectrum of uncertain yet plausible events to be explored” (p.258), and which are “descriptive, consistent, imaginative, structured, and not least, should lead to action” (p.258, emphasis added).

127 He also remarks that (p.259):

[Whilst the] scenario technique is a tool for helping one to think about the consequences of present decisions and to prepare for a changing environment... [a] science fiction (SF) story... can go further, not only because of its environment or breadth of plot but because an SF author is not afraid to break norms and established paradigms or cultures in order to explore human motivations and creations. In addition, characters are used in SF stories so that readers can project themselves into the stories. Such emotional involvement is absent from scenarios. It is certainly more difficult to control but probably more rewarding with respect to the future of actual human organisations and societies.

Therefore, a broader definition of science fiction can be proposed that does not refer just to science in the usual formal sense. Science fiction could also be a subject-centered story that includes emotional involvement and depicts future or alternative technological events where technology does not mean “technical feasibility” but instead a broader use of artificial items to sustain an activity or attain particular goals.
### Table Two: Science Fiction Pieces vs. Scenario Techniques
[adapted from Idier, 2000, p.263]

<table>
<thead>
<tr>
<th></th>
<th>Science Fiction Story (a piece of literature)</th>
<th>Scenario Technique (a piece of social science)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>VALUES</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High symbolic value</td>
<td>Technical value</td>
<td></td>
</tr>
<tr>
<td>Irrational components in construction</td>
<td>Normative rationality</td>
<td></td>
</tr>
<tr>
<td>Moral components</td>
<td>No moral components</td>
<td></td>
</tr>
<tr>
<td>Eros in science and technology</td>
<td>No eroticization</td>
<td></td>
</tr>
<tr>
<td>Expected ideological freedom</td>
<td>Fulfill a contract</td>
<td></td>
</tr>
<tr>
<td>May carry a dream</td>
<td></td>
<td>Often lobby/industry influenced</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CONSTRUCTION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No formal constraints</td>
<td>Constraints (normative scenarios), formal writing schemes</td>
<td></td>
</tr>
<tr>
<td>Wild impossibilities allowed</td>
<td>Plausibility required</td>
<td></td>
</tr>
<tr>
<td>Individual level prevails</td>
<td>Corporate or social level</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Several different short formalized stories</td>
<td></td>
</tr>
<tr>
<td>Focus of small groups of people</td>
<td>Collegial construction. Use of panels of experts, economical indicators</td>
<td></td>
</tr>
<tr>
<td>Should involve a dramatic tension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fictional character’s intelligence is lower than or equal to the writer’s</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TIME</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognizable present or past elements</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Often a straightforward extrapolation (but “atemporal” or “achronical” stories are allowed)</td>
<td>Only “near” future extrapolation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PUBLIC</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional involvement (projection)</td>
<td>No emotional involvement</td>
<td></td>
</tr>
<tr>
<td>Large public, mostly non-executive</td>
<td>Small public, executive</td>
<td></td>
</tr>
<tr>
<td>Early experience with technology</td>
<td>Mature experience with technology</td>
<td></td>
</tr>
</tbody>
</table>

In this chapter, I have examined the origins and applications of the term “thought experiment” and, in particular, found that, whilst thought experiments can be usefully “positioned” within the domain of hypothetical questions that employ subjunctive reasoning, they are not unique to either philosophers or scientists;[^128] and, in fact, as will be discussed later at some length,[^129] thought experiments simply formalize a number of

[^128]: Or, even, unique to chemists and physicists amongst those scientists; or, even, unique to mentally simulated “real” physical experiments amongst their possible mental simulations.

[^129]: For example, see Chapter Four.
the patterns of thinking and questioning that pervade all aspects of a normal, healthy mental life.

Also, having noticed significant differences between the applications of various thought experimental activities (and, also, the precise reasons for their application), it is essential to establish a set of precise descriptive terms that will allow important and significant distinctions to be made between the various activities I have identified.

Chapter Three begins this process with a discussion of analogy, and source-to-target mapping.
Chapter Three
Comparative Cognitive Processes

In this chapter I examine what I have called, for convenience, comparative cognitive processes. These processes are analogy, metaphor and models. It is important that we understand these as comparative cognitive processes, because they play a central role in thought experimentation. The fruits of this examination will become evident in later chapters.

Analogy

Often, progress in science begins with finding the right analogy. (Gopnik, 1996, p.485)

Analogy plays a significant rôle in human cognitive development, knowledge acquisition, creativity, general reasoning, problem solving, and decision making.¹

Involving “the transfer of relational information” from a known system² (the source) to a relatively unknown system (the target) by virtue of “there being some sort of [perceived] similarity between the source domain and the target domain” (Vosniadou and Ortony, 1989, p.6), analogies consistently take the form A : B :: C : D.³

The possible relationships between analogical sources and their targets lie along a continuum ranging from those from the same conceptual domain (“near” analogies),⁴ through those from close conceptual domains,⁵ to those from conceptually remote

---

¹ According to Holland, Holyoak, Nisbett, and Thagard (1986, p.297), there are four solution-oriented features essential to the ideal description of an analogy:
   (1) the goal: which is the “reason” for it;
   (2) the resources: which “enable” it;
   (3) the constraints: which “prevent” alternative pathways being followed; and
   (4) the outcome: which is the “result” of using the analogy.

² Barbour (1974, p.30) prefers to call it “the familiar system”; with ‘familiar’ meaning “better understood [familiar] rather than everyday [familiar]”.

³ Pronounced “A is to B in precisely the same way that C is to D”.

The convention is that the “C : D” slot is always filled by the analogy’s source, and the “A : B” slot by the analogy’s target, originated in Aristotle’s discussion on “distributive justice” in The Nicomachean Ethics (1980, V.3.A, pp.112-114).

The analogical problem solver “usually proceeds by finding the relationships between A and B and A and C [and] then, a mapping is made between these relationships to produce D” (Lee, 2002, p.6). Often, in everyday discourse, the “C : D” group is mentioned first; as in, for example, “It is easier for a camel to go through the eye of a needle, than for a rich man to enter into the kingdom of God.” (Matthew: XIX.24)

⁴ Which show “a striking similarity between the familiar and unfamiliar environments”; e.g., a racing bicycle and a recreational bicycle.

⁵ For example: Rugby Union Football and Rugby League Football; or baseball and softball.
domains ("far" analogies),\(^6\) where "the similarity between the known and unknown is muted and less obvious" (Chidambaram, 1999).

In the same way that the most productive metaphors communicate by virtue of achieving the best level of interactive mapping, driven by the optimum mix of similarity and difference between source and target,\(^7\) the best thought experiment designers are well aware that too much, or too little, similarity between the sources and targets of the selected analogies may not be understood in the way intended. (Oswick, Keenoy & Grant, 2002, p.298).\(^8\)

Vosniadou and Ortony (1989, p.7) identified two styles of analogy:

1. **Within-domain (or literal) analogy:** "the analogically related items are drawn from the same domain, or at least from conceptually very close domains".\(^9\)

2. **Between-domain (or metaphorical) analogy:** "the analogically related items are

---

\(^6\) For example: Australian Rules Football and chess; or Lacrosse and darts.

\(^7\) Yet, "everything is similar to everything else in an infinite number of ways, and everything is also dissimilar to everything else in an infinite number of ways" (Brewer, 1996, p.932); it just depends upon your perspective.

There are significant and insignificant differences: "after all, any two things have indefinitely many common properties, most of which are likely to be trivial; for example, this [page] and your shoes are both over a million miles from Venus". (Holyoak & Thagard, 1995, pp.22)

Staley (2002, p.87) warns that "analogy must be treated with caution": rather than being "law-like regularities, they are comparisons of different situations"; and there are dangers "that those who see only the similarities in two situations might be blindsided by the differences in those situations". Describing analogical reasoning as "[an activity] which draws conclusions about the nature of a single unknown thing from information about known thing or things which it to some extent resembles", Quinton (2000) strongly cautions against the extensive use of analogy because "[analogy] is a form of reasoning that is peculiarly liable to yield false conclusions from true premises".

\(^8\) The nearer the source and target domains are to one another, the more likely it is that the analogy will be drawn between structures; the greater the conceptual remoteness of the target and source domain from one another, the more likely it is that the analogy will be drawn between relationships.


When designing a new freeway system, designers could draw a near analogy from a closely related base domain, such as a freeway system in another city. However, they could also draw a far analogy from a more distant base domain, such as the human circulatory system. This distinction is important because the type of information mapped and transferred from near analogies is different from that transferred from far analogies.

When near analogies are drawn, both surface-level attributes (e.g., roads) and the relations among the attributes (e.g., the flow of cars through the freeway system) are mapped and transferred. However, when far analogies are drawn, few surface-level attributes can be mapped, leaving the mappings to occur between common relations. Because near analogies often fall into the category of "literal similarity", they represent smaller conceptual distances between the old and the new ideas and may be viewed as less original. Far analogies are considered the main drivers of truly innovative thought, serving as the basis for "mental leaps".

However, far analogies require the identification of similarities in the relational (not surface) structure between the base and the target domains, and when the target and base domains share few surface similarities (e.g., freeway systems and the human circulatory system), access is often difficult. Nonetheless, new concepts based on far analogies are likely to be more creative than those based on near analogies.

\(^9\) Vosniadou (1989, p.416) cites the analogy between a ceramic mug and a Styrofoam cup as an example of a within-domain analogy. Within-domain analogies "tend to focus on problem-solving tasks in which the problems to be solved are all of the same general kind (e.g., problems in computer programming or statistics)" (Vosniadou & Ortony, 1989, p.11).
drawn from conceptually different or remote domains”; they share this “domain incongruence” with metaphor (“the items juxtaposed in a metaphor must [also] belong to different conceptual domains”) and, because of this “domain incongruence”, most “between-domain analogies can be turned into metaphors” (Vosniadou, 1989, p.417).

Vosniadou goes on to argue that, whilst metaphors can easily be constructed from between-domain analogies (e.g., “Atoms are solar systems”, “Inflation is a disease” and “Illiteracy is a prison”, etc.), they can not be constructed from within-domain analogies such as a puppy : a dog :: a kitten : a cat, because a statement such as “A puppy is a kitten” is just nonsense.

According to Holland, et al. (1986, p.287), “the recognition and exploitation of some significant similarity between the target and the source” will generate new regularities in the target domain; and, yet, “by selectively highlighting abstract properties that they share”, they can also make source and target seem far more alike than they initially appeared.

Analogies enable scientists to discover, justify, establish, and efficiently organize their facts, “make new information more concrete and easier to imagine” (Duit, 1991, p.652), represent and communicate their observations and discoveries in a systematic way, recognize systematic regularities amongst their established facts,

---

10 Vosniadou (1989, p.415) cites the analogy between the solar system and the atom as an example of a between-domain analogy.

11 According to Klahr and Simon (1999, p.533), “analogy can be viewed as one method for changing the given problem space to another that is more effective.

12 Vosniadou and Ortony (1989, pp.8-9) attribute this gradual increase to increases in the extent to which the apparent similarity between the two concepts stands out from the perceptual background (i.e., there is an increase in their salience).

13 The extract from von Platen (1962) at Appendix Three provides an example how the unexpected emergence of an analogy can lead to the solution of an important scientific problem.

14 In order to communicate the significance of their discoveries, the analogies may artificially amplify the importance of core concepts and artificially reduce the importance of peripheral concepts.

15 Singer (2004, p.35) discusses recent discoveries of the function of a second type of receptor (the metabotropic receptor) for the neurotransmitter glutamate in the brain. (Previously only one type, the ionotopic receptor, was thought to be involved.) Having described ionotopic receptors, “which function
notice relationships between associated phenomena, and coherently represent all of these in an imaginative and illuminating way (Smith, 2002, p.244).

The coherence of an analogy depends upon the extent to which:

1. the analogy efficiently provides the "maximum information with the least cognitive effort" Rosch (1978, p.28); 

2. the mapping of correspondences from source to target is symmetrical;

like the lock on a gate" (the "gate" being opened by the binding of the glutamate molecules), she goes on to distinguish between functions of the two receptors using an analogy to a light switch:

Metabotropic receptors have a slower, more subtle effect than ionotropic receptors. If [ionotropic receptors] are an on-off switch, then metabotropic receptors are a dimmer, turning the strength of the signals up or down.

The fMRI study of neuroanatomical correlates of analogical thinking (Luo, et al., 2003) has demonstrated a significant distinction between analogical and deductive reasoning.

Clement’s work with experts’ spontaneous use of analogies to solve problems found “this naturally observed analogical reasoning process is more complex than the standard theories assume” (Millman and Smith, 1997, p.160-161):

First, some of the significant analogies used by experts to aid their thinking are novel (and hence constructed by them in the course of their problem solving) rather than familiar (and hence directly retrieved).

Second, experts often need to engage in considerable work to develop their understanding of both source and target before attempting any mapping.

Third, the processes of evaluating the appropriateness of an analogy are more complex than the mapping of preanalyzed symbolic features. Bridging analogies are sometimes created to facilitate evaluation.

At other times, experts engage in dynamic imagistic simulations on the source and target (i.e., they visualize some aspects of the source and target and then run a mental simulation that uses that image and some kind of supporting net of interpretations and purposes of the image in order to see what will happen).

Finally, experts frequently engage in cycles of analogical reasoning in which they generate, critique, and modify a series of analogies. One of the important hallmarks of the expert is persistence in this complex effort.

The speed and efficiency with which the source is identified, retrieved and mapped onto the target is critical; rather like Saint Jerome’s ideal of translating sense for sense, not word for word.

Obviously, the more familiar the source, the smaller the cognitive load upon the user, and the easier and more effective the analogical transfer; and, by contrast, the greater the cognitive load, the less there is left to scrutinise the mapping of the source onto the target.

According to the Oxford English Dictionary, mapping was first used in its cartographic sense in 1727. In 1900, the meaning extended to mathematics to denote "a correspondence by which each element of a given set has associated with it one element ([occasionally], one or more elements) of a second set”. In 1955, Chomsky further extended its meaning into linguistics to denote "the (degree of) correspondence between associated elements of different types or at different levels within a linguistic process or scheme". Bannet (1997) stresses that the sense of "what is here [being] projected onto what is there" — that is embedded within the representation of "analogy [as a process that] "maps" elements of a "source" domain onto a "target domain"" — is driven by the rigid, structured, mechanical thinking of "those who are currently trying to formalize analogy for computer intelligence" (p.658). Analogies often operate at more than one level, and the mapping of source onto target must precisely match for levels of organization, levels of explanation, levels of representation, levels of description, and levels of analysis.

In relation to "mapping" Elaismith and Thagard (2001, p.248) state that:

1) "a mapping is one to one if each element in the source corresponds to one element in the target",

2) "a mapping is structurally consistent if the objects of mapped relations are also mapped", and

3) "if an analogy has both of these properties, it is considered isomorphic".

Roth (2001, pp.55-57), uses a zoom lens metaphor: "by changing focus and by zooming, phenomena pertaining to different fields of attention become visible and are of different grain sizes and time scales"; with zooming allowing "analyses at multiple levels", and "different foci of analysis" widening or narrowing "the analytic frame" (or "focal area") and producing "associated changes in spatial and temporal scales". 
(3) the analogy, thus mapped, exclusively transfers the intended information from source to target;\(^{21}\) and

(4) the analogy serves its purpose.\(^{22}\)

However, there is often a problem with analogies. They tend to produce unintended mappings,\(^{23}\) essentially due to the fact that, ultimately, designers have no control over the way their users process the source of their analogy or map the correspondences from their source to its target.\(^{24}\)

---

\(^{20}\) Hayakawa (1949) constantly stressed that “the map is not the territory” (p.299) and, he said, if there’s ever a disparity between a map and the territory, then it was the map that must be abandoned:

Scientists have special ways of talking about the phenomena they deal with, special “maps” to describe the “territories” with which they are concerned. On the basis of these maps, they make predictions; when things turn out as predicted, they regard their maps as “true”. If things do not turn out as predicted, however, they discard their maps and make new ones; that is, they act on new sets of hypotheses that suggest new courses of action. Again they check their map with the territory. If the new one does not check, they cheerfully discard it and make still more hypotheses, until they find some that work. These they regard as “true”, but “true for the time being only”. When, later on, they find new situations in which they do not work, they are again ready to discard them, to re-examine the extensional world, and to make still more new maps that again suggest new courses of action. (p.287)

---

\(^{21}\) Analogy can also provide a means through which taboo subjects or otherwise dangerous issues may be raised with impunity. In a famous exchange (II Samuel, Chapter 12), the prophet Nathan used the analogy of a rich man stealing a poor man’s sheep to accuse the Israelite king, David, of serious breaches of at least three of the core moral precepts of his State’s religious code (viz., “Thou shalt not kill”, “Thou shalt not commit adultery”, and “Thou shalt not covet thy neighbour’s wife”).

---

\(^{22}\) Whilst some analogies are intended to be simply descriptive, others are intended to be explanatory. Clark (2001, p.122) makes an important distinction between the simple “description of a pattern of events and a real explanation of why the events unfold as they do”.

---

\(^{23}\) Where the user is less informed than the designer: a source comprised of Adolf Hitler, Ayatollah Khoumeini, and Charles Manson, intended to communicate “strict vegetarians”, would, in most users, generate an alternative notion of “infamous, deranged, psychopathic murderers”. Even the inclusion of another well-known vegetarian, the mean and vicious professional wrestler, Wladek (Walter) “Killer” Kowalski (of notorious “claw hold” fame) would be unlikely to alter this fact to any great extent.

Where the user is better informed than the designer: the designer might offer Bruno Bettelheim as a source to communicate “well-respected and innovative child psychologist”; however, in the well-informed user this would also map onto “unethical, serial child molester”.

---

\(^{24}\) Which is, of course, a user-centred, rather than a designer-directed activity.

An infamous example of incorrect (but non-analogical) mapping is when “NASA lost its $125 million Mars Climate Orbiter spacecraft as a result of a mistake that would shame a first-year physics student — failing to convert Imperial units to metric” (Hecht, 1999, p.6). The press release from NASA was quite explicit:
This is also evident in education:

Analogies are often used as a means for aiding learners in understanding an unfamiliar domain. However, the danger of using analogies is that unwarranted assumptions might be made about the target domain based on the source domain. Indeed, models of teaching with analogies suggest that teachers should indicate where the analogy breaks down in order to prevent students from drawing incorrect conclusions about the target domain. (Yanowitz, 2001, p.375)

From their study of medical students’ (inappropriate) extension of analogies encountered in earlier basic learning situations into complex situations, Spiro, Feltovich, Coulson & Anderson (1989) strongly recommended the use of multiple analogies, taken from different sources, to “capture correctly, yet manageably, the complexity of different concepts... [without inviting the] problems introduced by any single analogy, [and] without cancelling its beneficial effects” (p.499), and do this in a way that mitigated cognitive load.

At liftoff, the spacecraft will weigh 1,418 pounds (3,120 kilograms). It is 7.6 feet (21 meters) deep, and 5.4 feet (18 meters) wide. Power is provided by a single large solar array which is 18.6 feet (61 meters) long and 6.8 feet (22 meters) across. After cruising in space for 286 days, the spacecraft will be captured in an elliptical orbit around Mars. (NASA Press Release (September 11, 1998); taken from http://www.astronomynow.com/breaking/980914marsorb/index.html on 26 May 2003)

It is immediately obvious that NASA is seriously wrong: Instead of applying the Imperial-to-Metric conversion to Imperial measurement, or the Metric-to-Imperial conversion to Metric measurement, it has applied the Metric-to-Imperial conversion to the Imperial measurements, asserting that there are 2.2 pounds to the kilogram, and 3.3 metres to the foot.

The fact that some components were supplied by manufacturers working with metric specifications, and other components by manufacturers who worked only with Imperial measures, meant that this sort of error caused NASA’s precise specifications (which were accurately matched by the independent manufacturers) to create situations where installed components from different manufacturers were mutually incompatible.

Also, the data-processing software installed (that had been produced by “Metric manufacturers”) simply assumed that the digits in the data it was receiving from sensors (that had been produced by “Imperial manufacturers”) represented metric units, rather than Imperial units; as a result, the probe finished up “100 kilometres too close to Mars when it tried to enter orbit” (Hecht, p.6).

Duit, Roth, Komorek and Wilbers (2001) found that whilst analogies could be “powerful tools for guiding students from their pre-instructional conceptions towards science concepts”, the very same analogies could also “deeply mislead students’ learning processes” — and, consequently, “conceptual change... may be both supported and hampered by [precisely] the same analogy”.

Spiro, Feltovich, Coulson & Anderson (1989), p.498:

[The] simple analogies that help novices to gain a preliminary grasp of difficult, complex concepts may later become serious impediments to fuller and more correct understandings. Specifically, although simple analogies rarely if ever form the basis for a full understanding of a newly encountered concept, there is nevertheless a powerful tendency for learners to continue to limit their understanding to just those aspects of the new concept covered by its mapping from the old one. Analogies seduce learners into reducing complex concepts to a simpler and more familiar analogical core.

Spiro, Feltovich, Coulson & Anderson (1989, pp.527-528), stress that multiple analogies are not a set of "sequential" analogies (where "successive stages in a process are each represented by [a different] analogy"), but a synergistic aggregate that is processed as a unit:

The fundamental difference from our approach should be clear: Our multiple analogies are all applied to the same stage of a process, and all of the analogies in the set are partially correct — they are all "best" at conveying something about the topic domain. More generally, we claim that for most domains of any complexity the "pieces" (e.g., stages, sectors, aspects, etc.) into which the domain is divided will each need
According to Spiro, et al. (pp.525-529), multiple analogies can make very valuable contributions to understanding through functions such as:

(1) **Supplementation of an earlier analogy**: The new, additional analogy covers aspects of the target domain that were missed by earlier analogies.

(2) **Correction of an earlier analogy**: The new, additional analogy corrects aspects of the source domain that **mislead** users about aspects of the target domain, without altering the earlier analogy’s correct information.

(3) **Alteration of an earlier analogy**: Sometimes an incorrect element in an earlier analogy can be **modified**, rather than replaced altogether.

(4) **Enhancement of an earlier analogy**: Something in the earlier analogy is **refined** in a way that deepens the understanding of the target domain, without repairing the earlier analogy, as in (3), **or** changing its magnification, as in (5).

(5) **Magnification or elaboration of an earlier analogy**: The “magnification” of the source domain needs to be changed in order to correctly “capture” a particular aspect of the target domain.

(6) **Perspective shift through an additional analogy**: New, additional analogies provide new and different perspectives of a complex domain.29

---

27 Namely, **Roth’s zoom lens metaphor** (2001, pp.55-57).

28 Spiro, Feltovich, Coulson & Anderson (1989), p.526: Note that there are two possibilities following magnification: introduction of a **new** analogy for the magnified section of the old analogy and retention of the **old** analogy on a different scale...

Of course, when a more synoptic view is required, grain size may be decreased in a "reverse magnification". Also, a function related to magnification would be served by rotating the image of the source domain so that previously obscured elements are brought into view (e.g., as would happen if the first image involved the front of a thing and one needed to augment understanding by later looking at the back of the thing).

29 Spiro, Feltovich, Coulson & Anderson (1989):
It’s possible, further, to distinguish different kinds of analogy. Keynes (1921/1973, p.244) identified two:

(1) **Positive Analogy** based on similarity, and

(2) **Negative Analogy** based on difference, corresponding to an entity’s “essential” and “non-essential characteristics”.30,31

Hesse (1963, p.11) identified a highly significant third default “class of properties about which it is not yet known whether they are positive or negative analogies”:32

(1) **Positive Analogy**: involving features “that are identical or strongly similar”;

(2) **Negative Analogy**: involving features “known to be different or strongly dissimilar”; and

(3) **Neutral Analogy**: involving those features “for which there is no evidence yet as to similarity or dissimilarity”.33

A related use of new analogies for contributing to understanding involves the introduction of an analogy that addresses underlying causal mechanisms after an earlier analogy that was more concerned with the surface form of a situation. (p.526)

The employment of a single analogy for a complex concept may impede the acquisition of more advanced understandings of that concept and engender misconceptions [and it is clear that] access to a fuller and more immediate comprehension of conceptual complexities may be achieved by the systematic employment of integrated sets of multiple analogies. (p.528)

30 Keynes (1921/1973):

As the positive analogy measures the resemblances, so the negative analogy measures the differences between the two objects. (p.248)

[Using this terminology] we may term the class characteristics which are common to all [members of a statistical set] the positive analogy, and the class characteristics which are not common to all of them the negative analogy. (p.454)

31 Recognizing that many analogies were based on “partial knowledge”, rather than “complete knowledge” (p.248), Keynes further distinguished between:

1) “the total positive analogy”, where complete knowledge obtained;

2) “the known positive analogy”, where partial knowledge obtained;

3) “the total negative analogy”, where complete knowledge obtained; and

4) “the known negative analogy”, where partial knowledge obtained.

(Keynes’ categorization presents a problem, particularly because scientists never claim to have complete knowledge; they always claim to work with partial knowledge.)

32 Despite Hesse (1963) calling the advocate of her tripartite classification “Campbellian” (opposing the “Duhemist”) in her explanatory dialogue, and cites Campbell (1920/1957) as an important reference (p.146) for her overall discussion, there is no mention of anything like this sort of categorization in Campbell’s (1920/1957) discussion of analogy. Also, Keynes (1921/1973) does not mention Campbell.

33 Hesse (2001), p.300:

DNA models built of painted balls and metal struts are positively analogous to DNA molecules in spatial structure and connectedness, but negatively analogous in size, material, shape, and color of the constituents, etc. These models have a neutral analogy with molecules insofar as their further detailed properties are used to explore as yet unknown features of genetic material. The dividing line between these three sorts of analogy will of course shift as research goes forward — the better the model, the more of the neutral analogy will eventually be accepted as positive, whereas a poor model will become more and more negatively analogous.
Legal tradition, going back to Aristotle,\(^\text{34}\) demands that like cases are treated alike, and that different cases are treated according to their differences.\(^\text{35,36}\)

Thus, the prosecution produces a wide range of pertinent positive analogies,\(^\text{37}\) and uses them to argue that situation \(S\) is so extremely \(X\)-like and, because it's so much like \(X\) in so many ways, subject to laws \(A\), and \(B\), and \(C\).

\(^{34}\) Aristotle, in his *Nicomachean Ethics* (V.3: 1980, p.112), asserts that justice is both fair and equal, that injustice prevails when equals are treated unequally (or unequals treated equally):

The same equality will exist between the persons and between the things concerned; for as the latter — the things concerned — are related, so are the former; if they are not equal, they will not have what is equal, but *this is the origin of quarrels and complaints* — when either equals have and are awarded unequal shares, or unequals equal shares. (emphasis added)

\(^{35}\) Schick and Vaugn, 2003, p.27:

The American legal system is based on precedents. A precedent is a case that has already been decided. Lawyers often try to convince judges of the merits of their case by citing precedents. They argue that the case before the court is similar to one that has been decided in the past, and since the court decided one way in that case, it should decide the same way in this case. The opposing attorney will try to undermine that reasoning by highlighting the differences between the case cited and the current case. The person who wins such court cases is often determined by the strength of the analogical arguments presented.

Yet, unlike science, “what counts as an “answer” in the law is not clearcut” (Rissland, 1990, p.1962):

In law there is usually no unique right answer; rather there are reasonable alternative answers, more a matter of degree than of extremes. The answers are highly contextual, depend on goals and points of view, and change as the law evolves.

\(^{36}\) Courts are “obliged to explain new decisions in terms of their relation to past cases” (Sherwin, 1999, pp.1186-1187) in a unique, restricted set of circumstances:

[The appellate] judge faces a single case based on specific events in the lives of individuals, which are described in the record of the case. She has limited time to spend on the case and has no option but to decide in favor of one party or another. The facts before her are simplified both by the rules of evidence at trial and by the manner in which the [lower court] trial has been reported. She is expected to articulate reasons for her decision that operate at some intermediate level of generality. Parties, critics, and other judges will be disappointed if she decides either “because the facts are thus” or “because I believe this is best”. Finally, the judge is fallible — she makes factual, logical, and moral mistakes.

With lawyers completely ignoring the scientist’s quest of “trying to get it right”, and exclusively concentrating on “trying to win” by seeking support and confirmation for a preordained conclusion, Nussbaum (2003, pp.273-274) argues that, “a good litigator is not a fair arbiter of truth”; and because legal issues (unlike “pure moral argument”) are driven by the dictates of precedent and statute, “getting at the legal truth is not the same thing as getting at the moral truth”. Thus, argues Hampshire (2000), whilst one has the right to expect justice in terms of “fairness” of procedure in a court, one can never expect any sort of true, objective “fairness” in the decisions made by the court.

\(^{37}\) Hunter, 1995/2004:

Analogy plays a central role in the law. Each time a judge decides a case by reference to a precedent, or a law professor suggests a difficult hypothetical in class, or a student tentatively guesses at the answer to an exam problem, they draw analogies from a source case (the known precedent) to a target case (the undecided case). Though analogy is used in law in other ways — to understand statutes, as a fundamental feature of equality and justice, and so forth — analogical reasoning with precedents is one of the defining characteristics of legal reasoning in common law countries derived from the English tradition...

[Yet, unlike] science, in law there is no one “correct” or “more correct” answer. Of course, in science the (creative) process of discovery and (formal) process of justification are intertwined, and each are important. However, even after Kuhn, Popper, and Feyerabend, few are able to argue with a straight-face that there is no correct scientific solution to many — indeed most — scientific problems. This is commonplace in law: a good lawyer is almost always able to do this in legal cases. Thus in law, the evaluation of previously created analogies is not a fixed or formally resolvable matter, and relies in part on the compellingness of the analogy created.

According to Rissland (1990, p.1960), the “key ingredients of precedent-based reasoning [are things] such as making assessments of the relevance of precedents to new situations, distinguishing contrary cases, and drawing connections between relevant cases”.

In relation to legal precedent, the identification of a dissimilar precedent is just as important, if not more important than the identification of a similar precedent.
By contrast, the defence produces negative analogies, or disanalogies,\textsuperscript{38} which very clearly demonstrate, from this alternate perspective,\textsuperscript{39} that situation $S$ is so obviously so un-$X$-like that the prosecution’s case (that laws $A$, and $B$, and $C$ apply in these circumstances) must be immediately dismissed.

Brewer (1996, pp.1003-1006, 1014-1016) describes a case of theft of a steamboat passenger’s valuables from his cabin.\textsuperscript{40}

Driven by a need to treat like cases alike, and with only two relevant precedents ("that an innkeeper was strictly liable for the theft of boarders’ valuables", and "that a railroad company was not strictly liable to passengers for the theft of their valuables from open-berth sleeping car trains"), the court had to determine which of the precedents presented the strongest analogy — i.e., whether the steamboat was "sufficiently like an inn" or "sufficiently like a railroad" (p.1004) — before it could decide which law to apply to the case.

Through a series of analogical arguments, the judge established that the steamboat, in this case, was very much like an inn and very much unlike a railroad.\textsuperscript{41,42,43}

\textsuperscript{38} In rhetorical discourse, when words are being used in a sense other than its literal core meaning, the counterparts source-target relationships of the positive analogy are the tropes of metaphor, and metonymy or synecdoche; and the counterparts of the source-target relationships of the negative analogy (or the disanalogy) are the tropes of anomaly, and irony or paradox.

\textsuperscript{39} Which often involves thinking about “taken-for-granted phenomena in new, unusual, and unconventional ways” (Oswick, Keenoy & Grant, 2002, p.301).

\textsuperscript{40} The case was (1896) "Adams v. New Jersey Steamboat Co.". The appeal court had to decide "whether the steamboat owner was strictly liable to the passenger for the loss (it having been decided [in the lower court] that neither the steamboat owner nor the passenger was negligent)" (Brewer, p.1003).

\textsuperscript{41} The relevant sections of the judge’s argument appear at Brewer, pp.1004-1005, and pp.1013-1014.

\textsuperscript{42} In another case ("California v. Carney"); Brewer (1996, p.936), Peters (2004, pp.948-951) the issue was whether marijuana and associated paraphernalia seized in a "warrantless" search of a parked motor home (licensed for "vehicular travel" and parked in a "parking lot") could be used in evidence.

Given the established Fourth Amendment protection of one’s “house” from a search without a warrant — which (passed in 1791) asserts “The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrants shall issue, but upon probable cause, supported by oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized” — and the long-established principle that an “automobile”, due to its “mobility”, could be searched without a warrant, the critical question was:

\textit{Whether, for purposes of applying the warrant requirement of the Fourth Amendment, a motor home parked off the street was relevantly similar to a house, or instead relevantly similar to a car, because a warrant is (usually) required for the search of the former but not of the latter. (Brewer, p.936)}

Despite accepting that the “vehicle possessed some, if not many attributes of a home”, the judges did not accept that the motor home was "more properly treated as a residence" (Peters, 2004, p.949), and, instead "held that the motor home fell within the automobile exception because it was readily mobile and was in a place that indicated it was being used as a vehicle" (p.949).

Peters (p.950) describes the rationale behind the weight given to analogy and disanalogy:

\textit{Implicit in both the majority and dissenting opinions is the interplay between two factors — form and function — that suggest whether a structure qualifies for Fourth Amendment house protection. Under the first, or form factor, the greater the similarity of a structure to the archetypal home — "the brick bungalow or frame Victorian" — the more likely it will fall within the express protection of the Fourth Amendment. When a}
In their survey of the use of analogy in "the discovery, development, and evaluation of scientific theories", Holyoak and Thagard (1995, pp.185-189) identified sixteen analogies which met each of two significant criteria:

1. The analogy played an important role in some vital aspect of the scientist’s thinking in the discovery, development and/or subsequent defence of their original idea; and

2. (Regardless of whether the theory in question is still accepted today), the analogical thinking "contributed to a major theoretical advance".

structure deviates from the archetype, the Court looks to the function of the place. The majority found that Carney’s motor home was more motor than home, and, therefore, its primary function placed it within the automobile exception. The dissent on the other hand, thought the motor home was more home than motor. Carney’s Dodge [motor home] could accommodate a "wide range of private human activity" — sleeping, eating, storing personal effects — and was associated with a particular way of living. On this basis, the dissent concluded the motor home functioned more like a dwelling than a vehicle.

Érdi (1996, p.181) describes how cognitive scientists use analogies and disanalogies between brains and computers to elaborate the function and structure of the brain.

Millman and Smith (1997, pp.159-160) have identified Darwin’s simultaneous use of both analogy ("between artificial selection and natural selection") and disanalogy ("between artificial selection in domestication and species formation in the wild") between the same source and target, noting that "complex domains contain many potential features of similarity and difference".

And, setting aside the ethical debates on the use of animals in medical research, there are even more heated and intense disputes about the analogies and disanalogies that are offered to support or undermine "the claim that animal research is of immediate and direct relevance to human biomedical phenomena" (LaFollette & Shanks, 1996, p.108), and the extent to which "findings on laboratory animals are extrapolable to humans" (p.22) — or, even, "from one species to another" (p.viii).

The distinction between the initial use of analogy to discover, develop and evaluate a theory and the subsequent use to clarify and explain a particular theory, once developed, is very important.

Clark (2001, pp.121) distinguished between accounts which are "a ("mere") description of a pattern of events" and those which are "a real explanation of why the events unfold as they do".

Research by Holyoak & Thagard (1995, p.189) found that, even though (a) the kinetic theory of gases is routinely explained with an analogy to billiard balls in motion, and (b) the Rutherford-Bohr model of the atom is routinely explained with an analogy to the solar system, there was no evidence of any kind that either of analogy had played any part in the creation of their respective theory or model.

It seems that these "subsequent" analogies function rather like the mnemonic for the difference between dromedaries and Bactrian camels (which have one and two humps respectively). The initial letters **D** and **B** turned on their side indicates they type of camel; however, there is no evidence that the animals’ names were ever chosen for their initial letters having either one or two humps.
## Table Three: Analogy in Discovery, Development & Evaluation of Scientific Theories

[constructed from Holyoak and Thagard’s “The Analogical Scientist” (1995, pp.185-189)].

<table>
<thead>
<tr>
<th>ANALOGY</th>
<th>SCIENTIST</th>
<th>BRIEF DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sound &amp; water waves.</td>
<td>Roman architect Vitruvius (first century CE).</td>
<td>Vitruvius explained the acoustic properties of Greek amphitheatres by explicitly comparing “the sound of voices to water waves that can flow out and bounce back when obstructed, just as sound spreads and echoes” (p.186).</td>
</tr>
<tr>
<td>2. The Earth &amp; a small magnet</td>
<td>William Gilbert (in his <em>De Magnete</em>, published in 1600 CE).</td>
<td>Gilbert made a systematic comparison between the properties of the Earth (e.g., how it affects compasses), and “the properties of the small spherical magnets on which he had performed many experiments” (p.186). The consequence of this was the realization that the Earth not only acted like a magnet, but that it really was, in fact, a giant magnet.</td>
</tr>
<tr>
<td>3. The Earth &amp; the Moon.</td>
<td>Galileo (in <em>Dialogue Concerning the Two Chief World Systems</em>, published in 1630 CE).</td>
<td>Galileo supported his contention that Earth moved by comparing the Earth to the Moon, “both of which are spherical, dark, opaque, dense, and solid, with similar expanses of light and dark and of land and sea. Since the moon was known to move in an orbit, he argued it was reasonable to suppose that the earth does too.” (p.186)</td>
</tr>
<tr>
<td>4. The Earth &amp; a ship.</td>
<td>Galileo (in <em>Dialogue Concerning the Two Chief World Systems</em>, published in 1630 CE).</td>
<td>Galileo’s defended his theory that Earth moved against the view that “if a rock is dropped from a tower, it lands at the base of the tower, suggesting that the tower, and hence the earth, is not in motion” with an analogy between a tower and the mast of a ship. “He pointed out that a rock dropped from the top of the mast will fall and land at the base of the mast even though the ship is moving.” (pp.186-187)</td>
</tr>
<tr>
<td>5. Light &amp; sound.</td>
<td>Christiaan Huygens (in his <em>Treatise on Light</em>, published 1678 CE).</td>
<td>“Huygens used an analogy between light and sound in support of his wave theory of light. That theory was eclipsed for more than a century by Newton’s particle theory but was revived in the early nineteenth century by Thomas Young and Augustin Fresnel. These scientists exploited the analogy between light and sound to develop and defend a wave theory of light”. (p.187)</td>
</tr>
<tr>
<td>6. A Planet &amp; a projectile</td>
<td>Isaac Newton (in his <em>Principia</em>, published in 1687 CE).</td>
<td>In order to “help bring planetary motion within the scope of his theory of gravitation [Newton] compared a planet to a stone thrown upward from the earth with greater and greater force. He presented a diagram to show how with a great enough force the path of the stone would become the path of an object in orbit around the earth. Newton used this analogy to support his hypothesis that the orbits of the planets are governed by gravitational force.” (p.187)</td>
</tr>
<tr>
<td>7. Lightning &amp; electricity.</td>
<td>Benjamin Franklin (circa 1749 CE).</td>
<td>Franklin noticed many similarities between lightning and electricity: including, amongst other properties, “giving light”, “swift motion”, being “conducted by metals”, a “crack or noise in exploding”, “rendering bodies it passes through” and “firing inflammible substances” (p.184). Eventually &quot;this comparison became more than an analogy. Following Franklin’s work, the definition of electricity has expanded to include the natural phenomenon of lightning&quot;. (p.187)</td>
</tr>
<tr>
<td>8. Respiration &amp; combustion.</td>
<td>Antoine Lavoisier (during the 1770s CE).</td>
<td>As well as his &quot;oxygen theory of combustion&quot;, Lavoisier also “developed a theory of the role of oxygen in animal respiration. Much of his thinking was guided by an analogy between respiration and combustion, both of which involve a change of oxygen into carbon dioxide and a provision of heat.” (p.187)</td>
</tr>
<tr>
<td>9. Heat &amp; water.</td>
<td>Nicholas Leonard Sadi Carnot (1824 CE).</td>
<td>In his discussion of the motive power of heat, Carnot drew &quot;heavily on an analogy between heat and waterfalls. He argued that heat acts on substances, just as water acts on waterfalls, with the power depending in the former case on the amount of caloric (heat substance) and in the latter on the height of the waterfall.&quot; (p.187)</td>
</tr>
<tr>
<td>ANALOGY</td>
<td>SCIENTIST</td>
<td>BRIEF DESCRIPTION</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>-------------------</td>
</tr>
<tr>
<td>10. Animal and plant competition &amp; human population growth.</td>
<td>Charles Darwin (1838 CE).</td>
<td>Malthus’ <em>Essay on the Principle of Population</em> suggested to Darwin “that rapid population growth in the face of limited food and land could lead to a struggle for existence”. This, by analogy, was “a mechanism that could produce the evolution of species”, and “Darwin noticed the analogy between potential human strife (produced by population growth’s outstripping resources) and competition among animals and plants for survival”. (p.187)</td>
</tr>
<tr>
<td>11. Natural selection &amp; artificial selection.</td>
<td>Charles Darwin (in his <em>Origin of Species</em>, 1859 CE).</td>
<td>Darwin “compared natural selection to the artificial selection performed by breeders, who exploited the inherent variability in animals and plants to choose desired features”. And, argued Darwin, “just as artificial selection leads to different breeds”, so “natural selection leads to different species”. Darwin also extensively used this analogy “in developing explanations and in arguing for the acceptability of his overall theory”. (pp.187-188)</td>
</tr>
<tr>
<td>12. Electromagnetic forces &amp; continuum mechanics.</td>
<td>James Clerk Maxwell (1860s CE).</td>
<td>By constructing “a diagrammed mechanical model for electrical and mechanical forces, consisting of a fluid medium with vortices and stresses”, Maxwell was able to abstract “a general mathematical description that could be applied directly to electromagnetism.” (p.188)</td>
</tr>
<tr>
<td>13. Benzene Molecule &amp; a snake.</td>
<td>Friedrich Kekulé (1865 CE).</td>
<td>Kekulé stated that he owed his theory of the molecular structure of the benzene ring to “a reverie in which he saw a snake biting its own tail”. As with Newton (6), Maxwell (12), and Morgan (14), this is an example of “how visual representations can contribute to creative thinking using analogy” (p.186).</td>
</tr>
<tr>
<td>14. Chromosomes &amp; beaded strings.</td>
<td>Thomas Morgan and his colleagues (1915 CE).</td>
<td>In order to explain “the complex phenomena of inheritance”, they compared chromosomes to strings of beads corresponding to the various genes that led to inheritance: this “beaded-string analogy was most useful for describing how novel linkages could arise from crossover of chromosomes, just as new patterns of beads could arise from breaking and recombining the string”. (p.186)</td>
</tr>
<tr>
<td>15. Bacterial mutation &amp; a poker machine”.</td>
<td>Salvador Luria (1943 CE).</td>
<td>An analogy with poker machines, which pay out at random — mostly nothing, sometimes small amounts and, very occasionally, jackpots — allowed Luria to realize that “if bacteria become resistant because of gene mutations, then the numbers of resistant bacteria in different bacterial cultures should vary like the expected returns from different kinds of [poker] machines. This reasoning led to an experiment and theoretical model, for which he was awarded a Nobel prize.” (p.188)</td>
</tr>
<tr>
<td>16. The Mind &amp; a computer.</td>
<td>Alan Turing and many others (since the 1950s CE).</td>
<td>In attempting “to understand the nature of mind and thinking” the most fertile analogies have been those which compare thinking and computation. “Computational ideas have suggested hypotheses about the nature of mind that have led to much psychological and computational experimentation”; and, not only have “ideas about computation... influenced conceptions of the mind, [they] have in turn had an impact on the evolution of new types of computers”. (p.188)</td>
</tr>
</tbody>
</table>

Recently, researchers have studied a unique form of *indirect* counter-argument that emerges in adversarial encounters; a special case of *disanalogy*, first identified by Whaley and Holloway, now widely known as the “rebuttal analogy”. For example:

---


46 Rebuttal analogies take the form \[ \text{A : B} :: \text{C : D} \] and use comparisons between aspects of a relatively well-understood *comparison proposition (source)* and a relatively far-less-well-understood *stated proposition (target)*, to explain, elaborate or introduce some unknown (or poorly understood) aspect of the *stated proposition (or target domain)*. For example, Whaley (1997); Whaley & Holloway (1997); Whaley (1998); Whaley, Nicotera, & Samter (1998); and Colston (2000).
Besides, it's absurd that we only have an oral tablet to treat vomiting. It's like treating diarrhea with a suppository. (Brian Hecht, 1991)

Made “in response to another [real or hypothetical] person’s statement of the target proposition” (Colston, 2000, p.339), “[displaying] a structure specifically tailored for argumentative and persuasive purposes” (Whaley, 1997, p.162), using “situational irony”, and typically directed at an absent (real or hypothetical) third party, rebuttal analogies function simultaneously as both:

(a) an argument, “used to counter another’s claim by attacking... the credence of the opposition’s claim” (Whaley, p.163), by showing “the unsoundness of the opposed [target] proposition” (Colston, p.337), and

(b) a social attack, “used to counter another’s claim by attacking the character or competence of the other” and, at the same time “portray the person that made [the original claim] as ridiculous or at least worthy of derision” (Whaley, p.163), in order to emphasize the stupidity of anyone advocating such an

47 See Whaley & Holloway (1997, p.297) for complete citation details.

48 The following representative examples have been selected from the literature:

And finally, they [Japanese automakers] say all our problems are our fault. That's like blaming our Army and our Navy for Pearl Harbor because they weren't ready. (Lee Iacocca, 1992)

[See Whaley (1997, p.163) for complete citation details.]

Passing more comprehensive and more intrusive laws every year and then blaming the growing number of lawyers for the ensuing litigation is a little like starting a war and then blaming the casualties on the growing number of morticians. (R. Dooling, 1997)

[See Whaley, Nicotera, & Samter (1998, p.48) for complete citation details.]

Forcing Microsoft to include Netscape's competing software in our operating system is like requiring Coca-Cola to include three cans of Pepsi in every sixpack it sells. (Bill Gates, 1998)

[See Whaley, Nicotera, & Samter (1998, p.48) for complete citation details.]

We may disagree as to just how much a 1,000-year-old tree is worth, but it sure as hell isn't "zero". A 1,000-year-old tree is not "replaced" by five or six seedlings. That's like saying your grandmother is "replaced" by six sperm cells. (Paul Roberts, 1993)

[See Whaley & Holloway (1997, p.295) for complete citation details.]

49 According to Lucariello (1994), and Lucariello & Mindolovich (1995), "situational irony... is a condition of events opposite to what was, or might naturally be expected, or a contradictory outcome of events as if in mockery of the promise and fitness of things" (Lucariello, p.129). In recent times we have examples of stupid "smart bombs" and imprecise "precision bombing".

The comment that New Zealand Prime Minister, Sir Robert "Piggy" Muldoon (1921-1992) made on the apparent "brain drain" in the mass exodus of New Zealanders to Australia in the early 1980s — "New Zealanders who go to Australia raise the IQ of both countries" — is a fine example of the use of irony.

Colston's experiments found the extent to which the rebuttal analogy was ironic "predicted the degree to which speakers using those analogies were perceived as being argumentative, and as attacking their addressees" (p.350).

50 Colston (p.352):

Debates, quarrels, arguments, disputes, deliberations, differences of opinion, and other similar verbal contests pose potentially conflicting social demands of interlocutors... [and] speakers in these situations must simultaneously oppose or combat some person(s) and/or their opinion(s), as well as gain or maintain favor among witnesses or co-proponents. The nature of these social demands is thus uniquely suited to indirect language which can perform a variety of negative and positive pragmatic functions.
absurd proposition in a socially acceptable way (Colston, p.339).

However:

Data from recent investigations strongly suggest that persons perceive rebuttal analogy users as less polite, less likeable, less intelligent, less knowledgeable, and less ethical than the communicators using a literal translation of the message. Women appear particularly sensitive to the face attack inherent in rebuttal analogies. (Whaley & Holloway, p.302)

Whaley and Holloway argue (p.300) that one of the special qualities of rebuttal analogies is that their humour and irony allow "statements that most people might hesitate to make directly in public" to be freely made and to be treated as consistent with conventional argument discourse. Also, if the arguments embedded in rebuttal analogies were "translated into direct speech", they would mostly read as: "that argument is fallacious and foolish", "the speaker is wrong", "the speaker is not of sufficient character to speak to this issue", "the speaker is stupid", etc.: Rebuttal analogy is a socially sanctioned method of reaching the purpose of argument, performing an attack on an opponent, and adapting to the constraints of a mediated context. It can capture media attention, rebut argument, foster identification, promote the communicative skill of the speaker, and challenge the opponent — all in a concise, often humorous, manner. (Whaley & Holloway, p.301, emphasis added)

According to Sopory and Dillard, (2002, p.414) "research on... rebuttal analogy... has shown that analogy use compromises evaluations of communicators as well as [evaluations of the] quality of message arguments". Whaley (1997, p.163):

Implicit in rebuttal analogy is an evaluation not only of the opposing issue or argument, but also of the opposing arguer. The point of employing rebuttal analogy as argument is that it simultaneously addresses a second communicative function — the attack on the opposition's character or competence (i.e., "face"). That is, inherent in public argument is a motive to belittle the opposition. Rebuttal analogy not only suggests that one's idea may not have merit, but exploits identity issues by suggesting that the idea is really stupid.

As Colston indicates (p.339-340), a significant advantage of rebuttal analogies, and their embedded denigration of both the original proposal and its advocate(s), is that they require no special training in formal logic (or the conventions of argument) for a listener to draw the intended conclusion that the original proposition is absurd. Colston (p.340) labels this property the "absurdity comparison"; and remarks that, despite not requiring any special training in formal logic or the conventions of argument for a listener to get the point in these cases, a listener with any "knowledge of argumentative conventions" would identify the rebuttal analogy far more quickly — and, thus, an "absurdity comparison" is certainly not a disadvantage.

Whaley & Holloway (p.300):

What must not be overlooked is that rebuttal analogies are used strategically to create identification with audiences and are consciously performed, despite attempts to signal otherwise. The deceptive simplicity and movement toward the "common" [i.e., rather than the "proper"] makes rebuttal analogies rhetorically advantageous.

Rebuttal analogy is widely used by politicians, respected academicians, corporate heads and lay persons to refute the claims of the opposition during public oratory. However, recent theorizing and studies concerning rebuttal analogy dispute long-held notions concerning the nature of argument by analogy. For instance, investigations have failed to support traditional tenets that analogy use positively enrich perception of communicators and enhances the persuasive process. In fact, quite to the contrary, rebuttal analogy is regarded as a rather unconvincing and unethical form of argumentation and that individuals who employ rebuttal analogy are considered to be less polite, less competent, and less ethical than non-users.

Whaley, Nicotera & Samter's survey of the perception of rebuttal analogies by "African American" women (pp.54-55) clearly found that "African American women perceive communicators who employ rebuttal analogy as significantly less polite than communicators who utilize a non-analogy form of the same argument" and they "were also found to rate rebuttal analogy arguments as significantly less ethical than the non-analogy version".

These findings were consistent with earlier findings on the attitudes of "European American" women: "members of both groups perceive communicators who utilize rebuttal analogy to be impolite and also believe it is an unethical form of argumentation"; and this led Whaley, et al. to conclude that "rebuttal analogy may be particularly ineffective when attempting to persuade female audiences".

Whaley (1998, pp.355-356) reports that his own recent research had also shown that:

(1) "persons rated rebuttal analogy as a less effective (i.e., strong/convincing) and a less ethical argument than its non-analogy counterpart";

51 Whaley and Holloway argue (p.300) that one of the special qualities of rebuttal analogies is that their humour and irony allow "statements that most people might hesitate to make directly in public" to be freely made and to be treated as consistent with conventional argument discourse.

Also, if the arguments embedded in rebuttal analogies were “translated into direct speech”, they would mostly read as: “that argument is fallacious and foolish”, “the speaker is wrong”, “the speaker is not of sufficient character to speak to this issue”, “the speaker is stupid”, etc.: Rebuttal analogy is a socially sanctioned method of reaching the purpose of argument, performing an attack on an opponent, and adapting to the constraints of a mediated context. It can capture media attention, rebut argument, foster identification, promote the communicative skill of the speaker, and challenge the opponent — all in a concise, often humorous, manner. (Whaley & Holloway, p.301, emphasis added)

52 According to Sopory and Dillard, (2002, p.414) “research on... rebuttal analogy... has shown that analogy use compromises evaluations of communicators as well as [evaluations of the] quality of message arguments”. Whaley (1997, p.163):

Implicit in rebuttal analogy is an evaluation not only of the opposing issue or argument, but also of the opposing arguer. The point of employing rebuttal analogy as argument is that it simultaneously addresses a second communicative function — the attack on the opposition’s character or competence (i.e., “face”). That is, inherent in public argument is a motive to belittle the opposition. Rebuttal analogy not only suggests that one’s idea may not have merit, but exploits identity issues by suggesting that the idea is really stupid.

53 As Colston indicates (p.339-340), a significant advantage of rebuttal analogies, and their embedded denigration of both the original proposal and its advocate(s), is that they require no special training in formal logic (or the conventions of argument) for a listener to draw the intended conclusion that the original proposition is absurd. Colston (p.340) labels this property the “absurdity comparison”; and remarks that, despite not requiring any special training in formal logic or the conventions of argument for a listener to get the point in these cases, a listener with any “knowledge of argumentative conventions” would identify the rebuttal analogy far more quickly — and, thus, an “absurdity comparison” is certainly not a disadvantage.

Whaley & Holloway (p.300):

What must not be overlooked is that rebuttal analogies are used strategically to create identification with audiences and are consciously performed, despite attempts to signal otherwise. The deceptive simplicity and movement toward the “common” [i.e., rather than the “proper”] makes rebuttal analogies rhetorically advantageous.

54 Whaley, Nicotera & Samter (1998, pp.48-49):

Rebuttal analogy is widely used by politicians, respected academicians, corporate heads and lay persons to refute the claims of the opposition during public oratory. However, recent theorizing and studies concerning rebuttal analogy dispute long-held notions concerning the nature of argument by analogy. For instance, investigations have failed to support traditional tenets that analogy use positively enrich perception of communicators and enhances the persuasive process. In fact, quite to the contrary, rebuttal analogy is regarded as a rather unconvincing and unethical form of argumentation and that individuals who employ rebuttal analogy are considered to be less polite, less competent, and less ethical than non-users.

55 Whaley, Nicotera & Samter’s survey of the perception of rebuttal analogies by “African American” women (pp.54-55) clearly found that “African American women perceive communicators who employ rebuttal analogy as significantly less polite than communicators who utilize a non-analogy form of the same argument” and they “were also found to rate rebuttal analogy arguments as significantly less ethical than the non-analogy version”.

These findings were consistent with earlier findings on the attitudes of “European American” women: “members of both groups perceive communicators who utilize rebuttal analogy to be impolite and also believe it is an unethical form of argumentation”; and this led Whaley, et al. to conclude that “rebuttal analogy may be particularly ineffective when attempting to persuade female audiences”.

Whaley (1998, pp.355-356) reports that his own recent research had also shown that:

(1) "persons rated rebuttal analogy as a less effective (i.e., strong/convincing) and a less ethical argument than its non-analogy counterpart";
The "situational irony" of rebuttal analogies can also vividly highlight otherwise unrecognized irony in the stated proposition.\textsuperscript{56}

News that basketballer Michael Jordan was endorsing a personal line of hair shampoo would not immediately seem ironic;\textsuperscript{57} yet, the irony is immediately apparent in:

Michael Jordan is endorsing his own line of shampoo. \textit{Isn’t that a little like Heaven’s Gate Condoms?} (Jay Leno, 1997)\textsuperscript{58,59}

Colston (pp.338-339) extensively analyzed the external and internal processes of rebuttal analogies; and clearly demonstrated how and why rebuttal analogies simultaneously performed their unique function of “arguing” against a proposition and making a “social attack” on the proponent of that position:\textsuperscript{60}

(2) “subjects rated the communicators of messages containing rebuttal analogies as less polite than sources of identical messages with the rebuttal analogy translated to a non-analogy form”; and

(3) “rebuttal analogy adversely affects receivers’ perceptions of communicator politeness and likeability in argumentative exchanges based on, perhaps, perceived violations of fair argumentation practices”.

Moreover (p.356): Recent findings [by others] suggest [that] rebuttal analogy is a relatively weak, unconvincing argument form (conceivably because it only provides probable proof)... In addition, research suggests that the effects of rebuttal analogy may extend beyond communicators being deemed as unethical. Specifically, studies indicate that individuals who resort to verbal aggression are thought to lack argumentative skill. Associated with this deficiency may be the perception that for communicators to resort to such tactics they must lack the competence to proceed with an argument.

\textsuperscript{56} Listeners “[notice] the irony inherent in the stated relationships among the components of the [comparison] proposition of a rebuttal analogy” and “[notice] the analogous structure in the [stated] proposition” which, then, highlights “the irony in the [stated] proposition as well, along with the "faulty thinking" inherent in seriously advocating such a proposition” (p.340).

If there is some degree of irony in an advocate’s proposition, why is analogical comparison with a similar [comparison] domain needed to point out that irony? Wouldn’t listeners simply detect the irony in the [original] proposition? The likely answer is no. The person advocating the [stated] proposition failed to detect its irony. The speaker probably would not have stated the proposition in the first place had she [herself] noticed its irony. It is also possible to overlook the irony in many situations because of other salient information that is not ironic. (Colston, p.351)

\textsuperscript{57} Because Jordan was widely known to endorse a wide range of commercial products; and, like other celebrities, he was also known for endorsing products that were produced (or marketed) by himself.

\textsuperscript{58} See Colston (p.351) for complete citation details. This analogy exploits two specific, pertinent facts: (a) Jordan regularly shaved his head (and, thus, had no hair to shampoo), and (b) all male members of the Heaven’s Gate religious cult had been castrated (and, thus, had no sperm for a condom to contain).

\textsuperscript{59} Colston (p.352) argues that, along with their positive pragmatic functions (e.g., “winning over witnesses”), rebuttal analogies also have negative pragmatic functions (e.g., “attacking opponents”), which they achieve through their use of irony:

Even had the original statement [that Jordan was endorsing a personal line of shampoo] seemed ironic upon the first reading, the irony might not be the statement’s most obvious characteristic. Moreover, even if the statement were seen as fairly ironic, comparison with the greater irony in the [comparison] statement would likely enhance the perceived irony in the target.

It thus appears that rebuttal analogies serve to point out target ironies that were either not obviously ironic when first proposed, or at least seen as being only moderately ironic. The more blatant irony in the [comparison] domain serves then to overcome this obscurity and make the irony in the target domain stand out.

\textsuperscript{60} I have constructed figure one to display and represent Colton’s patterned analysis of the prototypical rebuttal analogy. The information it contains has been extracted, adapted and re-organized, with some terminological modifications, from the text of Colston (2000), p.338.
Colston (p.339) also demonstrated the practical utility of his analysis by employing it to elaborate the mechanism of the typical rebuttal analogy, using an imaginary case where the rebuttal analogy, “cutting taxes to boost the economy is like cutting the blood flow to your brain to improve your health”, has been made in response to a proposal advocating “tax cuts as an impetus to economic growth”.

---

61 I have constructed figure two to display and represent Colston’s example. The information it contains has been extracted, adapted and re-organized, with some terminological modifications, from the text of Colston (2000), p.339. Also, it has been carefully structured so that it is symmetrical with the prototype displayed in Fig.1.
Sometimes, those arguing against a particular thought experiment will produce a counter-thought-experiment; in some of these cases, the counter-thought-experiment can be usefully understood as a rebuttal analogy:
The counter-thought-experiment’s orientation

**Acceptance**

Whilst accepting the analogy within the original thought experiment, the counter-thought-experiment contends that the original thought experiment did not consider all of the knowledge relevant to the matter in question. The new, counter-thought-experiment attempts to over-ride the original thought experiment by presenting additional, relevant and superior grounds for accepting a new, alternative conclusion.

**Rejection**

The counter-thought-experiment contends that the analogy within the original thought experiment misrepresents the facts of the matter in question. The new, counter-thought-experiment attempts to undermine the original thought experiment by revealing some basic error in the ground of its analogy.

---

**Table Four: A Taxonomy of the Four Different Types of Counter-Thought-Experiment which Reveal Deficiencies in the Original Thought Experiment**

<table>
<thead>
<tr>
<th>THE COUNTER-THOUGHT-EXPERIMENT’S ORIENTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acceptance</strong></td>
</tr>
<tr>
<td>Whilst accepting the analogy within the</td>
</tr>
<tr>
<td>original thought experiment, the counter-</td>
</tr>
<tr>
<td>thought-experiment contends that the</td>
</tr>
<tr>
<td>original thought experiment did not</td>
</tr>
<tr>
<td>consider all of the knowledge relevant to</td>
</tr>
<tr>
<td>the matter in question. The new, counter-</td>
</tr>
<tr>
<td>thought-experiment attempts to over-ride</td>
</tr>
<tr>
<td>the original thought experiment by</td>
</tr>
<tr>
<td>presenting additional, relevant and superior</td>
</tr>
<tr>
<td>grounds for accepting a new, alternative</td>
</tr>
<tr>
<td>conclusion.</td>
</tr>
</tbody>
</table>

| **Rejection**                              |
| The counter-thought-experiment contends     |
| that the analogy within the original thought|
| experiment misrepresents the facts of the    |
| matter in question. The new, counter-thought-|
| experiment attempts to undermine the         |
| original thought experiment by revealing     |
| some basic error in the ground of its analogy.|

---

**THE COUNTER-THOUGHT-EXPERIMENT’S EFFECT**

**Constructive**

The counter-thought-experiment leaves us with a new conclusion to consider in place of the conclusion of the original thought experiment.

**Destructive**

The counter-thought-experiment leaves us no closer to a conclusion than we were before the original thought experiment’s analogy was introduced.64

**Counter-Analogy**

The analogy embedded within the counter-thought-experiment employs additional knowledge (that has been taken from a different source domain) from the original thought experiment in order to draw an alternate, superior, and far more coherent conclusion about the same target domain.62

**Disanalogy**

The analogy embedded within the counter-thought-experiment employs additional knowledge (that has been taken from the same source domain) as the original experiment in order to identify those attributes which the source and target ought to share, but do not share, in order to conclude that the original thought experiment’s analogy is incoherent.65

**Misanalogy**

Once the (incorrectly interpreted) facts of the matter in question in the original thought experiment’s analogy have been interpreted correctly, the analogy embedded within the original thought experiment is shown to be coherent.63

**False Analogy**

Once the (incorrectly interpreted) facts of the matter in question in the original thought experiment’s analogy have been interpreted correctly, the (superficially appealing) analogy embedded within the original thought experiment is shown to be incoherent.66

---

62 In this case, the counter-thought-experiments “operate on the contention that the original [thought experiment] simply does not represent all of the information relevant to the conclusion… [A counter-thought-experiment introduces] additional information [from a different source domain]… that suggests a new conclusion at variance with the conclusion suggested by the original” (p.236).

63 In this case, whilst the “implicit conclusion of the original analogy… is undermined by the revision [of that analogy in the counter-thought-experiment]”, it is also true that “the revised analogy [embedded within the counter-thought-experiment] is coherent… and supports the revised conclusion” (p.231).

64 Shelley (p.228) uses the technical term aporia, “the cognitive perplexity posed by a group of individually plausible but collectively inconsistent propositions” (Rescher, 1995), to label this irresolvable state of affairs.

65 In this case, “[the counter-thought-experiment’s] analogy supports mutually incoherent conclusions. As a result, the whole… analogy should be disregarded, much as a deduction based on inconsistent premises should be. Because it undermines any basis for a conclusion, the disanalogy is properly described as having a destructive effect” (p.234).

66 In this case, the counter-thought-experiment shows that “the analogy does not support the conclusion for which it was ostensibly presented in the first place”; and, from this, “the analogy [is shown] to be highly defective and therefore inconclusive” (p.230).
**Metaphor**

The essence of metaphor is understanding and experiencing one thing in terms of another. (Lakoff and Johnson, 1980, p.5)

Metaphors are productive in science.\(^\text{67,68}\)

For Aristotle, a skill in metaphor was both "a sign of natural genius" and an ability "to perceive resemblances" (Poetics 1459b: Russell & Winterbottom, 1989, p.80).\(^\text{69,70}\)

The conventional, "substitution" explanation of metaphor is far from satisfactory.\(^\text{71}\) It

---

67 Gentner & Jeziorski (1993, p.476) conducted research on the development and impact of the use of different sorts of comparisons (analogies, metaphors, etc.) in Western science; and, in particular, they investigated whether or not people evaluated comparisons in literature and science the same way:

Although analogy is preferred in science, literary metaphors are allowed to be rich, complex, and inconsistent, and to have many-to-one mappings and metonymies....

To see whether people use different criteria for scientific and literary comparisons, we asked subjects to rate scientific and literary comparisons for their clarity ("how easy is it to tell what matches with what") and their richness ("how evocative is the comparison; how much is conveyed by the comparison"). We also asked them to rate either the scientific explanatory value or the literary expressiveness of the comparisons. In judgments of scientific merit, clarity was considered crucial and richness was unimportant. In judgments of literary merit, both clarity and richness contributed, and neither was essential. These findings suggest a broader tolerance for non-clarified similarity in literary contexts than in scientific contexts.

68 See, for example, Jones (1983).

Paton, Nwana, Shave and Bench-Capon (1994), p.507:

Scientific metaphors are not ornamental properties of language that can be replaced by literal description. They play a central role in the conception, maintenance, and development of a scientific theory in a variety of ways, which includes:

- Supplying new terms for the theoretical vocabulary.
- Providing the linguistic context in which explanations can be made.
- Influencing the way(s) we interpret the world.
- Affecting the kinds of questions we ask.
- Suggesting new hypothetical entities and/or mechanisms.
- Facilitating communication, for example, between teacher and student.

Haack (1988), p.299:

*Metaphor* is a ladder that science aims to kick away!

The source for this is, obviously, Wittgenstein, (1922), p.189:

6.54: My propositions are elucidatory in this way: he who understands me finally recognizes them as senseless, when he has climbed out through them, on them, over them. (He must so to speak throw away the ladder, after he has climbed up on it.)

He must surmount these propositions then he sees the world rightly.

Whereof one cannot speak, thereof one must be silent.

69 He also expressed the view (The Art of Rhetoric, 1405a, Lawson-Tancred, 1991, p.220) that if you wanted to pay a compliment, you should take your metaphor from something better, and, if you wanted to disparage, you should take your metaphor from something worse:

If the orator wishes to ornament the subject, he must draw the metaphor from the best species of the same genus, and if to depreciate, from the worst.

70 Throughout this dissertation, the *metaphor* will be treated as a type of analogy; especially because a metaphor is "an analogy expressed in verbal form" (Dreistadt, 1968, p.97), and "the items juxtaposed in a metaphor must belong to different conceptual domains" (Vosniadou, 1989, p.417).

Lee (2002), p.6:

A subset of the basic form of an analogy \([A : B :: C : D]\) is the metaphor. Metaphors are used in speech to convey similarity between items. If we remove \(B\) and \(D\), we arrive at \(A : C\) which is illustrated by the example metaphor, "[Cigarettes are like time bombs" (Gentner, 1989, p.222).] Thus, a subset of the basic processes for analogical reasoning should be evident in a metaphor,

(N.B., Lee misquoted Gentner; it has been corrected in the above.)

71 Most grammars define a metaphor as a word or expression used in a sense other than its proper, normal sense; in some context that clearly allows this improper, abnormal sense to be immediately detected, and appropriately transformed, by the listener; as a direct substitute for some freely available,
does not explain the dynamics of the metaphorical process, or why some metaphors are successful and others not, or why there may be certain limitations on their use.

Metaphors create a special relationship of structured comparisons between two distinct systems, and communicate a specific perception of an entity in one domain (the “principal subject”) through a set of systematic, juxtaposed connotations with an entity in a second domain (the “subsidiary subject”).

Precisely equivalent, literal expression (implicit within this “substitution” theory of metaphor is the belief that a precise literal equivalent expression is always freely available); to communicate a meaning that can just as easily be clearly and unequivocally transmitted by that precise literal expression; to enrich, embellish and amplify the communication in some special way; and to make the communication more entertaining and diverting through the unexpected use of an unusually decorative style.

The listener must reverse the substitution and, using the literal meaning of the metaphor as a clue, decipher its special, encoded message; but, if a listener can't immediately reverse the substitution (and decipher the metaphor's special encoded message), the metaphor has failed.

Black, 1962, p.236-237: A memorable metaphor has the power to bring two separate domains into cognitive and emotional relation by using language directly appropriate to the one as a lens for seeing the other; the implications, suggestions, and supporting values entwined with the literal use of the metaphorical expression enable us to see a new subject matter in a new way. The extended meanings that result, the relations between initially disparate realms created, can neither be antecedently predicted nor subsequently paraphrased in prose. We can comment upon the metaphor, but the metaphor itself neither needs nor invites explanation and paraphrase. Metaphorical thought is a distinctive mode of achieving insight, [and is] not to be construed as an ornamental substitute for plain thought.

By contrast, in Breakfast of Champions, Vonnegut (1973/2000, p.53), speaks of a story written by the book’s principal character (Kilgore Trout, a science fiction writer), which featured a fascinating case of a literal message being misinterpreted as if it were metaphorical:

As for the story itself, it was entitled “The Dancing Fool”. Like so many Trout stories, it was about a tragic failure to communicate.

Here was the plot: A flying saucer creature named Zog arrived on Earth to explain how wars could be prevented and how cancer could be cured. He brought the information from Margo, a planet where the natives conversed by means of farts and tap dancing.

Zog landed at night in Connecticut. He had no sooner touched down than he saw a house on fire. He rushed into the house, farting and tap dancing, warning the people about the terrible danger they were in. The head of the house brained Zog with a golf club.

Various authors use different terms for the principal and subsidiary subjects.

<table>
<thead>
<tr>
<th>AUTHOR</th>
<th>PRINCIPAL SUBJECT</th>
<th>SUBSIDIARY SUBJECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richards (1936)</td>
<td>Tenor</td>
<td>Vehicle</td>
</tr>
<tr>
<td>Black (1962)</td>
<td>Principal Subject</td>
<td>Subsidiary Subject</td>
</tr>
<tr>
<td>Leatherdale (1974)</td>
<td>Topic Analogy</td>
<td>Imported Analogy</td>
</tr>
<tr>
<td>Sapir (1977)</td>
<td>Continuous Term</td>
<td>Discontinuous Term</td>
</tr>
<tr>
<td>Cooper (1986)</td>
<td>Principal Subject</td>
<td>Secondary Subject</td>
</tr>
<tr>
<td>Sapir (1977)</td>
<td>Continuous Term</td>
<td>Discontinuous Term</td>
</tr>
<tr>
<td>Gentner and Wolff (1993)</td>
<td>Target</td>
<td>Base</td>
</tr>
</tbody>
</table>

Unlike literal comparisons (e.g., “robbins are like birds”), all metaphors (e.g., “my lawyer is a shark”) are categorical, non-reversible assertions (Glucksberg, 2003, p.95):

The only circumstances under which a metaphor can be reversed is when the ground of the metaphor changes, as in “my surgeon was a butcher” (a negative comment) versus “my butcher is a surgeon” (a positive comment). (p.95)
According to Black (1962), no metaphor can ever be dismissed as a simple comparison between two objects; and, moreover, the “substitution” view is very seriously challenged by the wide range of common everyday expressions such as “He saw what I meant”, “She followed his argument”, “I got your point”, “They caught the bus”, “The sommelier let the wine breathe”, “The politician was spinning a tale”, that have no precise literal equivalents.

A metaphor uses sets of attributes (or implications), considered typical of the subsidiary subject to organize our view of the principal subject in some special way.

![Fig.3: The Metaphorical Process: Interaction between Principal and Subsidiary Subject](image)

It applies the subsidiary subject’s system of connotations to the principal subject in such a way that both speaker and listener reach a common understanding about the speaker’s unique perception and experiences of the principal subject.  

---

74 Black followed the earlier work of Richards (1936). Martin and Harré (1982) stress the extent to which Black, for purposes of his initial explanation, concentrated on metaphor as an interaction between words (and/or expressions), and how, eventually, he seemed to have entirely forgotten Richard’s more general view that metaphor was an interaction between two concepts.

Throughout this dissertation, for the purposes of clarity, I will limit the term ‘metaphor’ to words or expressions, and use the term ‘analogy’ to include the interaction between concepts.

75 Black was following the earlier argument of Richards (1936, pp107-108): Once we begin to “examine attentively” interactions which do not work through resemblances between tenor and vehicle, but depend upon other relations between them including disparities, some of our most prevalent, over-simple, ruling assumptions about metaphors as comparisons are soon exposed.

76 The use of “symbolic” attributes is an essential feature of metaphor. The extended meanings of these symbolic attributes are often quite arbitrary; and, because of their lack of logic, they must be remembered in isolation. Whilst snow and ice denote equally cold real world referents, the symbolic use of “snow” emphasizes whiteness (“white as snow”, “snowy hair”, etc.), whilst the symbolic use of “ice” emphasizes coldness (“cold as ice”, “icy stare”, etc.). Richards (1936, pp.108-109) speaks of this issue in a slightly different way:

The processes of metaphor in language, the exchanges between the meanings of words which we study in explicit verbal metaphors, are super-imposed upon a perceived world which is itself a product of earlier or unwitting metaphor, and we shall not deal with them justly if we forget that this is so.

77 For Blasko (1999, p.1675-1676), metaphors are “central to language and cognition”: The meaning of metaphor exists neither “in” the metaphor nor “in” the comprehender, but rather it involves a complex interaction involving messenger, message, and receiver all existing within a shared cultural context.
Nelkin (1993; 2001), who examined the negative impact of the metaphors used to disseminate genetic information to the public, stressed the importance of metaphor for communicating scientific information. Black (1962) identified two distinct classes of metaphor:

1. **Dynamic Metaphors**: the common type; which involves a special dynamic interaction; and

2. **Religious Metaphors**: "the gene is a sacred entity";

In its metaphorical construction, DNA shares striking qualities with the Christian soul. It seems relatively independent of the body, giving the body life, power and true identity. DNA, like the soul, bears the marks of good and evil. A man might look fine to the outside world, but despite external appearances, if he is evil or ill, it will be marked in his soul — or in his genes. And DNA, an invisible entity, seems to be immortal, containing in it everything needed to bring the body back. This is, of course, the theme of popular stories and films, such as Jurassic Park, in which the DNA is a powerful molecule with magical powers that can resurrect the dead...

By taking on the social and cultural functions of the soul, the DNA — a molecular structure — gains mystical power and iconic importance as a way to explore the most fundamental questions about human life, to explain the essence of human existence, to imagine immortality and to anticipate human fate. And endowing the gene with spiritual importance also empowers science and provides a foundation for its cosmic claims. (p.557)

3. **Fatalistic Metaphors**: "genes are destiny", "genes predict future fate"; and

The fatalistic metaphors of genetic destiny are not only a way to talk about health and disease, but also about blame, moral responsibility and appropriate social order. To locate human fate in the genes indicates a certain inevitability in the structure of existing social categories. They seem 'natural' and therefore 'right'. The great and famous, the successful and celebrated, are what they are because of their genes. Human life is what it has to be: it is 'mapped' and pre-ordained. (p.558)

4. **Commercial Metaphors**: "genes are commodities".

Nelkin (2001), p.556:
Metaphors are a prevalent and important vehicle of public communication, and they are especially important in conveying scientific information. Explaining and popularizing unfamiliar and frequently technical material can be done most effectively through images that are chosen for their richness of reference, their familiar meanings and their graphic appeal. By connecting different orders of reality, metaphors enable the translation of very complex scientific information in culturally meaningful ways.

But metaphors are more than an aid to explanation: repeated metaphors affect the ways we perceive, think and act, for they shape our understanding of events. They also structure our attitudes about public — and scientific — issues. By their choice of metaphors, scientists, their public relations strategists, and science writers, convey certain beliefs about the nature and importance of science and technology, and their limits, impacts and implications. Although people interpret scientific information and ascribe meaning to metaphors according to their personal experience and previous knowledge, metaphors are powerfully persuasive tools.

For more on the impact of specific metaphors on scientific thinking and the lay community’s understanding of science, and scientific issues in general, see Wilkins (1993), Maasen & Weingart (1995), Väliverronen E. & Hellsten (December 2002), and Wilson (2002), pp.246-268. For example, Wilson compares the applications of the well-established, precise scientific term, "greenhouse effect", against the far more ambiguous, far less precise, lay term "global warming" by US television weathercasters.

A recent article in New Scientist (Kingsland, 2004, p.43) attempting to explain the magnitude of the impact of "junk DNA" to lay people through a spam metaphor:

You’ve been spammed. You get back from holiday to find your inbox bursting with hundreds of messages with subject lines such as XXX Pics Included, Special Gift For You, and 75% Discount on Viagra. All useless, irritating junk. But consider this: living organisms are infected with enough of the biological equivalent of spam email to make the internet’s flood of junk pale into insignificance.

It is a little shocking to discover that every cell nucleus in your body is crammed with millions of years’ worth of genetic junk mail. It takes up the vast majority of the genome. Much of this junk is in the form of pieces of DNA called jumping genes, or transposons, that can move around genomes, copying and pasting themselves into our genetic inboxes and mailing themselves on to the next generation. They are so good at this that at least half of our genome comprises dead or barely suppressed transposons — some 4 million of them. Most of the remaining genome is a jumble of unrecognisable detritus that is probably the decayed remains of ancient transposons. Only around 2 per cent actually codes for the proteins we need. Biologists have long regarded themselves into our genetic inboxes and mailing themselves on to the next generation. They are so good at

Black (1962, pp.38ff) refers to an "interaction view of metaphor".
Chapter Three: Comparative Cognitive Processes — 63

(2) **Catachrestic Substitution Metaphors:**\(^{82}\) a rare type; which uses a special type of substitution in a very specific set of circumstances.

**Dynamic Metaphors**

On hearing Plautus’ famous metaphorical statement that "Man is a wolf", listeners immediately recognize an attempt to limit their view of humans to those attributes freely given to the “symbolic” wolf;\(^{83}\) yet, few ever notice that the constellation of notions we normally hold about a “symbolic wolf” have also been adapted on the basis of the attributes that it could possibly share with humans.\(^{84}\)

The dynamic metaphor produces a new way of looking at a very well known entity (the principal subject) by creating a relation of analogy with a set of attributes considered quite typical of another entity (the subsidiary subject).\(^{85}\)

---

\(^{82}\) *Catachresis* ("grammatical misuse").

\(^{83}\) By emphasizing a number of human qualities that can be thought of as somehow wolf-like and, simultaneously, ignoring an equal number that are not wolf-like at all, Plautus’ metaphor efficiently organizes our view of humans in a special way. Whilst the metaphor stops far short of actually classifying humans with wolves, connotations such as “dangerous”, “competitive”, “fierce”, “carnivorous”, “treacherous”, “scavenger”, “hunts in packs” immediately spring to mind.

\(^{84}\) Humans don’t have sharp teeth, bushy tails, long hairy coats, or have pads on their hands and feet; and, in this case, these particular wolf-like attributes are excluded from the metaphorical statement. Yet, if we were to say "the dingo is a wolf", many of the attributes we had automatically excluded from the symbolic subsidiary subject of wolf when it was interacting with the principal subject "man" would be automatically re-instated (and, of course, other attributes may well be excluded).

\(^{85}\) The metaphorical answer “He is a pig!” can be given in response to at least four questions: “Is your boss polite to you?”, “Does your son keep his bed-room tidy?”, “Does your brother-in-law have good table manners?”, and (to the teenage son of a policeman) "What does your father do for a living?". In each case we clearly have different aspects of the "symbolic" pig; which have immediately sprung to mind because of their interactions with a different principal subject — similarly, a green apprentice carpenter will be tend to be inexperienced, a green boyfriend will tend to be jealous, and a green thumb will tend to belong to one who is "horticulturally inclined" (Bache, 1980, p.188).

It must also be quite clear that:

1. no precise, literal expression can replace pig in any of the answers and, more importantly; and
2. if we were to replace pig with the detailed, long-winded set of sentences that would be needed to transmit the meaning of the comparative analogy, the extended translation would still fail to deliver the insight that had been transferred by the original metaphor because, as Searle remarked (1979, p.123) "metaphorical utterances... are intrinsically not paraphrasable".
Catachrestic Substitution Metaphors

Once again, principal and subsidiary subjects are locked in a mutual interaction.

Here, however, there is no word or expression to denote the principal subject.\(^6\)

\(^6\) It can be thought of as a “zero-word”, in the sense that there is a concept, but no name for it.
expression for the missing concept. And, if this *catachrestic* substitution successfully plugs the lexical gap, the new sense quickly attaches itself to the substituting word or phrase, and it becomes entrenched in the literal lexicon.

87 Thus, instead of the more usual process of forming a new word from a novel combination of prefixes, root elements and suffixes, it coins the new label for the referent by drawing an analogy with some other thing. Some of these metaphors are transient, such as "iron horse" eventually becoming "train"; others, such as "Iron Curtain", become entrenched (Cohen, 1998, p.12).

Black (1962, p.33) labels this substitution process *catachresis* ("grammatical misuse") — e.g., using "militate" for "mitigate", "decimate" for "devastate", "our mutual friend" for "our friend in common", "chronic" for "severe", "anachronism" for "anomaly", "alibi" for "excuse", etc. — and it is helpful to think of these as *catachrestic substitution metaphors*.

88 "Mapping", as in mapping the source onto the target in an analogy, is an example. According to the Oxford English Dictionary, the first English use of the word *lilac* was recorded in 1625CE. It was exclusively used to denote the shrub itself, and nothing more. Its meaning was extended to denote the specific colour of the flower of the lilac shrub in 1791CE, and to simply denote the colour in general was 1801CE (a similar process gave us our word for the colour *orange*). This *catachrestic substitution metaphor* has been so successful that the word lilac is now just as correctly (and un-metaphorically) applied to the colour as it is to the shrub.

When it became necessary to name the special class of planetary satellites — e.g., Prometheus and Pandora (Saturn); Cordelia and Ophelia (Uranus) — that lie on the margins of planetary rings, and act to confine ring particles to their correct orbit, the term *shepherd satellite* (also known as "shepherd moons") was chosen. Now we can speak of Promethius and Pandora as the *inner shepherd satellite* and the *outer shepherd satellite* respectively for Saturn’s F-ring, and of Cordelia and Ophelia as the *inner shepherd satellite* and the *outer shepherd satellite* respectively for Uranus’ epsilon-ring; with the inner shepherds accelerating those particles that are tending to move too far inwards (causing them to move outwards), and the outer shepherds slowing down those particles that are tending to move too far outwards (thus, causing them to move inwards).

Two further examples are:
(1) computer: in the 1940s, *computers* were not machines at all, they were human beings who performed *computations*, essentially with pencils and paper (especially those, usually female, who calculated the settings required for naval cannon).
(2) virus (as in computer virus).

Sutton (1993) examines a number of important issues connected with the selection and coining of scientific terminology.

89 Thus, "atheists can still hope for *angels of mercy*, despite their theology" (Cohen, 1998, p.10), and "one who refuses to believe in spirits may still complain of *ghosts* on his television screen or employ a *ghost* writer to assist him in his memoirs" (Bache, p.191).
Catachrestic substitution metaphors are very productive, often creating subordinate metaphors through the shifts of meaning that have been generated by the initial metaphor.⁹⁰

---

⁹⁰ When biologists chose the term “code” to represent certain genetic aspects of the DNA molecule, they were later able to expand their theoretical horizons on the basis that this genetic “code” could also be thought of as having “letters”, “words”, “sentences” and “punctuation”; all of which amplified the scope of their research, but would not have come to mind without the original “code” metaphor.

And, once scientists had developed the metaphor of a ship’s rope ladder to represent the peculiar structure of the DNA molecule, further descriptions were possible, such as:

> The structure of the DNA molecule has been likened to a twisted rope-ladder, the sides of which consist of sugarphosphate chains, the rungs of linked nitrogenous bases. The rungs consist of complementary base pairs linked by hydrogen bonds. (Uvarov, Chapman and Isaacs, 1979, p.110).

There is an important distinction between the “code” metaphor of DNA, and other uses of metaphor. A biology teacher, asked to explain something about the workings of the endocrine system, might make useful statements such as “the pituitary gland is the conductor of the endocrine orchestra”.

However, in this case, the teacher has a large number of other ways of thinking about, and describing, his theory of the endocrine system; and, even if he alone had discovered the endocrine system, it is most unlikely that his metaphor of the orchestra played any part whatsoever in his discovery.

The orchestra metaphor serves as a way of representing a difficult and obscure theory to his students in a more easily understood fashion, so that his particular point is far clearer and far easier to apprehend.

However, with DNA, the metaphor is not an epiphenomenon (i.e., an external, irrelevant device for explaining the theory). The metaphor was an indispensable component of the theory’s development.

So, our biology teacher, if asked a question relating to genetic theory, might reply: “Just for a moment, think of genes as if they are a sort of “code” that lies within each molecule”.

Here, our teacher has entered a different area, and has no other way of thinking about DNA. The “code” metaphor was so essential to the formation of our present concept of genes, it is inseparable from the concept itself; and, so, he really has no other way of describing it.
In the same way that false notions or, even, plain mistakes often lead to new ideas and results, there's often a very large chance element in the choice of possible subsidiary subjects; some work, some don't, and others work in a way not intended by the author.

91 Howerton (1963, p.192) draws a clear distinction between an experimental error ("previously unknown phenomena giving spurious results") — Merton's (1936) "unanticipated consequences" — and an experimenter's mistake. Two examples of "experimenter's mistakes" are:

1. the contamination in Fleming's laboratory leading to the isolation of penicillin, and
2. Charles Goodyear's accidental mixing of raw rubber with sulphur on his stove-top (which revealed the process of vulcanization).

In terms of science in general, it is important to distinguish between:

1. a discovery consequent upon an accident (such as Fleming's and Goodyear's);
2. an accidental (or serendipitous) discovery, such as Röntgen's discovery of X-Rays while researching cathode rays, and Bell's discovery of pulsars (originally called LGM, "little green men", by Bell) whilst trying to eliminate interference in radio telescope reception; and
3. the re-application of a previously discarded notion, such as the introduction of air-bags in cars. Air-bags were originally developed to save lives in air crashes; but a test involving an aeroplane crashing into a mountain showed that, in those circumstances, they were useless. However, someone in Detroit later recognized that the dynamics of a car crash were rather different.

There are also many examples of discoveries being later used in applications that were totally unintended by the discoverers: X-Rays (diagnostic radiography); laser beams (corrective eye surgery); the Jacquard loom (computer punched cards); microwaves (cooking appliances); and radar, originally conceived as a type of "death ray" that would incapacitate the pilot, disable the motors and detonate the bombs of an approaching aircraft (location devices or, even, a speed monitors).

92 "Vitamins" provide an excellent example of a concept going to unexpected places. Vitamins are "organic compounds which... when absent from the diet or not properly absorbed from the diet causes a specific deficiency disease" (Wagner and Folkers, 1964, p.6)

Casimir Funk, an organic chemist working at the Lister Institute in London, published a paper on beriberi in 1911 (Funk, 1911). In 1912, he published another paper surveying the literature on disorders that were, at that time, considered by many to have a dietary origin. He concluded that:

Beriberi, pellagra, rickets and scurvy were diseases of nutritional origin which could be prevented or cured by protective factors present in natural food. He proposed the existence of four such factors: an antiberiberi factor, an antipelagra factor, an antirickets factor, and an antiscurvy factor. (Wagner & Folkers, 1964 p.3)

As they were all amines and, obviously, vital (vs. lethal), Funk coined "vitamine" for these factors. In a 1914 paper (Die Aetilogie der Avitaminosen mit Besonderer Berücksichtigung der Physiologischen Bedeutung der Vitamine), he introduced avitaminosis — "primary avitaminosis" being dietary vitamine deficiency, and "secondary avitaminosis" being conditioned vitamine deficiency due to (a) malabsorption, (b) failure to meet increased nutritional requirements at times of special need (e.g., pregnancy), and (c) excessive excretion — as an alternative to what had been previously simply termed "deficiency disease".

Hopkins also published a paper in 1912 (later than Funk's) which described, in detail, a series of experiments in which control groups of rats were fed purified dietary mixtures. Other experimental groups were fed with "the addition of a minute ration of fresh milk" that, in his opinion:

confirmed the work of others in showing that animals cannot grow when fed upon so-called "synthetic" dietaries consisting of [unadulterated] mixtures of pure proteins, fats, carbohydrates, and salts [exclusively].

[However, my experiments, as described in this paper also show] that a substance or substances present in normal [viz. "impure"] foodstuffs (e.g. milk) can, when added to the [unadulterated] dietary in astonishingly small amount, secure the utilization for growth of the protein and energy contained in such artificial mixtures.

Hopkins (op.cit.) states "I have long known from my own observations that extremely small additions of [accessory food factors] were sufficient to induce growth", and his conclusion that "any effect of the [addition of the minute ration of fresh milk] upon appetite must have been secondary to a more direct effect upon growth processes"; thus, his "accessory food factors" were far more of a catalyst (and far less of a fundamental substance) than Funk's "vital amines". To save confusion between Hopkin's "accessory food factors", Funk's "vitamines", McCollum's "Fat-soluble A" and "Water-soluble B", Bottomley's "auximones", and Abderhalden's "nutramines", Drummond (1920) advocated the universal adoption of Funk's term on the condition that the final "e" was dropped (essentially because not all of this larger group were amines).
Metaphors in particular invite reification: the mental process of converting an abstract concept into a thing, and acting as if it had a substantial existence.

Schlinger (2003, p.24), identifies “intelligence” as an example of the reification of a discipline-specific construct, and speaks of “the error of [assuming] that just because the noun intelligence exists, that it refers to some underlying thing or entity”.

According to Star (1983), reification may also be an unintended consequence of the process of converting ill-structured problems into well-structured problems.

"Vitamin" identified a specific entity: something which, when absent from the diet, caused disease. Over time, the concept has turned into its polar opposite and, today, for most people, vitamins are considered to be performance enhancers (viz., something which, when present, creates robust health).

This view, no doubt, is largely due to the influence of Pauling’s (1968, p.265) claim that “mental disease” could be treated by “the provision of the optimum molecular environment for the mind”. Naming the practice “orthomolecular therapy”, Pauling advocated administering massive doses of ascorbic acid to schizophrenics.

This substance is, elsewhere, widely known as Vitamin C. It can be strongly argued, however, that it was a “chemical”, and not a “vitamin” that was being administered in such a large dose.

Pauling (awarded the 1954 Nobel Prize for Chemistry for research on the structure of molecules) was, obviously, still very attached to the word “molecule” in 1968 when he noted (p.271):

I might have described this [orthomolecular] therapy as the provision of the optimum molecular composition of the brain. The brain provides the molecular environment of the mind. I use the word mind as a convenient synonym for the functioning of the brain. The word orthomolecular [is the best way that I have found to express] the idea of the right molecules in the right amounts.

The term reification (lit. “thing-ification”, see Reber, 1985, pp.628-629; Hookway, 1993; and Labedz 2000), is often used interchangeably with Whitehead’s expression, the “fallacy of misplaced concreteness” (1929, p.11). It can also be termed either hypostatization or hypostasization.

When we reify objects we accept their existence as distinct things in the world, and this requires us to be able to count them and quantify over them...

Reber, 1985, pp.628-629:
From a purely rationalistic perspective [reification] is a cognitive/emotional act of children and other unsophisticated folk; in reality, it is one of the more seductive ways in which social scientists distort and misrepresent the status of many of their hypothetical entities and constructs.

In the same way, for example, that Sarbin (1964) has identified “anxiety”, and Szasz (1988) has identified “mental illness”, and Caldwell (1990) has identified “ego”, “id”, and “collective unconscious” as either reified metaphors or reified constructs.

Schlinger (2003), p.24:
When we speak of intelligence, we are not speaking of a thing or an entity that individuals possess and that determines their behavior. Intelligence is simply a descriptive term for an instance of behavior in a particular context or for a group of related behaviors (e.g., those called verbal, spatial, mathematical, etc.). We are fooled into believing that the term refers to a tangible entity because other words in our language, such as cat or table, have tangible referents. Intelligence, however, does not have a referent per se. Intelligence is like other constructs such as mind or personality in that the only objective referents are the behaviors that occasion the terms.

The reification of a metaphor (or construct) — thinking that simply because something can be named it actually has a substantial existence — is about as serious a mistake as that of claiming, on the basis of noticing that buttercups have 5 petals, delphiniums 8 petals, marigolds 13 petals, asters 21 petals, pyrethrum daisies 34 petals, other types of daisies 55 and 89 petals, etc., that the flowers in question have all understood, recognized, and intentionally striven to replicate a specific number from the Fibonacci series (0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, etc.).

[See also Ananthaswamy (2004) on the three sets of spirals found in cacti (compared of 3, 5, and 8 bumps from which the stickers protrude).]

Star (1983) pp.205-207:
Science involves the representation of chaos in an orderly fashion. When scientists (like everyone else) describe nature, their descriptions make the complicated turbulence of the world appear, at least in part, predictable and bounded. This process is central to all attempts to describe the world to oneself or others, as William James pointed out long ago. If you never sort the chaos of the world, it never becomes sensible...
Models

Models make **selected aspects** of an entity more intelligible, representing it some new, more effective way, by comparison with some other entity in some other domain. Like metaphors, models create relationships between the primary and secondary domains.

![Models: Interaction between Primary and Secondary Domain](image)

When compared with metaphor, models require a considerable level of factual knowledge of the secondary domain for their apprehension.

The secondary domain is chosen for its productive, distorting rôle: it selects, emphasizes, suppresses and/or organizes selected fundamental aspects of the primary domain in an innovative and significant way, which facilitates a certain new understanding of that primary domain. Yet, because a model uses selected aspects of a secondary domain, through analogy, to provide novel insights into aspects of a primary domain, each conclusion or inference that we draw from the workings of the model has to be monitored very carefully, for four basic reasons:

1. A model is a real world object in its own right. It has properties and attributes that are independent from its representation of the original.

---

Scientific work is complicated. Any set of scientific tasks involves multiple problems, qualifications, exigencies, demands and audiences. To work without getting lost in endless contingencies, scientists must draw boundaries and exclude some kinds of artefacts and complications from consideration.

In other words, part of doing science is transforming problems with many contingencies into those simple enough to work on. [According to Simon], computer scientists ordinarily distinguish between “well-structured” and “ill-structured” problems. Well-structured problems are those for which all possible contingencies can be programmed — no new contingencies arise as a result of the problem-solving process. Ill-structured problems, on the other hand, develop new and unpredictable contingencies in the course of solution.

In fact, the number of significant well-structured problems in the real world is almost nil. Scientists break ill-structured problems into pieces which they work on as if they were well-structured, in order to get the work done. Creating well-structured problems includes ignoring complexity: uncertainties in the environment, subjects’ reactions, unforeseen interaction effects. In order to actually do the research, lines and boundaries must be drawn around complications, implications, and exceptions. Goals, images and tasks simple enough to manage are developed...

The process of creating well-structured problems from ill-structured ones is an essential part of scientific work. However, *in conjunction with* the deletion of descriptions of this process from scientists' descriptions of their work, scientific "facts" become reified and their production histories lost.
(2) In terms of the aspects of the “original” that it represents, a model is always simpler, more organized than the “original”.

(3) Many of the “original’s” essential features have been discarded as irrelevant for the immediate purposes of the model’s representation.

(4) In terms of the details it includes, and its degree of correspondence to its “original”, a scale model is far less abstract in both its form and its degree of accurate representation than an analogue model, and an analogue model is far less abstract than an intellectual model.

As we move from scale model to analogue model, and from analogue model to intellectual model, the degree of comparative simplicity and abstraction increases; and, because there is a corresponding increase in the likelihood of each type of model introducing some distortion or irrelevant features of its own into the examination of the original, the need for some level of external validation, correction and confirmation also increases.

The secondary domain’s properties and attributes, which are better known, better understood, more organized, and more theoretically familiar (and, as a consequence, far easier to grasp), are used to communicate new information about selected aspects of a far less familiar, far more complex primary domain.98,99

98 In 2001, Bailer-Jones interviewed nine eminent scientists on the rôle and function of models in the philosophy and practice of science, and on how they perceived the inter-relationship between the technical terms “model” and “theory”. Although finding (2002, pp.297-298) that:

(1) “models are perceived as central to doing science”, and
(2) despite finding that “there is no uniform, and certainly no sharp, distinction between theory and model that is shared by scientific practitioners”, and that,
(3) although models are “no longer seen as merely heuristic devices or visual aids”, the “definitions of what a scientific model is are diverse”,
she also found considerable common ground amongst her informants (p.291); viz.:

(1) Modeling is widely thought to be central to doing science. This is a comparatively recent shift away from concern with theories.
(2) There is considerable diversity among definitions and descriptions of scientific models.
(3) Models are commonly characterized by simplifications and omissions with the aim to “capture the essence” of what is modeled.
(4) Models are thought to provide insight and they do not do this by purely matching data.
(5) While models have a range of features already specified, they are also expected to be subject to empirical testing.

99 And with all models, particularly with scale models, the “secondary domain” is far more accessible; this is by contrast with metaphors, where there’s a strong implication the listener’s already quite familiar with the primary subject, and the metaphor is simply being used to gain an additional, informative insight into that domain.
Whilst we must not take models too literally, we must always take them very seriously; and, to use models effectively, we must always proceed (at least temporarily) as though they were true.\textsuperscript{100}

Models must provide correlation rules\textsuperscript{101} which allow statements about the secondary domain to be translated into corresponding statements about the primary domain.\textsuperscript{102}

There are three major categories of model: (a) scale model, (b) analogue model, and (c) intellectual model.

**The Scale Model**

A scale model reproduces the dimensions of selected components of its “original” and the ratios between them according to a certain fixed scale. It employs a secondary domain — typically in the same medium, that has a known relationship with the primary domain — to illustrate some particular aspect (or clarify selected attributes) of the primary domain.

A scale model is tangible. It represents something that is tangible.\textsuperscript{103} And it always has a fixed relationship with its “original”.\textsuperscript{104}

\textsuperscript{100} Many important discoveries have been made when scientists commenced their work as if their theoretically postulated models of atoms, viruses, vitamins, hormones, and genes had actual, real world substantial existence. They proceeded as though each imaginary concept actually existed in precisely the form their theoretical speculation outlined; and, discarding any pretence of analogy, they proceeded with the view that the substantial, real world was exactly as they had theoretically described it.

\textsuperscript{101} Sokal & Sneath (1963) prefer the term “keys” for “correlation rules”.

\textsuperscript{102} Yet, even when correctly translated according to the correlation rules, inferences about the primary domain that flow from using the model, must always be carefully checked and independently calibrated against known (or predicted) real-world primary domain data. This ensures that any conclusion is about the primary domain, and is not just something peculiar to the secondary domain of the model.

\textsuperscript{103} For example, scale models of aircraft are used in wind tunnels to test wing design.

\textsuperscript{104} Thus, there are always correct and incorrect ways of interpreting any scale model's representation, which usually depend on the answers to five questions:

1. How many of the original’s features are included in the model?
2. What features stay the same: colour, shape, constituent materials, etc.?
3. What features are different: in particular, are normally associated factors being considered in isolation?
4. What is in relative proportions to the original: height, age, weight, etc.?
5. Have the ratios of the dimensions’ magnitudes to one another changed? (e.g., In many chemical and biological experiments, natural processes are artificially decelerated; changing the time dimension in proportion to all the other dimensions.)
Whenever scale models are used, mechanisms for external validation and correction must be constructed to ensure that any conclusions drawn from the model actually apply to the “original”; and are not just idiosyncratic features of the model.  

The Analogue Model

An analogue model describes specific relationships between selected components of the “original” by creating analogies with the relationships that are displayed by components in some other “secondary domain” of a totally different medium.

Analogue models provide a unified, easily remembered system of representing relationships.

---

105 It is essential to identify which parts of the model belong to the “model”, and which parts are there simply because they belong to the model as a real world entity.

For example, a 200:1 scale model of a mosquito would collapse under its own weight; but this would be a “model problem”, rather than a “mosquito problem”.

106 And, because they are isomorphic with the “original” for every incidence of every relationship demonstrated, analogue models are an ideal way to foster an understanding of complex networks of relationships.
Once again, because an analogue model’s connexion with its “original” is far less definite than a scale model’s, it is imperative that users are provided with “keys” to ensure that the “original” is always truthfully represented.\textsuperscript{107}

\textbf{The Intellectual Model}

Intellectual models are mental constructs,\textsuperscript{108} usually employing imaginary analogue (rather than scale) models, which attempt to account for real-world phenomena.\textsuperscript{109}

All thought experiments are intellectual models.\textsuperscript{110}

\textsuperscript{107} Consider the analogue model advanced to assist understanding of the behaviour of gases which suggests possible relationships between some theoretical activities of gas particles and some observable activities of billiard-balls. Achinstein (1964, p.332) reminds us that, despite thinking about gases in this useful way, “the physicist obviously supposes that molecules, not billiard balls, comprise gases”.

\textsuperscript{108} Hesse (1963, p.97) would rather call them “conceptual models”.

Intellectual models are never physically constructed; it is sufficient that they are imagined, and can be described.

\textsuperscript{109} With the so-called “billiard-ball theory of gases” — which suggested possible relationships between some theoretical activities of gas particles and some observable activities of billiard-balls — it is still very clear that, despite thinking about gases in this useful way, “the physicist obviously supposes that molecules, not billiard balls, comprise gases” (Achinstein, 1964, p.332).

\textsuperscript{110} Lawler (1996), p.243:
Each intellectual model has only an imaginary structure; and the model's creator may not be aware of certain real world practical inconsistencies, unless some kind of independent tests are available.

Many of the insights of natural scientists have been used, by analogy, to construct intellectual models to clarify the workings of other (scientific, moral or social) domains; and many theoretical insights have been gained from taking knowledge of newly created devices and projecting that knowledge onto the natural world.\textsuperscript{111}

\textit{Thinkable models are descriptions of things and relations simple enough for use as tools of thought and as the basis of thought experiments.}

\textsuperscript{111} Once the workings of engines and magnets were well understood, aspects of the Sun could be productively explained as if it were an engine, and certain aspects of the Earth as if it were a magnet.

See Paton et al. (1994) on using biologically oriented metaphors to describe aspects of computing.

Marakas and his colleagues (2000) discuss the specific implications of anthropomorphic descriptions of technology (where computers, for example, "read", "write", are "friendly", "catch and transmit viruses", etc.), and technomorphic descriptions of humans (where people are said to "interface" with one another, need "reprogramming", need to be "hardwired" to behave in certain ways, etc.).

The table below lists some of the intellectual models, derived, by analogy, from generally well-understood mechanisms, that have been used, with varying degrees of success, by scientists from time to time as vehicles for clarifying and explaining their particular points of view:

<table>
<thead>
<tr>
<th>THEORETICAL DOMAIN</th>
<th>KNOWN SECONDARY DOMAIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animal and human body</td>
<td>Machine (Descartes 1637)</td>
</tr>
<tr>
<td>Biological evolution</td>
<td>Social evolution</td>
</tr>
<tr>
<td>Bohr-Rutherford atom</td>
<td>Planetary systems</td>
</tr>
<tr>
<td>Brain nervous and system</td>
<td>Hydraulic mechanism (Descartes 1649)</td>
</tr>
<tr>
<td>Brain nervous and system</td>
<td>Electric device (Freud 1895)</td>
</tr>
<tr>
<td>Brain nervous and system</td>
<td>Computer (contemporary cybernetics)</td>
</tr>
<tr>
<td>Cloud chamber</td>
<td>Head on a glass of beer</td>
</tr>
<tr>
<td>Computer</td>
<td>Brain</td>
</tr>
<tr>
<td>Darwin's natural selection</td>
<td>Malthus' idea of struggle within society</td>
</tr>
<tr>
<td>DNA's genetic &quot;information&quot;</td>
<td>Secret encoded messages used by spies</td>
</tr>
<tr>
<td>DNA's molecular structure</td>
<td>Rope ladder</td>
</tr>
<tr>
<td>Electric current</td>
<td>Fluid</td>
</tr>
<tr>
<td>Kinetic theory of gases</td>
<td>Container of billiard-balls in motion</td>
</tr>
<tr>
<td>Light waves</td>
<td>Water waves</td>
</tr>
<tr>
<td>Organism</td>
<td>Society</td>
</tr>
<tr>
<td>Society</td>
<td>Organism</td>
</tr>
</tbody>
</table>
An intellectual model constructs an analogy of relation between a comparatively well-known secondary domain and a comparatively unknown primary domain, using the secondary domain as a "fact filter" to control our view of the primary domain:

![Fig.10: The Mechanism of the Intellectual Model](image)

Although the "secondary domain" may be, in the real world, far more complex than the "primary domain", the "secondary domain" is chosen because it is well-ordered and far better known.

My examination of the comparative cognitive processes of analogy, metaphor and models in this chapter has largely served the purpose of providing resources for the appreciation of how these processes figure in the functioning of thought experiments, which will be discussed in later chapters; however, an immediate benefit can be derived, at this point, in the transition to the concerns of Chapter Four.

As I emphasized at the start of this dissertation, I have been drawn to examine the cognitive aspects of thought experimentation, with the intention of developing a broad understanding of the wide range of activities that pass for thought experimentation. Working on the assumption that thought experimentation involves common cognitive
processes, I move on to consider, in the next chapter, a contextual framework for thought experimentation, grounded in human need and activity, and extending the cognitive orientation of the dissertation.
Chapter Four
Understanding: A Context for Thought Experimentation

In the last chapter I examined the comparative cognitive processes of analogy, metaphor and models with a view to elaborating certain key aspects of thought experimentation. In this chapter I want to provide a contextual framework for understanding thought experimental activity: namely, understanding itself.

Understanding is woven into human need and activity in ways that no doubt require significant analysis; mainly because human understanding is, in general, taken for granted, and is not commonly subjected to scrutiny by itself. My argument in this chapter is that certain aspects of understanding are crucial to our appreciation of thought experimentation; specifically, because they “drive” thought experimental activity. These aspects are: explanation, the need for control, and counterfactual thinking.

Explanation

According to Brewer et al. (1998, p.120), “an explanation is an account that provides a conceptual framework for a phenomenon (e.g., fact, law, theory) that leads to a feeling of understanding in the reader/hearer”;¹ and this “explanatory conceptual framework goes beyond the original phenomenon, integrates diverse aspects of the world, and shows how the original phenomenon follows from the framework”.²

¹ Trout (2002, p.230) believes that the essential feature of good explanations is their “systematic tendency to produce increasingly accurate theories”.

² Suppe (1998a, p.230) clearly distinguishes “explanations why” and “explanations how”: Most philosophical accounts of explanation tacitly or explicitly assume all explanations are equivalent to explanations why. But that is false since many explanations how cannot be reduced to explanations why. Indeed, many explanations why prove to be a sub-class of how possible explanations. Plate tectonics is a kinematic theory and so does not purport to explain why the plates move as they do.

For Tuchanska (1992 p.107), an explanation consists of two "crucial elements": the "existing knowledge that is modified" and the "modifications of [that] knowledge".

An act of explanation is an epistemic operation, an act of modifying knowledge in two senses. First, since it is an act of elaborating a body of knowledge (a theory) it leads to an enrichment of existing knowledge of the world. Second, when in the result of an explanation we learn why previous (theoretical) knowledge did not succeed in explaining the phenomena in question, the explanation leads to enrichment of metatheoretical knowledge.

For Friedman (1974, p.18-19), scientific explanations increase “our overall understanding of the world”:
Our total picture of nature is simplified via a reduction in the number of independent phenomena that we have to accept as ultimate.
According to Brewer, et al. (p.125) there are five types of explanation: causal, or mechanical;³ functional;⁴ teleological;⁵ intentional;⁶ or formal or mathematical.⁷

According to Brem and Rips (2000, pp.575-576), whilst “their roles are qualitatively different... both explanations and evidence are essential to our understanding and evaluation of claims”.⁸

They identified six advantages of explanations:⁹

(1) “[their] structure improves our [speed and] ability to comprehend, use, and recall information”;
(2) “[they] provide us with a quick-and-dirty method for vetting claims”;¹⁰
(3) “an explanation increases our ability to detect, as well as our confidence in, patterns of correlation or covariation”;
(4) “once patterns are detected, explanations promote generalizability”;¹¹
(5) “[their] promise of a stable connection between cause and effect improves our ability to abstract and transfer knowledge”; and
(6) “[they] help us understand a problem”.¹²

Scientific understanding is a global affair. We don't simply replace one phenomenon with another. We replace one phenomenon with a more comprehensive phenomenon, and thereby effect a reduction in the total number of accepted phenomenon. We thus genuinely increase our understanding of the world.

³ “Why did the light not come on? — Because it was not plugged in” (p.125).
⁴ “Why is there rubber on the end of pencils? — So that you can erase mistakes” (p.125).
⁵ “Why do fish have gills? – So they can take in oxygen” (p.125).
⁶ “Why did you cut off the light? — Because I intended to go to sleep” (p.125).
⁷ “Why does this emission line of hydrogen have this frequency? — Because of Balmer’s formula” (p.125).

Yet, as they remark (p.574):

Many people seem unable to make the conceptual distinction between the explanations of mechanisms underlying claims and the evidence that helps us determine whether those claims and mechanisms really hold. If this is so, overconfidence may arise not only because an explanation influences the search for and interpretation of evidence, but also because people believe that their explanation is evidence.

⁹ (1) to (5) are taken from Brem & Rips (2000), pp.575-576. (6) is taken from p.596.

¹⁰ Brem & Rips (p.575):

If we can elaborate a causal mechanism without running into internal inconsistencies or violating background knowledge, the probability that the claim is true may rise. Alternatively, if no satisfactory explanation is available, the perceived likelihood of a claim may decline.

¹¹ Brem & Rips (p.575):

Without an explanation to establish a reliable connection between cause and effect, we have no reason to believe patterns will persist except that they have done so in the past.

¹² Brem & Rips (p.595):
According to Brewer, et al., the characteristics of scientific explanations are that they “provide a theoretical/conceptual framework for a phenomenon”; 13 “go beyond the original phenomenon”; “integrate a range of phenomena”; “show how the original phenomenon follows from the framework”; “provide a feeling of understanding”; 14 and “must be, in principle, testable” (p.121).

And, whenever a phenomenon is explained, “there is a natural human tendency to evaluate the quality of the explanation”.

Read and Marcus-Newhall (1993, pp.429-431) offer five mutually interactive principles which, all things being equal, underpin all explanatory coherence:

(1) **Strength**: an explanation more coherent than its competitors is preferred; 15

(2) **Breadth**: an explanation that accounts for more of the evidence is preferred; 16

(3) **Parsimony**: the simplest explanation is preferred; 17

(4) **Explicability**: an explanation that can be explained is preferred; 18 and

(5) **Support**: an explanation supported by an analogy to another system with the

---

13 The function of which is “to reduce the arbitrariness of the phenomenon” (p.120).

Faye (1999), p.62:

Explanation is an intentional act of communication [designed to inform a questioner] about what he does not understand by providing some missing information, by making something probable, or by making abstract issues concrete.

14 Lakatos, (1963), p.131:

I recall Karl Popper distinguishing three levels of understanding. The lowest was the pleasant feeling of having grasped the argument. The medium level was when one could repeat it. The top level was when one could refute it.

15 The more coherent an explanation is — when compared with any competing explanations — the more it is preferred.

16 That is, the explanation that “explains more facts” (p.431).

17 That is, the explanation “requiring the fewest assumptions” (p.431).

18 "For example, an explanation in terms of someone's goals should be better if one can explain why they have those goals” (p.430).
same causal structure is preferred.¹⁹

Brewer and his colleagues offer eight important attributes (pp.121-125) which they consider “relevant to the issue of the judged quality of explanations” (pp.121):²⁰

1. **Empirical accuracy**: the degree to which empirical evidence is consistent with the explanation.

2. **Scope**: the degree to which an explanation accounts for a greater range of phenomena.

3. **Consistency**: the degree to which an explanation is internally consistent.

4. **Simplicity**: the degree to which an explanation is simple.²¹

5. **Plausibility**: the degree to which an explanation is consistent with the individual’s “larger background beliefs... about matters of theory and data” (p.122).

6. **Precision**: the degree to which an explanation produces “precise quantitative predictions” (p.124).

7. **Formalisms**: the degree to which an explanation is expressed in a quantitative, mathematical form.²²

8. **Fruitfulness**: the degree to which an explanation provides “guidance for future research” and is able to “open up new areas of research” (p.124).²³

Gopnik (1998) offers a valuable explanation for humans’ natural “explanation-seeking curiosity”.

---

¹⁹ Research has shown that “subjects gave higher ratings to explanations for unexpected events when subjects were reminded of previously encountered analogous unexpected events with analogous causal explanations” (Read & Marcus-Newhall, p.443).

²⁰ They note that, however, “the criteria used by scientists to evaluate explanations have changed over time” (p.124).

²¹ That is, assuming that factors like “empirical accuracy and scope are held constant”, an explanation is considered to be simpler “if it accounts for a phenomenon with fewer assumptions or hypotheses than a rival explanation” (p.122).

²² That is, by contrast with those explanations which are “expressed in a more qualitative, discursive form” (p.124).

²³ Trout (2002, p.230) adds a further dimension to the discussion with his assertion that the *best* explanations are always those that display a “tendency to produce increasingly accurate theories”; and, from this, that “only explanations capable of sustaining theoretical progress are good explanations”.
From intentional psychology and child developmental psychology — especially those recent findings which “[suggest] that children are sensitive to the underlying causal structure of the world, and seek to form causal representations at a much earlier age than we had previously supposed” (p.104) — she argues that all humans possess a specific cognitive system, which she labelled the “theory-formation system”.

And, because “the distinctive phenomenology of the theory-formation system impels us to action as well as to knowledge, it reflects a sort of theory-drive”(p.102).

This “drive system” is evident in the fact that “we not only know an explanation when we have one, we want explanations, and we are satisfied when we get them”; and this satisfaction involves a set of experiences that “goes beyond the merely cognitive” (p.109).

Moreover, the drive of this theory-formation system is so strong that she postulates (p.102) that it is an “evolutionarily determined drive”, and she proposes that:

**EXPLANATION : COGNITION :: ORGASM : REPRODUCTION**

From our phenomenological point of view, it may seem to us that we construct and use theories in order to achieve explanation or have sex in order to achieve orgasm.

From an evolutionary point of view, however, the relation is reversed, we experience orgasms and explanations to ensure that we make babies and theories.

---

24 In an aside, Gopnik (1998, p.112) remarks that, despite the possibility that scientists, deeply embedded within the “complex social organization of science”, might well “engage in something like theory use and theory formation without explanatory phenomenology” (or the additional possibility that “a whole group of scientists scattered over many places comes upon the correct theoretical answer to a question without any single scientist experiencing any phenomenology at all”), it is “[both] striking [and] comforting… to see the phenomenology of explanation persist even in relatively sophisticated and socially organized scientists. The scientists in the recent Mars probes almost without exception described their joy and excitement by saying it felt like being a little kid again. None of them said it felt like getting a raise.”

25 This view is highly resonant with the widely accepted views of Daniel Berlyne (1954, 1971, 1978) on the powerful motivating force of **curiosity**, and the extent that it drives us to seek answers.

Explanation is a goal-directed human activity. It depends on what is relevant or important to the explainer, it satisfies a special kind of explanation-seeking curiosity, it answers “why” questions. (Gopnik, 2000, p.300)

Gopnik also notes (1998, p.110) strong similarities with the Zeigarnik effect (see Zeigarnik, 1927/1938, and Savitsky, Medvec & Gilovich, 1997).

26 Or, as she expressed slightly differently, elsewhere: “my hypothesis is that explanation is to theory formation as orgasm is to reproduction — the phenomenological mark of the fulfillment of an evolutionarily determined drive” (Gopnik, 2000, p.300).

27 Gopnik (1996, p.498] makes it very clear that she is referring to the **male** orgasm.
The Need for Control

It is widely accepted that a sense of control is a “a robust predictor of physical and mental well-being” (Skinner, 1996, p.549); and, aside from the question of the value of having a sense of control in various life circumstances, the literature distinguishes between four significant dimensions of control (Skinner, p.551):

(1) **objective control, subjective control, and experiences of control**;  
(2) **agents of control, means of control, and ends of control**;  
(3) **retrospective control and prospective control**; and  
(4) **specific control and general control**.

---

Both experimental and correlational studies have shown that across the life span, from earliest infancy to oldest age, individual differences in perceived control are related to a variety of positive outcomes, including health, achievement, optimism, persistence, motivation, coping, self-esteem, personal adjustment, and success and failure in a variety of life domains.

29 For example:  
What is more adaptive in the face of insurmountable odds: Should a person maintain the perception of control, even if it is an illusion, or should a person admit the uncontrollability of the situation and give up? (Skinner, p.549)

30 Skinner, p.551:  
The most fundamental distinction in the literature on control is between **actual control**, or the objective control conditions present in the context and the person, and **perceived control**, or an individual’s beliefs about how much control is available...  
The reverse combination, in which people have high perceived control in objectively uncontrollable or chance-determined situations, is studied as [Ellen Langer’s] illusion of control...  
As opposed to actual conditions (objective control) or beliefs (subjective control), the experience of control refers to a person’s feelings as he or she is interacting with the environment while attempting to produce a desired or prevent an undesired outcome...  
Experiences of control are products of external conditions (e.g., the degree of contingency between actions and outcomes), subjective interpretations (whether a success is believed to indicate ability or luck), and individual actions. Prototypical experiences of this sort are referred to in the literature on causal reasoning as “generative transmission”, in which an individual intentionally exerts effort toward a goal and can feel the energy of the effort transmitted into the environment to produce the outcome. In the control area, these experiences are sometimes referred to as feelings of efficacy or experiences of mastery.

31 Skinner, p.552:  
Ends refer to the desired and undesired outcomes over which control is exerted, agents refer to the individuals or groups who exert control, and means refer to the pathways through which control is exerted.

32 Skinner, p.555:  
The distinction between retrospective and prospective control refers to the time orientation, that is, whether subjective control describes the past, present, or future. Control perceived in the present is described in time-neutral terms. (These labels can suggest that subjective control reflects mental processes, through the use of such terms as estimates, judgments, representations, and evaluations, or that it reflects cognitive constructions, through the use of terms such as beliefs, convictions, understanding, and sense of control.)  
Interpretations of past control are sometimes simply referred to as "retrospective control" but can also be used more narrowly to refer to beliefs about the causes of past outcomes, with terms such as explanations and attributions. Future control can be labeled prospective control, anticipatory control, or simply expectations of control or control expectancies.

33 Skinner, p.555:  
Control beliefs can be arrayed along a continuum from the extremely situation-specific to the extremely general or global. At the specific pole are control beliefs that are relevant only to certain episodes, interactions, or behaviors, such as being able to lift weights of a certain number of pounds or solve subtraction problems involving a certain number of digits. At the general or global pole are beliefs that span all outcomes and areas in life; these beliefs may be considered almost worldviews. In between are beliefs that focus on specific domains of life, such as health, work, school, marriage, and peer relationships.
Chapter Four: Understanding: A Context for Thought Experimentation — 83

Brickman and his colleagues (1982) identified four patterns of responsibility in the way people approached trying to help others (or, even, help themselves):

1. The *moral* model: "in which people are attributed responsibility for both creating and solving their problems" (p.370);34

2. The *compensatory* model: “in which people are not blamed for their problems but are still held responsible for solving these problems” (p.371);35

3. The *enlightenment* model: “in which people are blamed for causing their problems, but not believed to be responsible for solving them” (p.373);36 and

4. The *medical* model:37 “in which people are not held responsible for either the origin of their problems or the solution to their problems” (p.372).38

---

34 Because only the subject can produce change, all that’s needed for success is *proper motivation*. Success is one’s own doing, lack of success means that one was not in the right frame of mind. People are at most cheerleaders for one another. (p.371)

35 Because only the subject can produce change, all that’s needed for success is *sufficient power*: Actors see themselves and are seen by others as deprived or as suffering, not from their own deficiencies, but from the failure of their social environment to provide them with goods and services to which they are entitled... To solve their problems, and perhaps to compel an unwilling social environment to yield the resources necessary to solve them, actors must be assertive. In this they may need the help of peers or subordinates. If they receive training, it is designed to empower them to deal more effectively with their environment. (p.372)

36 Although usually unaware of this fact, subjects are uniquely responsible for the problem. They must accept “a strikingly negative image of themselves and, in order to improve, to accept a strong degree of submission to agents of social control” (p.373). For success, all that's needed is personal discipline.

   Actors see themselves as guilty or sinful, or at least responsible, by their past behavior, for suffering or a problem that they must endure in the present. It is their own impulses — to eat, drink, lie, cheat, steal — that are out of control. To control these impulses people must submit to the stern or sympathetic discipline provided by agents who represent the authoritative moral (and if necessary physical) force of the community. Since the solution to these problems lie outside the person, the solution can be maintained only so long as the relationship with this external authority or spiritual community is maintained. (pp.373-374)

37 Brickman, et al. (pp.372-373) explain their choice of the adjective "medical":

   [Within] the practice of modern medicine... patients are collections of organs that can malfunction or become infected. Drugs, surgery, and other treatments can be aimed directly at the distressed organs, ignoring the patient as a person. Neither the illness nor the treatment is the person's responsibility...

   The medical model in our sense refers not only to cases in which people are thought to be victims of disease but to all cases in which people are considered subject to forces that were and will continue to be beyond their control. Thus a [B.F.] Skinnerian view of determinism, that human behavior is determined by rewards and punishments in a way that makes it foolish to blame people for their problems or give them credit for solutions, is also a version of our medical model.

   Under the medical model, individuals with problems see themselves, and are seen by others, as ill or incapacitated. They are expected to accept this state, which in turn involves exempting them from their ordinary social obligations and imposing on them the responsibility for seeking and using expert help. Unless the illness is chronic or terminal, people are also expected to try to get well. The other actors believed necessary to bring about this change are experts (e.g., doctors) who have been trained to recognize what the problem is and to provide what service or treatment is available. Even when the solution is largely one that the person can or must carry out themselves, such as bed rest, the responsibility for prescribing this solution and for judging whether it has been successful rests with the expert.

   The advantage of the medical model for coping is that it allows people to claim and accept help without being blamed for their weakness...

38 Because the subject is "not responsible for their problem and cannot be expected to take care of it by themselves", and are allowed to "claim and accept help without being blamed for their weakness" (p.373), all that is needed for their success is *treatment delivered by the appropriate expert*. 
Table Five: The Consequences of Attribution of Responsibility
in Four Models of Helping and Coping
[adapted from Brickman, Rabinowitz, Karuza, Coates, Cohn & Kidder (1982), p.370]

<table>
<thead>
<tr>
<th>Level of attribution to the self of responsibility for the problem</th>
<th>Level of attribution to the self of responsibility for the solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH</td>
<td>Moral Model</td>
</tr>
<tr>
<td>Perception of Self</td>
<td>Lazy</td>
</tr>
<tr>
<td>Actions Expected of Self</td>
<td>Striving</td>
</tr>
<tr>
<td>Others besides self who must act</td>
<td>Peers</td>
</tr>
<tr>
<td>Actions expected of others</td>
<td>Exhortation</td>
</tr>
<tr>
<td>Implicit view of human nature</td>
<td>Strong</td>
</tr>
<tr>
<td>Potential pathology</td>
<td>Loneliness</td>
</tr>
<tr>
<td>LOW</td>
<td>Compensatory Model</td>
</tr>
<tr>
<td>Perception of Self</td>
<td>Deprived</td>
</tr>
<tr>
<td>Actions Expected of Self</td>
<td>Assertion</td>
</tr>
<tr>
<td>Others besides self who must act</td>
<td>Subordinates</td>
</tr>
<tr>
<td>Actions expected of others</td>
<td>Mobilization</td>
</tr>
<tr>
<td>Implicit view of human nature</td>
<td>Good</td>
</tr>
<tr>
<td>Potential pathology</td>
<td>Alienation</td>
</tr>
</tbody>
</table>

Overall, a view has emerged in the literature that the highly necessary subjective perception of being in control depends upon knowing the answers to two questions:

(1) “Am I responsible, or am I not responsible for this event?”, and

(2) “Am I able, or am I not able to I fix this problem?”

The search for answers to these eternal questions is the basis of the ancient ritual practice of ‘divination’. the asking questions of oracles.

---

39 In a similar vein, Frazier, Berman & Steward (2002, p.207) identified three dimension of control: (1) Past control (“Could I have prevented this?”); (2) Present control (“What can I do now?”); and (3) Future control (“Can I keep this from happening again?”).

40 Cicero (I.v.9: 1923, p.233) defined ‘divination’ as “the fore-telling and the fore-seeing of those things which are thought to happen by chance”.

---
Cicero (I.xviii.34: 1923, p.263) distinguished between two types of divination: *divination by artifice*,\(^{42}\) and *natural divination*.\(^{43,44}\)

Although divination is dismissed as irrational *pseudoscience*, it is also, obviously, a structured, theory-driven ritual activity that has been intentionally used by vast numbers of people, from vastly different civilizations (with vast individual differences in their levels of education, religious belief, and social circumstances), over many thousands of years, in order to meet a specific need, consistently felt across different cultures, different geographical locations, different languages and different historical eras, to gain a particular sort of knowledge (and, thus, control), that has been broadly described by Evans-Pritchard (1937, pp.69-70) as the answer to two core questions:

"*Why me?*” and "*Why now?*”

As it seems that we all have a natural propensity to ask these sorts of questions of ourselves and others, it’s not surprising that thought experiments happily ask precisely the same sort of questions.\(^{45}\)

Thought experiments involve an imagined situation, within a specific domain (D),

\(^{41}\) The sorts of questions asked: ascribing responsibility for an illness — because an illness caused by witchcraft requires a different treatment from that caused by, say, a rotten food; similarly, a medical registrar needs to determine whether a young woman’s abdominal pain is caused by, say, appendicitis, pre-examination nerves, indigestion, food poisoning, hernia, mittelschmerz, menstrual cramp, ovarian cyst, appendicitis, an ectopic pregnancy, a phantom pregnancy, colic, diverticulitis, or Crohn’s disease before any treatment can be delivered — the fertility of one’s wife, whether one should marry a particular woman, whether one should marry at this time, strategic and military questions, the likely success of a business venture, the likely outcome of doing Y, the likely outcome of not doing Y, the likely success of a hunting venture, how another person will act in response to a particular set of circumstances, why did X happen? why did X happen now? Could X have been avoided? Is Z inevitable? etc. (Evans-Pritchard, 1937, pp.261-262; Shchutskii, 1979, pp.191-192).

\(^{42}\) *Artificiosa divinatio*; "which is allied with art", and performed by “those... who, having learned the known by observation, seek the unknown by deduction”.

\(^{43}\) *Naturalis divinatio*; "which is devoid of art", and performed by “those... who, unaided by reason or deduction or by signs which have been observed and recorded, forecast the future while under the influence of mental excitement, or of some free and unrestrained emotion”.

Noting that “this condition often occurs to men while dreaming and sometimes to persons who prophesy while in a frenzy” (which he terms *furor oraculorum*), he also clearly states that “oracles given by “equalized lots”" (like Chinese "oracle bones") are *divination by artifice*, unlike “those [oracles which are] uttered under the impulse of divine inspiration”, which are a form of *natural divination*.


\(^{44}\) In this context, when compared with natural “creative insights” (such as Kekulé and the benzene ring), it is very clear that thought experiments are an *artifice*.

\(^{45}\) Albeit in a “scientific” setting.
where certain elements are considered significant (and others irrelevant);\(^{46}\) where certain facts are known, certain regularities have been established and in which a specific event \((E)\) takes place.\(^{47}\)

Three different sorts of question are asked in thought experiments; each serving a different orientation, application and purpose:

1. **Prefactuals**\(^{48}\) are future-oriented and concerned with what might happen:

   "What will be the outcome in \((D)\) if \((E)\) occurs?"

2. **Counterfactuals**\(^{49}\) are past-oriented and concerned with what might have been different:\(^{50}\) "Given that \((O)\) has already occurred, what might have been the outcome in \((D)\) if \((A)\) had happened instead of \((E)\)?"\(^{51,52}\)

---

\(^{46}\) The matter of what is relevant and what is irrelevant is, often, designer-specific and may not be understood by, or shared with the user.

\(^{47}\) This event \((E)\) may involve either an entity or an action (which may be either present or absent). Also, the intellectual activity is either future- or past-oriented; and, if the activity is past-oriented, it is accepted as a matter of fact that specific outcome \((O)\) has already occurred.

\(^{48}\) No doubt influenced by a need to distinguish these sorts of thought experiment from both **semifactuals** and **counterfactuals**, Sanna (1998, p.635) first used the valuable, symmetrical term **prefactual** in 1996 to label the act of "mentally simulating alternative possible outcomes... before an event" which, he strongly feels, is a "pervasive and ubiquitous human tendency":

   Contemplating an important business meeting with a valued client, tentatively preparing for an upcoming date, or anticipating the imminence of a golf match against a vaunted rival are but a few examples of common situations that can elicit simulations of alternative possible outcomes.

\(^{49}\) Consequent upon Chisholm’s (1946, p.289) description of a "contrary-to-fact conditional" (on the same page he speaks of "counter-factual information"), the term **counterfactual** ("contrary to established fact") was introduced by Goodman in 1947 ("The Problem of Counterfactual Conditionals").

   His original ideas have been further discussed by Parry (1957), Goodman (1957a), Cooley (1957), Goodman (1957b), and Lewis (1973/2002).

   Hume (1748/1999, (7), p.146) also referred to a counterfactual case when defining causation;:

   ...we may define a cause to be an object, followed by another, and where all objects, similar to the first, are followed by objects similar to the second. Or in other words, where, if the first object had not been, the second never had existed.

\(^{50}\) Stalnaker (1981), expanding on a suggestion made by Frank Ramsey, describes counterfactual thought experiments as a process which involves "[adding] the antecedent (hypothetically) to your stock of knowledge (or beliefs), and then [considering] whether or not the consequent is true" (p.43).

\(^{51}\) Thorson and Sylvan, 1982, pp.540-541:

   "What would have happened if?" sorts of questions are scientifically significant since it is through the analysis of counterfactuals that cause is distinguished from mere correlation.

   Pratchett, Cohen & Stewart, 1999, p.48:

   What should a really good historian tell you about the assassination of Abraham Lincoln? A very average historian would name the murderer and the date. A good historian would tell you what make of gun the assassin used, how he concealed it under his coat, why he hatched the plot to begin with. But an excellent historian would tell you what could have happened had Lincoln only been injured.

\(^{52}\) Cowan and Foray (2002, p.542) argue that counterfactual conditionals belong to a group of six mutually interchangeable, equivalent types of scientific statement. They demonstrate their point with the following statements (based on distinctions made by Nelson Goodman in his (1984) fourth edition of *Fact,*
(3) Semifactuals\textsuperscript{53} are past-oriented, and concerned with what might have remained the same: “Even if (E) had not happened, would (O) still have occurred in (D)?”

| Table Six: The Typical Questions Examined by Each Type of Thought Experiment |
|-----------------------------|---------------------------|---------------------------|
|                             | PREFACTUAL                | COUNTERFACTUAL             | SEMIFACTUAL               |
| + Action                    | What would happen to Michael if you tell the police what you know? | If Newton and Leibniz had cooperated with each other, what would mathematics look like today? | Even if the goalie had moved left could he have intercepted a ball travelling at such a speed? |
| - Action                    | What would happen to Michael’s neighbours if you remain silent about his illegal activities? | If the A bomb had not been used on Japan, would we still know as much today about the effects of full-body exposure to radiation on human beings? | Even if the goalie had not anticipated a left-side kick, would he still have been quick enough to save the goal? |
| + Entity                    | What would happen if the watercolour was hung in bright sun-light? | If his own son had also been molested, would the inspector still have stopped the man from shooting the priest? | Even if you had your umbrella with you, would you still have got wet in such a fierce downpour? |
| - Entity                    | What would happen if the plant was totally deprived of sunlight? | If the inspector had not been a Roman Catholic, would he still have dismissed the complaints brought against the priest? | Even if you had not forgotten your umbrella would John still have given you a lift home from the station? |

Although scientists and philosophers operate in different domains and pursue entirely different goals,\textsuperscript{54} they approach their problems in a strikingly similar fashion:

\textit{Fiction and Forecast}) all derived from a typical, standard assertion that ”the number of patents filed will increase by 10% for every 25% increase in R&D spending”:

1. the counterfactual conditional: “If R&D spending were 25% higher than it is, patenting would be 10% higher”;
2. the scientific law: “The ratio of the change in R&D spending to the change in patents is 2.5:1”;
3. the dispositional statement: “Patenting has the tendency to increase (in the ratio 1/2.5) as R&D spending increases”;
4. the factual conditional: “Since patenting increased by 10%, R&D spending must have increased by 25%”;
5. the statement regarding possible worlds: ”In a world like ours but with 25% more R&D spending, we would observe 10% more patenting”; and
6. the causal statement: “Changes in patenting activity are caused by (\textit{inter alia}) changes in R&D spending”.

\textsuperscript{53} The term semifactual was also introduced in Goodman’s 1947 paper. His original concept has been subsequently developed and expanded by Cohen (1985), Barker (1991), and McCloy and Byrne (2002).

Whilst the assertion “If Princess Diana had married a more compassionate man than Prince Charles, she would have been happier” is a typical counterfactual, the assertion “Even if Princess Diana had married the Dalai Lama, she still would have eventually become discontented” is a semifactual.

Barker (1991, p.1) informs us that, in the case of semifactuals:

Most writers have singled out \textit{even} as the crucial word. Hence semifactuals have been known as “\textit{even-if}-conditionals”.

Because counterfactual and semifactual statements could both well be false in a factual sense, Goodman distinguished between the two on the basis that counterfactual statements \textit{affirm} ”that a certain connection obtains between antecedent and consequent”, whilst semifactual statements \textit{deny} ”that a certain connection obtains between antecedent and consequent” (p.115).

\textsuperscript{54} Schick and Vaugn (2003):

Scientists try to explain the causal relations among events, while philosophers try to explain the logical relations among concepts. (p.17)
Table Seven: Socratic & Scientific Five-Step Method for Analysing Concepts & Events
[Based on Schick and Vaugn, 2003, pp.15-17]

<table>
<thead>
<tr>
<th>Five Step Method</th>
<th>FIVE STEP (SOCRATIC) METHOD FOR ANALYSING CONCEPTS</th>
<th>FIVE STEP (SCIENTIFIC) METHOD FOR ANALYSING EVENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Identify a problem or pose a question.</td>
<td>Asks &quot;How is it possible for concept X to apply?&quot; &quot;What makes X apply to something?&quot; &quot;What is the logical relationship between X and Y?&quot;</td>
<td>Asks &quot;How is it possible for event X to occur?&quot; &quot;What makes X occur?&quot; &quot;What is the causal relationship between X and Y?&quot;</td>
</tr>
<tr>
<td>2. Propose a hypothesis.</td>
<td>Specify the necessary or sufficient conditions for applying the concept in question. Try to identify the features shared by all and only those things to which the concept applies.</td>
<td>Specify the necessary or sufficient conditions for the event's occurring. Try to identify the features shared by all and only those things that cause X.</td>
</tr>
<tr>
<td>3. Derive a test implication.</td>
<td>Asks &quot;What if the hypothesis were true?&quot; &quot;What does it imply?&quot; &quot;What does it commit us to?&quot; Test implications have the following form: If hypothesis H is true, then concept X should apply in this situation.</td>
<td>Asks &quot;What if the hypothesis were true?&quot; &quot;What does it imply?&quot; &quot;To what does it commit us?&quot; Test implications have the following form: If hypothesis H is true, then event X should occur in this situation.</td>
</tr>
<tr>
<td>4. Perform the test.</td>
<td>Determine whether the concept applies in the situation envisioned.</td>
<td>Produce the situation in the laboratory or look for it in the field.</td>
</tr>
<tr>
<td>5. Accept or reject the hypothesis.</td>
<td>If the concept applies in the situation envisioned, there is reason to believe that it is true. If it doesn't apply, there is reason to believe that it is false. In that case, either reject the hypothesis or go back to step (2) and revise it.</td>
<td>If the event occurs in the situation specified, there is reason to believe that the hypothesis is true. If it doesn't apply, there is reason to believe that it is false. In that case, either reject the hypothesis or go back to step (2) and revise it.</td>
</tr>
</tbody>
</table>

Thought experiments "test claims about the logical relations between concepts by helping us determine whether the claims hold universally" (p.41).

However, it’s also true that "inventing a thought experiment involves a creative leap of the imagination that cannot be dictated by a set of formal rules". In particular, "the most difficult part of performing a thought experiment is deriving the test implication because there is no formula for deriving one" (p.37).  

While philosophers are in the business of trying to identify the necessary or sufficient conditions for the application of concepts, scientists are in the business of trying to identify the necessary or sufficient conditions for the occurrence of events. (p.16)

Scientists explain how it’s possible for an event to occur by specifying the causally necessary or sufficient conditions for its occurrence.

Philosophers, on the other hand, explain how it’s possible for a concept to apply by identifying logically necessary or sufficient conditions for its application. (p.17)

And, yet, as Dennett (1993, p.4) explains, "sometimes an impossibility in fact is theoretically more interesting than a possibility in principle".

Thought experiments are constructed and polished according to this five-step method.

55 Schick and Vaugn (2003): The German philosopher Edmund Husserl called thought experiments "free fancies" because the situations involved are often produced by the free play of the imagination. But even though thought experiments can be fanciful, they are not frivolous, for as Husserl recognized, "fiction is the source from which the knowledge of 'eternal truths' draws its sustenance". To determine whether a conceptual claim is true, we have to determine whether it holds in all conceivable situations. And to determine that, we have to go beyond the actual to the possible. (p.37)

Some thought experiments describe situations that are physically impossible. That is not necessarily a strike against them, however, for their more fantastic aspects may not be relevant to their outcome. Thought experiments examine the logical relations between concepts, and abstracting from physical reality is sometimes necessary to throw those relations into proper relief. Of course, the more outlandish a thought
If a philosophical theory "[explains] how it's possible (or impossible) for a concept to apply by identifying the conditions for applying it", then a thought experiment "[tests] such theories by determining whether these conditions are necessary or sufficient for the application of the concept" (Schick & Vaugn, p.36); i.e., if a situation can be imagined where:

1. the concept applies but the conditions are not met then the conditions are not necessary conditions\(^\text{57}\) for the application of the concept; and
2. the conditions are met but the concept does not apply then the conditions are not sufficient conditions\(^\text{58}\) for the application of the concept.

According to Schick and Vaughn (p.41), thought experiments can be trusted, specifically because of the "conceptual competence" that all humans share:

Having a concept gives us the ability to make accurate judgments about its applicability, even in imaginary situations\(^\text{59}\).

---

\(^{57}\) Necessary conditions are those conditions in the absence of which an event can not take place . . . They are necessary in the sense that (i) they produce outcomes that are an inevitable consequence of the way that things naturally are, and (ii) the contrary is impossible. Clouds are a necessary condition of rain; because, if clouds are absent, the event in question (rain) will certainly not occur.

\(^{58}\) Sufficient conditions are those in the presence of which an event is certain to take place . . . With rain, even if the necessary condition (clouds) is present, there's no guarantee rain will occur. Whilst we know it can't rain without clouds, we also know that there are often clouds without rain. *Ipso facto*, clouds are not, in themselves, a sufficient condition for rain to occur. The necessary condition (presence of clouds) must be combined with other factors that are also necessary conditions (e.g., clouds exceeding a certain saturation point, air temperature being more than freezing point etc.) in order to create a sufficient condition for rain to occur. Thus, in the case of rain, the sufficient condition can be said to be a conjunction of more than one necessary condition.

\(^{59}\) Schick and Vaughn (2003, p.41):

We acquire a concept by being given a definition of it or by being shown examples of it. In either case, once we have a concept, we have the ability to apply it to things we have never before encountered. If we have the concept of the letter *A*, for example, we can apply it to typefaces we have never seen before. A thought experiment is like a newly encountered typeface. Just as we can trust our judgment to determine whether the concept of the letter *A* applies to a letter in a new typeface, so we can trust our judgment to determine whether a particular concept applies to the situation described in a thought experiment.

Of course, the more flourishes the letters in a typeface have, the more difficult it will be to determine whether a letter is an *A*. Similarly, the more outlandish the thought experiment, the more difficult it will be to determine whether the concept in question applies. So not all thought experiments have equal evidentiary value. Some are more persuasive than others.

To have a concept is to be able to apply it correctly. But we may be able to apply a concept without being able to state the criteria we use in applying it. For example, we may be able to identify a grammatical sentence without being able to state the rules of grammar. In such a case, we have an intuitive understanding of grammar even though we do not have a theoretical understanding of it. In attempting to identify the conditions for applying a concept, we are trying to transform our intuitive understanding into a theoretical one. That is, we are trying to make explicit what is implicit in our understanding of a concept. Because having the ability to apply a concept correctly doesn't necessarily give us the ability to state the conditions for applying it, different people may have different theories about what those conditions are. But because we
Chapter Four: Understanding: A Context for Thought Experimentation

Counterfactual Thinking

If god had not made yellow honey, they would say that the fig was much sweeter. Xenophanes (6th Century BCE)\(^\text{60}\)

The modern interest in the psychological study of the all-pervasive cognitive activity, counterfactual thinking,\(^\text{61}\) began with Kahneman and Tversky’s (1982) study of people’s habitual use of mental models involving “If only…” thinking when trying to estimate probabilities in uncertain circumstances.\(^\text{62}\)

Based on a view that “in general, frequent events are easier to recall or imagine than infrequent ones”, earlier work (1982, p.164)\(^\text{63}\) found that, when estimating probabilities, people used an “availability heuristic”\(^\text{64}\) (viz., “the ease with which instances or associations could be brought to mind”) which seemed to be an index of “the frequency of classes or the likelihood of events”.

Tversky and Kahneman found a strong cognitive bias in reasoning towards information that far more readily comes to mind, is more salient and is, thus, more mentally available.\(^\text{65}\)

Given the narrative of a death in a motor accident, and asked to (counterfactually)

\(^{60}\) Taken from Barnes (1987) p.94.

\(^{61}\) Literally, “contrary to established fact”. Roese & Olson (1995a) p.vii:

Counterfactual thinking is something familiar to nearly everyone. Even if they have not previously heard the term counterfactual, people instantly recognize it, once it has been defined for them, as something with which they are intimately acquainted. Few indeed have never regretted some action nor yearned to have avoided some circumstance. But it is the childlike wonder with which we gaze upon “what might have been”, into realms of possible, alternative worlds, which truly underlies the excitement of counterfactual research. What if Kennedy had survived his assassin’s bullets into a second term in the White House? What if the Nazis had triumphed over the western democracies in the Second World War? What if your parents had never met? There is something at once obsessively compelling and oddly unsettling about confronting such unrealities that might well have been.

\(^{62}\) For example see Appendix Five and Appendix Six.

\(^{63}\) This experimental work had taken place circa 1973.

\(^{64}\) Tversky & Kahneman (1982, p.163) clearly state that their availability heuristic includes the notion of association: “one may estimate probability by assessing availability, or associative distance”.

\(^{65}\) Sherwin (1999), p.1191:

We focus quickly on information of this kind, at the expense of more remote or abstract considerations that also bear on accurate calculation. To give just one example from [Tversky and Kahneman’s] studies, people commonly underestimate the time necessary to complete a project. The best explanation for this error is that they concentrate their attention on the steps they themselves will need to perform and, as a result, fail to consider the probability that external events may cause a delay. The actor’s own tasks are salient, while background risks are relatively obscure.

imagine how things could have been otherwise, subjects consistently supposed that, of all possible permutations, if the most unusual condition was altered that the accident would not have occurred.

In the case of the accident, they assumed subjects had been trying to undo the “past outcomes by altering their causal antecedents” (Trabasso & Bartelone, p.904); and found unusual events to be more easily converted (more mutable) than usual events.

Noting that “there appear to be many situations in which questions about events are answered by the operation that resembles the running of a simulation model” — and based on further examination of subjects’ “mental undoing of past events” (Kahneman & Tversky, 1982, p.202) — they postulated a “simulation heuristic”, which attributed levels of causation according to the ease with which scenarios (that would undo the event in question) were brought to mind.

Based on this work, Kahneman and Miller (1986) developed their “Norm Theory” which proposed that, given a set of circumstances, subjects produce a set of counterfactual

---

Participants were asked to imagine what could have prevented the death of a fictional agent, Mr. Jones. Participants read a story in which Mr. Jones leaves his office at a certain time, takes a particular route, stops at an amber light at an intersection, and resumes his journey when the light turns green. Mr. Jones is killed when a speeding truck, driven by a teenaged boy who is on drugs, strikes his vehicle in the intersection. Across two story versions, Kahneman and Tversky varied whether the time of departure or the route taken was normal or unusual. Following the reading of the narrative, the participants were asked to complete the thoughts of a family member in response to an “If only. . .” stem. The participants’ completions were analyzed according to which one of four conditions (time of departure, route taken, the stopping at an intersection, or the state of the truck driver) was mentioned first.

Not a single subject mentioned that if Mr. Jones had come to the intersection two or three seconds earlier he would have gone through safely.

Trabasso & Bartelone (2003), p.904 (emphasis added):
[Kahneman & Tversky] assumed that the conditional part of a counterfactual requires conversion of what was done (e.g., taking a regular route) to an alternative that was not done (e.g., taking of an unusual route) in order to undo the consequence (e.g., Mr. Jones’s not being killed). They assumed that the conversion of an unusual condition to its normal alternative was easier to accomplish than the conversion of a normal condition to its unusual alternative. The difference in the ease of conversion was the reason for the observed preference in selection of unusual over normal conditions as causes.

68 Roese & Olson (1995b), p.12:
In order for an antecedent to be judged as causal, it must be mutable and it must undo the outcome.

69 Later (e.g., p.202) although they speak of the “mental simulation” as being a scenario, it is clear from the context of their paper that they are referring to a thought experiment — rather than one of Kahn’s scenarios — and, in fact, in the cases they offer, they are referring to counterfactual thought experiments. See also, Dougherty, Gettys & Thomas (1997).

70 Kahneman & Tversky (1982), p.202:
To test whether event A caused event B, we may undo A in our mind, and observe whether B still occurs in the simulation. Simulation can also be used to test whether A markedly increased the propensity of B, perhaps even made B inevitable. We suggest that a test of causality by simulation is involved in examples such as “You know very well that they would have quarreled even if she had not mentioned his mother.”
alternatives\textsuperscript{71} with which they then used evaluate the original situation.\textsuperscript{72}

They found that "unusual events evoke normal alternatives but that normal events do not evoke unusual alternatives and, as a result, unusual events are more likely targets for undoing" (Trabasso & Bartelone, p.905).\textsuperscript{73}

Counterfactual thinking has a significant preparative\textsuperscript{74} and emotional\textsuperscript{75,76} function and,

\textsuperscript{71} It is essential to recognize that "the norm theory conception of availability focuses on the relative ease of one's imagining, not the events themselves, but alternatives to the events... [and] according to norm theory, events that have easily imaginable alternatives are more likely to be mentally mutated than events for which no such alternatives exist" (Seelau, Seelau, Wells & Windschitl, 1995, p.61).

\textsuperscript{72} Kahneman and Miller (1986, p.136):

...each stimulus selectively recruits its own alternatives and is interpreted in a rich context of remembered and constructed representations of what it could have been, might have been, or should have been. Thus, each brings its own frame of reference into being...

In addition... events are sometimes compared to counterfactual alternatives that are constructed ad hoc rather than retrieved from past experience.

\textsuperscript{73} Since this, researchers have investigated the rôle of "norms" in a wide range of decision making; especially on the extent to which certain features are considered to be far more mutable than others. For example, speaking generally, \textit{effects} are considered to be more mutable than \textit{causes} (it's more natural to say that a girl is big for her age, than young for her size); \textit{acts of commission} are considered to be more mutable than \textit{acts of omission} — Feigenson (1999-2000, p.764) argues that this natural propensity to recognize an "active versus passive distinction" (with jurors far more likely to find \textit{acts of commission} "causally and legally responsible" than \textit{acts of omission}) can be exploited by skilled, intelligent counsel in the way they represent circumstances and argue their cases, on the grounds that "just about any behaviour can be characterized as either an act or an omission" (e.g., being silent is an act, whilst refraining from speaking is an omission) — on the basis that it is easier to simulate the absence of a factual action than the presence of an action that was not in fact performed; \textit{acts of oneself} are considered to be more mutable than \textit{acts of another}, and so on.

\textsuperscript{74} Roese & Olson (1995c), pp.170-171:

[Counterfactual] thoughts may serve a preparative function; that is, they may illuminate means by which individuals may prepare for the future, and accordingly, improve their lot in the future. By manipulating alternatives to past actions, individuals can scrutinize and come to understand past mistakes as well as past triumphs, synthesizing them into prescriptions that may facilitate success in the future.

The mechanism underlying this function is based on the causal information contained in counterfactual propositions. Counterfactual conditionals are essentially causal statements. For example, the judgment that \textit{X} caused \textit{Y} may directly follow from the counterfactual inference that if \textit{X} had not been present, \textit{Y} would not have occurred (i.e., deleting antecedent \textit{X} undoes outcome \textit{Y}). If Bob fails an exam, then realizes that he would have passed had he only worked through the study guide, he has identified a causally potent antecedent action. This action (working through the study guide), if performed in the future, may permit the avoidance of future failure. Thus, the mechanism underlying the preparative function is based on the counterfactual identification of a causally potent antecedent action, which in turn triggers an expectancy of the consequences of that action in the future. This realization should then heighten intentions to perform that action, which may then influence the behavioral manifestation of that action. To the extent that the original causal inference was at least partly correct, subsequent performance will be enhanced.

\textsuperscript{75} Roese & Olson (1995c), pp.171-172:

Counterfactual thoughts may also serve an affective function; that is, they may, under some circumstances, be used to make individuals feel better. This affective function would be based on a contrast-effect mechanism. That is, a given outcome will be judged more favorably to the extent that a less desirable anchor is salient, a contrast perhaps leading to feelings such as relief or joy. For example, Jennifer was bruised when she fell from her bicycle. Although in pain, she may think, "At least I didn't break anything...." By generating a counterfactual that is evaluatively worse, her actual state of affairs may seem less negative by contrast. Accordingly, people may strategically generate thoughts of worse alternatives in order to make themselves or others feel better.

\textsuperscript{76} Counterfactual thinking can just as easily amplify negative emotional responses (e.g., the self-generated \textit{negative} emotions of regret, disappointment, shame, guilt, and distress) as positive responses (e.g., the self-generated \textit{positive} emotions of elation, satisfaction and feelings of fortunateness); and, it is well established that "individuals' affective responses to their actual outcomes are intensified to the extent that they can easily imagine how that outcome might not have occurred" (Boninger, Gleicher and Strathman, 1994, p.297).
in recent times, the study of counterfactual thinking has increasingly engaged the interest of scholars in a wide range of domains such as philosophy, psychology, cognitive psychology, history, political science, economics, social psychology, law, organizational theory, marketing, and epidemiology.

And, also, where would sports commentators be without counterfactual questions?

Counterfactual thinking seems to stem from one of two factors:


78 Including Fillenbaum (1974), who argued (p.45) that the counterfactual assertion “if he had caught the plane he would have arrived on time” not only strongly implies that “he did not catch the plane” but, also, strongly suggests the negation of the consequent proposition: viz., that “he did not arrive on time”.

By contrast, he argues (pp.48-49) that, whilst “he did not arrive on time because he did not catch the plane” presupposes “if he had caught the plane he would have arrived on time”, the converse does not necessarily apply — i.e., “if he had caught the plane he would have arrived on time” does not necessarily mean that “he did not arrive on time because he did not catch the plane” — for two reasons:

(a) within the statement “if he had caught the plane he would have arrived on time... there is no absolute requirement that the person in question must have been late”, and

(b) it is entirely possible that the additional statement “if he had caught the plane he would have arrived on time, which he did anyway because he caught a fast train” is entirely true.


Those which are essentially the products of imagination but (generally) lack an empirical basis; and those designed to test hypotheses by (supposedly) empirical means, which eschew imagination in favour of computation. In the case of the former, it is the tendency to rely for inspiration on hindsight, or to posit reductive explanations, which leads to implausibility. In the case of the latter, it is the tendency to make anachronistic assumptions.


82 Including, for example, Cowan & Foray (2002).


87 See Randerson (2002), and Haydon, Chase-Topping, Shaw, Matthews, Friar, Wilesmith & Woolhouse (2003).

88 For example: “Who would have won a world title fight between Cassius Clay and Muhammad Ali?” Or, even better, “If “Babe” Ruth — champion pitcher with the Boston Red Sox (1914-1919) and champion batter with the New York Yankees (1920-1934) — had been a track and field athlete, would he have been better at the javelin than the hammer throw?” (For examples of counterfactual questions taken from real-life sports broadcasts, see Sanna, Parks, Meier, Chang, Kassin, Lechter, Turley-Ames & Miyake, 2003.)
(1) contrast effects: where one’s judgment becomes more extreme in proximity to some benchmark; and

(2) causal inferences: which are implied independently of the contrast effects.

As discussed earlier, counterfactuals are involved in the sort of thought experiment that is past-oriented, concerned with what might have been different, and asks “What might have been the outcome if X had happened instead of (historically factual) Y?”

Cederman (1996, pp.248-256) describes the counterfactual as an attempt to establish a causal link between the action of an independent variable (the antecedent) and its dependent variable (the consequent), by demonstrating that the actual historical outcome O would not have occurred in the absence of the historical antecedent X.

Tetlock and Belkin (1996b, p.6) argue that “counterfactual reasoning is unavoidable in any field in which researchers want to draw cause-effect conclusions but cannot perform controlled experiments”; e.g., world politics where “the potential causes are simply too numerous and too interrelated”.

Researchers do this by “randomly [assigning] "subjects" to treatment conditions that differ only in the presence or absence of the hypothesized cause” and attempt to “justify [their] claims that a given cause produced a given effect by invoking counterfactual arguments about what would have happened in some hypothetical world in which the postulated cause took on some value different from the one it assumed in the actual world”.

---

89 Roese (2000), p.278:
For example, a cup of coffee feels hotter, by contrast, if one has just been eating ice cream. In the same way, a factual outcome may appear worse if a more desirable alternative outcome is salient and better if a less desirable outcome is salient.

90 Roese (2000), p.278:
By virtue of their conditional structure and implicit reference to a parallel factual statement, counterfactual propositions exemplify the logic of Mill's method of difference. For example, consider the counterfactual statement that “If only Jean had bought the insurance as she had intended, she would not have been in such a dire financial situation after the fire”. This alternative scenario is implicitly connected to the parallel facts that Jean did not buy insurance and Jean was in a dire financial situation. If the counterfactual world in which Jean does buy insurance is identical in all other respects to factuality, then the only thing that can account for the difference in financial situations between the counterfactual and the factual is Jean’s decision regarding the insurance purchase. Therefore, the above counterfactual yields the conclusion that Jean’s decision was causally related to her financial situation. To the extent that this counterfactual is salient, Jean might be seen to be more blameworthy for her financial situation.

91 Penrose (1994, p.240) considers counterfactuals to be “things that might have happened, although they did not in fact happen”.

92 According to Roese and Olsen (1995b, pp.11-12) counterfactuals are intimately related to causal influence:
All counterfactual conditionals are causal assertions. Causal assertions may almost always be expressed as conditional propositions. However, not all conditionals are causal: Some conditionals assert simple correlations or rules (e.g., The proposition, "If today is Thursday, then tomorrow is Friday," does not mean...
According to Fearon (1991, pp.171-172), where only one case is available for analysis, those wanting to test whether a proposed cause actually does (or did) produce the effect in question will often create imaginary counterfactual cases in order to increase the number of cases available for testing their theories.93

[There are two reasons] for counterfactual analysis. Firstly, it is a logical necessity when asking questions about causation to pose “but for” questions94 and try to imagine what would have happened if our supposed cause had been absent. For this reason we

[In these circumstances] our analyst begins with one case and at least one explanatory variable, which means negative degrees of freedom. Legitimate causal inferences cannot be made on the basis of negative degrees of freedom, so the analyst wishing to assess a causal hypothesis or to assess the relative weights of different causes has no choice but to add or create more cases: either a counterfactual case (or cases) that never actually existed or actual cases.

Put otherwise, the analyst, in explaining why some particular event $E$ occurred, cannot help but explain why $E$ occurred rather than some other possible outcome or outcomes. These other possible outcomes define the range of variation that the analyst accounts for, and this range is treated differently in different research traditions.

[Walker (1991, p.234) defines degrees of freedom as "the number of variables defining the state of a system (e.g., pressure, temperature) which may be fixed at will". In the specific context of his far more statistical representation, Fearon defines it this way: "degrees of freedom are the number of cases minus the number of explanatory variables minus one" (p.172).]

Chwieroth’s (2002) paper specifically examines the problems associated with a domain where only one case is available (i.e., a specific American “presidency”), and the difficulties this presents for scholars hoping to test their tentative hypothesis with a comparison of cases (in order to satisfy the demands of Mill’s method of difference). There was only one Carter, and only one Clinton.

Thus, argues Chwieroth (p.308) “it is difficult, if not impossible, to employ the comparative method or any type of statistical techniques” unless a “counterfactual case strategy” is used to expand the sample of cases available for study.


Thinking about actual and imaginary exceptional cases is indispensable if we wish to avoid mistaking accidental regularities for regularities which reflect a deeper truth about the world. And because the world does not provide us with easily accessible instances of all the combinations there might be, thought experiment — the contemplation of a well-described imaginary scenario in order to answer a specific question about some non-imaginary situation — is an indispensable technique: in philosophy, in science, and in ordinary reasoning...

94 These are the sine qua non ("without which not") questions of the law (e.g., Spellman & Kincannon, 2001). Deane, J. said this in March v. E & MH Stramare Pty Ltd [(1991) 171 CLR 506-537]:

The mere fact that something constitutes an essential condition (in the "but for" sense) of an occurrence does not mean that, for the purposes of ascribing responsibility or fault, it is properly to be seen as a "cause" of that occurrence as a matter of either ordinary language or common sense. Thus, it could not, as a matter of ordinary language, be said that the fact that a person had a head was a "cause" of his being decapitated by a negligently wielded sword notwithstanding that possession of a head is an essential precondition of decapitation. Again, the mere fact that a person makes a gift of money to another is not, in any real sense, a "cause" of the damage sustained by that other person when his agent negligently loses the money notwithstanding that the loss would not have occurred "but for" the original gift. (p.523, emphasis added)
are obliged to construct plausible alternative pasts on the basis of judgements about probability; and these can be made only on the basis of historical evidence. Secondly, to do this is a historical necessity when attempting to understand how the past “actually was” — precisely in the Rankean\(^95\) sense, as we must attach equal importance to all the possibilities which contemporaries contemplated before the fact. And greater importance to these than to an outcome which they did not anticipate. (Ferguson, 1999, p.87)

A major difficulty with thought experiments in general, and with counterfactual thought experiments in particular, is that there are no formally accepted criteria “for accurately gauging the risk of Type I or Type II errors associated with the hypothesized causal factor” (Chwieroth, 2002, p.307).\(^96\)

**Characteristics of Counterfactuals**

According to Roese and Olson (1995b), counterfactuals share certain characteristics:

1. They all seek to identify “a causally potent antecedent”.

2. They take a factual outcome and *mutate* one (or more) factual antecedent(s) and, then, examine the consequences of that alteration.\(^97\)

3. They are concerned with either present or past outcomes; never future

---

95 Refers to Leopold von Ranke (1795-1886CE):

The work of Ranke and his followers revealed that historians could draw very different lessons from the world of science. Ranke was suspicious of the way in which previous historians and philosophers had sought to pluck universal historical laws out of the air (or at best out of books by other historians and philosophers). It was his belief that only through properly scientific methods — meticulous and exhaustive research in the archives — could one hope to arrive at any understanding of the universal in history. This was the reason for his early pledge to write history “wie es eigentlich gewesen” (“as it actually was”) and his repeated stress on the uniqueness of past events and epochs. “Historicism” — the movement which Ranke is often credited with having founded — was about understanding particular phenomena in their proper context.

96 Statisticians are very aware of two serious sorts of error:

(a) **Type-I Error**, the error of rejecting something that should have been accepted (e.g., finding an innocent person guilty, incorrectly rejecting authorized users of a computer network as imposters (Moulton, 1983), etc.): a “false negative”); and

(b) **Type-II Error**, which is the error of accepting something that should have been rejected (e.g., finding a guilty person innocent, incorrectly accepting imposters as authorized users of a computer network (Moulton, 1983), etc.): a “false positive”.

Of the two sorts, scientists generally consider Type-I Error to be the most serious; and, as a matter of routine, their research is very strongly oriented to avoid it. Although the concept appeared earlier (Neyman & Pearson, 1928/1967, p.3; and Pearson & Neyman 1930.1967, p.100) the first time the errors were specifically designated Type-I and Type-II was in 1933 (Neyman & Pearson, 1933/1967, p.190).

97 According to Wells and Gavanski (1989, p.161), “an event will be judged as causal of an outcome to the extent that mutations to that event would undo the outcome”:

For an event to be judged as causal, it must be psychologically mutable. For example, following a suicide in which a man leapt from a window, people would not cite the presence of gravity as a cause of his death. Although it is true that an absence of gravity would have undone the outcome, the presence of gravity is an immutable characteristic of life on Earth. Most events are not either mutable or immutable but instead vary in their degree of mutability…
outcomes.  

(4) Because their antecedents always refer to events that did not occur, they are necessarily false.

(5) As the specified events did not occur, their antecedent is always false; however, the consequent may or may not be true. If the consequent is false, the antecedent is said to “undo” the factual outcome.

(6) Whilst their antecedents are always false, they are all causal assertions related to a factual state of affairs (e.g., “If only I were taller, I would be happier”).

(7) They tend to employ changes to episodic specifics but not changes to general laws.

Also they may be of one or more of six kinds, distinguished on the basis of:

(1) Structure:
   (a) **Additive**: adds a new antecedent not present in historical fact.  
   (b) **Subtractive**: removes a historically factual antecedent.

(2) Direction:
   (a) **Upward**: involve alternatives that are better than historical fact.
   (b) **Downward**: involve alternatives that are worse than historical fact.

---

98 Although some use “anticipated counterfactual” to describe the process of generating potential future outcomes, this is clearly a terminological error; they are **prefactuals**, not counterfactuals.

99 Some of this, of course, is driven by the ease which things are **mutable** (i.e., mentally changeable). Some things naturally seem to be far more mutable than others (which may, even, seem to be immutable): for example, **dynamic factors**, such as a weigh-lifter’s **effort**, seem to be far more mutable than **static factors**, such as a weigh-lifter’s **body-weight**, whilst other things connected with natural laws, such as **gravity**, seem to be quite immutable (e.g., you would not try to “undo” an aeroplane crash by supposing that the plane fell upwards).

100 Generally used to undo a situation of failure. They concentrate on an option that might ensure success. These are **far more creative** than subtractive counterfactuals.

101 Generally used to undo a situation of success. These merely remove a previous option, and make no suggestion about what might be better. They are **far less creative** than additive counterfactuals.

102 Very often combinations are used:
   (1) **Additive-Upward**: “[Even though I thought it was wrong] I could have dialed the telephone number that Fred had written down.”
   (2) **Subtractive-Upward**: “I should not have assumed that all telephone numbers start with a 9.”
   (3) **Additive-Downward**: “I could have chosen to study Mandarin Chinese and be confronted with an entirely different writing system.”
   (4) **Subtractive-Downward**: “If I had not studied as hard as I did, I would have done even worse in my test.”
(3) **Focus:**

(a) *Internal:* Focus is on oneself and/or one's own actions.\(^{104}\)

(b) *External:* Focus is on others and/or the actions of others.\(^{105}\)

**Counterfactual Types**

Despite the wide range of counterfactual arguments being advanced in distinctly different ways, driven by different needs, and selective to different "scholarly goals", Tetlock and Belkin (1996b, pp.6-7) have detected five distinct patterns amongst them:

---

\(^{103}\) In most cases, both *upward* and *downward* counterfactuals will *undo* the original outcome; however, for example, whilst a *downward* counterfactual of studying less before an important examination will *undo* the historical outcome (of, say, 51%), the new outcome will be worse. Yet, an *upward* counterfactual (of studying more before the important examination) will not only *undo* the historical outcome, but will produce a better outcome.

\(^{104}\) Whilst focussing on oneself can promote feelings of control, other self-oriented counterfactual thoughts can promote quite the opposite: either condemning one’s character ("If only I were...") "If only I weren't..."), causing *shame*, or one’s behaviour ("If only I had..." "If only I hadn't..."), causing *guilt*.

\(^{105}\) Whilst a focus on others ("If only the test was easier, I would have got a higher mark") certainly can undo the historical outcome, it contributes nothing to enhance perceived sense of control.
### Table Eight: Five Distinct Types of Counterfactual Thought Experiment

[Based on Tetlock & Belkin (1996b)]

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Idiographic case-study counterfactuals</td>
<td>These &quot;highlight points of indeterminacy at particular junctures in history&quot;.</td>
</tr>
<tr>
<td>2. Nomothetic counterfactuals</td>
<td>These &quot;apply well-defined theoretical or empirical generalizations to well-defined antecedent conditions&quot;.</td>
</tr>
<tr>
<td>3. Joint idiographic-nomothetic counterfactuals</td>
<td>These &quot;combine the historian's interest in what was possible in particular cases with the theorist's interest in identifying lawful regularities across cases, thereby producing theory-informed history&quot;.</td>
</tr>
<tr>
<td>4. Computer-simulation counterfactuals</td>
<td>These &quot;reveal hitherto latent logical contradictions and gaps in formal theoretical arguments by rerunning &quot;history&quot; in artificial worlds that &quot;capture&quot; key functional properties of the actual world&quot;.</td>
</tr>
<tr>
<td>5. Mental-simulation counterfactuals</td>
<td>These &quot;reveal hitherto latent psychological contradictions and gaps in belief systems by encouraging people to imagine possible worlds in which causes they supposed irrelevant seem to make a difference, or possible worlds in which causes they supposed consequential seem to be irrelevant&quot;.</td>
</tr>
</tbody>
</table>

---

106 *Idiographic* (‘concerned with individual cases’); not *Ideographic*, the adjective derived from *ideogram*.

107 Tetlock & Belkin (1996b, pp.6) comment on how this serves to "[remind] us of how things could easily have worked out differently and of how difficult it is to apply abstract hypothetico-deductive laws to concrete cases".

108 The term *nomothetic* means "concerned with general laws". According to Sloman (2000, pp.414-415) the terms *idiographic* and *nomothetic* are:

- *idiographic* to the study of particular cases (e.g., persons, social groups, works of art), *nomothetic* to the search for general laws or theories which will cover whole classes of cases. Thus HISTORY and GEOGRAPHY, in so far as they are concerned with the study of particular events, persons and PLACES, are *idiographic* subjects, whereas some economists would claim that, since they formulate ECONOMIC LAWS, ECONOMICS is a *nomothetic* science.

- According to Schutz (1954, p.257):

  It has been maintained that the social sciences are *idiographic*, characterized by individualizing conceptualization and seeking singular assertory propositions, whereas the natural sciences are *nomothetic*, characterized by generalizing conceptualization and seeking general apodictic propositions.

- Also, Grant (2002, p.98) distinguished between *nomothetic* (universal, predictive) and *idiographic* (historical, retrodictive) discovery operations as follows:

  - Non-historical sciences are *nomothetic*, seeking to establish general laws for indefinitely repeatable events and processes. Historical sciences, such as evolutionary biology, are *idiographic*, aiming to discover individuals and explain singular, nonrecurrent historical events — the evolution of a trait, the extinction of a lineage, the occurrence of a species at a certain location — each of which resulted from the combination of historical contingency and the sum total of the propensities acting at the moment in question, and each of which is therefore necessarily unique...

  - The ontological differences between past and future pose *nomothetic* and *idiographic* sciences with different obstacles to be overcome. *Nomothetic* sciences are primarily predictive, and they must therefore accommodate the objective indeterminacy of the future in their discovery operations: There is no way to know which number will be rolled on the next throw of a fair die, not because of lack of understanding, but because the outcome of any roll is truly (objectively) undetermined...

  - However, far from being hindered by their inability to employ rationally frequency-based probabilities, *idiographic* sciences have access to a larger, more diverse, and potentially more powerful arsenal of tests to determine the relative strength of opposing explanatory hypotheses: Because the events of interest have already occurred, *idiographic* sciences can search for evidence that the events happened and that evidence can be drawn from an abundance of sources...

- Therefore, instead of modeling the probability of error, *idiographic* inference relies on congruence of evidence.

109 Tetlock and Belkin (1996b, pp.6) comment that these serve to "[remind] us that deterministic laws may have been at work that were invisible to the original historical actors as well as to contemporary scholars who insist on a radically idiographic focus on the particular".
**Robustness Criteria**

Given “that scholars use counterfactual arguments for a variety of distinct, albeit interrelated, purposes”, in pursuit of “diverse goals” (“from hypothesis generation to hypothesis testing, from historical understanding to theory extension”) it is not surprising that “there is no single answer to the question of what counts as a good counterfactual argument” (Tetlock & Belkin, 1996b, p.16).

However, despite the fact that “different investigators will inevitably emphasize somewhat different criteria in judging the legitimacy, plausibility, and insightfulness of specific counterfactuals” (pp.16-17), Tetlock and Belkin feel that, wherever possible, “advocates of conflicting hypotheses [should] embrace at least some common standards for judging the plausibility of each other's counterfactual claims” (p.17).

They offer (p.17-18) “six normative criteria for judging counterfactual arguments” which are “most likely to contribute to the ultimate social-science goals of logically consistent, reasonably comprehensive and parsimonious, and rigorously testable explanations that integrate the idiographic and the nomothetic” and “appear to command substantial cross-disciplinary support”.

Work by Lebow (2000, pp.581-585) and Chwieroth (2002, p.300-309) has subsequently expanded the group of six criteria to eleven.

1. **Logical Clarity**: The independent and dependent variables (the hypothesized antecedent and consequent) must be specified and circumscribed.\(^{111}\)

2. **Cotenability (or Logical Consistency)**:\(^{112}\) The principles linking the antecedent and

---

\(^{110}\) Tetlock and Belkin (1996b, pp.16):

> A counterfactual that is idigraphically incisive (advances our understanding of a particular case) might be nomothetically banal (devoid of interesting theoretical implications) and vice versa. A counterfactual grounded in an elegant computer simulation might blow a gaping logical hole in an influential theoretical argument but tell us precious little about the actual world it supposedly simulates. A counterfactual that stimulates us to think of new hypotheses might run afoul of the received wisdom on what counts as a trivial or influential cause.

\(^{111}\) Chwieroth, 2002, p.304:

> All causal claims should identify as clearly as possible the *explanandum* (the consequent in counterfactual cases), the *explicans* (the antecedent), and the principle connecting the two. This ensures that researchers manipulate only one cause at a time, thereby highlighting the causal relationship.

\(^{112}\) The first two criteria of *logical clarity* and *cotenability* are so widely accepted by all scholars that Tetlock and Belkin feel they "are perhaps the best candidates for the status of universal minimum standards" (p.18).

The notion of "cotenability" of counterfactuals was introduced by Goodman (1947, p.120):

> In Goodman's account, a counterfactual assertion is judged true if (1) the counterfactual antecedent, when joined with appropriate theories and facts, implies the consequent; and (2) the counterfactual antecedent is "cotenable" with the facts or "initial conditions" used to draw the inference, meaning that if the antecedent had actually occurred, the initial conditions could also have occurred. (Fearon, 1991, p.193)
the consequent must be:

(a) clearly specified,

(b) cotenable with each other, and

(c) cotenable with the antecedent.\(^{113}\)

3. The enabling counterfactual should not undercut the antecedent: Just as "the antecedent [must be] consistent with the connecting principle", the "enabling counterfactual" must also be "tenable with the antecedent" (Chwieroth, 2002, p.304).\(^{114}\)

4. Historical consistency (or the "Minimal-Rewrite Rule"): Antecedents should require the alteration of as few well-established historical facts as possible.\(^{115,116}\)

5. Theoretical Consistency: Connecting principles must be consistent with the well-established theoretical generalizations relevant to the antecedent-consequent link.

6. Statistical Consistency: Connecting principles must be consistent with the well-established statistical generalizations relevant to the antecedent-consequent link.

7. Projectability:\(^{117}\) The testable implications of the connecting principles must be

---

\(^{113}\) (Chwieroth, pp.303-304)

Counterfactuals consist of connecting principles that link the antecedent to the consequent. The antecedent should not undercut any of the principles linking it to the consequent.

\(^{114}\) Lebow, 2000, p.582:

Counterfactuals may require other counterfactuals to make them possible... Researchers need to specify all important enabling counterfactuals and consider their implications for the consequent.

Lebow provides this example (p582):

...for Richard Nixon to have been president at the time of the Cuban missile crisis, he would have had to have won the 1960 election, and that would have required significant changes in the political context at home and possibly abroad. These changes might have had significant implications for both American and Soviet foreign policy.

\(^{115}\) According to Tetlock (2002):

Advocates of the minimal-rewrite standard argue that counterfactuals are worthy of serious historical attention only to the degree they: (a) start with the real world as it was otherwise known before asserting the counterfactual; (b) do not require us to unwind the past to rewrite vast stretches of history; (c) do not unduly disturb what we otherwise know about the original actors and their beliefs and goals.

Chwieroth (2002, p.302) summarizes "four selection rules that scholars have used to judge what constitutes a minimal rewrite" which can be used to defend one’s selection of a particular antecedent:

1. when something else nearly happened that could have led to an alternative world (the close-call counterfactual),

2. when an individual or event deviates from the norm (the norm-restoring counterfactual),

3. the final point at which history could have taken an alternative trajectory (the last-chance counterfactual), and

4. when the antecedent is one of the alternative courses of action that the actors actually considered (the reasonableness principle).

\(^{116}\) Tetlock and Belkin (1996b, p.18) found this criterion "is carefully observed by most idiographic researchers" (Type 1), but is "frequently ignored by many nomothetic researchers" (Type 2).

\(^{117}\) Projectability was first used by Goodman (1947, p.128) to discuss the problem of determining "what predicates are projectible from known to unknown cases". Tetlock & Belkin (1996b, p.30):
"teased out" to determine the extent to which they are consistent with additional real-world observations.\footnote{118, 119, 120}

8. The "Conjunction Fallacy"\footnote{121} must be avoided: The conjunction fallacy is "[the

Goodman draws a sharp distinction between coincidental generalizations that just happen to be true at a particular time and place (and are therefore unprojectable) and robustly law-like generalizations that hold up over a range of circumstances and permit projection into the past and future. An example of a merely coincidental generalization is "All the coins in my pocket yesterday were silver." Nothing follows from this observation — certainly not "If this penny were in my pocket yesterday, it would be silver." The counterfactual fails because "if this penny were in my pocket yesterday," we would simply assume that the original generalization — "all the coins in my pocket yesterday were silver" — was false. By contrast, a robustly lawlike generalization — such as that oxygen is a necessary but not sufficient condition for fire — inspires confidence when we move either backward in time (if there had been no oxygen, the Great Fire of London would not have occurred) or forward in time (if we cut off any future fire's source of oxygen, the fire will expire).\footnote{Chwieroth (2002), pp.303-304: Since counterfactuals rely on theoretical generalizations that link the antecedent and consequent, the factor that determines if these theories are robust enough to support counterfactual inference is whether they can predict what will occur in new cases... A counter-factual inference... [will only be deemed robust] if the theory utilized to support it can tell us something about the yet unobserved future.}

Tetlock and Belkin (1996b, pp.30-31) argue that, because the evaluation of theories and counterfactuals are "inextricably entangled", it is often difficult to determine the soundness of the theories used to provide the "lawlike generalizations" upon which counterfactuals are based. Also, it is often difficult to determine whether "these lawlike generalizations are robust enough to support counterfactual inferences". Most social-science generalizations... take the form of either contingent generalizations (under this set of boundary conditions, \(x\) causes \(y\); under that set, \(x\) causes \(z\)) or statistical generalizations (\(x\) increases or decreases the likelihood of \(y\)) or contingent statistical generalizations. From Goodman's perspective, however, whether the generalization is bounded or unbounded by moderator variables and whether the generalization is deterministic or probabilistic, it is subject to the same acid test of scientific legitimacy: namely, its projectability or its ability to predict what will happen in new, hitherto unobserved cases. The same causal principles that allow us to retrodict the past should allow us to predict the future. Indeed, the strong Popperian form of this argument asserts that we should take counterfactual claims seriously if and only if the lawlike generalizations supporting the claims yield falsifiable forecasts... King, Keohane, and Verba... urge scholars to search aggressively for the observable implications of their causal constructs by regularly asking themselves, "If my argument is correct, what else should be true?" Counterfactuals that are devoid of testable implications in the actual world leave us marooned in hypothetical worlds of our own subjective making. Projectability, from this vantage point, stands as the preeminent criterion for judging the value of counterfactual speculation.\footnote{Tetlock and Belkin (1996b, p.18) observe that the criteria of \textit{theoretical consistency}, \textit{statistical consistency} and \textit{projectability} "are more widely acknowledged among nomothetic than among idiographic researchers (although a significant contingent of idiographic researchers do apply these standards in their own work)".}

Tversky and Kahneman, who coined the term \textit{conjunction fallacy}, define it as the error of failing to recognize that a conjunction of events "cannot be more probable than one of its constituents" (1983, p.294). Suppose there is a group of 100; comprised of 75 basketballers and 50 females. These are "two well-defined classes" (p.304); and, thus any "error in judgement" is a "fallacy" rather than just being a "misunderstanding"; such as thinking that a "physical education teacher" was not a "teacher", or that an "insect" was not an "animal". In this case, the conjunction fallacy is one of predicting that a specific individual is more likely to be a female basketballer than a female alone. According to Lebow, "the laws of statistics indicate that the probability of any compound counterfactual is exceedingly low" (2000, p.583, emphasis in original).\footnote{Tversky and Kahneman (p.304) defend their unusual choice of the term "\textit{fallacy}" thus: We have described violations of the conjunction rule in direct tests as a fallacy. The term \textit{fallacy} is used here as a psychological hypothesis, not as an evaluative epithet. A judgment is appropriately labeled a fallacy when most of the people who make it are disposed, after suitable explanation, to accept the following propositions: (a) They made a non-trivial error, which they would probably have repeated in similar problems, (b) the error was conceptual, not merely verbal or technical, and (c) they \textit{should} have known the correct answer or a procedure to find it. (Ibid)

Kanwisher (1989), who examines the impact of the conjunction fallacy (pp.654-656), defines a "\textit{fallacy}" (in this context) as "an argument or assumption... [that] is held with a conviction much stronger than warranted by the available evidence"; especially if those beliefs are "opposed by powerful counterevidence or unanswered counterarguments" (p.653).}

9. Recognition of the Inteconnectedness of Causes and Outcomes: To what extent should “causes that, because of their location in complex systemic networks of causation, do not have effects that can be conceptually isolated” be excluded from consideration? (Tetlock & Belkin, 1996b, p.21). Also, much of the problem circulates around the vexing question of precisely what is a “cause”.

10. “Second-Order Counterfactuals” must be considered: This involves the identification of “the most likely course of events that could unravel [the] consequent or negate its value as an outcome” (Lebow, 2000, p.584).

122 Tetlock and Belkin (1996b, p.21) see this as a problem of the need to resolve "the tension between the desire of methodologists to "hold other things equal" and the insistence of latter-day Leibnizians that once we tamper with one element from the past, we have to trace through the causal implications for all other elements, in effect creating a full-fledged alternative world for each counterfactual".

According to Lebow (2000, p.584) in order for a counterfactual to be a "good counterfactual", it "must specify what else might change as a result of a hypothesized antecedent, and [it] must consider how the most important of these changes might interact and influence the probability of the consequent".

123 When discussing the difference between conceivable causes and miracle causes, Fearon (1996, p.41) argues that whenever we say that “A caused B”, we are saying much more than "if A had not occurred, B would not have occurred"; we are really saying that "if A had not occurred, B would not have occurred and the world would be otherwise similar to the world that did occur".

Conceivable causes are "factors that could actually have been different, according to the best of our knowledge about how the social and physical worlds work" (Fearon, 1996, p.41), and miracle causes (the term suggested by Lewis, 1973/2002, pp.75-77) are those which are not required to be "actually, historically possible" (p.61) and which "imagine their counterfactual occurrence as resulting from an intervention from outside the system (the hand of God as it were)" (Fearon, 1996, p.60) — and, according to Lebow (2000, p.585), they have the decided advantage of not being: required to meet any real-world tests. The value of such a counterfactual is based entirely on its ability to provoke or, better yet, to compel researchers to think about issues and problems they would not otherwise address, or to look at them in a new light. For a field where careful, technical work is increasingly valued over imagination, miracle-world counterfactuals can refocus our attention on important, big questions.

In relation to miracle causes, Weber remarks that the social scientists’ “low probability events” happen frequently in world politics, and reports that Tetlock once said that “what today appears impossible, tomorrow appears overdetermined”; which, in Weber’s view, “says more about the state of our understanding than about the state of the world” (1996, p.281).

Fearon (1996, p.56) argues that whenever we attempt to explain why event B occurred “we explain why B occurred rather than some other alternative or set of alternatives” (emphasis in original). He stresses the importance of understanding that “explanation takes place relative to a "contrast space" of alternatives” and how the extent to which an explanation is considered satisfactory is influenced by “how this space is implicitly imagined”:

Invariably, … when we try to explain why some event B occurred, we implicitly imagine a contrast space in which B is absent and the rest of the world is similar to the world in which B is present. To show that A caused B, the relevant counterfactual to make plausible is not “if A had not occurred, B would not have occurred”, because this admits a contrast space that includes worlds where B did not occur and the rest of the world was entirely dissimilar to the world that did occur. Instead, the relevant counterfactual should be, “if A had not occurred, B would not have occurred and the world would be otherwise similar”. The goal is to explain the presence or absence of B against a fixed, actual “back-ground”, rather than B in the context of all conceivable backgrounds. (pp.56-57)

124 Whilst one may strongly believe a chosen antecedent will produce the "desired consequent", one must also recognize that “subsequent developments may return history to the course from which [it] was diverted by the antecedent" (Chwieroth, 2002, p.307).
11. **Proximity**: Lewis (1973/2002, pp.84-91) stresses the importance of distance between “actual” and “possible worlds”.\(^{125}\),\(^{126}\)

Whilst all of these criteria can be used to assess a counterfactual’s robustness (or plausibility) and, simultaneously, identify a counterfactual that should be eliminated as implausible, they certainly can not validate a counterfactual.\(^{127}\)

In addition to prefactual thought experiments,\(^{128}\) counterfactual thought experiments,\(^{129}\) semifactual thought experiments,\(^{130}\) there are at least another four sorts of hypothetical question that employ subjunctive reasoning:\(^{131}\)

1. **Prediction, Forecasting and Nowcasting**: attempt to project the known present

---

\(^{125}\) Lewis (1973/2002) p.84:
I believe that things could have been different in countless ways; I believe permissible paraphrases of what I believe; taking the paraphrase at its face value, I therefore believe in the existence of entities that might be called "ways things could have been". I prefer to call them "possible worlds".

\(^{126}\) Fearon (1996), p.66:
Consider only thought experiments in which the hypothetical antecedent and consequent are close together in time and are separated by a small number of causal steps...

A great many counterfactuals that are, unfortunately, of great interest will be deemed unassessable by this [proximity] criterion. This does not mean that they should never be posed — such counterfactuals may serve as valuable "spotlights" directing our attention to more local and assessable counterfactuals. Rather, exploring counterfactual claims that fail the proximity criterion is unlikely to yield any very defensible judgment on the causes of the event in question.

\(^{127}\) Most of these criteria apply to possible-world counterfactuals only and, therefore, have no relevance to "miracle-world counterfactuals".

Tetlock & Belkin (1996b, p.18) describe an additional test, based on the “the semantics of possible worlds” notions of Lewis and Stalnaker that could possibly be used “for judging the truth or falsehood of counterfactual claims”:

To test a proposition of the form "if \( p \), then \( q \)," possible-worlds semantics directs us to do three things:

1. identify the set of possible worlds \( \{ p \} \) in which the counterfactual antecedent, \( p \), is true;
2. identify that possible world \( \{ p' \} \) that is "closest" to the actual world;
3. determine whether that possible world \( \{ p' \} \) falls in the intersection of the set of possible worlds in which \( p \) is true, \( \{ p \} \), and the set of possible worlds in with \( q \) is true, \( \{ q \} \).

We should judge "if \( p \), then \( q \)" to be true if and only if the "closest" \( p \) falls in the intersection of \( \{ p \} \) and \( \{ q \} \). In other words, we should judge "if \( p \), then \( q \)" to be true if and only if the closest world in which \( p \) is true is also a world in which \( q \) is true.

This logical calculus provides an elegant framework for evaluating counterfactual claims. It assumes, however, a vastly more sophisticated knowledge of the causal workings of the world than social scientists currently possess (or are likely to possess anytime in the next century). We need to partition the universe of possible worlds into overlapping sets, to locate the actual world in the universe of possible worlds, and to quantify the "distance" between the actual world and each possible world. Not surprisingly, none of our contributors could implement this test. We differ from Lebow and Stein [1996] in that we distinguish the Lewis-Stalnaker approach from the minimal-rewrite rule. Even if the antecedent \( p \) is an historically implausible miracle cause (Fearon [1996]), the closest world in which \( p \) is true could still be a world in which \( q \) is true, in which case the counterfactual violates the minimal-rewrite rule but passes the Lewis-Stalnaker test.

\(^{128}\) Speculating on future outcomes.

\(^{129}\) Speculating on the possible outcomes of a different past.

\(^{130}\) Speculating on the extent to which things might have remained the same, despite a different past.

\(^{131}\) This is, of course, ignoring Kahn’s scenarios — which, from Vazsonyi’s (1982) perspective, are a composite aggregate of a number of thought experiments amalgamated into a single activity.
into the unknown and speculated future.\(^\text{132}\)

(2) \textit{Retrodiction} (or \textit{Postdiction}):\(^\text{133}\) involves step-by-step movements backward in time from the present into the speculated past.

(3) \textit{Backcasting}: involves descriptions of a specific future and, then, step-by-step movements backward in time from the future to the present, in order to reveal the mechanism through which that speculated future could be attained.

(4) \textit{Hindcasting}: involves running a forecast model after an event has happened in order to test whether the model’s simulation is valid.

\textbf{1. Prediction}

Scientific prediction takes two forms (Sarewitz, D. & Pielke, 1999, p.123):

(1) “The elucidation of invariant — and therefore predictive — principles of nature”;\(^\text{134}\) and

(2) “[Using] suites of observational data and sophisticated numerical models in an effort to foretell the behavior or evolution of complex phenomena”.\(^\text{135}\)

Whilst the activity of \textit{nowcasting},\(^\text{136}\) defined as “a detailed description of the current weather along with forecasts obtained by extrapolation up to 2 hours ahead”, is essentially concerned with describing the current state of affairs, it is common practice

\footnotesize
\(^{132}\) Although they perform different social and scientific functions, the only difference between the qualitatively identical activities of predicting, forecasting and nowcasting is the distance of the speculated future from the present moment occupied by the user.

\(^{133}\) These terms are identical and are interchangeable. Each was coined using \textit{prediction} as the model. According to the Oxford English Dictionary \textit{retrodiction} was first used in 1895; \textit{postdiction} (sometimes, \textit{post-diction}), first used in 1940, and seems to be increasingly popular in the US literature.

It is also clear that retrodiction and prediction are intertwined; a “correct” theory should not only predict that a fox will leave prints of a certain shape, size and depth in the snow as it passes, it should also retrodict that prints of a certain shape, size and depth in the snow have probably come from a fox.

\(^{134}\) Sarewitz & Pielke (1990), p.123: Prediction is central to the process of science — it is fundamental to the scientific method. Scientists test their ideas by comparing predictions based on theory to actual events in nature or the laboratory... Prediction is thus a crucial means of testing and confirming hypotheses.

\(^{135}\) Sarewitz & Pielke (1990), p.123: These activities are made possible by rapid advances in computer and data acquisition technologies. Scientists and policy makers alike have come to view this second type of prediction as a powerful tool for helping to guide political decisions and resolve societal problems. These predictions of complex phenomena seek to ascribe time, place and characteristics to events. Such efforts are distinct from the development of more general scientific insights about the future (often referred to as analogies, scenarios, foresight activities, projections, or sensitivity analyses). Scientific insight tells us that floods are more likely to occur on flood plains than on hillsides; scientific predictions seek to tell us which flood plain, on what day, and to what extent.

\(^{136}\) \textit{Nowcasting} (obviously based on \textit{forecasting}) is also known as \textit{very-short-term forecasting}; thus, also indicating a \textit{very-short-term}, \textit{mid-range}, and \textit{long-range forecasting} continuum.
to extend the term “to cover very-short-range forecasting up to 12 hours ahead” (Browning, 1982, p.ix).

Murphy, and Brown (1982) describe a large range of specific applications for meteorological *nowcasting* over wide range of user demands.\(^{137}\)

### 2. Retrodiction (or Postdiction)

In 1972, Kahnemann and Tversky identified a “representativeness heuristic”, a human propensity to believe that similar causes had similar effects.\(^{138}\)

It is certain that such a belief is embodied in the practice of retrodiction. Retrodiction involves a series of step-by-step movements backwards in time, in as many stages as required, from the present into the unknown and speculated past, in

\(^{137}\) Murphy, and Brown (1982, pp.4-7) divided their survey of nowcasting needs into five sectors (the following summarizes their findings):

1. **Agriculture**:
   - (a) wind and precipitation forecasts for effective seeding and spraying from aircraft;
   - (b) precipitation forecasts to minimize damage to seedlings;
   - (c) minimum temperature, dewpoint, cloud cover, and wind speed forecasts to protect crops from frost;
   - (d) maximum temperature forecasts to reduce adverse effects of high temperatures on crops and livestock;
   - (e) humidity and cloud cover forecasts to prevent fungal disease crop losses;
   - (f) hail forecasts to minimize damage to livestock and greenhouses;
   - (g) precipitation, temperature, and dewpoint forecasts to avoid during- and after-harvest losses due to crops rotting in the field;
   - (h) precipitation forecasts to minimize losses in drying raisins; and
   - (i) humidity forecasts to reduce costs and losses resulting from poor conditions for drying tobacco.

2. **Construction**:
   - (a) precipitation and wind speed forecasts to avoid damage to finished work (e.g. concrete) and minimize costs of protecting exposed surfaces, structures, and work sites; and
   - (b) precipitation, wind speed, and high/low temperature forecasts to schedule work in an efficient manner.

3. **Energy**:
   - (a) temperature, humidity, wind, cloud, etc. forecasts to optimize procedures related to generation and distribution of electricity and gas;
   - (b) forecasts of thunderstorms, strong winds, low temperatures, and freezing precipitation minimize damage to lines and equipment and to schedule repairs.

4. **Transportation**:
   - (a) ceiling height and visibility, winds and turbulence, and surface ice and snow forecasts minimize risk, maximize efficiency in pre-flight and in-flight decisions and other adjustments to weather-related fluctuations in traffic;
   - (b) forecasts of wind speed and direction, as well as severe weather and icing conditions along flight paths facilitate optimal airline route planning;
   - (c) forecasts of snowfall, precipitation, and other storm-related events allow truckers, motorists, and public transportation systems to avoid damage to weather-sensitive goods, select optimum routes, prevent accidents, minimize delays, and maximize revenues under conditions of adverse weather.

5. **Public Safety & General Public**:
   - (a) rain, snow, wind, and temperature forecasts assist the general public in planning activities such as commuting, recreation, and shopping;
   - (b) forecasts of temperature/humidity extremes (or significant changes) alert hospitals, clinics, and the public to weather conditions that may seriously aggravate certain health-related illnesses;
   - (c) forecasts related to potentially dangerous or damaging natural events (e.g., tornados, severe thunderstorms, severe winds, storm surges, avalanches, precipitation, floods) minimize loss of life and property damage; and
   - (d) forecasts of snowstorms, surface icing, visibility, and other events (e.g. floods) enable highway maintenance and traffic control organizations to take appropriate actions to reduce risks of traffic accidents and protect roads from damage.

order to establish the ultimate cause of a specific event.\textsuperscript{139}

The simplest view is that:

\textbf{retrodiction : events that have happened :: prediction : events that have not yet happened} \textsuperscript{140}

3. Backcasting

Backcasting is a process which attempts “to understand what a [desirable] future situation… could look like and how it could be attained” (Höjer and Mattsson, 2000, p.629).\textsuperscript{141,142}

\textsuperscript{139} Examples of this type of activity are: the medical speculation on whether clinical depression (the well-verified prelude to many cases of pancreatic cancer) is a \textit{paraneoplastic} (unconnected with the cancer) \textit{syndrome}, rather than a \textit{prodrome} (precursor) that is a legitimate part of the illness pathway (Levenson, Bemis & Presberg, 1994); the "reverse engineering" of unknown objects, such as the Antikythera Mechanism; and the problem a corpse presents to a homicide detective and forensic chemist, "Did they die of "natural" causes?", "If not, what caused their death?", "How were they killed?", "who killed them?", "Why were they killed?". "Is anyone else in imminent danger?", etc.).

\textsuperscript{140} For example, Sternberg and Kalmar (1997), p.200.

Given that \textit{retrodiction} is a process in which "past observations, events and data are used as evidence to infer the process(es) the produced them" and that \textit{diagnosis }["involves going from visible effects such as symptoms, signs and the like to their prior causes" (Einhorn & Hogarth, 1982, p.24), the essential balance between prediction and retrodiction could be characterized as:

\textbf{retrodiction : diagnosis :: prediction : prognosis}

regardless of whether the prognosis is of the course of the disease in the absence of treatment, or of the application of a specific treatment regimen to a specific disorder in a particular patient.

Einhorn & Hogarth, 1982, p.24:

We consider diagnostic inference to be based on causal thinking, although in doing diagnosis one has to mentally reverse the time order in which events were thought to have occurred (hence the term "backward inference"). On the other hand, predictions involve forward inference; i.e., one goes forward in time from present causes to future effects. However, it is important to recognize the dependence of forward inference/prediction on backward inference/diagnosis. In particular, it seems likely that success in predicting the future depends to a considerable degree on making sense of the past. Therefore, people are continually engaged in shifting between forward and backward inference in both making and evaluating forecasts. Indeed, this can be eloquently summarized by Kierkegaard's observation that, "Life can only be understood backwards; but it must be lived forwards".

Sternberg and Kalmar (1997) examined whether, in the case of "\textit{real-world stimuli}", prediction was easier than retrodiction, or retrodiction easier than prediction.

It could be argued that prediction should be easier than postdiction, because we are more accustomed to making predictions than to making postdictions, and more familiar tasks tend to be easier than less familiar ones. (p.186)

Alternatively, it could be argued that postdiction should be easier than prediction, because the events to be judged have already occurred, and hence we are dealing with facts rather than forecasts; moreover, whereas past experience clearly bears upon likelihoods of past events, it may or may not bear upon likelihoods of future events. (p.186)

They found prediction was relatively less difficult than retrodiction; and that although "prediction items are solved more rapidly than [retrodiction] items, [they] also are more susceptible to error" (p.200).

In everyday life, we frequently wonder about both the future and the past. Will I be a has-been in a year? Will my lover and I still be together in five years? Did people in my line of work always have to worry so much about job security, or is this worry a recent development? Was this painting I bought for $20 really done by Rembrandt? (p.186)

Prediction and [retrodiction] are important parts of everyday lives. Probably scarcely a day goes by that we do not make at least some kind of guess about what will happen in the future, or what might have happened in the past. Indeed, observations would suggest that there are few people who do not consider likely futures for themselves, and often, likely pasts for the people they meet. (p.201)

\textsuperscript{141} Höjer and Mattsson (2000, p.630) characterize backcasting as an attitude to research that asks:

...how can we possibly attain a state that has been identified as desirable? In other words, standing at a desirable end-point, we need to work back to check the physical and social feasibility of the route that got us there (in fact, this is the origin of the term "backcasting").
Starting at some specified future moment and, then, working step-by-step, through time, from the future to the present moment, backcasting attempts to uncover a feasible pathway which links that desired future to the present reality.

Backcasting\textsuperscript{143} is not concerned with predicting the future:

The major distinguishing characteristic of backcasting analyses is the concern, not with likely energy futures, but with how desirable futures can be attained. It is thus explicitly normative, involving 'working backwards' from a particular future end-point to the present to determine what policy measures would be required to reach that future.\textsuperscript{144}

4. Hindcasting

Hindcasting is an operation that tests simulations by running a set of past historical data through a simulator and measuring the extent to which the simulator's output matches known historical fact.\textsuperscript{145}

Robinson (2003, p.842) stresses the importance of recognize that backcasts, which are concerned with “how desirable futures can be attained”, are driven by the commonsense understanding that “even if the future were predictable... the most likely future may well not be the most desirable”; and, therefore, that "unlike predictive forecasts, backcasts are not intended to reveal what the future will likely be, but to indicate the relative feasibility and implications of different policy goals".

To the degree that the future is not already determined but remains to be created, then the search for the most likely future (i.e., the best prediction) is not only often misguided (since we are usually wrong) but actually counterproductive. This is so because the most likely future may not be the most desirable, and thus what are needed are not techniques that converge on likelihood but techniques that reveal the possibility, and test the feasibility and impacts, of alternative futures. The focus thus shifts from prediction and likelihood to feasibility and choice. This approach, of course, has significant implications for the way forecasts are prepared and used, and by whom. (Robinson, 1988, p.326)

\textsuperscript{142} According to Robinson (1982, p.337-338), "backwards-looking analysis" was first used by Amory Lovins in 1974 in examining possible benefits from renewable energy resources such as wind-generated electricity and solar heating. Lovins' famous 1976 paper "Energy Strategy: The Road Not Taken?" is an excellent example of his "backwards-looking analysis".

\textsuperscript{143} For Robinson backcasting is a \textit{method}, for Dreborg (1996, p.818)] it is an \textit{approach}, and for Höjer and Mattsson (2000, p.629) it is an \textit{attitude}.

The term \textit{backcasting} was coined by Robinson; and his backcasting approach is very similar to the \textit{anticipatory scenarios} of Ducot and Lubben (1980), and Bunn and Salo (1993).

For some (e.g., Bloomfield, 1985), the term \textit{backcasting} denotes a different sort of activity: "backcasting is the name given to the technique of running a simulation model backwards in time (retrodiction) in order to try and test its assumptions and parameter values" (pp.636-637).

For those sharing this view, backcasting can also be conducted by, say, (p.637-638) going back as far as 1900 and then "starting with initial values for the state variables in 1900, the model is run until 2100".

However, in terms of this dissertation, it is best to relegate whatever is unique to Bloomfield's activity (i.e., whatever is distinct from \textit{forecasting}) to the \textit{hindcasting} category.

\textsuperscript{144} Jansen (1994, p.503) makes the following distinction between backcasting and forecasting:

Within the framework of technological development, "forecasting" concerns the extrapolation of developments towards the future and the exploration of achievements which can be realized through technology in the long term. Conversely, the reasoning behind "backcasting" is: on the basis of an interconnecting picture of demands which technology has to meet in the future — "sustainability criteria" — to direct and determine the process that technology development must take and possibly also the pace at which this development process must be put into effect.

Backcasting [is] both an important aid in determining the direction technology development must take and in specifying the targets to be set for this purpose. As such, backcasting is an ideal search toward determining the nature and scope of the technological challenge which is posed by sustainable development, and it can thus serve to direct the search process toward new — sustainable — technology.

\textsuperscript{145} Bem (1972, pp.27-31) emphasizes:
In 2003, Chen and colleagues\textsuperscript{146} “trained” a computer using the data of the surface temperature of the oceans from the last 20 years. Then, using data that had been collected on the surface temperature of the oceans for the period 1857 to 2003, they went through a \textit{hindcasting} exercise and discovered that their simulation not only accurately predicted every \textit{El Niño} event for the last 148 years, it also identified the (up to 2 years) looming foreshadow of every single one of those \textit{El Niño} events.\textsuperscript{147}

Impressive as such an example might be, we certainly don’t need computers to run more humble versions of hindcasting. I would suggest that we often resort to such thinking when a recently revealed fact, hitherto concealed from us, makes us realize that we can understand and make sense of the past by “reframing” the past in the light of that new fact. Which brings me back to my starting point in this chapter: understanding as context for thought experimentation.

Even this brief account of counterfactual thinking should suffice to illustrate its ubiquitous nature. And we are driven to it because of our need to understand our situation and our world; because explanation is deeply satisfying; and because we have a need for control. No wonder thought experimentation is such a pervasive activity.

In the next chapter, I move on to consider what I call the “performance” aspects of thought experimentation. Thought experimentation is, after all, something we do, and we need to appreciate not just why we do thought experiments (as this chapter has shown), but what it’s like to do them.

\textsuperscript{146} See Chen, Cane, Kaplan, Zebiak & Huang (2004), and the associated article Anderson (2004).

\textsuperscript{147} What is significant here is not only has the hindcasting process demonstrated that the computerized simulation models can predict the onset of \textit{El Niño} climatic events from changes in the temperature of the ocean's surface temperature that occur up to two years earlier — meaning that we now have at least 2 years’ lead time — but the results also imply that \textit{El Niño} events seem to be the effects of some causal regularity; and, therefore, are not due to simple chance, or to some other “chaotic” event.
Chapter Five: Performance Aspects of Thought Experimentation

The aim of this chapter is to examine the “performance” of thought experiments. Conceived as a kind of journey, a thought experiment takes us somewhere new; and, in this process, a number of things need to be considered. First, there is a narrative dimension to this process. In other words, we engage with the “story” of the thought experiment (and thus should consider the creator of the thought experiment as a narrator). Second, we are receiving and assimilating a kind of cognitive artifact that affects and extends our cognitive abilities. Third, we acquire a certain kind of knowledge by performing the thought experiment: knowledge by acquaintance. Finally, performance leads to a sense of personal “ownership” of that knowledge.

The Narrative Dimension of Thought Experiments

Nersessian (1993, p.292) stresses the importance of the creation of an appropriate narrative as part of the design and construction of a thought experiment:

The original thought experiment is the construction of a dynamical model in the mind by the scientist who imagines a sequence of events and processes and infers outcomes. She then constructs a narrative to describe the setting and sequence in order to communicate the experiment to others...

Nell (1988), speaks of the reader’s sensation of “being lost in a book”, and observes that we “tell stories all the time” and constantly judge the stories of others.\footnote{Nell (1988), p.52: [It seems that] the criteria by which narrative products are judged form part of the intellectual equipment of almost every adult (and most children) because the production and consumption of narrative are near-universal human experiences.}

He distinguishes between the \textit{ludic} readers, who read for pleasure\footnote{\textit{Ludic}, from Latin \textit{ludo}, “I play”. Nell argues (pp.7-10) that there are three necessary conditions for ludic reading; and, if any of the three are absent “ludic reading is either not attempted or fails”:

(1) \textit{Reading Ability}: “whether as a precondition for ludic reading or as a consequence of it, ludic readers are articulate, well informed, and endlessly curious [and] seem by and large to be skilled readers who rapidly and effortlessly assimilate information from the printed page”.

(2) \textit{Positive Expectations}: “the expectation that reading will be a pleasurable experience [and will produce] the kind of consciousness change that narrative produces so readily”.

(3) \textit{Correct Book Choice}: because “my good read” is another reader’s rubbish, it is essential that readers are able to select what is, for them, “an appropriate book”. Nell observes that “[ludic] readers develop great skill in selecting the kind of book that promises [them] a "good read"”.

Nell argues that if all three conditions are present, and if “reading is more attractive than the available alternatives, reading begins and is continued as long as the reinforcements generated are strong enough to withstand the pull of alternative attractions”. The “reinforcers” are of two sorts:} — and whom, by...
virtue of the “absolute control [they] exercise over their reading with regard to pace, content, initiation, and duration” (p. 9), can put down their reading as soon as they lose interest — and those who read either for “adult education” or “professional training” (p. 20).

According to Ryan (1998, pp. 518-519), from the perspective of their story and discourse, all narratives:

(1) are “constructed by language”;

(2) are “constructed… through a performative force that is granted to their narrative by cultural convention”;

(3) are not “representations (mimesis) of the actual world”;

(4) are “autonomous realities [that are] called into being through the unrestricted creative power of fictional language”.

(1) **Physiological**: these involve “a series of physiological changes in the reader mediated by the autonomic nervous system, such as alterations in muscle tension respiration, heart beat, electrical activity of the skin, and the like” which are “by and large unconscious and feed back to consciousness as a general feeling of well-being”.

(2) **Cognitive**: The cognitive changes experienced by the reader “are numerous and profound”. The very act of “reading changes the focus of attention from self to environment”; as a consequence of “the heavy demands reading makes on conscious attention, the reader is effectively shielded from other demands, whether internal or external”, and “at the same time, the intense attention brought to bear by the entranced reader may have the effect of transfiguring both book and reader”.

And, mediating between all these factors “is the reading process itself, in which meaning is extracted from the symbols on the page and formed into inner experience”, which involves questions of how a reader extracts meaning from a text and “what he or she does with it once it has been extracted”.

Graesser, Singer and Trabasso (1994, p. 372) clearly distinguish between narrative texts and expository, persuasive, descriptive texts; i.e., the type of text that “is normally written to inform the reader about new concepts, generic truths, and technical material”:

The typical reader [of an expository text] does not have extensive background knowledge about the topics in expository texts, so [these] readers generate fewer inferences than they generate during the comprehension of a narrative text.

It is clear, then, that almost none of the readers of thought experiments are ludic readers; it is as though they have been compelled to read, and the reader–narrative dynamic is considerably different.


Fictional language can be referential without entering into a mimetic relation to the real world. ...reference to an entity does not presuppose its language-independent existence.

A wide range of theoretical distinctions have been made between aspects of the typical narrative (Bortolussi & Dixon, 2003, p. 98): for example, discourse (“the telling”) vs. story (“the told”); fabula (“events in their chronological order”) vs. sjuzet (“events in the order in which they are presented in the narrative”); story (“narrative of events with an emphasis on chronology”) vs. plot (“narrative of events with an emphasis on causality”) [Carter (2003, p. 6) understands story as “what is presented” and plot as “how it is presented”]; what is told vs. how it is told; the content vs. the expression; the narrated vs. the narrating; and the fiction vs. the narration.

These six features have been derived and adapted from Ryan’s summary (1998, pp. 518-519) of the theoretical assumptions made by Dolezel about the structure of fictional worlds.

With the only “limits of the fictionally possible [being] the limits of the expressible, or imaginable” (Ryan, 1998, p. 518).
(5) consistently represent possible worlds that are both incomplete in their nature and ontologically different from the real world.\(^9\)

(6) can be expressed in “referential [language] without entering into a mimetic relation to the real world” (Ryan, 1998, p.519).\(^{10}\)

A narrative implies a narrator; and the following table, based on Ryan (1995), classifies six possible types of narrator:

---

Nell draws attention to how it is the individual reader’s acceptance of the author’s constructed “fictive reality” — i.e., the “willing suspension of disbelief for a moment” that was spoken of by Samuel Taylor Coleridge — which “[allows] the reader or listener or audience to attribute reality to something that the author claims to be true... but which the reader knows to be perfectly false” (p.54).

Because it is impossible for the human mind to think up an object (much less a world) in all of its properties, every fictional world presents areas of radical indeterminacy (i.e. ontological gaps).

\(^{10}\) Ryan (1998, p.519):
Fictional language can be referential without entering into a mimetic relation to the real world. ...reference to an entity does not presuppose its language-independent existence.
Table Nine: Six Types of Narrator
[based on Ryan, 1995]

<table>
<thead>
<tr>
<th><strong>Illocutionary Levels</strong>&lt;sup&gt;11&lt;/sup&gt; (identify different narrative acts)&lt;sup&gt;12&lt;/sup&gt;</th>
<th><strong>Heterodiegetic Narrators</strong>: those who tell about others.</th>
<th><strong>Homodiegetic Narrators</strong>: those who tell about themselves.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Extradiegetic Narrators</strong>: those who narrate but are not, themselves, narrated.</td>
<td><strong>Intradiegetic Narrators</strong>: those who become narrator by being introduced and then being &quot;quoted&quot;.</td>
<td></td>
</tr>
</tbody>
</table>

| **Ontological Levels**<sup>13</sup> (identify different reference worlds) | **Extraontological Narrators**: those who describe the reference world of their non-factual story from the outside, without ever becoming part of that world. *This is how the narratives in all types of thought experiment represent non-actual worlds.* The narrator, who is also the author, remains located in the actual world and maintains a "foreign" point of view towards the non-actual world. | **Intraontological Narrators**: those who describe the reference world of their non-factual story from the "insider" perspective of one of its members. *This is how fictional narratives represent non-actual worlds.* The narrator, who is distinctly different from the author, pretends to be a member of the "foreign", non-actual world; and describes it in the real mode. |

According to Ryan (2001) additional distinctions must be made, because:

1. Across the board, authors and narrators are not necessarily identical:
   - (a) in works of non-fiction, *the author = the narrator*;
   - (b) in works of fiction, *the author ≠ the narrator*; and

2. The narrator fulfills one or more of three entirely different functions:<sup>14</sup>
   - (a) *creative (or self-expressive) function*: "the activity of shaping and encoding the story";<sup>15</sup>

---

<sup>11</sup> The terms are derived from *diegesis*, "narrative"; diegetic means "pertaining to diegesis".

<sup>12</sup> According to Ryan (1995), this four-part classification permits the unequivocal identification of four distinct types of narrator:
   - (1) extradiegetic-heterodiegetic: the standard third-person narrator.
   - (2) extradiegetic-homodiegetic: the standard first-person narrator.
   - (3) intradiegetic-heterodiegetic: those characters who gossip or tell stories as fiction.
   - (4) intradiegetic-homodiegetic: those characters who narrate their own story.

<sup>13</sup> According to Ryan (1995), this additional level of unequivocal descriptive precision allows an extra, most valuable level of distinction to be made amongst narrators; for example:
   - (1) Scheherazade, when narrating the story "Ali Baba" in the 1001 Nights (a character who is telling a story as fiction) is described as an *intradiegetic-extraontological-heterodiegetic* narrator.
   - (2) If the story of "Ali Baba" stands alone — i.e., as an independent story totally removed from the 1001 Nights — the narrator is not Scheherazade, but some impersonal, genderless, omniscient third person, who is described as an *extraontological-heterodiegetic* narrator.
   - (3) A character who is narrating his/her own story as true fact is an *intradiegetic-intraontological-homodiegetic* narrator.
   - (4) A character who is narrating another’s story as true fact is an *intradiegetic-intraontological-heterodiegetic* narrator.

<sup>14</sup> The three of which are, according to Ryan (2001), “the building blocks of the act of narration”.

<sup>15</sup> Ryan (2001):
(b) transmissive (or performative) function: the expression of "the product of the creative function [which leaves] the privacy of the mind [in] the form of either oral performance or written inscription";\textsuperscript{16} and

(c) testimonial (or assertive) function: the assertion that the statements made are to be "taken as truthful representations of facts".\textsuperscript{17,18}

The following table, derived from Ryan (2001) “shows the various possible modes of participation in a narrative transaction for a given individual in a real-life situation”:

<table>
<thead>
<tr>
<th>Different types of agent</th>
<th>Three Distinct Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Creative</td>
</tr>
<tr>
<td>1 Witness vouching for the truth of someone else’s testimony in a trial.</td>
<td>–</td>
</tr>
<tr>
<td>2 Actor performing a narrative monologue; narrator of “Peter and the Wolf”; Scribe taking dictation from a blind poet.</td>
<td>–</td>
</tr>
<tr>
<td>3 Well-known broadcaster narrating a TV documentary; Religious figure taking dictation from God.</td>
<td>–</td>
</tr>
<tr>
<td>4 Writing a play; Telling oneself imaginary stories.</td>
<td>+</td>
</tr>
<tr>
<td>5 Writer of a TV documentary.</td>
<td>+</td>
</tr>
<tr>
<td>6 Oral storyteller (of tales or jokes); Novelist.</td>
<td>+</td>
</tr>
<tr>
<td>7 Historian; Narrator of personal experience; Designer of a Thought Experiment.</td>
<td>+</td>
</tr>
</tbody>
</table>

The degree of spatial, temporal and emotional immersion\textsuperscript{19} promoted by a narrative

The creative function covers, for instance, such phenomena as control of rhetorical devices, speed, stance, self-presentation, chronological rearrangement, alternation between diegetic and mimetic narration, or economy versus disgressivity.

\textsuperscript{16} Ryan (2001):

A description of the transmissive function will ask: is the narrative written or oral; what is the channel of communication; in what genre is the narrator narrating; to what extent are the generic norms respected?

\textsuperscript{17} In the case of non-fiction these facts are objective facts; and, in the case of fiction, they are "largely accurate statements" that are made about "an event in the history of the fictional world".

\textsuperscript{18} Ryan (2001):

The testimonial function subsumes questions of reliability, of source of knowledge, of sincerity, and of authority.

\textsuperscript{19} Cowles (1998, p.361):

In studying the child’s experience of trance, researchers have identified a common trance state outside of hypnosis. Guille and Boersma (cited (p.365) as “Guille, M. & Boersma, F, (1992). Fairy tales as a trance experience: Possible therapeutic uses. American Journal of Clinical Hypnosis. 34 (4). 245-253”) detailed the child’s experience of being deeply absorbed in hearing a fairy tale and suggested that it is, in fact, a trance state. They explicated trance as an altered state of awareness that allows the subject to focus, using the unconscious without external sources impinging on the experience. Children who become so absorbed in the
has been shown to be far more significant to “fictive reality” than the life-likeness of the fiction. The immersion may be of three kinds:

1. **Spatial immersion**: the sense of being at the scene described;
2. **Temporal immersion**: the burning desire to know what happens next; and
3. **Emotional immersion**: the personal attachment to the characters, and the sense of participating in their human experience.

Gerrig (1993, p.2) emphasizes that readers are not passive; they perform a narrative, and to the extent that they do so, they are transported.20

According to Gerrig (pp.10-11) the features of being transported are:

1. Someone (“the traveler”) is transported
2. by some means of transportation
3. as a result of performing certain actions.
4. The traveler goes some distance from his or her world of origin
5. which makes some aspects of the world of origin inaccessible.
6. The traveler returns to the world of origin, somewhat changed by the journey.

To which another two features could be added:

---

20 Perhaps this is due, in part, to the fact that — unlike, say, a photograph — a narrative is extended in time.
(7) The traveller becomes a “celebrity” of the moment; and, upon their return, is constantly asked for accounts of his or her experience.

(8) The returned traveller may be quarantined for the sake of the community because of contamination they picked up during their journey.\(^{21}\)

The following diagram, taken from Nell (1988, p.65), has been slightly adapted to highlight the pathway of thought experimental narratives:

---

\(^{21}\) As an example of ostracism for holding bizarre views, consider the case of the highly successful, rapidly expanding business of the specialized motor car dealership — centred on Australian champion racing car driver Peter Brock, in partnership with General Motors Holden — capable of generating annual turnovers in excess of five million Australian dollars in 1984 (Tuckey, 1987, pp.165, citing the 1984 dollar value), which had totally collapsed to zero in 1987 because of Brock’s controversial, unilateral decision to insist that his team’s vehicle modifications include a device called the “Peter Brock Energy Polarizer”.

The Energy Polarizer was a small box perhaps 10cm long, 5cm wide, and 4cm deep (photo: Tuckey, facing p.189), which used “micro-print instructions, magnets, and crystals to re-align and distribute the energy produced by the magnets” (p.186).

According to the bizarre theories of Melbourne chiropractor, Eric Dowker, this polarizer would “[harness] orgone energies to improve the engines” (p.166).

A later statement, issued by Brock in 1985, attributed it all to a re-alignment of molecules:

> WHAT IT DOES? A high technology energy device which creates a “polarized” or “ordered” molecular arrangement as distinct from the normal “random” structure. This alters the behaviour and characteristics of materials and components in the vehicle. (Tuckey, p.197)

GMH’s engineer, Ray Borrett, had taken several of these devices apart in 1985 and found nothing apart from “four magnets wrapped in tinfoil set in an epoxy resin” (Tuckey, p.196).

No scientific test of any kind, including tests designed to “determine performance in exhaust and noise emissions, interior noise levels, ride, steering, handling, acceleration, performance under load... and fuel economy”, was ever able to demonstrate any advantage of having the Brock/Dowker device attached to any vehicle at any time (Tuckey, p.227).

General Motors Holden eventually issued a bland comment in late 1987, just prior to the partnership finally ending, obviously emanating from and directed by its legal advisors, that they had found that “there was ‘no technical merit’ in the polarizer” (Tuckey, p.232).
Chapter Five: Performance Aspects of Thought Experimentation

Narrator

Audience Size?

Personal Audience: all group members are known to narrator

Unseen Mass Audience

Genre Options?

One Self Alone

One Other (e.g., relating day’s events to one’s flat-mate)

Impersonal Audience: group is physically present (but without individual identities)

Unseen Mass Audience

Narrative

Mechanical Storytelling

Story Medium?

Oral Storytelling

Electronic Storytelling

Daydream

Oral Storytelling

Elaborated Storytelling

Electronic Storytelling

Informal Narrative

Shared Narrative

Formal Narrative

Mechanical Storytelling

Fig.11: The Thought-Experimental Narrative’s Pathway: in the Context of Other Types of Narrative (Adapted from Nell, 1988, p.65)
**Cognitive Artifacts**

Hutchins (1999, p.126) defines *cognitive artifacts* as “physical objects made by humans for the purpose of aiding, enhancing or improving cognition”; and, because they extend and strengthen our mental powers, *cognitive artifacts* are a “means to augment human intellect” which help us “apply [our] native sensory, mental and motor capabilities” to complex problem situations (Engelbart, 1962, pp.1-2).\(^{22}\)

Norman (1993), commenting that they “are themselves artificial objects that can be perceived and studied” and that, “because they are artificial, created by people, they can take on whatever form and structure best serves the task of the moment” (p.51), further “relaxes the definition of cognitive artifacts to include mental as well as material elements” (Hutchins, 1999, p.126).

Norman further separated *cognitive artifacts* into two divisions (pp.52-53):

1. *(experiential artifacts)*, which “provide ways to experience and act upon the world”;\(^{23}\) and
2. *(reflective artifacts)*, which “provide ways to modify and act upon representations”.\(^{24}\)

---

\(^{22}\) Norman describes them as “tools that aid the mind” (p.4); and Clifton and Pang (1997, p.564), describe them as things which “display or represent information in such a way as to improve human cognitive performance”. As Bisdorff (1999, p.381) indicates, the ‘artifact’ in *cognitive artifact* does not carry the “negative connotations” of the French usage; it carries the positive Anglo-Saxon sense of “a formal construction supposed to enhance the cognitive abilities of a person”.


\(^{23}\) Norman (1993), p.52 (emphasis added):

Experiential artifacts allow us to experience events as if we were there, even when we are not, and to get information about things that would be inaccessible, even if we were present. A telescope gives us information about something distant in space. A movie or recording lets us experience events distant in time and space. Instruments, such as the gas gauge of an automobile, give us information about states of equipment that would otherwise be inaccessible. **Experiential artifacts thus mediate between the mind and the world.**

\(^{24}\) Norman (1993), p.52 (emphasis added):

Reflective artifacts allow us to ignore the real world and concentrate only upon artificial, representing worlds. In reflection, one wants to contemplate the experience and go beyond, finding new interpretations or testing alternative courses of action. The process can be both powerful and dangerous. The power comes from the ability to make new discoveries. The danger occurs whenever we fool ourselves into believing that the representation is the reality.

When we concentrate only upon the information represented within our artifacts, anything not present in the representation can conveniently be ignored. In actuality, **things left out are mostly things we do not know how to represent, which is not the same as things of little importance.** Nonetheless, things not represented fall in importance. They tend to be forgotten or, even if remembered, given little weight.

Engelbart (1962), p.14:

The important thing to appreciate here is that a direct new innovation in one particular capability can have far-reaching effects throughout the rest of your capability hierarchy. A change can propagate up through the capability hierarchy; higher-order capabilities that can utilise the initially changed capability can now reorganize to take special advantage of this change and of the intermediate higher-capability changes. A change can propagate down through the hierarchy as a result of new capabilities at the high level and modification possibilities latent in lower levels. These latent capabilities may previously have been unusable in the hierarchy and become usable because of the new capability at the higher level.
On this analysis, thought experiments are certainly *reflective artifacts*.

Norman (1993, p.78) has also argued that cognitive artifacts can be understood from two mutually exclusive perspectives:

(1) The *personal* perspective: which concentrates on “the impact the [cognitive] artifact has for the individual”; \(^{25}\) and

(2) The *system* perspective: which concentrates on “how the [cognitive] artifact + person, as a system, is different from the cognitive abilities of the [individual] person alone”. \(^{26}\)

---

**Knowledge by Description vs. Knowledge by Acquaintance**

The difference between the Reason of man and the Instinct of the beast is this, That the beast does but know, but the man knows that he knows.

*(John Donne, 1628)*

In relation to thought experiments in general, it is important to briefly consider the epistemological question: “What kind of knowledge are they producing?”

In the late nineteenth century, philosophers interested in the knowledge of things \(^{28}\) began to examine the difference between *knowing things* and *knowing about things* (Martens, 1993). In 1865, John Grote distinguished between what he described as

---

\(^{25}\) Norman (1993, p.78):

From *the personal perspective*, cognitive artifacts don’t make us smarter or make us have better memories; they change the task[, not the individual].

Engelbart stresses that being able to sort out what might, otherwise, remain a “tangle of thoughts” (p.13) allows you to “integrate your new ideas more easily, and thus harness your creativity more continuously” (p.9); which facilitates rapid, flexible change and can “[provides] direct aid to an individual in comprehending complex situations, isolating the significant factors, and solving problems” (p.8).

\(^{26}\) As Engelbart (p.18) observed, their various constituents act together synergistically.

Norman (1993, p.78):

From *the system perspective*, the person + [cognitive] artifact is more powerful than either alone. Performance of the *system* of person + [cognitive] artifact is indeed enhanced, but that of the individual person is not.

In other words, it is the system that is smarter; not the individual.

\(^{27}\) From a sermon preached at St. Pauls, Easter-Day, April 1628 (Gardner & Healy, 1967, p.336). [Spinoza (*Ethics*, II. Proof xliii) also held the view that whatever we know, we know that we know.]

\(^{28}\) Lazerowitz (1937) defines knowledge as “[possessing] cognitive idubitability, such that anyone directly knowing that \(p\) could entertain no *rational doubt* that \(p\)” (p.403).
"knowledge of acquaintance" and "knowledge-about".29

In a similar fashion, Helmholtz (1868/1881/1962, pp.178, 179, 182) clearly distinguished between das Kennen, the knowledge that was comprised of "mere familiarity with phenomena", and das Wissen, "the knowledge of [phenomena] which can be communicated by speech" (p.182).

Stressing that the Kennen sort of knowledge could not "compete with" the Wissen sort of knowledge,30 he argued that, despite the fact that it might be of "the highest possible degree of precision and certainty" (p.179), the Kennen kind of knowledge can not be expressed in words, "even to ourselves".31

Adopting Grote's terminology, and agreeing there were two fundamental kinds of knowledge, William James (1890, I, pp.218-223) further developed the distinctions made by Grote and Helmholtz:

I am acquainted with many people and things, which I know very little about, except their presence in the places where I have met them. I know the color blue when I see it, and the flavor of a pear when I taste it; I know an inch when I move my finger through it; a second of time, when I feel it pass; an effort of attention when I make it; a difference between two things when I notice it; but about the inner nature of these facts or what makes them what they are, I can say nothing at all. I cannot impart acquaintance with them to any one who has not already made it himself I cannot describe them, make a blind man guess what blue is like, define to a child a syllogism, or tell a philosopher in just what respect distance is just what it is, and differs from other forms of relation. At

29 Grote noted that these distinctions were made in many languages. He cited Greek (γνῶναι and εἰδέναι), Latin (noscere and scire), German (kennen and wissen), and French (connaître and savoir) as examples (James, 1890, p.221; Martens, 1993, p.237). Grote's "knowledge OF acquaintance" is far better known today as "knowledge BY acquaintance" following Russell's decision to change the preposition in a paper read to the Aristotelian Society on 6 March 1911 (Russell, 1910-1911).

30 Helmholtz (1868/1881/1962, pp.178-179):
Besides the knowledge which has to do with Notions, and is, therefore, capable of expression in words, there is another department of our mental operations, which may be described as knowledge of the relations of those impressions on the senses which are not capable of direct verbal expression. For instance when we say that we "know" a man, a road, a fruit, a perfume, we mean that we have seen, or tasted, or smelt, these objects. We keep the sensible impression fast in our memory, and we shall recognise it again when it is repeated, but we cannot describe the impression in words, even to ourselves. And yet it is certain that this kind of knowledge (Kennen) may attain the highest possible degree of precision and certainty, and is so far not inferior to any knowledge (Wissen) which can be expressed in words; but it is not directly communicable, unless the object in question can be brought actually forward, or the impression it produces can be otherwise represented — as by drawing the portrait of a man instead of producing the man himself.

31 Helmholtz (1868/1881/1962, pp.178, 179, 182):
[It] is not directly communicable, unless the object in question can be brought actually forward, or the impression it produces can be otherwise represented — as by drawing the portrait of a man instead of producing the man himself.

Helmholtz (1878/1971, pp.390-391) revisits the distinction between knowledge that is expressible in language and that which is not in a slightly different context.
most, I can say to my friends, Go to certain places and act in certain ways, and these objects will probably come. (p.221)

Because all thought experiments demand their performers immerse themselves in the designer's constructed narrative, it's immediately obvious that the designers, along with James, are saying "Imagine that you are standing at X, and acting in a Y fashion, and tell me what you observe", in order to provide performers with a direct experience of an (otherwise) inexplicable matter by intentionally generating a "knowledge by acquaintance" within each performer's mind.

According to Hayner (1969, p.425), Bertrand Russell was extremely critical of the equivocal nature of the word know, and believed the equivocation rose from the failure to distinguish the notion of "knowledge by description" from "knowledge by acquaintance".

According to Russell (1912/1997), only those things that we experience directly, and have an awareness of, are known by acquaintance (e.g., such as the fact that I have a toe missing from a mishap chopping wood).

Other things can only be known by description (e.g., that the micro-surgeons at the Prince of Wales Hospital have the skills and equipment to re-attach the severed toe).

However, within Russell's arguments there are strong implications that he considers "knowledge by acquaintance" to be the foundation for all "knowledge by description".

---

32 Sainsbury (1995b), who describes how Russell used the concept of knowledge by description to "[explain] how one could know things with which one had no acquaintance", clearly distinguishes (1995a) between acquaintance and description:

We are not acquainted with Sir Walter Scott, so we know him only by description, for example as the author of Waverley. By contrast, we can know of one of our experiences 'by acquaintance', that is, without the intermediary of any definite description.

33 In the sense that whilst it can be completely known (i.e., it is both comprehensible and intelligible) and experienced, it can not be described or otherwise unravelled.

34 Using Russell's term "knowledge by acquaintance", rather than James' "knowledge of acquaintance".

35 Parker (1945b, p.458) holds the strong opinion that the term "knowledge by description" is misleading; he advocates using "knowledge merely by description".

36 Lazerowitz (p.403) prefers "direct knowledge" and "indirect knowledge" for "knowledge by acquaintance" and "knowledge by description" respectively.

The pursuit of knowledge by acquaintance is always susceptible to what James (1980, I, pp.196-197) labelled "the psychologist's fallacy": the psychologists' tendency to confuse their analyses of subjective experience with the nature of reality ("the great snare of the psychologist is the confusion of his own standpoint with that of the mental fact about which he is making his report").

37 Martens (1993), p.240:

Knowledge about things is essentially propositional knowledge, where the mental states involved refer to specific things. This propositional knowledge can be more or less complete, can be justified inferentially and
Knowledge by description has always been an essential part of science, especially in relation to those things that must be taken on trust.\(^{38}\)

Knowledge by description is also transmitted by testimony.\(^{39}\)

The most important aspect of testimony is that it is credible.

One way that scientific testimony becomes credible is that enough information is supplied to replicate the event; another is to provide credible witnesses to attest to the veracity of one’s description of events.\(^{40}\)

on the basis of experience, and can be communicated. Knowing things, on the other hand, involves experience of things. This experiential knowledge provides an epistemic basis for knowledge about things, and in some sense is difficult or impossible to communicate, perhaps because it is more or less vague.

\(^{38}\) Ravenshear (1999), p.63:

In the best scientific work, even as in the worst, much must be taken upon trust; on the authority of the competent observer, skilled instrument maker, or original investigator. The Chemist, in establishing the existence of a new compound, or defining its properties, relies to a great extent upon the determinations of others as to the atomic weights, formulas, densities, specific heats, boiling-points, refractive indices and other coefficients of the auxiliary substances or reagents he employs. In much of his apparatus — weights, balances, polarimeters — he relies upon the work of the instrument maker. The Physicist in like manner employs all such results as are in general repute — tabulated densities, temperature-coefficients, elasticities, weights, resistances. The Astronomer makes use of the observations of his predecessors, as well as of his contemporaries in distant observatories. The acceptance of observations and descriptions in this manner is still more marked in geology, zoology, botany. The disposition of rocks in various countries and the occurrence of minerals; the kinds and distribution of plants and animals; all are to any given systematise largely matters of report.

It might indeed seem that, in physical science, time and opportunity alone are needed to enable a man of sufficient energy and capacity to do the work of a hundred observers. But in psychology it is far otherwise; for, whatever his capacity, one man knows only one mind. Reliance upon others in physical science may be merely unavoidable, but for Psychology it is essential.

\(^{39}\) Lackey (1999), p.471:

We often talk about knowledge being transferred or transmitted via testimony. This suggests two things: (1) that hearers can acquire knowledge via the testimony of others; and (2) that speakers must themselves have the knowledge in question in order to pass it to their hearers. In this way the picture we have of testimonial knowledge is like a chain of people passing buckets of water to put out a fire. Each person must have a bucket of water in order to pass it to the next person, and moreover there must be at least one person who is ultimately acquiring the water from another source. Similarly, each person in the chain of transmitting knowledge that \(p\) must know that \(p\) in order to pass it to the next person, and moreover there must be at least one person in the chain who ultimately acquired knowledge that \(p\) from another source, e.g., sense perception, introspection, reason, and the like.

\(^{40}\) The following is part of Oersted’s original announcement of his discovery of electromagnetism in a pamphlet (Experimenta circa effectum conflictus electrici in acum magneticum) privately distributed on 21 July 1820. [The translation, by Rev. J.E. Kempe, first appeared in the Journal of the Society of Telegraph Engineers, vol. V (1876); taken from Shamos, 1959/1987, p.123]

The first experiments on the subject which I undertake to illustrate were set on foot in the classes for electricity, galvanism, and magnetism, which were held by me [at the University of Copenhagen during the winter of 1819-1820]. By these experiments it seemed to be shown that the magnetic needle was moved from its position by the help of a galvanic apparatus [viz., a battery], and that, when the galvanic circuit was closed, but not when open, as certain very celebrated physicists in vain attempted several years ago. As, however, these experiments were conducted with somewhat defective apparatus, and, on that account, the phenomena which were produced did not seem clear enough for the importance of the subject, I got my friend Esmarch, the king’s minister of justice, to join me, that the experiments might be repeated and extended with the great galvanic apparatus which we fitted up together. A distinguished man, Wleugel, knight of the Danish Order, and president of our Pilot Board, was also present at our experiments as a partner and a witness. Besides these there were as witnesses at these experiments that most excellent man, decorated by the king with the highest of honors, Hauch, whose acquaintance with natural science has long been celebrated; that most acute man Reinhardt, professor of natural history; Jacobsen, professor of medicine, a man of the utmost sagacity in conducting experiments; and the most experienced chemist, Zeise, doctor of philosophy. I have indeed somewhat frequently carried out by myself experiments relating to the matter proposed, but the phenomena which it thus befell me to disclose I repeated in the presence of these most learned men.

This is the issue with the motto of the Stock Exchange of London, Dictum Meum Pactum, “My word is my bond”, and the traditional absence of documentary support for the verbal transactions made by its
Ownership

In his Two Treatises of Government, Locke (II.V.45-51; 1965, pp.341-344) stresses that because we own ourselves and our labour we are strongly inclined to sense ownership over the things that we create, shape or produce.41

Pierce, Kostova and Dirks (2001, p.304) believe that psychological ownership “arises from certain processes of association of the individual with the target [and] through these processes individuals become psychologically tied to the target, and the target becomes part of their extended self”.

They also claim (p.300) that “the roots of psychological ownership can be found in three main motives: (1) efficacy and effectance,42 (2) self identity,43 and (3) having a place [in which to dwell]”.44

Pierce, Kostova and Dirks (2003, p.86), observe that “although ownership is generally experienced as involving person-object relations, it can also be felt toward nonphysical entities such as ideas, words, artistic creations,45 and other people”.

They remark that “feelings of ownership toward various objects have important and potentially strong psychological and behavioral effects”,46 and define psychological members. It’s interesting, however, that in certain parts of contemporary USA, where organizations feeling the need for a motto have “Latinized” the English “My word is my bond” and use verbum ("the words that I write on a page") rather the original dictum ("the words that come out of my mouth").

And, of course, the entire issue of “unsubstantiated testimony” is the major reason one takes an oath in court; because, setting aside potential supernatural beings, there was no-one else present at the time.

41 It is clear that ownership “operates both as a formal (objective) and a psychologically experienced phenomenon” Pierce, Rubenfeld and Morgan (1991, pp.124), and that a sense of psychological ownership has “positive consequences... [and] does make a difference” Vandewelle, van Dyne and Kostova (p.223).

42 Pierce, Kostova and Dirks (2001, p.300):
Being the cause through one’s control or actions results in feelings of efficacy and pleasure and also creates extrinsic satisfaction as certain desirable outcomes are acquired.

43 Pierce, Kostova and Dirks (2001, p.300):
Possessions also serve as symbolic expressions of the self since they are closely connected with self-identity and individuality.

44 Pierce, Kostova and Dirks (2001, p.300):
Ownership and the associated psychological state can also be explained in part by the individual’s motive to possess a certain territory or space — to have a “home” in which to dwell.

45 Pierce, Kostova and Dirks (2001, p.302):
Creation involves investing time, energy, and even one’s values and identity. Academics, for example, invest all of these into their research and, hence, may feel very strong ownership toward the outcome of their scholarly pursuits.

46 James (1890, I, p.291) commented on the very close connexion between “me” and “mine”:
...a man’s Self is the sum total of all that he CAN call his, not only his body and his psychic powers, but his clothes and his house, his wife and children, his ancestors and friends, his reputation and works, his land, and
ownership as "the state in which individuals feel as though the target of ownership or a piece of that target is “theirs”“.

Pierce, Kostova & Dirks (2003, pp.86-87) identify four distinguishing features of the state of psychological ownership:

1. it “manifests itself in the meaning and emotion commonly associated with my or mine and our. Psychological ownership answers the question “What do I feel is mine?” and its conceptual core is a sense of possession toward a particular target (e.g., the products of one’s labor, toys, home, land, or significant others).”

2. it “reflects a relationship between an individual and an object (material or immaterial in nature) in which the object is experienced as having a close connection with the self, becoming part of the “extended self”.”

3. it “is complex and is composed of a cognitive and affective core. It is a condition of which one is aware through intellectual perception. It reflects an individual’s awareness, thoughts, and beliefs regarding the target of ownership. This cognitive state, however, is coupled with an emotional or affective sensation. Feelings of ownership are said to be pleasure producing per se and are accompanied by a sense of efficacy and competence. The affective component becomes apparent in the feelings that arise when others lay claim to objects for which one feels a sense of personal ownership (e.g., "Those ideas are mine!") or collective ownership shared with a group (e.g., "That garden space is ours!").”

4. “people can legally own an object (e.g., automobile or home) yet never claim the possession as their own... under these conditions individuals simply fail to claim the object as "theirs" because they do not find personal meaning in the object’s symbolic properties, which is necessary for claiming something as "mine".”

I contend that the successful performance of a thought experiment generates a powerful sense of psychological ownership. Having undergone the journey of the narrative, having receiving and assimilated the thought-experimental cognitive artifact, and having acquired a particular knowledge by acquaintance, our tendency is to feel a psychological ownership of that knowledge. And we would tend to encourage others to take the same journey if we wanted them to understand the matters addressed by the thought experiment for themselves; for the only way they can

yacht and bank account. All these things give the same emotions. If they wax and prosper, he feels triumphant; if they dwindle and die, he feels cast down.

and, as well, he (p.293) also observed that: although it is true that a part of our depression at the loss of possessions is due to our feeling that we must now go without certain goods that we expected the possessions to bring in their train, yet in every case there remains, over and above this, a sense of the shrinkage of our personality, a partial conversion of ourselves to nothingness, which is a psychological phenomenon by itself.
become acquainted with the knowledge is to experience the thought experiment by performing it for themselves — and then they, too, can own the knowledge, as do all of the other performers.

Of course, before anybody can "own" the knowledge generated by a thought experiment, somebody has to conceive, construct and design the thought experiment. That is the subject of the next chapter.
Chapter Six
Constructing, Designing, and Polishing Thought Experiments

The human mind is the laboratory of the new physics. (Wolf, 1990, p.312)

While I have stressed the extent to which thought experiments employ common cognitive processes, and are driven by human needs, there is nevertheless a sense in which explicit, polished, and published thought experiments are relatively uncommon and noteworthy. It is important, therefore, to examine the processes through which someone might go about constructing, designing, and polishing a thought experiment, with a view to having it enter into a particular domain of discourse, and be influential in that domain.

Nersessian (1993, pp.295-297) found that most scientific thought experiments display eight characteristics:

1. “By the time a thought experiment is communicated it is... as packaged and polished as a real-world experiment is when it is published”.2

2. They are presented as a narrative,3 which “[functions] in much the same way as other narratives”.4

3. Thought-experimental narratives are selective in what they contain, and what they exclude5,6

---

1 That is, having already independently conceived the thought experimental situation in the first place.

2 According to Nersessian (p.296), the presentation of a “polished” experiment makes “it an effective means of getting comparable mental models among the members of a community of scientists”.

3 Nersessian (1993, p.296):
   Occasionally, presentations of thought experiments also include some form of visual illustration, so my use of the term "narrative" is to be taken as broad enough to encompass these.

4 And, according to Nersessian (1993, p.296), these “thought-experimental narratives are... a central form of effecting conceptual change within a scientific community”.

5 Wilkes (1988), p.7:
   It cannot be stressed too strongly that all experiments (in real life and in thought) are for a particular purpose, testing a particular factor, and [the issue of what background conditions are relevant] is relative to this... [and it is also true that] we know, or can assume, that we have before us everything that is thought to be relevant for assessing the outcome of the experiment.

6 According to Nersessian (1993, p.296), “[this] aids in focusing attention on the salient dimensions of the model and in recognizing the situation as prototypical”:
   The thought-experimental narrative depicts abstractions. For example, certain features of objects that would be present in a real-world experiment are not included, such as the color of rocks and the physical characteristics of observers. That is, there has been a prior selection of the pertinent dimensions on which to focus that evidently derives from experience in the world. From experience we believe, e.g., that the color of a rock does not effect its rate of fall. Such information is customarily excluded from real-world experimental
(4) The narrative “has the character of a simulation”.7

(5) The reader is invited to “imagine a dynamic scene”, which “unfolds in time and follows a specific causal sequence”. 8

(6) Regardless of however bizarre or fantastic the simulated scene might be,9 there is nothing bizarre or fantastic in the way the simulation unfolds.10

(7) In its final polished form, the thought experiment always "works"; and, in this final form it "is usually so compelling that even in those cases where it is possible to carry it out, the reader feels no need to do so".11

(8) Whilst thought experiments can be reconstructed as arguments, “the argument is not evident until after the thought experiment has been constructed and executed”.12

The typical thought experiment emerges from an extensive interaction of experimental design, narrative development, the comments and criticism of peers, input from subject narratives as well. This facilitates the reader’s recognition of the situation as prototypical, i.e. as representing a class of experimental situations...

Extremely colorful [thought experimental] narratives may include highly specific details [that] usually serve to reinforce crucial aspects of the experiment. For example, in one version of the chest, or "elevator" experiment, Einstein depicts the physicist as being drugged and then waking up in a box. This colorful detail serves to reinforce the point that the observer could not have known before entering the chest if he were falling in outer space or sitting in a gravitational field. It also reinforces the condition that the observer cannot know whether or not there are gravitational sources around.

7 Nersessian (1993, p.297): Unlike the fictional narrative, however, the context of the scientific [thought-experimental narrative] makes the intention clear to the reader that the situation is one that is to represent a potential real-world situation.

8 One then "[follows] through a sequence of events or processes as one would in the real world": The function of the narrative form of presentation of a thought experiment is to guide the reader in constructing a structural analog of the situation described by it and to make inferences through simulating the events and processes depicted in it. So, as with other forms of discourse models, the operations carried out in executing the thought experiment are performed not on propositions but on the constructed model. (p.297)

9 For example, “being in a chest in outer space” (Nersessian, 1993, p.295).

10 According to Nersessian (1993, p.297): The assumption is that if the experiment could be performed, the chain of events would unfold according to the way things usually take place in the world.

11 According to Mach (1926/1976, p.137): The outcome of a thought experiment, and the surmise that we mentally link with the varied conditions can be so definite and decisive that the author rightly or wrongly feels able to dispense with any further tests by physical experiment.

12 Nersessian (1993, p.297, emphasis added): Exhibiting the soundness of a thought experiment by reconstructing it as an argument can perform an important rhetorical function. However, real-world experimental outcomes can be recast in argument form as well, but no one would argue that the experiment can be replaced by the argument. In similar fashion, we need to differentiate between the reasoning that is done with the thought experiment and that which is done with the reconstruction of it. Thought experimenting is a complex form of reasoning that integrates various forms of information — propositions, models, and equations — into dynamic mental models [and links] the conceptual and the experiential dimensions of human cognitive processing...
matter experts, feedback from actual users.

It may well involve the input of, say, twenty colleagues and the audiences of, say, eight dedicated seminars, the designer’s own periodic reflections, whatever intricate and complex rebuttals the designer may constructed over, say, a 36-month period to overcome challenges identified during the polishing process.

Thus, the extended interaction between the designer and the scientific community at large that produces the ultimate, polished version of the thought experiment follows the typical pathway through which all scientific notions develop:
According to Nersessian, the polishing of thought experiments is a complex process:

We are not shown the tinkering that went into setting it up and refining it.

The reader/listener rarely, if ever, gets a glimpse of failed thought experiments or avenues explored in constructing the one presented to them.\(^{13}\) (p.296)

Whilst nothing has been specifically written about the polishing process, there are...

---

\(^{13}\) The same applies to "physical" experiments: "published scientific conclusions tend to present results as faits accomplis, without mention of production or decision-making processes" (Star, 1983, p.208).
many significant parallels between the polishing of thought experiments and the testing and validation of Programmed Instruction materials:

(1) Each enterprise is driven by the view that the only measure of its product’s value is the outcome it generates in users’ minds;

(2) The products are usually created by a single designer;

(3) The products are designed from a view that all (appropriate) users will possess the specific pre-requisite knowledge, experience and skill levels demanded in all of the necessary domains;

(4) The products are designed to produce a specific result in all (appropriate) users, and are constantly polished and re-polished in order to achieve that end;

(5) The products are always constructed top-down by their designers;

(6) The products are performed and understood bottom-up by their users;

(7) The products are designed so that “the [user’s] activity of choosing [seems to be] genuine, but [their] alternatives [are, in fact,] not” (Black & Tweedale, 2002, p.295);

(8) Each enterprise is driven by a clear, implicit understanding that initial designs all fail for some reason or other (Markle, 1992, p.235);

(9) Each enterprise is driven by the view that any faulty output, misunderstanding, or misinterpretation is 100% a design fault, it is never a user error;

(10) The products are artifacts, not entities, and from their capacity to produce change, they are tools (or agents of change), rather than just objects;

(11) In terms of the changes they generate, the products have an equal capacity

---

14 See, for example, Basescu (1963), Cook (1964), Espich & Williams (1977), Markle (1992), and Nersessian (1993).

15 The designer’s intentions are irrelevant.

16 This, of course, brings with it all sorts of “ego involvement” and “exaggerated familiarity” with the product (Basescu, p.119).

17 Artifact (lit. “produced by skill”), entity (lit. “thing”).
to create (or augment) or destroy (or diminish);\(^\text{18}\)

(12) Each enterprise is driven by the view that its final, polished, written product must stand alone, as written, without any need for additional instruction, coaching, clarification or explanation; and

(13) The products are delivered in a precise, well-polished written form, comprised of a very specific, fixed sequence of words which are read silently by each individual user (or recited, precisely as written, by some other person).

The following observations have been gleaned from the literature review:

(1) The designer must have a clear notion of the typical user’s knowledge.\(^\text{19}\)

(2) The source must be analysed to ensure that it is not:
   
   (a) Inconceivable;\(^\text{20}\)
   
   (b) Unimaginable;\(^\text{21}\) or
   
   (c) Too bizarre\(^\text{22}\) or too fantastic;\(^\text{23}\) or

\(^{18}\) The nature of the changes they generate are dictated by their application alone; and have no connexion whatsoever with the natural character of the artifact itself (which, of course is outcome-neutral).

\(^{19}\) The designer must determine the levels of previously acquired general and specialized knowledge, thought experimentation skills, and problem solving experience that will be assumed to exist in the typical target user before commencing to design the experiment (Markle, 1992, p.174).

\(^{20}\) If the narrative is either unintelligible, through (resolvable) under-description, or is totally incoherent, the intended source can not be conceived; and, therefore, can not be imagined.

\(^{21}\) In this case, whilst the source can be mentally apprehended, it can not be imagined; e.g., Descartes’ example of changing from a 9,999-sided figure to a 10,000 sided figure.

\(^{22}\) An example of a “bizarre scenario” thought experiment is the 1994 Australian film, Lucky Break, starring Gia Carides and Anthony LaPaglia (they later married each other, in ‘real life’), which posed the fascinating hypothetical question: “if an otherwise attractive young woman, shunned because of the calipers she must wear, and the crutches she must use (because of childhood polio), breaks both legs, and is in plaster from hip to toe, and strangers assume that she is “perfectly normal” with her legs broken in an unfortunate skiing accident, will she be treated differently, especially by a handsome man she has fancied from a distance?”.

The answer, of course, was a resounding “Yes”!

The bizarre aspect of this otherwise intriguing and extremely realistic film is the way in which the heroine’s leg is broken: she is on the ground floor of a shopping mall and has her legs broken by an enormously fat man falling on her legs from four stories above her. Even more bizarre, the fat man has toppled over the fourth floor balcony in the midst of his performance in a Luciano Pavarotti impersonation contest that was being held on the causeway of the fourth floor of the shopping mall. The bizarre scene that greeted the ambulance staff: a young woman with leg irons and crutches, in excruciating pain, squashed beneath a grotesque, greasy-black-dyed-haired, immensely fat man in a black tie and an ill-fitting dinner suit, forlornly mopping his sweating brow with a white handkerchief.

Through the extraordinarily bizarre “breaking the leg” episode, we are, effectively, given “a leg is broken” (the extraordinary nature of the breakage is never mentioned again); when, if it had been broken in more mundane circumstances, there might have been many distracting peripheral issues.
(d) Equivocal.\(^{24,25}\)

(4) The source must be scrutinized to "[ensure] that the [particular] intuitions the thought experimenter wishes to invoke [will be] universally shared by his audience"\(^{26}\) (Cohnitz, 2002b, p.2).

(5) The source must be closely examined to ensure that it does not communicate a message entirely different from that intended by the designer.\(^{27}\)

For example, specifically because of the intrusive efficiency, the persuasive power, and the non-linguistic cross-cultural appropriateness of the visual

\(^{23}\) The extent to which the situation in any thought experiment is so far removed from "reality" that it can only be considered to be a total fantasy is, also, an index of the degree to which anything is possible; and, in such a case, there could be no reliable "transfer" back to any identifiable real world circumstances.

\(^{24}\) Supposing the source is a set containing Adolf Hitler, Charles Manson and Ayatollah Khoumeini, is the theme of the source that rises in the minds of users "strict vegetarians" or "psychopaths"?

\(^{25}\) In these three situations, it is a case of the source not communicating anything at all; essentially because it has a zero-meaning.

\(^{26}\) Cohnitz (2002b), p.2:

Some might find a certain scenario perfectly conceivable, whereas others are convinced that the case described is surely impossible, and still others might claim they do not know what to say at all about such a counterfactual situation.

Essentially, this is where the designer’s claim for the source is unsubstantiated.

Imagine that Ahmed is a Turkish-Cypriot and that Patrick wants to insult him. Patrick knows little about Turkey, except that the Turks were brave opponents for the ANZACs at Gallipoli, and that Greek Archbishop Makarios was the scourge of Turkish Cypriots.

So, Patrick decides to serve him Greek coffee. He makes the coffee the traditional Greek way, sintering the coffee on and off the fire on the kitchen stove in a special Greek coffee pot. He takes a little Greek coffee cup, a glass of water, and some of the beautiful, delightful Greek sweet loukoumia on a plate, and places them all on a tray, and takes the tray to Ahmed.

Patrick is astonished at Ahmed’s instant response: a broad, happy, friendly smile.

What Patrick does not know is that, in Ahmed’s view, he has been very friendly; for, from Ahmed’s perspective, he brought Turkish coffee and the Turkish sweet, rahat lokum (lit. "throat’s ease"), known in English as Turkish Delight.

\(^{27}\) The source might not map onto the target the way claimed; e.g., "[many argue] that the story of the violinist [Appendix Eight] is just not analogous to pregnancy and abortion" (Wiland, 2000, p.466).
communication embedded within the famous sign,

that is widely and actively employed to deliver the message:\textsuperscript{28}

\textit{Smoking Cigarettes is Prohibited in this Location}\textsuperscript{29}

Based on the fact that:

(a) the “burning coal” is at the end of the cylinder, and

(b) the dark patch extends along something like a quarter of the cylinder,

even a cursory examination of the sign indicates that it very emphatically, very forcefully and quite unequivocally delivers an entirely different message:

\textit{Don’t Light your Cigarettes at the Filter-Tip End}\textsuperscript{30}

(6) The narrative must be checked for\textit{ unsound} argument (Gendler, 2000, p.22);

\textsuperscript{28} That is, given the world-wide convention that whatever it is that is represented graphically within a red, bisected circle is forbidden at the location in question.

\textit{(In relation to this specific image, I have been unable to determine the name of the artist/designer.)}

\textsuperscript{29} From a community health point of view, it is unfortunate that if this really was the case, and the sign\textit{ actually} said “Abandon all cigarettes, ye who enter here!”, to the extent that it was obeyed by all who entered the area, it would also, through the same cognitive mechanism, deliver the subliminal command “Smoke Now!” to (those-up-to-that-point-obediently-abstinent) smokers leaving the area.

\textsuperscript{30} The designer fails to communicate ABC; but, rather than communicating nothing, the alternative message of XYZ (mutually exclusive of ABC) is sent. It is XYZ that is mapped onto the target. This is far worse than transmitting a zero message (which results in no experiment being performed); here, the experiment “works”, but supports an entirely different conclusion from that intended.
i.e., either:

(a) independently unsound argument;\(^{31}\) or

(b) situationally unsound argument\(^{32,33}\) (Gendler, 2000, p.22).

(7) To ensure technical accuracy, the designer must consult subject matter experts (Espich & Williams, 1977, p.20).\(^{34}\)

(8) The experiment must be checked by one “who has not had as much exposure to the [experiment] as [the designer]” (Basescu, p.119),\(^{35}\) as early as possible in order to eliminate compounding of unnecessary errors.\(^{36}\)

(9) The narrative must be checked for grammar, terminology, language, spelling,

\(^{31}\) Gendler (2000, p.22):
An argument is independently unsound if it is unsound for reasons having nothing to do with the thought-experimental scenario.

\(^{32}\) Gendler (2000, p.22):
An argument is situationally unsound if its use depends on appeal to a principle that cannot justifiably be employed in the way that the author wishes. This may be (a) because something about the imaginary scenario implies that the ordinary criteria for application of a particular concept are unavailable or, (b) because the ordinary justification for appeal to a particular principle is missing.

Gendler (p.24) explains further in relation to the “famous violinist” experiment (Appendix Eight):
Although we can imagine there being such a machine, if we try to reason about how we would react to the possibility of being hooked up to it, we will inevitably make judgments on the basis of assumptions that would not be relevant under such circumstances.

\(^{33}\) Alternatively, the experiment may not support the designer’s conclusion (Wiland, p.466). For example, from a view that those responsible for such a deception would communicate the fact to those bright enough to uncover the coded message, a claim is made that the “Shakespeare” who wrote the famous plays was not a real person, based on the following well-established facts:

(1) “Shakespeare” was born on 23 April 1564.

(2) The authorized King James version of the Bible was published in 1611.

(3) It was published in the year that “Shakespeare” was 46 years old.

(4) Psalm 46 appears in the King James Bible.

(5) The 46th word from the beginning of Psalm 46 in the King James Bible is “shake”.

(6) The 46th word from the end of Psalm 46 in the King James Bible is “spear”.

\(^{34}\) This must be done as soon as possible in the designing process, because the situation described may not behave in the “real” world in the way the designer has supposed (e.g., the fact that the blue or purple amethyst turns yellow when exposed to heat is only known by an expert; a non-expert would assume that, like all other gemstones, it does not change colour).

Once the experiment is in its final form, it is essential to consult the relevant subject matter expert again, to ensure that any simplifications made during the polishing process have not inadvertently introduced technically inaccurate descriptions (Espich & Williams, p.110).

\(^{35}\) Also, it should be kept in mind that “argumentation, with its employment and study of argument forms, is not a set of principles or tools which leave the user unaffected. Reasoning, specifically evaluating arguments, affects the evaluator” (Tindale, 1999, p.101).

\(^{36}\) Because the polishing process may have inadvertently omitted an important step (or steps), designers must routinely check their amended versions against earlier, less polished versions (meaning these versions should be archived, rather than destroyed as obsolete) and, from time to time, designers must also check each sentence of their narrative in isolation, in case the constant re-edits over time have inadvertently made an individual sentence incoherent (Espich & Williams, p.101).
simplicity of expression, descriptive coherence, ease of communication, punctuation, and overall precision (Espich & Williams, p.99). 37

(10) The narrative must flow smoothly from item to item from beginning to end (Espich & Williams, p.100); 38 and provide neither far too much information, 39 nor far too little (Basescu, p.120). 40

(11) It is better to test on a subject "who is a little slower than the average [user], since he will probably stumble more often throughout the [experiment] than would most of the [typical users]" (Espich & Williams, p.113). 41

37 Counter-productive imprecision and incoherence may result from bad word choice. Typical problem choices lie in words or expressions that display the following features:

(1) **Ambiguity**: such as:
Australian doctor to (recently arrived from US) patient: "Are your bowels regular?"
Patient (tentatively, expecting some bad news): "Yes . . . I'm sure they're regular bowels."

As Galbraith (1977, p.236) remarked of followers of Keynes:
Like the Bible and *Das Kapital*, the *General Theory* is often deeply obscure. And, as with Marx and the Bible, this helped greatly to win converts. I am not reaching here for paradox. If you must struggle, really struggle, to understand a book, as with the Bible, Marx, and Keynes, you feel a commitment to the conclusions. After so much pain, the light. And, as again with the Bible and Marx, there were enough contradictions, enough ambiguities so that the discriminating reader could very often find just what he wanted to believe. Ambiguity, too, wins disciples.

(2) **Amphiboly or amphibology** (grammatical ambiguity): such as "the dog is not allowed to run outside" (is the dog kept inside, or is the dog allowed outside, but not permitted to run?), or "she sees more of her children than her husband" (does she see her children more often than her husband does, or does she see her husband less often than she sees her children?).

(3) **Equivocation**: in the strictest sense, equivocation means that a word or expression has one of two equal meanings (e.g., the famous motto of Winchester College, *Manners Makyth Man*, can be taken to mean either that manners are what typefy and separate men from beasts, or that a man is made through a boy's acquisition of manners). The pharmacist's slogan "We dispense with care" is also equivocal. Words such as *scale* and *sound* each have so many separate meanings that their particular intended meaning may need to be specifically emphasized in certain contexts.

(4) **Contranyms** (or *Janus words*): words with at least two contradictory meanings; e.g., *cleave* (to join together, and to render apart), *copy* (the original material used in print or artwork [advertising copy], and an imitation or reproduction of that material [photocopy]), *draw* (to pull an open curtain shut, and to pull a closed curtain open), *dress* (to put things on, or take things off; a woman might put some feathers on when dressing herself, but she'll always remove all of the feathers when dressing a chicken), *dust* (to add dust to something, and to remove dust from something), *strike* (to hit in golf, and to miss in baseball), *table* (to propose legislation in the British Parliament, and to set legislation aside in the U.S. Congress), etc.

38 This must take place in steps that are neither too large nor too small; as "too large a step size will lose the [user], and too small a step size may drive him to distraction" (Espich & Williams, p.101).

39 And, thus, rather than being *overdetermined*, the answer simply lies within the experiment.

40 Which would mean that the experiment is impossible to conduct. (In the realm of Programmed Instruction these two errors are known as "overprompting" and "undercuing".)

41 One particular advantage of this approach is that "many times the [designer] discovers that the smallest items have distracted the [user] and thrown him off the track. [Users] don't always think as the [designer] imagines they will" (Espich & Williams, p.113).

This may also highlight "words or concepts... which are not yet in the [routine target user's] repertoire". If a test-user has difficulty with the way things have been expressed, the designer must isolate the nature of the difficulty (Basescu, p.122); e.g., was the term, concept or expression that caused the user's problem *unknown, equivocal, or misleading*. 
(12) Testing is to find out what is wrong with the experiment (Basescu, p.122).

(13) Designers must neither conduct nor attend the testing (Basescu, 121-122); but, they (not the testers) must debrief test-subjects (Basescu, p.116).

(14) Because the extended polishing process may have caused an important step (or steps) to have been inadvertently omitted, the designer must constantly check amended versions against previous versions (Espich & Williams, p.101).

Finally, given Gopnik's notion that humans possess a "theory-formation system" (see the discussion in Chapter Four) that drives, and rewards, explanation-seeking behaviour, and that performers complete the task set by the designer in their own mind, it is imperative that the designer of a thought experiment is aware that the user’s successful performance not only generates the sense of euphoria described by Gopnik, but also a sense of psychological ownership over the experiment’s outcome. Thus, there is, or at least there should be, an onus on the designer of a thought experiment to

In many cases, rather than adding more and more stipulations, it may well be far more efficient to abandon the current experiment altogether and start all over again (Basescu, p.120).

42 This is, of course, is based on the sensible, pragmatic “[assumption] that [the test-user] made a sincere attempt to understand the material and failed for some good reason” (Markle, 1992, p.226).

Espich & Williams (p.113) suggest that it is productive to offer the test as an instance of engaging the test-subject’s assistance to write a better experiment.

Each test-subject should be told something like “Do your best to follow the instructions and the narrative exactly as written. This is not a test of your intelligence. I am testing the experiment as written, I am not testing you” to each test-subject (Basescu, p.122).

Thus, always, it is the experiment that fails; not the subject. Espich & Williams, p.114-115):

It should be impressed upon [subjects] that they are [performing] a draft of a [thought experiment], one that is still in the developmental stage, and that it is the [experiment] that is being tested, not them. [Subjects] should understand that they act merely as advisers telling the [designer] what is wrong with the [experiment as presented] and as a bench mark against which he will measure how well the [experiment] performs.

43 This means, of course, the test-subject is entirely dependent upon (a) the thought experiment’s narrative-as-written-by-that-designer, and (b) the thought experiment’s narrative-as-understood-by-the-individual-user, rather than an amended, interactive understanding that is consequent upon the designer’s additional instructions, elaborations, explanations or promptings (Basescu, p.121-122).

44 In particular, this allows the designer to discuss the user’s problems directly with the user (Espich & Williams, p.115) and check the test-user’s response to the experiment (and, even, the relevance of their response) in order to determine whether additional polishing is necessary (Espich & Williams, p.101). The designer may have incorrectly presupposed the possession of important, pertinent information. Or, even, for the ease of writing, the designer omitted some significant thing from the narrative (Espich & Williams, p.110) — this usually happens when the designer has found it just too hard to write about.

The subject’s performance may also reveal that the narrative’s level of organization, level of explanation, level of representation, level of description, or level of analysis must be adjusted (higher or lower) in order to ensure that the desired mapping from source to target takes place.

45 Also, from time to time, the designer must check each individual sentence in isolation, in case the constant re-edits have inadvertently made one or more sentences incoherent (Espich & Williams, p.101).

46 It seems to me that it is almost as though humans possess an "explanation-seeking module".
consider the psychological effect on the performer. I could say rather more about this, but it will have to await a further project.

For now, let us turn to the final chapter, where I hope to demonstrate that thought experiments are belief attenuating engines.
Chapter Seven
Thought Experiments as Belief Attenuating Engines

Time is intimately bound up with the content of human experience in that the past and future are reflected in the present. The past preconditions the present and is responsible for its taken-for-granted nature; the future is embedded in the present in terms of expectations, possibilities, and strivings...

In positing that the past and the future are connected to the present, it is obvious that not all aspects of the past and the future are important in defining the present, only those that are relevant to the phenomenon being studied whether at the individual, group, or organizational level of analysis...

[In order] to grasp the essential nature of a phenomenon it is important to understand how its existence at any point in time is a reflection of both the past and the anticipation of the future as they come together at any single moment in time. (George and Jones, 2000, pp.659-660)

I believe that the most productive way to understand thought experiments is to treat them as a special sub-set of what I have earlier termed hypothetical questions that employ subjunctive reasoning.

My extensive survey of the literature covering both hypothetical questions that employ subjunctive reasoning and thought experiments has revealed two patterns that seem to have been overlooked by other researchers and commentators: namely, their orientation in time, and their temporal direction and sense. In my view, descriptions based on these patterns allow unequivocally precise distinctions to be made amongst all types of thought experiment.

These qualitative, pattern-based distinctions that I have made, which have been made on the basis of a thought experiment’s qualitative arrangement of features —i.e., rather than on the quantitative basis of its application (as most others have done)¹ — have the extra advantage of facilitating a far wider, and far more objective discussion of the function of individual experiments.²

---

¹ For example, Brown’s (1993, pp.33-48) taxonomy, comprised of “destructive thought experiments”, “constructive thought experiments” (including “mediative”, “conjectural”, and “direct” thought experiments) and “Platonic thought experiments”.

² Essentially because most of the available taxonomical descriptions of thought experiments are based on some aspect of their function (and, so, tend to display the “begging the question” fallacy).
**Orientation in Time**

The first of characteristic pattern that thought experiments display is their orientation in time. They are either:

1. **Antefactual speculations**: those experiments which speculate about what might have happened prior to a specific, designated event, or
2. **Postfactual speculations**: those experiments which speculate about what may happen subsequent to (or consequent upon) a specific, designated event.³

This vital and useful distinction between *antefactual speculations* and *postfactual speculations* is the key to the construction the following concept map, which clearly elaborates the mechanism of thought experiments:

---
³ I have intentionally chosen the terms *antefactual* and *postfactual* for four reasons:

1. The Latin words *antefactum* (lit., “before the fact”) and *postfactum* (lit., “after the fact”) are widely used in legal circles; they denote “a thing done before” (or “a previous act or deed”), and “a thing done afterwards” (or “a subsequent act or deed”) respectively.
2. The concepts of “before the fact” and “after the fact” are widely and clearly understood and are free of any controversy.
3. The terms are symmetrical with others (viz., *prefactual*, *semifactual* and *counterfactual*) that are already extensively used in the thought experiment literature.
4. The meaning of (and the distinction between) the terms *antefactual* and *postfactual* is unequivocally clear within the context of thought experiments.

[2013 Note: It was from this work that the categories that now appear in the Wikipedia article on thought experiments were taken; not the reverse, LBY.]
Figure 3: Thought Experiments: A Concept Map


**Temporal Direction and Sense**

As well as their antefactual and postfactual orientations, thought experiments manifest another characteristic pattern: their movement in time in relation to “the present moment standpoint” of the individual performing the experiment.

In order to unequivocally identify this feature, and facilitate the further precision of its description, I draw an analogy with a convention of geometrical vector analysis, which clearly distinguishes between two closely related concepts: direction and sense.

In vector analysis, in relation to a line drawn to represent the particular course taken by a moving body from its original position to its final position:

(a) The line’s direction is its orientation: e.g., horizontal (vs. vertical).

(b) The line’s sense is one of the only two possible movements along that line, with each movement the reverse of the other: e.g., left-to-right (vs. right-to-left).

The identification of this second characteristic allows another useful distinction to be made, by analogy, amongst the sets of antefactual and postfactual thought experiments, on the basis of these additional features:

(1) Their temporal direction: are they past-oriented or future-oriented?

(2) Their temporal sense:

   (a) in the case of past-oriented thought experiments, are they examining the consequences of temporal “movement” from the present to the past, or “motion” from the past to the present? or,

   (b) in the case of future-oriented thought experiments, are they examining the consequences of temporal “movement” from the present to the future, or “motion” from the future to the present?

These characteristic features — viz., temporal direction and sense, and the orientation in time — allow the following table to be constructed, which identifies the four different, symmetrical patterns which exist amongst the seven types of thought experimental activity I have discussed.

---

4 See, for example, Weinreich, 1998.
Table Eleven: Temporal Symmetry Among Thought Experiments

<table>
<thead>
<tr>
<th>Orientation in time</th>
<th>Temporal Direction</th>
<th>Temporal Sense</th>
<th>Style of Thought Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Future-Oriented</td>
<td>Forwards</td>
<td>Prefactual; Prediction; Forecasting; Nowcasting.</td>
</tr>
<tr>
<td></td>
<td>Backwards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antefactual</td>
<td>Past-Oriented</td>
<td>Forwards</td>
<td>Counterfactual; Semifactual; Hindcasting.</td>
</tr>
<tr>
<td></td>
<td>Backwards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postfactual</td>
<td>Future-Oriented</td>
<td>Forwards</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Backwards</td>
<td></td>
<td>Backcasting.</td>
</tr>
<tr>
<td></td>
<td>Past-Oriented</td>
<td>Forwards</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Backwards</td>
<td></td>
<td>Retrodiction; Postdiction.</td>
</tr>
</tbody>
</table>

From this, it can be seen that a further set of patterned comparisons emerge:

Table Twelve: Enantiomorphic Symmetry Among Thought Experiments

<table>
<thead>
<tr>
<th>Antefactual</th>
<th>Postfactual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefactual thought experiments (and prediction, forecasting and nowcasting) are symmetrical with counterfactual and semifactual thought experiments (and hindcasting).</td>
<td>Backcasting is symmetrical with Retrodiction/Postdiction</td>
</tr>
<tr>
<td>Both involve the same temporal sense of a movement forwards in time</td>
<td>Both involve the same temporal sense of a movement backwards in time</td>
</tr>
<tr>
<td>Prefactual thought experiments begin their journey in the objective present.</td>
<td>Retrodiction/Postdiction begins its journey in the objective present.</td>
</tr>
<tr>
<td>Counterfactual and semifactual thought experiments begin their journey in the speculated past.</td>
<td>Backcasting begins its journey in the speculated future.</td>
</tr>
<tr>
<td>Prefactual thought experiments start with the objective present.</td>
<td>Retrodiction/Postdiction starts with the objective present.</td>
</tr>
<tr>
<td>Counterfactual and semifactual thought experiments begin with a speculated past.</td>
<td>Backcasting starts with a speculated future.</td>
</tr>
<tr>
<td>Prefactual thought experiments terminate in a speculated future.</td>
<td>Retrodiction/Postdiction terminates in a speculated past.</td>
</tr>
<tr>
<td>Counterfactual and semifactual thought experiments terminate in a speculated present (which may or may not be identical with the objective present).</td>
<td>Backcasting terminates in the objective present.</td>
</tr>
</tbody>
</table>

The symmetry exposed in Tables Eleven and Twelve indicate that these temporal features are quite real; and, in and of themselves, highly significant.
A further significant value of these features is that they allow significant distinctions to be made between the different styles of thought experiment of a nature that can facilitate discussion on the particular functions of specific thought experiments without the embedded “begging the question” problems that are concomitant with, for example, Brown’s “destructive”, “constructive”, and “Platonic” taxonomy (which presupposes such functions), as the following diagrams demonstrate:

1. The Prefactual Thought Experiment

This type of thought experiment speculates on future outcomes; and asks the question “What will be the outcome if...?”

![Diagram of Prefactual Thought Experiment]

Fig.14: The Prefactual Thought Experiment

2. The Counterfactual Thought Experiment

This type of thought experiment speculates on the possible outcomes of a different past; and asks the question “What might have happened if...?”

---

5 Questions such as “What would happen to Michael if you tell the police what you know?”, “What would happen to Michael’s neighbours if you remain silent about his illegal activities?”, “What would happen if the watercolour painting was hung in bright sun-light?”, “What would happen if the plant was totally deprived of sunlight?”, etc.

6 Questions such as “If Newton and Leibniz had cooperated with each other, what would mathematics look like today?”, “If the A bomb had not been used on Japan, would we still know as much today about the effects of full-body exposure to radiation on human beings?”, “If his own son had also been molested, would the inspector still have stopped the man from shooting the priest?”, “If the inspector had not been a Roman Catholic, would he still have dismissed the complaints brought against the priest?”, etc.
3. The Semifactual Thought Experiment

This type of thought experiment speculates on the extent to which things might have remained the same, despite a different past; and asks the question "Even though X, would Y have still occurred?"\(^7\)

---

\(^7\) Questions such as "Even if the goalie had moved left could he have intercepted a ball travelling at such a speed?", "Even if the goalie had not anticipated a left-side kick, would he still have been quick enough to save the goal?", "Even if you had your umbrella with you, would you still have got wet in such a fierce downpour?", "Even if you had not forgotten your umbrella would John still have given you a lift home from the station?", etc.

Semifactual questions, such as "Even if Jill had used her puffer before going out in the cold evening air, would she have still had an episode of asthma after smoking the joint down by the river?" are an important part of clinical medicine.
4. **Prediction, Forecasting and Nowcasting**

This type of thought experiment attempts to project the circumstances of the present into the future; the only difference between prediction, forecasting and nowcasting of the three is the distance of that speculated future from the present.\(^8\)

---

**Fig.17: Prediction, Forecasting and Nowcasting**

5. **Hindcasting**

This type of thought experiment involves running a forecast model after an event has happened in order to test whether the model’s simulation is valid.

---

**Fig.18: Hindcasting**

\(^8\) For brevity, the different activities of *prediction*, *forecasting*, and *nowcasting* are treated together because the three manifest precisely the same pattern of activity.
6. *Retrodiction or Postdiction*\(^9\)

This activity involves moving backwards in time, step-by-step, in as many stages as are considered necessary, from the present into the speculated past.

The goal of retrodiction or postdiction (or post-diction) is to establish the ultimate cause of a specific event.\(^10\)

---

7. *Backcasting*

This type of thought experiment involves the description of a definite, specific future situation and, then, moving backwards in time, step-by-step, in as many stages as are considered necessary, from the future to the present.

The goal of backcasting is to reveal the mechanism through which the specified future could be attained.

---

\(^9\) These terms are synonymous and can be used interchangeably. Each was coined using *prediction* as the model. According to the Oxford English Dictionary *retrodiction* was first used in 1895, and *postdiction* (or, sometimes, *post-diction*) was first used in 1940.

\(^10\) The “reverse engineering” of an unknown object, and the problem a corpse presents to a homicide detective and forensic chemist (e.g., the sequence “Did the individual die of “natural” causes?”, “If not, what caused their death?”, “How were they killed?”, etc.) are examples of this type of thought experiment.
In my extended discussion of thought experiments I have established a descriptive, concise, unequivocal and all-inclusive “generic” label for all thought experiment-like activities: "hypothetical questions that require subjunctive reasoning".

I have also created a concept map that unequivocally distinguishes between these activities on the basis of the questions they ask (viz., prefactual, semifactual or counterfactual) and their orientation in time (viz., antefactual vs. postfactual).

I have also created a set of seven diagrams that further distinguish between these activities on the basis of their temporal direction and sense.

From this, it is now possible to construct a clear definition of thought experiments, provided four particular characteristics of thought experiments are clearly identified.

The first is that all thought experiments exhibit a feature which Dear (2001, pp.138-139) also attributes to the experiments of Francis Bacon: they describe a sequence of actions that are supposed to be testing a set of circumstances for which the outcomes were already known before the test in question was performed.\footnote{This is rather like the philosophers’ (and scientists’) equivalent of the barristers’ aphorism “Don’t ask a question if you don’t already know the answer.”}

\footnote{Perhaps the only exception to this rule is Frank Stockton’s “The Lady, or the Tiger?” (1882) thought experiment (see Appendix One), in which a man must choose between two equiprobable alternatives. Stockton makes certain that his narrative supports each alternative choice with equal strength; and, therefore, the performer is really performing an experiment; i.e., rather than just running an algorithm. Yet, of course, the entire thought experiment \textit{must} be algorithmic — and, therefore, produce the same output with all users — up to the moment the man opened door on the right (otherwise any two subjects would be performing a different experiment).

And, because it is such an exceptional case, Stockton’s “The Lady, or the Tiger?” could well be dismissed as being one of Mackie’s (1976, p.175) “puzzle cases”; and, therefore, not a “genuine” thought experiment at all.}

I.

\begin{tabular}{|c|c|c|c|c|c|}
\hline
\textbf{REPORT} & \textbf{Describe each step; and, also, the Logical Antecedent} & \textbf{Current status quo} & \textbf{Future-Directed Postfactual Speculation} & \textbf{Create specific Consequent} & \textbf{~ Backcasting ~} \\
\hline
\end{tabular}
In giving examples, Bacon sometimes proposes tests the outcomes of which he already knows:

Try an experiment with burning glasses in which (as I recall) the following happens: if a burning glass is placed (for example) at a distance of a span [i.e. nine inches] from a combustible object, it does not burn or consume as much as if it is placed at a distance of (for example) a half-span, and is slowly and by degrees withdrawn to the distance of a span. The cone and the focus of the rays are the same, but the actual motion intensifies the effect of the heat. (Bacon, New Organon, Book II, aph.13, item 28)

The universality of this description of an experiment is part of its very effectiveness. By describing a trial the outcome of which Bacon claims to know, from the warrant of personal experience (“as I recall”), he tells the reader about something that happens in nature without actually tying it down to a specific event, an occasion on which this was tried with this outcome. Presenting experience in such a manner served to bypass, at least rhetorically, the difficulties that would arise if Bacon’s argument had depended on taking his word for an historical event that lacked corroborating witnesses (recall, too, that Bacon was a lawyer). By telling you what happens rather than what happened, and by giving an account in the form of instructions as to what to do to produce this claimed effect, Bacon can create the illusion of having revealed to his reader a fact about the natural world, one that can then be used to undergird a philosophical argument about the nature of heat.

From the understanding that a thought experiment is a bespoke artifice, specifically designed to mechanically produce a particular result, a thought experiment is best described as a device: a device for augmenting the intellect. By "augmenting man’s intellect" we mean increasing the capability of a man

---

13 From a view that "theories need not accommodate cases outside their scope", Suppe (1998a) argues that laboratory experiments "use experimental control to minimize the influence of factors not taken into account by the theory under test"; and "exclude observations contaminated by intrusion of identifiable extraneous factors". Thus, theories are "counterfactual descriptions of how real phenomena would behave were they isolated from influence by factors not taken into account by the theory". It takes a logician to appreciate how complex and sophisticated thought experimentation actually is. What it involves is not simply the drawing of an appropriate conclusion from a given fact; rather, it exploits the higher-level consideration that a particular thesis (be it fact or mere supposition) carries a certain conclusion in its wake. (Rescher, 1991, p.32)

Nersessian (1993) stresses the fact that "by the time a thought experiment is communicated it is in a [highly] polished form" (p.296).

14 The important advantage of this choice of terminology is that the word device is equivocal; and, therefore, along with its simple meaning of “a mechanical thing” it also, simultaneously, suggests connotations of “a contrivance”, “a scheme”, “a cunning plan”, “a rhetorical ploy”, and “a trick”. A further advantage is that it indicates that the thought experiment has been “devised”, rather than just having been “created”.

15 And, of course, another name for a device that augments human intellect is cognitive artifact.
to approach a complex problem situation, gain comprehension to suit his particular needs, and to derive solutions to problems. Increased capability in this respect is taken to mean a mixture of the following: that comprehension can be gained more quickly; that better comprehension can be gained; that a useful degree of comprehension can be gained where previously the situation was too complex; that solutions can be produced more quickly; that better solutions can be produced; that solutions can be found where previously the human could find none. And by "complex situations" we include the professional problems of diplomats, executives, social scientists, life scientists, physical scientists attorneys, designers — whether the problem situation exists for twenty minutes or twenty years. We do not speak of isolated clever tricks that help in particular situations. We refer to a way of life in an integrated domain where hunches, cut-and-try, intangibles, and the human "feel for a situation" usefully coexist with powerful concepts, streamlined terminology and notation, sophisticated methods, and high-powered electronic aids. (Engelbart, 1962, p.1; 1963, p.1).

The second characteristic common to all thought experiment-like activities is that they involve an interaction between an antecedent and a consequent: 16 either speculating on the nature of a possible consequent of a specified antecedent, or on the possible antecedent of a specified consequent.

The third characteristic common to all thought experiment-like activities is that they are designed by people who ask others to conduct them; 17 and, thus, the core notion of a performance seems critical. 18

16 This clearly seems to be the case; because you can remove any of the other features, and you will still have a thought experiment. Moreover, there is also the ever-present sense that the thought experiment harnesses an inevitable chain of events that relentlessly work to produce an inevitable consequent (i.e., given the occurrence of the antecedent "trigger").

17 It is significant that the designers themselves do not "perform" thought experiments. They create experiments in order to achieve a specific result; and implicitly assume that all performers will, as a consequence of their performance, assent to the designer's (pre-decided) conclusions. However, it is often the case that a given performer will reach a different conclusion for a wide range of performer-centred reasons (other than non-compliance, that is) or, even, find the experiment, as described, inconceivable.

18 Gendler (forthcoming, pp.4-5) takes the fundamental position on thought experiments that the core notion is one of "performing a thought experiment", and the derivative notion is that of "being a thought experiment"; and, on the basis that:

(1) "thought-experimental reasoning involves reasoning about a particular set of circumstances (which may be specified in more or less detail), described at a greater level of specificity than that of the conclusion",

(2) "the reasoner's mode of access to the scenario is via imagination rather than via observation", and

(3) "contemplation of the scenario takes place with a specific purpose" (i.e., "the confirmation or disconfirmation of some hypothesis or theory"),

she provides the following definition:

...to perform a thought experiment is to reason about an imaginary scenario with the aim of confirming or disconfirming some hypothesis or theory, and that to perform a scientific thought experiment is to reason
Finally, the fourth characteristic is that the designer may explicitly stipulate the domain in which the thought experiment takes place or, at the other extreme, may simply leave the domain unspecified, and presume that the performer has a commonsense understanding of that domain that totally matches the designer’s understanding.\(^{19}\)

From this, the following productive, all-encompassing definition can be produced:

\[\text{A thought experiment is a device with which one performs an intentional, structured process of intellectual deliberation in order to speculate, within a specifiable problem domain, about potential consequents (or antecedents) for a designated antecedent (or consequent).}\]

Having established this definition, it is now also possible to identify and describe the function of the device itself; and if a thought experiment could be considered to be like the physics students’ “black box”,\(^{20}\) there are two separate issues to be decided:

(1) “What is the function of a thought experiment?“; and

(2) “What is the internal structure and mechanism through which a thought experiment realizes that function?”

Once all the different styles of thought experiment are placed together, it becomes

---

\(^{19}\) And, rather than simply being polar opposites, there is a wide ranging continuum of degrees of domain specification (or non-specification) in between these extremes..

\(^{20}\) The black box (so-called because it is 100% opaque) containing a complex of electrical components and an array of wires connecting those components in particular sequences. The black box also has pairs of terminals on its outer surface. The students task is to use a standard set of instruments (ammeter, voltmeter, etc.) and determine:

(1) the function of the box as a whole (“What does the box do in its present configuration?”), and

(2) the precise internal circuit arrangement, within the box, that serves that function (“How does the box do it?”).
immediately apparent that a single pattern emerges.

Given that, in all cases, thought experiments:

(1) are devices constructed to produce particular outcomes;

(2) involve intentional, structured intellectual deliberative acts on the part of the individual performing the experiment;

(3) involve the transformation of a designated antecedent into its potential logical consequent, within the particular circumstances of the antecedent’s specified setting.\(^{21}\)

And, moreover, regardless of whether specifically designed to establish a new theory, refute a prevailing theory, challenge a prevailing theory, supplement deficiencies in a prevailing theory,\(^{22}\) correct some aspect of prevailing theory\(^{23}\) or, even, modify a prevailing theory,\(^{24}\) all thought experiments are offered from an embedded belief and a conscious intention that all performers will replicate the pre-determined outcome.

This observation that thought experiments have a pre-determined outcome has another benefit: it unequivocally decides, once and for all, the controversial issue of whether thought experiments are arguments and makes the question irrelevant. It is now clear that thought experiments are highly polished exercises in coercion and can never operate as _demonstrative_ arguments.

Given this fact — and adopting the simple convention that either _knowing nothing about a certain matter_ or _believing that everything is random in a certain domain_\(^{25}\) is

\(^{21}\) Or, it involves the transformation of a designated consequent into its potential logical antecedent, within the particular circumstances of the consequent’s specified setting.

\(^{22}\) By supplying some additional, missing aspect (without altering any of the otherwise correct aspects that are already present).

\(^{23}\) By correcting misleading aspects of the current concept (without altering any of the correct aspects).

\(^{24}\) For example, Coué’s Thought Experiment (Appendix Two) could have been destroyed if he had, inadvertently, asked a steeplejack (who would be easily able to imagine performing the task) to perform it; so, rather than abandoning the concept altogether, the prevailing theory could now speak of “imagining a normal person” — this allows an incorrect element to be slightly modified, rather than fundamentally changed or replaced altogether.

\(^{25}\) On the basis that many people have quite an unrealistic view of randomness — and often reject truly random sequences as being far too ordered, and not random enough — Lisanby & Lockhead (1991) differentiated between _subjective_ randomness and _genuine_ randomness (upon which they which they bestowed the bizarrely tautologous title of _stochastic randomness_).
the equivalent to having a “zero belief” — thought experiments can be immediately understood as a process of attenuating a belief.

For example, the performance of a thought experiment might establish a new belief (i.e., a zero-belief changes to a plus-belief), or it might destroy an old belief (i.e., a plus-belief changes to a zero-belief), or it might eliminate an old belief by installing another, mutually exclusive belief in its place (i.e., a plus-belief changes to a minus-belief), or it might reinforce an old belief (i.e., a plus-belief changes to a doubleplus-belief), or it might challenge an old, strongly held belief (i.e., a doubleplus-belief changes to a plus-belief), and so on.

Given the mechanical, structured and coercive nature of the thought experimental process, it is clear that the devices known as thought experiments function as Belief Attenuating Engines, and operate as follows:

![Belief Attenuating Engine: Input and Output](image)

Or, more precisely:

---

26 In the same way that linguists, when they discuss sequences such as one horse, two horses, one dog, two dogs, one tree, two trees, one sheep, two sheep, systematically explain that the word sheep in “two sheep” has a “zero-suffix”.

27 The belief can also be a belief in the sense of long-held, deeply felt, non-negotiable values.

28 This terminology is suggested by the conventions of the bureaucratic language introduced in George Orwell’s novel, *1984* (1949), viz., “Newspeak”.

29 Orwell’s terminology would, I think, demand “unbelief” for what I have termed “zero-belief”; viz., the active belief that something is very definitely not the case (rather than, what I have called a minus-belief, which is the polar opposite of the plus-belief). Despite the fact that the terminology here and below is my own, the concept is 100% Orwell’s.

30 And, as a consequence, individuals who have performed thought experiments have a tendency to make immediate-post-experiment statements such as “I am now much more certain of X”, “I am now much less certain of X”, “I am now 100% certain that X is the case”, “I am now 100% certain that X is not the case”, “I am now 100% certain that it is impossible for X to be the case”, “I can now see how X may be possible”, “I am now 100% certain that it is impossible to conceive X”, etc. — all of which testify to the presence of “corrected beliefs”.
Thus, whenever the output from the Belief Attenuating Engine is 100% consistent with the "original belief", the original belief remains unchanged; but, when an anomaly is detected in the output, this leads to an attenuation of the "original belief".

Therefore, I argue, the embedded assumption within the thought experiment process is that thought experiments function as a corrector of erroneous beliefs; given, of course, the convention that:

(1) a "zero-belief" can be an "erroneous belief"; and,

(2) an "unchanged belief" that is tested and found to have no need for attenuation, has only been tested because there has been an "erroneous belief" that it is not 100% correct.

Although this representation is helpful, it still does not clarify and elaborate the thought experiment process to the maximum extent possible.

Using Coué’s "walking the plank" thought experiment\(^1\) as an example — where Coué decided to use the concept of fear over-riding knowledge to challenge the (then prevailing) view that the "will" was more powerful than the "imagination" — and given that a thought experiment is a goal-directed device,\(^2\) it is useful to describe thought

---

\(^1\) See Appendix Two. This was chosen as an example of a thought experiment specifically because it lay outside the normal thought-experimental literature; and, therefore, it could be examined without any previously embedded commitment to any particular theoretical orientation or descriptive protocol.

\(^2\) It is goal-directed in three ways:

1. only teleological explanations are able to explain thought experiments;
2. a thought experiment is like an extrusion die — it is expected that, given an "input" matching the "input" intended by the designer, each and every performer will produce the same "output"; and
3. its very construction is dictated by the predetermined result it seeks — in precisely the same way that a corrupt interviewing committee spends a long time establishing the questions it will ask in order to select a predetermined candidate. Or, the same as an Australian Cabinet Minister arranging for a member of his/her own party to ask a "Dorothy Dix" question in "Question time" so that he/she can produce a pre-prepared "answer" (which is, really, the equivalent of a press release, but is made under parliamentary privilege).

While Fittler was making his [retirement] announcement at a press conference, selectors were meeting nearby to choose the City and Country teams, ostensibly as a guide to the make-up of the NSW team to play Queensland in the season’s first State of Origin match on May 26. But it seemed...
experiments as circumstances where the designer has identified a particular “challenge” which, then, in turn, determines the nature of entire thought experimental process:

So, firstly, having established that a thought experiment is device for feeding an erroneous belief (plus a specific challenge) into the Belief Attenuating Engine, it’s obvious that the Belief Attenuating Engine generates an output:

And secondly, this “input” and “output” can be understood as feeding the challenge into the Belief Attnuating Engine and generating an anomaly:

---

33 Which, of course, is “designer-internal” and is never revealed to the performer; because it is the “hidden” nature of the “challenge” that motivates the entire thought experimental process.

34 That is, the form and content of the “black box” (viz., the Belief Attenuating Engine).

35 This is, of course Fig.21 again, inserted here for the reader’s convenience.
The precise internal workings of the *Belief Attenuating Engine* are dictated by:

1. the precise nature of the “attenuated belief” which the thought experiment is designed to produce;
2. the precise nature of the “erroneous belief” which the thought experiment is intended to attenuate;
3. the precise nature of the “anomaly” that will impel the “erroneous belief’s” transformation into the desired “attenuated belief”; and
4. the precise nature of the “challenge”.

Obviously, each of these factors is experiment-specific.

Despite this, it is entirely possible to represent the internal structure and working mechanism of the *Belief Attenuating Engine* — i.e., the device that “runs” the thought experiment — in a qualitative manner that represents the conventions\(^{36}\) that are the same for all thought experiments:

---

\(^{36}\) Curtis (1994), p.419: 
The scientific paper, the printed communication to a learned journal, is a literary artefact, a form of discourse with a conventional style of exposition and a conventional rhetorical and narrative strategy.
Thomas Kuhn (1964/1977; 1970) has always strongly argued that the true function of thought experiments in science is that of engineering paradigm shifts.

Using the ‘Belief Attenuating Engine’ explanation that I have woven together, the rôle of thought experiments in one of Kuhn’s paradigm shifts can now be clearly understood as a two-part process of:

(1) belief attenuation; and

(2) creation of a new analogical arrangement (very often involving source-to-target mapping from an entirely new source domain) through which, to use Hesse’s terminology, the number of neutral analogies is significantly reduced.

Finally, moving step-by-step through this sequence, and using Coué’s "walking the plank" thought experiment as an example, we can follow the thought-experimental pathway from beginning to end:
Chapter Seven: The Belief Attenuating Engine — 157

1. *Erroneous Belief*: That the will is more powerful than one’s imagination.\(^{37}\)

2. *Challenge Identified*: A person frozen with the fear of anticipation.


4. *Analogue Selected*: using Hesse’s (1963, pp.9-11) tripartite division, certain relevant features (e.g., being able-bodied, having normal sense of balance, not being deaf), were implicitly selected as relevant.

   Certain (otherwise relevant) features (e.g., prevailing wind, height and weight of the plank walker) were implicitly rejected as irrelevant to the experiment.

   Others (e.g., gender, religion, race, and I.Q. of the plank walker were not even considered at all (and were, therefore, ignored).

5. *Narrative*:\(^{38}\) This projects the performer to the precise moment of the antecedent. The story of the plank on the ground, walking along the plank, and then the plank being hoisted to the height of the cathedral’s towers.\(^{39}\)


7. *The normal natural laws, causal processes and causal links of the “real world”*: Because no alternative circumstances have been specified, because no alternative circumstances have been described such that require a "natural law" to be amended in some way, because no alternative circumstances have been described such that some natural variable is made constant (or some natural constant made a variable), and because no alternative circumstances have been described such that some naturally present feature is to be removed (or some naturally absent feature is to be introduced), the normal natural laws, causal

---

\(^{37}\) The specific manifestation of this erroneous belief is embodied in the communiqué issued after the Plenary Session of the Second International Congress on Experimental Psychology, held in London in 1892 (Anon, 1892, pp.586-587), which stated that, in the collective opinion of those present, “therapeutic suggestion” (especially, “hypnotic suggestion”) is only effective to the extent that it is driven by the powers of a subject’s “will”.

\(^{38}\) “Thought-experiment narratives are like fictional stories in being narratives containing claims that mention individual items in ways that would have existential import if taken as fact” (Cross, 1995, p.354).

\(^{39}\) Irvine (1991), pp.160-161:

The reason that thought experiments so often can be misleading is that behind almost every thought experiment will lie a large number of unquestioned auxiliary assumptions, assumptions which are assumed to be true but which, if false, would overturn the result in question.
processes and causal links of the “real world” prevail in this experiment.

8. **Consequent**: The moment you step out towards the plank you are frozen with fear (the designer’s pre-determined result of performing the experiment).

9. **Anomaly**: This could not and would not happen if the will was more powerful than the imagination. According to the precepts of the “original belief” this *should not take place*; but it has. Thus, the “original belief” is “erroneous”.

10. **Attenuated Belief**: One’s imagination is more powerful than one’s will.

From this, it is clear that the “antecedent-caused consequent-moment” drives the entire belief attenuating engine.

In other words, the extended pathway from the origin to the antecedent has been constructed so that — given the consequences of the natural laws, causal processes and causal links of the “real world”, given the precise particulars of scenario contained in the narrative, given the features that the narrative selected, rejected, or ignored in drawing analogies from the chosen antecedent domain, and given the nature of the challenge made to the supposed erroneous belief within the narrative — the antecedent will unequivocally produce the predetermined consequent, in the same way that a rod is produced by impelling nonferrous metal through an extrusion die.

And at this stage the thought experiment, precisely like a catalyst, disappears and “falls away” leaving the experimental product in its wake; and, of course the user also automatically has, from this, a persuasive anecdote to relate.

The following table shows the above sequence represented in a standard format that can be used to display the workings of most thought experiments:
## Table Thirteen: Standard Format for Representing the Thought-Experimental Pathway

<table>
<thead>
<tr>
<th>STAGES</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a. Erroneous Belief</td>
<td>That the will is more powerful than the imagination.</td>
</tr>
<tr>
<td>1b. Specific Manifestation of Erroneous Belief</td>
<td>“Therapeutic suggestion” (especially “hypnotic suggestion”) is only effective to the extent that it is driven by the powers of a subject’s “will” (from the Communiqué issued after Plenary Session of the 1892 Second International Congress on Experimental Psychology: Anon (1892), pp.586-587).</td>
</tr>
<tr>
<td>2. Challenge Identified</td>
<td>A person frozen with the fear of anticipation.</td>
</tr>
<tr>
<td>3. Antecedent Domain Chosen</td>
<td>Fear of heights.</td>
</tr>
<tr>
<td>4a. Analogue Features Selected</td>
<td>Being able-bodied, having normal sense of balance, not being deaf.</td>
</tr>
<tr>
<td>4b. Analogue Features Rejected</td>
<td>Prevailing wind; height and weight of the plank walker.</td>
</tr>
<tr>
<td>4c. Analogue Features Ignored</td>
<td>Gender, religion, race, and I.Q. of the plank walker.</td>
</tr>
<tr>
<td>5. Narrative</td>
<td>The story of the plank on the ground. You walking along the plank, and then the plank being hoisted to the height of the cathedral’s towers.</td>
</tr>
<tr>
<td>6. Antecedent</td>
<td>The moment you attempt to step out on the plank.</td>
</tr>
<tr>
<td>7a. Natural Laws</td>
<td>The normal natural laws of the “real world” prevail in this experiment.</td>
</tr>
<tr>
<td>7b. Causal Links</td>
<td>The normal causal links of the “real world” prevail in this experiment.</td>
</tr>
<tr>
<td>7c. Causal Processes</td>
<td>The normal causal processes of the “real world” prevail in this experiment.</td>
</tr>
<tr>
<td>8. Consequent</td>
<td>The moment you step out towards the plank you are frozen with fear (the designer’s pre-determined result of performing the experiment).</td>
</tr>
<tr>
<td>9. Anomaly</td>
<td>This could not and would not happen if the will was more powerful than the imagination. According to the precepts of the “original belief” this should not take place, but it has. Thus, the “original belief” is “erroneous”.</td>
</tr>
<tr>
<td>10. Attenuated Belief</td>
<td>The imagination is more powerful than the will.</td>
</tr>
</tbody>
</table>

My results in this chapter are rather novel. I hope that I have provided not just a useful answer to the question, “What is a thought experiment?” but also a set of graphic tools that will be helpful to researchers in the field. And I’d certainly like to believe that I have brought some cognitive insights, and order, to the topic.
Chapter Eight: Conclusion

... those who paint landscapes place themselves in a low position on the
plain in order to consider the nature of the mountains and the high places and
place themselves high atop mountains in order to study the plains...

Niccolo Machiavelli (1469-1527)¹

Once upon a time, in a faraway land, a rich, powerful and tyrannical King decided to
have his portrait painted for posterity. He summoned the three best artists in his
Kingdom, and said: “He that paints the best portrait will be given enough riches to
support himself and his family forever.”

The artists smiled.

“But”, the King continued, “the other two will be put to death.”

Their smiles vanished.

“And, by the way”, said the King, almost as an afterthought, “it must be realistic.”

The three artists trembled; for the King was not an ordinary man . . .

He was a hunchback; his left leg, damaged in a fall from a horse when he was nine,
was seven inches shorter than his right; his right arm, deformed since birth, was five
inches shorter than his left; his left eye, blinded in a hunting accident, was permanently
closed; and half of the tip of his right index finger had recently been bitten off by his pet
monkey. . .

The first artist presented a realistic portrait, showing every grotesque feature of the
King, in excruciating detail, illuminated by the bright noonday sun.

He was executed by a very angry King.

The second, learning something from the first, presented a fanciful portrait of a
handsome superman, with every feature healthy, strong and symmetrical; intending to
suggest that this was the King’s true inner nature.

¹ From the dedication to Lorenzo de Medici in Macchiavelli’s Prince (Bondanella & Musa, 1979, p.78):
And although I consider this work unworthy of your station, I am sure, nevertheless, that your humanity
will move you to accept it, for there could not be a greater gift from me than to give you the means to be
able, in a very brief time, to understand all that I, in many years and with many hardships and dangers, came
to understand and to appreciate.

I have neither decorated nor filled this work with fancy sentences, with rich and magnificent words, or
with any other form of rhetorical or unnecessary ornamentation which many writers normally use in
describing and enriching their subject matter; for I wished that nothing should set my work apart or make it
pleasing except the variety of its material and the seriousness of its contents.

Neither do I wish that it be thought presumptuous if a man of low and inferior station dares to debate and
to regulate the rule of princes; for, just as those who paint landscapes place themselves in a low position on
the plain in order to consider the nature of the mountains and the high places and place themselves high atop
mountains in order to study the plains, in like manner, to know well the nature of the people one must be a
prince, and to know well the nature of princes one must be of the people.
He, too, was executed by an angry King.

The third, having carefully considered the fate of both his colleagues, produced an immense, lavish portrait that realistically depicted the King with all his deformities. The King was overjoyed with this realistic portrayal, and showered immense riches on the artist, built him a beautiful castle, and gave him the hand of his youngest daughter in marriage.

So, what was the essential difference between the work of the first and third artists? The third artist portrayed the King in a hunting scene, standing in the forest, with his weight firmly resting on his good leg, and his left leg perched on the trunk of a fallen tree, aiming his hunting rifle at a distant stag; giving a very different significance to his closed left eye, humped back, and short left leg. And, with the rifle butt pulled firmly into his right shoulder, and his right index finger squeezing the trigger, neither the King’s short arm nor his missing fingertip were noticed.

I hope that I have been just as realistic in my account of thought experiments.²

As I began my literature survey it became obvious that the entire field was in great disorder; and, although it seemed that each scholar had a gut-level intuitive response that allowed them to say that this is a thought experiment,³ and that that is not, it appeared that none of them could provide an adequate definition. It seemed odd that so many scholars were speaking and writing so emphatically about an entity that could not eventually be isolated and spoken about; yet, because this general understanding of thought experiments could be motivated, but not described, it began to appear that the individuals in question:

(1) were either just slavishly following somebody who had told them that this particular entity was an instantiation of a thought experiment, or

(2) actually were able to distinguish between the things which are thought

² For the sake of clarity, in the account that follows I have (anachronistically) applied the precise terminology that I had established by the end of my extensive literature search, and consequent upon many long post-search deliberations.

³ This was, of course, from the perspective of their idiosyncratic orientation; however, my point still stands because these individual differences proved to be irrelevant, as all of them seemed to be able to display this talent with respect to their own theoretical domain.
experiments and the things which are not thought experiments through the application of some (externally inexpressible) internalized standard.  

I had encountered a similar type of problem some years ago, whilst at the Australian National University, when attempting to clarify traditional Chinese medical concepts for conventional (western) medical practitioners. From this experience, I knew that the only way that this “inner prototype” might possibly be exhumed is through an examination of specific, “classic” thought-experimental activities; and, from this, my experience told me, there was a good chance that a pattern would emerge. (Nothing of any value emerged from a study of gedankenexperiments; mainly because the mental gymnastics involved with conducting these imaginary enterprises in, say, quantum physics were so question-beggingly connected with whatever the experiment in question was supposed to elaborate.)

A general examination of scientists’ application of thought experiments found that scientists almost exclusively used prefactual thought experiments. As I moved into other domains that routinely employed thought experiments, I discovered that perhaps 90% of the thought experiments in use were counterfactual thought experiments. This led on to an examination of what philosophers had to say about counterfactuals and, from this, what social psychologists had discovered about counterfactual thinking; which led me, in particular, to the field of medicine, where semifactual thought experiments are an essential part of everyday professional life.

---

4 In other words, they had evolved an inner prototype of “thought-experiment-ness” in much the same way that we learn, as children, to establish an inner prototype of “green-ness” by being told endlessly that “This is true green”, “This is quite green”, “This is rather greenish”, “This is not green at all”, and so on, until we suddenly know (by acquaintance) precisely what green is; although we certainly can’t describe it to, say, a blind man, we can certainly identify what is green and what is not.

5 Provided, of course, that I concentrated on the qualitative task of “[focussing] on process rather than [the quantitative task of focussing on] product: [on] the how of theorizing, rather than the what of theory content” (Folger & Turillo, 1999, p.742).

6 Also, despite having studied applied physics (as a therapy radiographer and a chemical engineer) at tertiary level for at least 6 years, I found the entire speculative domain of quantum physics unintelligible and counter-intuitive; particularly because the concepts that I was struggling to understand, in order to identify what the author was claiming about a specific thought experiment, were being explained by the thought experiment itself.

I strongly suspect that the massive cognitive load imposed by the task of imagining the precise details of the thought experiment’s fanciful narrative — and continuously keeping all of the extraordinary sets of thought-experiment-specific (internally consistent) conventions in mind — meant that I had no cognitive resources left over with which I could actually perform (i.e., rather than set up) the experiment.

7 For example, law, history, and political science.
It was only when I recognized that the work of a homicide detective (or a forensic chemist) was a thought-experimental activity, that I came to see that the only way to understand the literature (and the orientation of the various scholars that had contributed collectively and severally to that literature) and establish precise sets of descriptive mechanisms and a well-calibrated distinctive terminology, was to examine every sort of thought-experimental activity possible; and, as soon as I did this, everything began to fall into place.  

Once the literature survey was complete, and the descriptive mechanisms had been established and tested, and the necessary terminology had been created and calibrated, important, simple and unequivocal distinctions could be made between:

1. hypothetical questions that employ subjunctive reasoning in general, and thought experiments in particular;

2. fanciful hypothetical questions in general, and highly polished thought experiments in particular;

3. thought experiments in general, and Mach’s *gedankenexperiments* in particular;

4. Mach’s *gedankenexperiments* in particular, and real, physical experiments in general;

5. those *gedankenexperiments* that were proxies for real, physical experiments,

---

8 Including prediction, forecasting and nowcasting, retrodiction (or postdiction), backcasting, and hindcasting, as well as prefactual, counterfactual, and semifactual thought experiments.

9 This approach, in itself is not so unusual. The notion of expanding one’s field of interest in order to make sense of an otherwise inexplicable narrower domain, has many precedents.

Danesi (2002, p.156) cites a famous puzzle, constructed by Niccolò Fontana Tartaglia (circa 1499-1557 CE), that specifically highlights a similar problem:

_A man dies leaving 17 camels to be divided amongst his heirs in the proportions 1/2, 1/3, 1/9. How can this be done?_

The core of this seemingly intractable problem is that, in order to divide the 17 camels in the manner specified, at least one of the camels must be split; which, of course, would kill it.

The innovative solution offered by Tartaglia is to borrow an extra camel; something which not only makes the mathematics simple, but saves bloodshed, and leaves all of the bequeathed 17 camels alive.

So, given this artifice, from the (now) 18 camels we subtract 9 camels (or one half), 6 camels (one third) and, finally, 2 camels (one ninth); which leaves us with:

(a) the _intact_ legacy of 17 camels, which has been distributed according to the testator’s wishes; and

(b) the _remainder_ of 1 camel, which is returned _unharmed_ to its original owner.

10 That is, simple in the sense of being non-complex; not simple in the sense of being immediately obvious to even the most superficial observation by the least interested observer.
and those that were not;

(6) thought experiments in the so-called “hard” sciences, and thought experiments in the social sciences;

(7) prefactual thought experiments, counterfactual thought experiments, and semifactual thought experiments;

(8) antefactual thought experiments, and postfactual thought experiments;

(9) past-oriented thought experiments, and future-oriented thought experiments;

(10) thought experiments which examine movement backwards in time, and those which examine movement forwards in time;

(11) descriptions of thought experiments based on their structures, and descriptions of thought experiments based on the types of problem they address; and

(12) analogies, and metaphors and models of different types.

It is highly significant that all of these systematic patterns, integrated frameworks, and overarching symmetrical understandings were developed “bottom-up”, from an interaction with the literature; and were not developed “top-down” from any sort of a prior expectation. Yet, despite all of this, this particular approach has also produced a perspective, a set of understandings, and a constellation of descriptive mechanisms that can be motivated “top-down”.

It is also significant that this approach has produced, amongst other things a productive, all-encompassing definition of thought experiments (“a thought experiment is a device with which one performs an intentional, structured process of intellectual deliberation in order to speculate, within a specifiable problem domain, about potential consequents (or antecedents) for a designated antecedent (or consequent”), a concept map describing thought experiments, and a set of seven diagrams that isolate the unique features of different sorts of thought experimental activities. In addition, it has produced the additional understanding of what is shared by all thought experiments: that a thought experiment provides “knowledge by acquaintance”, and is a powerful device for attenuating beliefs.
IN the very olden time there lived a semi-barbaric king, whose ideas, though somewhat polished and sharpened by the progressiveness of distant Latin neighbors, were still large, florid, and untrammeled, as became the half of him which was barbaric. He was a man of exuberant fancy, and, withal, of an authority so irresistible that, at his will, he turned his varied fancies into facts. He was greatly given to self-communing, and, when he and himself agreed upon anything, the thing was done. When every member of his domestic and political systems moved smoothly in its appointed course, his nature was bland and genial; but, whenever there was a little hitch, and some of his orbs got out of their orbits, he was blander and more genial still, for nothing pleased him so much as to make the crooked straight and crush down uneven places.

Among the borrowed notions by which his barbarism had become semified was that of the public arena, in which, by exhibitions of manly and beastly valor, the minds of his subjects were refined and cultured.

But even here the exuberant and barbaric fancy asserted itself. The arena of the king was built, not to give the people an opportunity of hearing the rhapsodies of dying gladiators, nor to enable them to view the inevitable conclusion of a conflict between religious opinions and hungry jaws, but for purposes far better adapted to widen and develop the mental energies of the people. This vast amphitheater, with its encircling galleries, its mysterious vaults, and its unseen passages, was an agent of poetic justice, in which crime was punished, or virtue rewarded, by the decrees of an impartial and incorruptible chance.

When a subject was accused of a crime of sufficient importance to interest the king, public notice was given that on an appointed day the fate of the accused person would be decided in the king's arena, a structure which well deserved its name, for, although its form and plan were borrowed from afar, its purpose emanated solely from the brain of this man, who, every barleycorn a king, knew no tradition to which he owed more allegiance than pleased his fancy, and who ingrafted on every adopted form of human thought and action the rich growth of his barbaric idealism.
When all the people had assembled in the galleries, and the king, surrounded by his court, sat high up on his throne of royal state on one side of the arena, he gave a signal, a door beneath him opened, and the accused subject stepped out into the amphitheater. Directly opposite him, on the other side of the enclosed space, were two doors, exactly alike and side by side. It was the duty and the privilege of the person on trial to walk directly to these doors and open one of them. He could open either door he pleased; he was subject to no guidance or influence but that of the aforementioned impartial and incorruptible chance. If he opened the one, there came out of it a hungry tiger, the fiercest and most cruel that could be procured, which immediately sprang upon him and tore him to pieces as a punishment for his guilt. The moment that the case of the criminal was thus decided, doleful iron bells were clanged, great wails went up from the hired mourners posted on the outer rim of the arena, and the vast audience, with bowed heads and downcast hearts, wended slowly their homeward way, mourning greatly that one so young and fair, or so old and respected, should have merited so dire a fate.

But, if the accused person opened the other door, there came forth from it a lady, the most suitable to his years and station that his majesty could select among his fair subjects, and to this lady he was immediately married, as a reward of his innocence. It mattered not that he might already possess a wife and family, or that his affections might be engaged upon an object of his own selection; the king allowed no such subordinate arrangements to interfere with his great scheme of retribution and reward. The exercises, as in the other instance, took place immediately, and in the arena. Another door opened beneath the king, and a priest, followed by a band of choristers, and dancing maidens blowing joyous airs on golden horns and treading an epitaphic measure, advanced to where the pair stood, side by side, and the wedding was promptly and cheerily solemnized. Then the gay brass bells rang forth their merry peals, the people shouted glad hurrahs, and the innocent man, preceded by children strewing flowers on his path, led his bride to his home.

This was the king’s semi-barbaric method of administering justice. Its perfect fairness is obvious. The criminal could not know out of which door would come the lady; he opened either he pleased, without having the slightest idea whether, in the next instant, he was to be devoured or married. On some occasions the tiger came out of one door, and on some out of the other. The decisions of this tribunal were not only fair, they were positively determinate: the accused person was
instantly punished if he found himself guilty, and, if innocent, he was rewarded on
the spot, whether he liked it or not. There was no escape from the judgments of
the king’s arena.

The institution was a very popular one. When the people gathered together on
one of the great trial days, they never knew whether they were to witness a bloody
slaughter or a hilarious wedding. This element of uncertainty lent an interest to the
occasion which it could not otherwise have attained. Thus, the masses were
entertained and pleased, and the thinking part of the community could bring no
charge of unfairness against this plan, for did not the accused person have the
whole matter in his own hands?

This semi-barbaric king had a daughter as blooming as his most florid fancies,
and with a soul as fervent and imperious as his own. As is usual in such cases, she
was the apple of his eye, and was loved by him above all humanity. Among his
courtiers was a young man of that fineness of blood and lowness of station
common to the conventional heroes of romance who love royal maidens. This royal
maiden was well satisfied with her lover, for he was handsome and brave to a
degree unsurpassed in all this kingdom, and she loved him with an ardor that had
enough of barbarism in it to make it exceedingly warm and strong. This love affair
moved on happily for many months, until one day the king happened to discover
its existence. He did not hesitate nor waver in regard to his duty in the premises.
The youth was immediately cast into prison, and a day was appointed for his trial
in the king's arena. This, of course, was an especially important occasion, and his
majesty, as well as all the people, was greatly interested in the workings and
development of this trial. Never before had such a case occurred; never before had
a subject dared to love the daughter of the king. In after years such things became
commonplace enough, but then they were in no slight degree novel and startling.

The tiger-cages of the kingdom were searched for the most savage and
relentless beasts, from which the fiercest monster might be selected for the arena;
and the ranks of maiden youth and beauty throughout the land were carefully
surveyed by competent judges in order that the young man might have a fitting
bride in case fate did not determine for him a different destiny. Of course,
everybody knew that the deed with which the accused was charged had been done.
He had loved the princess, and neither he, she, nor any one else, thought of
denying the fact; but the king would not think of allowing any fact of this kind to
interfere with the workings of the tribunal, in which he took such great delight and
satisfaction. No matter how the affair turned out, the youth would be disposed of,
and the king would take an aesthetic pleasure in watching the course of events, which would determine whether or not the young man had done wrong in allowing himself to love the princess.

The appointed day arrived. From far and near the people gathered, and thronged the great galleries of the arena, and crowds, unable to gain admittance, massed themselves against its outside walls. The king and his court were in their places, opposite the twin doors, those fateful portals, so terrible in their similarity.

All was ready. The signal was given. A door beneath the royal party opened, and the lover of the princess walked into the arena. Tall, beautiful, fair, his appearance was greeted with a low hum of admiration and anxiety. Half the audience had not known so grand a youth had lived among them. No wonder the princess loved him! What a terrible thing for him to be there!

As the youth advanced into the arena he turned, as the custom was, to bow to the king, but he did not think at all of that royal personage. His eyes were fixed upon the princess, who sat to the right of her father. Had it not been for the moiety of barbarism in her nature it is probable that lady would not have been there, but her intense and fervid soul would not allow her to be absent on an occasion in which she was so terribly interested. From the moment that the decree had gone forth that her lover should decide his fate in the king's arena, she had thought of nothing, night or day, but this great event and the various subjects connected with it. Possessed of more power, influence, and force of character than any one who had ever before been interested in such a case, she had done what no other person had done — she had possessed herself of the secret of the doors. She knew in which of the two rooms, that lay behind those doors, stood the cage of the tiger, with its open front, and in which waited the lady. Through these thick doors, heavily curtained with skins on the inside, it was impossible that any noise or suggestion should come from within to the person who should approach to raise the latch of one of them. But gold, and the power of a woman's will, had brought the secret to the princess.

And not only did she know in which room stood the lady ready to emerge, all blushing and radiant, should her door be opened, but she knew who the lady was. It was one of the fairest and loveliest of the damsels of the court who had been selected as the reward of the accused youth, should he be proved innocent of the crime of aspiring to one so far above him; and the princess hated her. Often had she seen, or imagined that she had seen, this fair creature throwing glances of admiration upon the person of her lover, and sometimes she thought these glances
were perceived, and even returned. Now and then she had seen them talking together; it was but for a moment or two, but much can be said in a brief space; it may have been on most unimportant topics, but how could she know that? The girl was lovely, but she had dared to raise her eyes to the loved one of the princess; and, with all the intensity of the savage blood transmitted to her through long lines of wholly barbaric ancestors, she hated the woman who blushed and trembled behind that silent door.

When her lover turned and looked at her, and his eye met hers as she sat there, paler and whiter than any one in the vast ocean of anxious faces about her, he saw, by that power of quick perception which is given to those whose souls are one, that she knew behind which door crouched the tiger, and behind which stood the lady. He had expected her to know it. He understood her nature, and his soul was assured that she would never rest until she had made plain to herself this thing, hidden to all other lookers-on, even to the king. The only hope for the youth in which there was any element of certainty was based upon the success of the princess in discovering this mystery; and the moment he looked upon her, he saw she had succeeded, as in his soul he knew she would succeed.

Then it was that his quick and anxious glance asked the question: "Which?" It was as plain to her as if he shouted it from where he stood. There was not an instant to be lost. The question was asked in a flash; it must be answered in another.

Her right arm lay on the cushioned parapet before her. She raised her hand, and made a slight, quick movement toward the right. No one but her lover saw her. Every eye but his was fixed on the man in the arena.

He turned, and with a firm and rapid step he walked across the empty space. Every heart stopped beating, every breath was held, every eye was fixed immovably upon that man. Without the slightest hesitation, he went to the door on the right, and opened it.

Now, the point of the story is this: Did the tiger come out of that door, or did the lady?

The more we reflect upon this question, the harder it is to answer. It involves a study of the human heart which leads us through devious mazes of passion, out of which it is difficult to find our way. Think of it, fair reader, not as if the decision of the question depended upon yourself, but upon that hot-blooded, semi-barbaric princess, her soul at a white heat beneath the combined fires of despair and jealousy. She had lost him, but who should have him?
How often, in her waking hours and in her dreams, had she started in wild horror, and covered her face with her hands as she thought of her lover opening the door on the other side of which waited the cruel fangs of the tiger! But how much oftener had she seen him at the other door! How in her grievous reveries had she gnashed her teeth, and torn her hair, when she saw his start of rapturous delight as he opened the door of the lady! How her soul had burned in agony when she had seen him rush to meet that woman, with her flushing cheek and sparkling eye of triumph; when she had seen him lead her forth, his whole frame kindled with the joy of recovered life; when she had heard the glad shouts from the multitude, and the wild ringing of the happy bells; when she had seen the priest, with his joyous followers, advance to the couple, and make them man and wife before her very eyes; and when she had seen them walk away together upon their path of flowers, followed by the tremendous shouts of the hilarious multitude, in which her one despairing shriek was lost and drowned!

Would it not be better for him to die at once, and go to wait for her in the blessed regions of semi-barbaric futurity?

And yet, that awful tiger, those shrieks, that blood!

Her decision had been indicated in an instant, but it had been made after days and nights of anguished deliberation. She had known she would be asked, she had decided what she would answer, and, without the slightest hesitation, she had moved her hand to the right.

The question of her decision is one not to be lightly considered, and it is not for me to presume to set myself up as the one person able to answer it. And so I leave it with all of you: Which came out of the opened door — the lady, or the tiger?

**********

In 1885, Stockton published a sequel to this tale:

Stockton, F.R.,
*The Discourager of Hesitancy: A Continuation of “The Lady, or the Tiger?”*
Appendix Two: Emile Coué’s Thought Experiment

Since the Second International Congress on Experimental Psychology held in London in 1892, the prevailing view on the relationship between the will and the imagination (especially in relation to “suggestion”) had been that, whilst “self-suggestion” had great significance in all forms of psychotherapy, it only had an effect because of the influence of the powers of the subject’s own will (Anon, 1892, pp.586-587).

The French pharmacist, Emile Coué, thought very differently; and he strongly argued that the subject’s imagination was far stronger than his/her will.

In support of his view, he offered the following thought experiment (Coué, 1922, pp.7-8):

If we open a dictionary and look up the word “will” we find this definition: “The faculty of freely determining certain acts”. We accept this definition as true and un-attackable, although nothing could be more false. This will that we claim so proudly, always yields to the imagination. It is an absolute rule that admits of no exception.

"Blasphemy! Paradox!" you will exclaim. "Not at all! On the contrary, it is the purest truth," I shall reply.

In order to convince yourself of it, open your eyes, look round you and try to understand what you see. You will then come to the conclusion that what I tell you is not an idle theory, offspring of a sick brain but the simple expression of a fact.

Suppose that we place on the ground a plank 30 feet long by 1 foot wide. It is evident that everybody will be capable of going from one end to the other of this plank without stepping over the edge. But now change the conditions of the experiment, and imagine this plank placed at the height of the towers of a cathedral. Who then will be capable of advancing even a few feet along this narrow path? Could you hear me speak? Probably not. Before you had taken two steps you would begin to tremble, and in spite of every effort of your will you would be certain to fall to the ground.

Why is it then that you would not fall if the plank is on the ground, and why should you fall if it is raised to a height above the ground? Simply because in the first case you imagine that it is easy to go to the end of the plank, while in the second case you imagine that you

---

1 That is, Emile Coué of “Every day, in every way, I’m getting better and better” (‘Tous les jours, à tous points de vue, je vaux mieux en mieux’) fame.

2 Coué’s once controversial view, that the imagination is stronger than the will, is now universally accepted as a matter of established fact.

---

By developing specific tools and skills, cultural groups extend the cognizable worlds of their members in different directions. (Sperber, 1982, p.155)
cannot do so.

Notice that your will is powerless to make you advance; if you imagine that you cannot, it is absolutely impossible for you to do so. If tilers and carpenters are able to accomplish this feat, it is because they think they can do it.

Vertigo is caused by the picture we make in our minds that we are going to fall. This picture transforms itself immediately into fact in spite of all the efforts of our will, and the more violent these efforts are, the quicker is the opposite to the desired result brought about.

********

It seems highly likely that this thought experiment has its origins in a comment by Francis Bacon (1677, p.168):

795. Experiment Solitary, touching the force of Imagination. Imitating that of the Sense.

Those effects, which are wrought by the percussion of the Sense, and by things in Fact, are produced likewise in some degree by the Imagination: Therefore if a man see another eat soure or acide things, which set the Teeth on edge, this object tainteth the Imagination; so that he that seeth the thing done by another, hath his own Teeth also set on edge. So if a man see another turn swiftly and long, or if he look upon Wheels that turn, himself waxeth Turn sick. So if a man be upon a high place, without Rails, or good hold, except that he be used to it, he is ready to fall; for imagining a fall, it putteth his spirits into the very action of a fall. So many upon the seeing of others Bleed, or Strangled, or Tortured, themselves are ready to faint, as if they bled, or were in strife.
Appendix Three: von Platen and the Demiurge

The following is the relevant section of von Platen’s (1962) account of the source of his synthetic diamond innovation (pp. 118-120):

**A MULTIPLE PISTON, HIGH PRESSURE, HIGH TEMPERATURE APPARATUS**

*Baltazar von Platen*  
*Götgatan 16, Stockholm*

I. 'Demiurge and the Diamonds' — a Hypothesis

Let me begin by telling a story.

It was a fine autumn day, and I had just entered the University of Lund in the south of Sweden. As I passed the School of Botany, I saw that one wall was covered with Virginia creeper. Its leaves were a beautiful red. Every autumn brought the change from green to red, and passers-by would pause at the display of colour in sudden admiration. I was one of these passers-by, and had no idea that my pleasure in this seasonal beauty would later play an important role in my life, that it would reveal the clue to the problem of achieving very high pressures in large volumes of the order of a thousand cubic centimetres or more. I had no suspicion that later this sight would show me the way to construct a machine which makes diamonds, the machine now owned by ASEA.¹

A couple of years later I was studying at the Royal Institute of Technology in Stockholm. Although botany was not one of our studies, Tarras Sallforst,² a friend of mine, was very interested in the subject. One day I happened to describe the autumn in Lund, and the colours, and I asked him why the leaves of the creeper turned red in the autumn as they died.

"Your question reminds me of a hypothesis — or call it a speculation — which a philosopher described to me not long ago", answered Tarras, and continued:

"He was not a biologist himself, and certainly didn’t want his thoughts on the subject to be described by so scientific a word as ‘hypothesis’. He just called it an interesting speculation, and I think you will find it interesting too. But there is something else I must say first. . . ."

"Perhaps you know", Tarras continued, "that the natural force which solved the problems of life on this earth was symbolically represented in antiquity by the spirit called Demiurge or

---

¹ von Platen’s footnote:  
ASEA: Allmänna Svenska Elektriska Aktiebolaget (Swedish General Electric Company Ltd.)”

² von Platen’s footnote:  
Former General Director of Swedish Industrial Organization Committee, *obit* 1960.
Demiurgos. God himself gave Demiurge the task of solving these problems, and to help him He gave him divine understanding. He made Demiurge into the world’s greatest inventor. The five senses, sight, hearing, touch, taste and smell, are his inventions, and so is the human brain — perhaps the most wonderful of all. Demiurge has made thousands upon millions of inventions, all necessary to support life.”

“And I imagine he invented something useful for the leaves of that Virginia creeper too”, I said.

“Of course. Now I will tell you what the philosopher told me.

“You asked why these leaves turned red in the autumn when they die. Well, our philosopher supposed that they do this not because they are dying, but because they don’t want to die. Each autumn, when they feel that they have not much longer to live, a physico-chemical process takes place in them which gives them those beautiful colours. We know that certain molecules in the leaves disintegrate, and that this must happen sooner or later — that is, either before the leaves die, or after. The fact that it happens before they die was taken by my philosopher friend to be a typical example of adaptability in the world of living organisms. He assumed that there must be some good reason for this to happen, and came to regard the process as a weapon which Demiurge had given the leaves and taught them how to use in the unequal struggle against Death’s advance. This seems to be a paradox, because with this weapon the leaves actually help their enemy Death to carry out part of his work of destruction — namely, the disintegration of molecules. During the last part of their lives, the leaves co-operate with the enemy by doing a job that he would otherwise have had to do soon after their death. He would have to do it anyway, because it is part of his task of bringing about the dissolution of the leaf’s organism.

“But this co-operation with Death, the philosopher speculated, was nothing more nor less than a way of tricking him. The leaves had to meet Death half way; but Death did not understand that in teaching them this, Demiurge was giving them an extra couple of weeks of life, and giving us the beautiful blaze of autumn.”

Not being a biologist, I allowed this speculation of Tarras’ philosopher to become an accepted scientific fact in my mind, as though it were one of the real inventions of Demiurge.

The years passed. I was fully occupied with other work and thought very little about the

---

1 Gerson (1995):
The ancient Greek word means “craftsman” or “artisan”. Plato, in the Timaeus, uses the word for the maker of the universe. Plato says of this maker that he is unreservedly good and so desired that the world should be as good as possible. The reason why the world is not better than it is is that the demiurge had to work on pre-existing chaotic matter. Thus, the demiurge is not an omnipotent creator. Early Christian philosophers were quick to claim that the demiurge represented pagan philosophy’s anticipation of the God of revealed religion.

2 Floridi (2003, p.464): Plato’s Demiurge is not an omnipotent God, who produces the universe out of nothing, but a smaller god, who moulds a pre-existing reality according to reason.
problem of making diamonds. Then, between Christmas of 1930 and the New Year, I happened to be visiting a friend, John Tandberg, now a professor and well-known physicist in Sweden. On a table there was a publication which I glanced at quite by chance. It showed the function between pressure and temperature during the crystallization of carbon into diamond. Discussing the subject with him, he was doubtful whether steel could stand up to such enormous pressures.

Some hours later when I was thinking over what he had said, I happened to remember what Tarras had told me about the leaves. In my mind's eye I saw a mass of red leaves with Demiurge in the background. All at once a thought struck me and an association of ideas sprang to life. It was utterly unexpected, for up to then I had done no work at all on the diamond problem, though I had often been tempted to get to grips with it. Suddenly I saw how Demiurge's principle for prolonging the life of the leaves could be applied to a machine for making diamonds. One had merely to convert botanical facts into mechanical ones, and the parallel between the corresponding details seemed to me to be quite complete. It was as though Demiurge himself had told me a great secret, in words like this:

“You must show as much care for your diamond machine as I showed for my leaves when I taught them how to meet Death by doing some of his work for him. You know your machine must fall to pieces when it is destroyed — when it is killed — by the enormous pressure. And the form of these pieces will be wrong, since the dead steel has not been invested with the wit and knowledge that I gave to the leaves. Your machine will be destroyed in the same way as my leaves died before they learned what to do from me. You must let your machine go to meet Death by dividing it up into pieces, but you must give these pieces the shape that the machine itself would have wished, had it been a living organism like the leaves. Then it will stand much higher pressures, and live longer. And then you will be able to make diamonds.”

The association of ideas had now led to a hypothesis.

As we know, a hypothesis is an assumption we make to help us understand certain facts more easily or more thoroughly. We may not, however, make just any assumption at all. We must show logically that it is likely to be correct. The likelihood that a hypothesis is correct is derived not directly from scientific facts, but from a philosopher's speculations. This made it necessary to describe my own hypothesis together with the circumstances of its origin, otherwise it might have seemed abrupt and obscure.

We also know that a more general way to understand facts, for example, thoughts or phenomena, is to combine them logically into a system or several systems. These may themselves be combined into super-systems. A hypothesis always implies the building of a system, or an attempt to do so. This process of combination or synthesis is constantly being carried on, intuitively and subconsciously, by most human beings. It is a form of
understanding. When it is performed consciously and the system is brought towards the point of completion, it becomes science, and with it the view of a complex of problems becomes clearer and the perspective widens. And so the process helps us to find the right path, or the direction of that path, towards the result we wish to achieve. This may be diamonds, or millions of other concrete or abstract things.

There is little doubt that the phenomenon upon which our present hypothesis rests has been known for many years in other branches of technology. If this is so, there can be no doubt at all that when it was first observed it was at once fitted into one or several systems. None of these, however, happened to be any that could lead us to our goal — that is, the making of diamonds. My hypothesis fitted the phenomenon into an entirely different system, which proved to be one of those that led to the right direction, since it led to the construction of ASEA's diamond machine and the production of diamonds with this machine in February, 1953.

In other words, my hypothesis forced me to think along a certain line which proved to be one of the correct ones. I say "one of the correct ones" because we know that quite apart from the present hypothesis, other lines have led to the same goal — man-made diamonds.

Hazen (1999, pp.61-64) speaks of von Platen's role in the history of synthetic diamonds:

The ASEA effort [to produce synthetic diamonds] started with the dreams of one man, Baltzar von Platen — called a "genius maniac" by some — was the kind of crazy inventor stereotyped in popular movies and dime-store novels...

Some colleagues speculated that his idiosyncrasies may have arisen, in part, from a facial deformity — a large disfiguring birthmark on his left chin that embarrassed him greatly. But whatever the cause, he marched to a very different drummer. Baltzar von Platen's most famous and curious invention, a portable thermal refrigerator that produced ice with a gas flame, epitomized his quirky brilliance. He used an ammonia refrigerant that was vaporized by a blue-hot gas flame at one end of his contraption; the ammonia gas flowed through the system to condense in refrigeration coils at the other end. Before the universal availability of electricity, von Platen's thermal refrigerators, manufactured and marketed by Electrolux, were a godsend that transformed rural life in many parts of Europe and America...

In this freewheeling spirit von Platen designed his incredible high-pressure, diamond-making system. He began thinking about synthetic diamonds as early as 1930, when he read an article describing the extreme conditions required to convert graphite to diamonds. He knew that the tremendous temperatures and pressures necessary to make diamonds were too much for any device made of steel. He concluded that to make diamonds you had to accept as inevitable the destruction of your machine in the process. But, he emphasized, if you accept the fact that your machine's destruction is inevitable, that knowledge allows you to approach the design in a radical new way: build the device as if it were already broken!
Many years after the ASEA effort, Baltzar von Platen contributed a strange, rambling chapter on his diamond-making machine to the 1962 monograph Modern Very High Pressure Techniques, which is for the most part a rather dry, technical book. However, in prose atypical of scientific exposition, von Platen revealed the mystical source of his inspiration, which was found in a bit of ancient mythology regarding the origin of brilliantly colored autumnal foliage. According to the myth, the spirit-philosopher Demiurge convinced leaves to begin the inevitable process of death and disintegration while still attached to their tree, rather than succumb to death by falling to the ground in a fresh and green state. By beginning the irreversible decay process before falling from the branches, leaves provide some beauty and benefit for others and, in the process, cheat death out of two or three weeks.

This quaint story evidently came forcefully to mind as Baltzar von Platen contemplated the diamond problem. “All at once a thought struck me and an association sprang to life. It was utterly unexpected, for up to then I had done no work at all on the diamond problem, though I had often been tempted to get to grips with it. Suddenly I saw how Demiurge’s principle for prolonging the life of the leaves could be applied to a machine for making diamonds. One had merely to convert botanical facts into mechanical ones, and the parallel between the corresponding details seemed to me quite complete.”

In von Platen's extraordinary mind, the myth contained the philosophy around which to build an experiment: If destruction is inevitable, then turn that destruction to your advantage. “You know your machine must fall to pieces when it is destroyed — when it is killed — by the enormous pressure. And the form of these pieces will be wrong, since the dead steel has not been invested with ... wit and knowledge.... You must let your machine go to meet Death by dividing it up into pieces, but you must give these pieces the shape that the machine itself would have wished, had it been a living organism like the leaves. Then it will stand much higher pressures, and live longer. And then you will be able to make diamonds.”

This is not typical experimental protocol, and von Platen's metaphors do not speak persuasively to most working scientists. Nevertheless, although it is difficult to see how a high-pressure invention could follow from Demiurge's principle, von Platen's apparatus successfully generated extraordinary temperatures and pressures, and his idea continues to serve as the basis for many of today's most successful high-pressure machines.
Appendix Four: From an Episode of Frasier

This particular scene involves Frasier Crane, his brother Niles (both of whom are professional psychiatrists), and their ex-policeman father Martin.

Niles, currently going through the last stages of his divorce (from wife Maris), is terrified that his brother Frasier (who is to give a deposition under oath on the following day), will reveal Niles’ deep (but as-yet-undisclosed feelings) for his father’s physiotherapist, Daphne.

**Martin**: Niles, now get a grip! Maris's people are just fishing. We're the only one's who know how you feel about Daphne.

**Niles**: Today! Tomorrow, everyone will know, including Daphne. And this is how she's going to find out. I wanted to whisper it to her over a candlelight dinner, not have it read back to her by a court stenographer!

**Martin**: What are you talking about? Frasier's not going to tell!

**Niles**: He's gonna be under oath! What if they ask him point blank?

**Martin**: Well, he'll just say no. Right, Fras?

[Fraser is silent, unsure.]

**Martin**: [worried] Fras?

**Frasier**: Well, it's not that simple! You heard what he said, I'll be under oath.

**Niles**: Oh my God. I can't breathe! I can't breathe!

[He staggers, gaping like a landed fish. Frasier grabs his neck and bows him over.]

**Frasier**: Niles, here! Put your head between your legs! I'll get you some water!

[He rushes to the kitchen.]

**Martin**: Niles, you'll be all right. Just take a couple of deep breaths.

**Niles**: If I could take deep breaths it would mean I could breathe.

[In the kitchen, Frasier fills a water glass. Martin comes in.]

**Martin**: You're really not going to go down there tomorrow and blow it for him, are you?

**Frasier**: Dad, there is nothing more I'd like to do than to help my brother. But I would like to find a way to do it without violating my ethics!

**Martin**: But Maris is responsible for this divorce! Niles's feelings for Daphne has nothing to do with it! Now, if you cover for Niles you won't be doing anything wrong.

**Frasier**: Dad, we are talking about perjury! When is that ever acceptable?

---

1 This episode, “To Tell The Truth” (*Frasier*, Season 6, Episode 15), was first aired in the U.S.A. on 18 February 1999. It was written by Rob Hanning. The transcript was taken from [http://www.geocities.com/Hollywood/Derby/3267/615.html](http://www.geocities.com/Hollywood/Derby/3267/615.html) on 14 October 2002.
Martin: Oh, you want an example? Fine! Let's say, uh, what if there was a comet hurtling towards the earth—

Frasier: Oh, for God's sake!

Martin: And you were the only person who could save the earth, but the only way to do it is by lying under oath. Would you do it then?

Frasier: Who am I lying to, the comet?

Martin: Oh, just answer the question!

Frasier: All right, I suppose in certain extreme cases—

Martin: So, then you'd lie?

Frasier: To save mankind from a talking comet, yes!

Martin: But you won't lie for Niles.

Frasier: Oh, for God's sake, you make me sound like some sort of insensitive lout who's not aware that his brother's out there in pain! [takes a sip from the glass]

Martin: Isn't that Niles's water?

Frasier: I'm just checking to see it's not too cold!
Many country and western songs manifest counterfactual thinking. A superb example appears in the hit song, *Put Yourself in My Shoes*: “if you put yourself in my shoes/You'd have some sympathy/And if I could only put myself in your shoes/I'd walk right back to me”

**Put Yourself in My Shoes**

(© 1989 Clint Black/Hayden Nicholas/Shake Russell)

Your mind is made up you won't even try
You didn't even cry this time
You say that we could never see eye to eye
And one of us just must be blind

We have our differences
We're still the same
See what we want to see
But you take a second look
And maybe things wouldn't seem the same
If you could see what you mean to me

Put yourself in my shoes
Walk a mile for me
I'll put myself in your shoes
Maybe then we'd see
That if you put yourself in my shoes
You'd have some sympathy
And if I could only put myself in your shoes
I'd walk right back to me

You're gonna keep walkin' and
You're gonna pass me by
You say you don't even care
But I could always recognize a real good-bye
And I know your heart's not there
We've had our differences
We're still the same
Hear what we want to hear
Now I'm head over heels in the lost and found
It's a cryin' shame
I thought we made the perfect pair

Put yourself in my shoes
Walk a while walk that mile for me
I'll put myself in your shoes
Maybe then we'd see
That if you put yourself in my shoes
You'd have some sympathy
And if I could only put myself in your shoes
I'd walk right back to me

I'd walk right back to me
I'd walk right back to me

***************
Appendix Six: A Counterfactual Headline

Front page of the Sydney *Daily Telegraph*, Wednesday 3 July 2002.
Appendix Seven: “Exchange of Bodies” Thought Experiments

A modern “exchange of bodies” thought experiment:¹

Suppose that there were some process to which two persons, A and B, could be subjected as a result of which they might be said — question-beggingly — to have exchanged bodies. That is to say — less question-beggingly — there is a certain human body which is such that when previously we were confronted with it, we were confronted with person A, certain utterances coming from it were expressive of memories of the past experiences of A, certain movements of it partly constituted the actions of A and were taken as expressive of the character of A, and so forth; but now, after the process is completed, utterances coming from this body are expressive of what seem to be just those memories which previously we identified as memories of the past experiences of B, its movements partly constitute actions expressive of the character of B, and so forth; and conversely with the other body.

A version from John Locke:²

For should the Soul of a Prince, carrying with it the consciousness of the Prince's past Life, enter and inform the Body of a Cobler, as soon as deserted by his own Soul, every one sees, he would be the same Person with the Prince, accountable only for the Prince's Actions: But who would say it was the same Man?

¹ Taken from Williams (1970), p.161.
An ancient Chinese version:³

Gong Hu of Lu and Qi Ying of Zhao fell ill, and both asked Bian Qiao⁴ to treat them. Bian Qiao did treat them and when they had both recovered told them:

“You have just been suffering from diseases which attacked your organs from outside, and which of course medicine and the needle⁵ can cure. But you also have diseases which were born with you and have grown with the growth of your bodies; would you like me to treat them for you?”

“First tell us what makes you think so.”

“Your ambition is greater than your energy”, said Bian Qiao to Gong Hu, “so that you are capable of forming plans but seldom come to decisions. Qi Ying’s energy is greater than his ambition, so that he rarely thinks ahead and comes to grief by acting irresponsibly. If I exchange your hearts you will benefit by the equalising of ambition and energy.”⁶

Then Bian Qiao gave the two men drugged wine, and they lost consciousness for three days. He cut open their breasts, pulled out their hearts, exchanged them, put them back, and applied a magic medicine. When they woke up they were as well as before.

The two men took their leave and went off home. Thereupon Gong Hu returned to Qi Ying’s house and took possession of wife and children, who did not recognise him. Qi Ying likewise returned to Gong Hu’s house and took possession of his wife and children, who also did not recognise him. So the two families went to law against each other, and called on Bian Qiao to explain. Bian Qiao explained the cause, and the litigants were satisfied.

---

³ Taken from “The Book of Tang” section of the Chinese Daoist text, the Lie.zi. The text, which contains fragments from as early as the fifth century BCE, reached its final form circa 370CE. This translation is taken from Graham (1990, pp.106-107). I have changed the Wage-Giles transcription to modern Pin-Yin (Graham, mistakenly, transcribes the character for his family name as P’ien, which would be Pian in Pin-Yin).

⁴ Bian Qiao (the Chinese characters can also be pronounced Bian Qüe) is the most famous Chinese doctor of all time. He is often termed “The Hippocrates of China” by those with Eurocentric views. Skilled in herbal medicine, acupuncture, and surgery he is renowned for his use of herbal anaesthetics.

From the texts that mention him, in various places at various times, he is obviously a composite character constructed from a number of stories from different historical eras; obviously, what the Chinese communists used to term “an accumulation of the folk wisdom of the masses”.

⁵ Namely, herbal medicine and acupuncture respectively.

⁶ Translator’s Note: “According to Chinese conceptions, mental functions are located in the heart instead of the brain; on the other hand the Qi (breath, vitality, energy) is spread over the whole body.”
Appendix Eight:
Foot’s 1967 “Tramcar Problem” and Thomson’s 1971 “Famous Violinist Problem”.

These thought experiments, which centre on the issue of the difference between actively killing and passively letting die,\(^1\) form a significant part of the ethical debate on abortion.

Suppose that a judge or magistrate is faced with rioters demanding that a culprit be found for a certain crime and threatening otherwise to take their own bloody revenge on a particular section of the community. The real culprit being unknown, the judge sees himself as able to prevent the bloodshed only by framing some innocent person and having him executed. Beside this example is placed another in which a pilot whose aeroplane is about to crash is deciding whether to steer from a more to a less inhabited area. To make the parallel as close as possible it may rather be supposed that he is the driver of a runaway tram which he can only steer from one narrow track on to another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed. In the case of the riots the mob have five hostages, so that in both the exchange is supposed to be one man's life for the lives of five. The question is why we should say, without hesitation, that the driver should steer for the less occupied track, while most of us would be appalled at the idea that the innocent man could be framed. It may be suggested that the special feature of the latter case is that it involves the corruption of justice, and this is, of course, very important indeed. But if we remove that special feature, supposing that some private individual is to kill an innocent person and pass him off as the criminal we still find ourselves horrified by the idea. The doctrine of double effect offers us a way out of the difficulty, insisting that it is one thing to steer towards someone foreseeing that you will kill him and another to aim at his death as part of your plan. Moreover there is one very important element of good in what is here insisted. In real life it would hardly ever be certain that the man on the narrow track would be killed. Perhaps he might find a foothold on the side of the tunnel and cling on as the vehicle hurtled by. The driver of the tram does not then leap off

---

\(^1\) Thomson (1976/1986), p.78:
Morally speaking it may matter a great deal how a death comes about, whether from natural causes, or at the hands of another, for example/ Does it matter whether a man was killed or only let to die? A great many people think it does: they think that killing is worse than letting die. And they draw conclusions from this for abortion, euthanasia, and the distribution of scarce medical resources. Others think it doesn’t, and they think this is shown by what we see when we construct a pair of cases which are so far as possible in all other respects alike, except that in one case the agent kills, in the other he only lets die. So, for example, imagine that

1) Alfred hates his wife and wants her dead. He puts cleaning fluid in her coffee, thereby killing her.

and that

2) Bert hates his wife and wants her dead. She puts cleaning fluid in her coffee (being muddled, thinking it’s cream). Bert happens to have the antidote to cleaning fluid, but he does not give it to her; he lets her die.
and brain him with a crowbar. The judge, however, needs the death of the innocent man for his (good) purposes. If the victim proves hard to hang he must see to it that he dies another way. To choose to execute him is to choose that this evil shall come about, and this must therefore count as a certainty in weighing up the good and evil involved. The distinction between direct and oblique intuition is crucial here, and is of great importance in an uncertain world. Nevertheless this is no way to defend the doctrine of double effect. For the question is whether the difference between aiming at something and obliquely intending it is in itself relevant to moral decisions; not whether it is important when correlated with a difference of certainty in the balance of good and evil. Moreover we are particularly interested in the application of the doctrine of the double effect to the question of abortion, and no one can deny that in medicine there are sometimes certainties so complete that it would be a mere quibble to speak of the “probable outcome” of this course of action or that. It is not, therefore, with a merely philosophical interest that we should put aside the uncertainty and scrutinise the examples to test the doctrine of the double effect. Why can we not argue from the case of the steering driver to that of the judge? (Foot, 1978/1967, pp.23-24)

Thomson (1976/1986 and 1985/1986 passim), who uses the US term “trolley” as a substitute for tramcar (thus “the trolley problem”), discusses a number of ways in which Foot’s original example might be embellished.

Thomson (1971/1986, p.2) noting that much of the inadequate debate on abortion gets lost within the issue of whether or the foetus is a person or, simply, a mass of tissue, circumvents that misleading and irrelevant issue by “[granting] that the fetus is a person from the moment of conception”, and addressing what she felt was the only issue involved: whether the mother or the foetus had the “stronger and more stringent… right to life”:

A newly fertilized ovum, a newly implanted clump of cells, is no more a person than an acorn is an oak tree. But I shall not discuss any of this. For it seems to me to be of great interest to ask what happens if, for the sake of argument, we allow the premise. How, precisely, are we supposed to get from there to the conclusion that abortion is morally impermissible? Opponents of abortion commonly spend most of their time establishing that the fetus is a person, and hardly any time explaining the step from there to the

---

2 In British and Australian English a tramcar is a street railway, whilst a trolley is a bus driven by electricity supplied by overhead cables; and in US English the terms are “trolley-car” and “trolley-bus” respectively.
impermissibility of abortion. Perhaps they think the step too simple and obvious to require much comment. Or perhaps instead they are simply being economical in argument. Many of those who defend abortion rely on the premise that the fetus is not a person, but only a bit of tissue that will become a person at birth; and why pay out more arguments than you have to? Whatever the explanation, I suggest that the step they take is neither easy nor obvious, that it calls for closer examination than it is commonly given, and that when we do give it this closer examination we shall feel inclined to reject it.

I propose, then, that we grant that the fetus is a person from the moment of conception. How does the argument go from here? Something like this, I take it. Every person has a right to life. So the fetus has a right to life. No doubt the mother has a right to decide what shall happen in and to her body; everyone would grant that. But surely a person's right to life is stronger and more stringent than the mother's right to decide what happens in and to her body, and so outweighs it. So the fetus may not be killed; an abortion may not be performed.

It sounds plausible. But now let me ask you to imagine this. You wake up in the morning and find yourself back to back in bed with an unconscious violinist. A famous unconscious violinist, He has been found to have a fatal kidney ailment, and the Society of Music Lovers has canvassed all the available medical records and found that you alone have the right blood type to help. They have therefore kidnapped you, and last night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract poisons from his blood as well as your own. The director of the hospital now tells you, "Look, we're sorry the Society of Music Lovers did this to you — we would never have permitted it if we had known. But still, they did it, and the violinist now is plugged into you. To unplug you would be to kill him. But never mind, it's only for nine months. By then he will have recovered from his ailment, and can safely be unplugged from you." Is it morally incumbent on you to accede to this situation? No doubt it would be very nice of you if you did, a great kindness. But do you have to accede to it? What if it were not nine months, but nine years? Or longer still? What if the director of the hospital says, "Tough luck, I agree, but you've now got to stay in bed, with the violinist plugged into you, for the rest of your life. Because remember this. All persons have a right to life, and violinists are persons. Granted you have a right to decide what happens in and to your body, but a person's right to life outweighs your right to decide what happens in and to your body. So you cannot ever be unplugged from him." I imagine you would regard this as outrageous, which suggests that something really is wrong with that plausible-sounding argument I mentioned a moment ago.

In this case, of course, you were kidnapped; you didn't volunteer for the operation that plugged the violinist into your kidneys. Can those who oppose abortion on the ground I mentioned make an exception for a pregnancy due to rape? Certainly. They can say that persons have a right to life only if they didn't come into existence because of rape; or they
can say that all persons have a right to life, but that some have less of a right to life than others, in particular, that those who come into existence because of rape have less. But these statements have a rather unpleasant sound. Surely the question of whether you have a right to life at all, or how much of it you have, shouldn’t turn on the question of whether or not you are the product of a rape. And in fact the people who oppose abortion on the ground I mentioned do not make this distinction, and hence do not make an exception in case of rape.

Nor do they make an exception for a case in which the mother has to spend the nine months of her pregnancy in bed. They would agree that would be a great pity, and hard on the mother; but all the same, all persons have a right to life, the fetus is a person, and so on. I suspect, in fact, that they would not make an exception for a case in which, miraculously enough, the pregnancy went on for nine years, or even the rest of the mother’s life.

Some won’t even make an exception for a case in which continuation of the pregnancy is likely to shorten the mother’s life; they regard abortion as impermissible even to save the mother’s life. Such cases are nowadays very rare, as many opponents of abortion do not accept this extreme view. All the same it is a good place to begin: a number of points of interest come out in respect to it... (Thomson, 1971/1986, pp.2-4)
Appendix Nine: Consequences of the Free-Trade Agreement

The following article was specifically written by Don Henry (2003), the Executive Director of the Australian Conservation Foundation, to highlight significant “unanticipated consequences” (Merton, 1936), or “unintended consequences” (Norton, 2004) of the proposed free-trade agreement between Australia and the United States of America.

**Free-Trade Clause would be a Dangerous Weakening of the Law**

*Don Henry, Sydney Morning Herald, (27 October 2003)*

Flash forward to the future. It is 2006. A US waste management company is operating a waste treatment plant on the outskirts of Sydney. Nearby residents begin to complain of severe headaches and other strange illnesses. Investigations soon trace the cause to toxins treated at the plant.

The NSW Environment Protection Authority immediately shuts the plant. All states ban the treatment of the toxins that were processed at the US plant.

The Australian public is relieved but the US company is less than impressed. Based on a free-trade agreement (FTA) signed by the US and Australian governments two years before, the US company sues the Australian government for $US300 million in compensation for potential lost earnings.

Rather than have the case heard in a drawn-out court hearing in Australia, the company chooses to take the dispute to a special international tribunal which deals with FTA investment disputes.

After a brief, secret hearing, the tribunal decides in favour of the US company. It finds the Australian government is in breach of obligations under the FTA that require it to pay compensation to US investors if their profits are affected by government laws. The compensation award, $US70 million, is not quite as much as it had wanted but it is sizeable. The Australian government is shocked but there is little it can do because the tribunal's decision is binding and there is no mechanism for appeal.

Sound far-fetched? Australia is negotiating an FTA with the US which could result in such a scenario. Early indications from the secret trade negotiations suggest the Australia-US FTA is set to include an investment chapter based on the infamous Chapter 11 of the North American Free Trade Agreement (NAFTA). An Australian-US FTA incorporating a NAFTA-style investment chapter will provide US corporations with unprecedented rights to seek millions, or possibly billions, of dollars in compensation should Australian laws breach FTA rules protecting their investments.
Ordinarily, only government parties to an international agreement have the right to enforce the agreement, but US corporations may soon have the right to sue Australian governments directly.

US corporations are not afraid of using these new rights to protect their investments, particularly from laws designed to protect the environment and public health. For example, in 1997, the Canadian Government imposed a ban on the import and interstate transport of MMT, a fuel additive containing manganese. The ban was imposed because of public health concerns. Ethyl Corporation, a US chemical company which produces MMT, sued the Canadian Government, arguing the ban was an expropriation of its investments and was therefore illegal under NAFTA. The claim was for $US251 million in compensation. The Government eventually settled the case by reversing its ban on MMT and paying $US13 million in legal fees and compensation to Ethyl Corporation.

It is highly likely the Australian-US FTA will include an investment chapter that would permit US companies to bring similar legal actions against the Australian Government. The US Government is pushing for the inclusion of a NAFTA-style investment chapter, and the Australian Government has quietly slipped a similar investment chapter into the recently signed Australian-Singapore FTA.

These new rights for US corporations go far beyond those enjoyed by Australians. In most cases, Australian law recognises the right to compensation only when property is taken by government. The FTA will extend the right to compensation for US companies when an Australian law seeks merely to regulate the use of property.

This is a dramatic departure from a legal principle that was designed to protect private property interests from unjust government acquisitions, while also balancing the need of Australian governments to freely regulate the use of property in the public interest. The FTA rules threaten this balance.

Also of concern is the fact that Australian courts are not likely to be the forum in which the FTA compensation claims are heard. Instead, disputes would probably be heard by an international arbitration tribunal. Under international trade arbitration rules judges are appointed by the parties - with huge potential for bias. The hearings are often closed to the public and there is no mechanism for appeal. This is hardly the kind of decision-making body that should be granted the power to override matters of Australian public policy and make compensation awards for millions of dollars of taxpayers' money.

Fortunately, the FTA with the US has not been finalised. The Australian Government still has the opportunity to say no to an NAFTA-style investment chapter being included.
Appendix Ten: “The Monk Problem”

In their paper “Conceptual Integration Networks” (1998), Fauconnier and Turner examine a cognitive operation they term “blending”. In the process of explaining this concept they discuss at some length the ramifications of what they call “the riddle of the Buddhist monk”.

The following displays the evolution of this problem from Duncker’s original into the form presented by Fauconnier and Turner.

Here is Duncker’s (1945, p.56) original version:

There exist conditions, therefore, under which what is to be read off becomes sufficiently evident in the given situation. Only where this is not the case, does a naive person look around for proofs. The original function of a proof is to resolve something not directly evident into something which is directly evident or into a chain of directly evident steps.

An instructive example for this: On a mountain trip, on which descent was by the same path as had been the ascent of the previous day, I asked myself whether there must be a spot on route at which I must find myself at exactly the same time on the descent as on the ascent. It was of course assumed that ascent and descent took place at about the same time of day, say from five to twelve o’clock.— But without further probing, I could arrive at no conclusive insight. Since then, I have put this question to dozens of naive persons as well as of intellectuals, and have observed with great satisfaction that others had the same difficulty. Let the reader himself ponder a bit.— Certainly there exist several approaches to an evident solution. But probably none is, I might almost say, more drastically, evident than the following view of the situation: let ascent and descent be divided between two persons on the same day. They must meet. Ergo . . . With this, from an unclear, dim condition not easily surveyable, the situation has suddenly been brought into full daylight. The answer becomes evident, inspection sufficiently exact.

Gardner, (1961a, p.168, 170) posed this problem:

One morning, exactly at sunrise, a Buddhist monk began to climb a tall mountain. The narrow path, no more than a foot or two wide, spiraled around the mountain to a glittering temple at the summit.

The monk ascended the path at varying rates of speed, stopping many times along the way to rest and to eat the dried fruit he carried with him. He reached the temple shortly before sunset. After several days of fasting and meditation he began his journey back along the same path, starting at sunrise and again walking at variable speeds with many pauses along the way. His average speed descending was, of course, greater than his average climbing speed.

Prove that there is a spot along the path that the monk will occupy on both trips at precisely the same time of day.

Gardner, (1961b, p.154) offered this solution:
Appendix Ten: "The Monk Problem" — 192

A man goes up a mountain one day, down it another day. Is there a spot along the path that he occupies at the same time of day on both trips? This problem was called to my attention by psychologist Ray Hyman of the General Electric Company, who in turn found in a monograph entitled "On Problem Solving," by the German Gestalt psychologist Karl Duncker. Duncker writes of being unable to solve it and of observing with satisfaction that others to whom he put the problem had the same difficulty. There are several ways to go about it, he continues, "but probably none is . . . more drastically evident than the following. Let ascent and descent be divided between two persons on the same [sic] day. They must meet. Ergo…. With this, from an unclear dim condition not easily surveyable, the situation has suddenly been brought into full daylight."¹

Koestler, (1964, pp.183-185) offered this version:²

Let me take a more trivial example: a famous brain-teaser:

One morning, exactly at sunrise, a Buddhist monk began to climb a tall mountain. The narrow path, no more than a foot or two wide, spiralled around the mountain to a glittering temple at the summit.

The monk ascended the path at varying rates of speed, stopping many times along the way to rest and to eat the dried fruit he carried with him. He reached the temple shortly before sunset. After several days of fasting and meditation he began his journey back along the same path, starting at sunrise and again walking at variable speeds with many pauses along the way. His average speed descending was, of course, greater than his average climbing speed.

Prove that there is a spot along the path that the monk will occupy on both trips at precisely the same time of day.

I used to amuse myself putting this to various friends — scientists and others. Some chose a mathematical approach; others tried to 'reason it out' — and came to the conclusion that it would be a most unlikely coincidence for the monk to find himself at the same time of day, on the same spot on the two different occasions. But others—who evidently belonged to the category of visualizers — saw the solution in a manner for which the following description of a young woman without any scientific training is typical:

I tried this and that, until I got fed up with the whole thing, but the image of that monk in his saffron robe walking up the hill kept persisting in my mind. Then a moment came when, super-imposed on this image, I saw another, more transparent one, of the monk walking down the hill, and I realized in a flash that the two figures must meet at some point some time — regardless at what speed

¹ Notice the significant difference between Gardner’s “quote” and Duncker’s original version.

² In a footnote (1964, p.710), Koestler cites Scientific American, June 1961 as his source. However, he also states that, “in fact the problem originates with Carl [sic] Duncker”.
they walk and how often each of them stops. Then I reasoned out what I already knew: whether the monk descends two days or three days later comes to the same; so I was quite justified in letting him descend on the same day, in duplicate so to speak.

Now it is, of course, quite impossible for the monk to duplicate himself, and to be walking up the mountain and down the mountain at one and the same time. But in the visual image he does; and it is precisely this indifference to logical contradiction, the irrational dream-like telescoping of the two images into one, which leads to the solution.

We could call the double image of the monk, or Einstein’s traveller riding on a ray of light, a concretization of abstract problems as it sometimes occurs in dreams...

Hayes, (1978, pp.177-179) offered this version:

Problems are presented to us by the outside world. This outside problem environment defines the gaps we try to bridge when we solve problems. But to solve a problem, we must first construct an adequate internal representation of the problem — that is, we must understand it. Sometimes, because of our perspective, our egocentrism, it is hard for us to distinguish between our representation of the problem and the problem as it exists in the outside world, but it is important to do so.

A problem may be represented in many different ways. In some representations, the problem may be very easy to solve while in others it may be extremely difficult. Our choice of representations, then, can make a considerable difference in our problem-solving success.

Before proceeding, try to solve the problem shown [below]:

Once there was a monk who lived in a monastery at the foot of a mountain. Every year the monk made a pilgrimage to the top of the mountain to fast and to pray. He would start out on the mountain path at 6 A.M., climbing and resting as the spirit struck him, but making sure that he reached the shrine at exactly 6 P.M. that evening. He then prayed and fasted all night. At exactly 6 A.M. the next morning, he began to descend the mountain path, resting here and there along the way, but making sure that he reached his monastery again by 6 P.M. of that day.

That evening as he was hastening to a much needed dinner, he was stopped by the monastery’s visiting mathematician, who said to him, "Do you know, I suddenly realized a very curious thing. Every time you make your pilgrimage there is always some point on the mountain path, perhaps different on each trip, that you pass at the same time when you are climbing up as when you are climbing down." "What!" snorted the monk, annoyed. "Why, that’s ridiculous! I walk at all manner of different paces up and down the path. It would be a great coincidence if I should pass any spot at the same time of day going up as coming down. The
idea that such a coincidence might happen time after time surpasses belief!" The mathematician, who had a touch of fiendishness in his soul, smiled sweetly and said, "Bless you, Brother, not only should you believe it, but if you will just think about it in the right way, it's obvious." He then locked himself in his cell, confident that he had spoiled the monk's dinner and probably his night's sleep as well.

Now, what the mathematician claimed is, in fact, true. If it does not seem obvious to you, it is probably because of the way you are representing the problem to yourself. The problem is written so that you will think of the monk going up the mountain one day and coming down again the next. But suppose you change that representation. Imagine instead of one monk who ascends the mountain on one day and descends on the next, that there are two monks, one of them ascending the mountain on the same day that the other is descending. At some point on the mountain path, the two monks meet, and must therefore be at the same place at the same time.

[The figure below (from p.179)] shows another way to represent the Old Monk problem which some find easier to understand than the one just given.

In this figure the monk's altitude on the mountain (shown vertically) is graphed against the time of day (shown as increasing from left to right). The monk's progress up the mountain on the first day is shown by the line which moves from the lower left corner to the upper right corner. His progress down the mountain on the next day is shown on the same
graph by the line which moves from the upper left corner to the lower right corner. Clearly, these two lines must cross somewhere. (Convince yourself of this by drawing other possible ways to ascend and descend the mountain on the graph, but remember, do not draw any trips which make time run backward.) The point where the lines cross shows the altitude which the monk reached at the same time on his two trips.

Mayer (1983, pp.75-76) offered this version:

Another example [of the use of imagery in problem solving], attributed to the Gestalt psychologist Duncker, concerns the following monk problem:

A monk began to climb a mountain at sunrise. He reached the temple at the top as the sun was setting and meditated all night. At sunrise of the next day, he came down the mountain, following the same path, but moving at a faster rate, of course. When he reached the bottom he proclaimed: "There is one spot along this path that I passed at exactly the same time of day on my way up the mountain as on my way down." Can you prove that the monk is correct?

If you try to represent this problem algebraically or even in words, you will have a difficult time. However, if you visualize the problem as [shown above], you will see that there must be a point at which the time of day is the same for the ascending and descending trips. Another way to state the problem is to say there are two monks, one at the bottom going up and one at the top going down, and to ask whether they will meet, or be at the same place at the same time. The problem is solved when you can think of the ascending and descending trips occurring simultaneously on the same day.

---

3 Mayer states that his diagram is based on Hayes’ (1978) diagram.
These results suggest that translation of problem information to a visual representation may involve a sort of assimilation and that integrated visual diagrams may be useful tools in certain types of problem solving.


Consider a classic puzzle of inferential problem-solving:

A Buddhist monk begins at dawn one day walking up a mountain, reaches the top at sunset, meditates at the top for several days until one dawn when he begins to walk back to the foot of the mountain, which he reaches at sunset. Making no assumptions about his starting or stopping or about his pace during the trips, prove that there is a place on the path which he occupies at the same hour of the day on the two separate journeys.

Our demonstration of the power of blending is likely to be more effective if the reader will pause for a moment and try to solve the problem before reading further. The basic inferential step to showing that there is indeed such a place, occupied at exactly the same time going up and going down, is to imagine the Buddhist monk walking both up and down the path on the same day. Then there must be a place where he meets himself, and that place is clearly the one he would occupy at the same time of day on the two separate journeys.

The riddle is solved, but there is a cognitive puzzle here. The situation that we devised to make the solution transparent is a fantastic one. The monk cannot be making the two journeys simultaneously on the same day, and he cannot "meet himself." And yet this implausibility does not stand in the way of understanding the riddle and its solution. It is clearly disregarded. The situation imagined to solve the riddle is a blend: it combines features of the journey to the summit and of the journey back down, and uses emergent structure in that blend to make the affirmative answer apparent. Here is how this works........

---

4 They cite Arthur Koestler’s *The Act of Creation* (1964) as their source.
Bibliography


Backhaus, W., "How to compare color sensations in different brains" [Commentary on Palmer], *Behavioral and Brain Sciences*, Vol.22, No.6, (December 1999), pp.944-945.


Benson, P.J., "Color: How you see it, when you don’t", [Commentary on Palmer], Behavioral and Brain Sciences, Vol.22, No.6, (December 1999), pp.945-946.


Bernstein, M. (with new material by Barker, W.J.), The Search for Bridey Murphy Murphy, Bantam Books, (London), 1990. [This is a reprint of the second revised edition of the 1956 book, originally published in 1965]


Brauner, T., Hasle, P. & Øhrstrøm, P., "Ockhamistic Logics and True Futures of Counterfactual Moments", pp.132


Byrne, A., "Subjectivity is no barrier" [Commentary on Palmer], *Behavioral and Brain Sciences*, Vol.22, No.6, (December 1999), pp.949-950.


Byrne, R.W., "So much easier to attack straw men" [Commentary on Heyes], *Behavioral and Brain Sciences*, Vol.21, No.1, (February 1998), pp.116-117.


Cansino, S.

Carr, D., "Place and Time: On the Interplay of Historical Points of View",

Carranza, I.E., "Log In..."


Damper, R.I., "Thought Experiments can be Harmful", Abstract for Conference on Model-Based Reasoning (MBR '01), May 17-19, Pavia, Italy. [Taken from http://www.bib.ecs.soton.ac.uk/data/6807/pdf/abstract.pdf on 26 March 2002]


De Rose, K., "Can It Be That It Would Have Been Even Though It Might Not Have Been?", *Philosophical Perspectives*, Vol.13, (October 1999), pp.385-413.


Dennett, D.C., personal correspondence (22 September 2003).


Bibliography — 218


Bibliography — 224


Galton, F., "Composite Portraits, Made by Combining Those of Many Different Persons Into a Single Resultant Figure", The Journal of the Anthropological Institute of Great Britain and Ireland, Vol.8, (1879), pp.132-144.


Gooding, D.C., "Thought experiments", Routledge Encyclopaedia of Philosophy, [Taken from http://www.rep.routledge.com/philosophy/cgi-bin/article.cgi?it=Q106 on 29 May 2002]


Bibliography — 230


Bibliography — 231

Griffin, D. & Tversky, A., "The Weighing of Evidence and the Determinants of Confidence", pp.230-249 in Gilovich, T.,

pp.199-206. [Plus commentaries pp.207-221 (Manier, E., pp.207-214; Kitts, D.B., pp.214-216; and Coleman, W.,
pp.216-221)].


Grünbaum, A., "Temporally Asymmetric Principles, Parity between Explanation and Prediction, and Mechanism versus
Teleology", pp.57-96 in Baumrin, B. (ed.), The Philosophy of Science, the Delaware seminar, Volume 1, 1961-1962,


Gutting, G., "'Rethinking Intuition': A Historical and Metaphilosophical Introduction", pp.3-13 in DePaul, M. & Ramsey, W.


pp.293-301.

Hacking, I., "Do Thought Experiments Have a Life of Their Own? Comments on James Brown, Nancy Nersessian and David
Philosophy of Science Association, Volume Two, Symposia and Invited Papers, Philosophy of Science Association, (East
Lansing), 1993.

Hacking, I., "Experiment" [Chapter 9], Representing and Intervening: Introductory Topics in the Philosophy of Natural


1945.

Haegeman, L., "Conditional Clauses: External and Internal Syntax", Mind and Language, Vol.18, No.4, (September 2003),
pp.317-339.


Haidt, J., Koller, S.H. & Dias, M.G., "Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?", Journal of Personality


Hakfoort, C., "Nicolas Béguelin and his Search for a Crucial Experiment on the Nature of Light (1772)", Annals of Science,


(August 2003), pp.66-73.


Bibliography — 232


Hanson, N.R., "Is there a logic of scientific discovery?", pp.20-24, in Feigl, H. & Maxwell, G. (eds.), Current Issues in the Philosophy of Science, Holt, Reinhart & Winston, (New York), 1961 [includes "Comments" by Feyerabend, P.K. (pp.35-39) and "Rejoinder to Feyerabend" by Hanson, N.R. (pp.40-42].


Hardin, C.L., "Color relations and the power of complexity" [Commentary on Palmer], Behavioral and Brain Sciences, Vol.22, No.6, (December 1999), pp.953-954.


Hochberg, J., "Perception as purposeful inquiry: We elect where to direct each glance, & determine what is encoded within & between glances" [Commentary on Barsalou], Behavioral and Brain Sciences, Vol.22, No.4, (August 1999), pp.619-620.


Howard, H., "If not functionalism, then what? Eliminative materialism?" [Commentary on Palmer], Behavioral and Brain Sciences, Vol.22, No.6, (December 1999), pp.955-956.


Hurford, J.R., "Individuals are abstractions" [Commentary on Barsalou], *Behavioral and Brain Sciences*, Vol.22, No.4, (August 1999), pp.620-621.


Bibliography — 242


Krishnamurthy, P. & Sivaraman, A., "Counterfactual Thinking and Advertising Responses",

Krippner, S.C., "Conflicting Perspectives on Shamans and Shamanism: Points and Counterpoints",


MacLennan, B., "Neurophenomenological constraints and pushing back the subjectivity barrier" [Commentary on Palmer], *Behavioral and Brain Sciences*, Vol.22, No.6, (December 1999), pp.961-963.


McCarthy, I. E. & Weigold, E., *A real "thought" experiment for the hydrogen atom*, Flinders University of South Australia, Institute for Atomic Studies, School of Physical Sciences, (Bedford Park), 1981.


McKenna, F., "It won't happen to me: unrealistic optimism or illusion of control?", British Journal of Psychology, Vol.84, No.1, (February 1993), pp.39-50.


Bibliography — 264


Pepperberg, I.M., "To see or not to see, that is the question: Designing experiments to test perspective-taking in nonhumans" [Commentary on Heyes], Behavioral and Brain Sciences, Vol.21, No.1, (February 1998), pp.128-129.


Saunders, B., "One machine among many" [Commentary on Palmer], Behavioral and Brain Sciences, Vol.22, No.6, (December 1999), pp.969-970.


Slezak, P., "They Laughed at Galileo, Didn’t They? (Or, Should Philosophy of Science be Rated X Too?)", Paper presented at Conference of Australasian Association of History, Philosophy and Social Studies of Science, University of Melbourne, August 2001.


Smythies, J., "Consciousness and introspection: How we get to know the inner world" [Commentary on Palmer], Behavioral and Brain Sciences, Vol.22, No.6, (December 1999), pp.971-972.


Spellman, B.A. & Mandel, D.R., "When Possibility Informs Reality: Counterfactual Thinking as a Cue to Causality", Current Directions in Psychological Science, Vol.8, No.4, (August 1999), pp.120-123.


Sun, C.C., *As the Saying Goes: An Annotated Anthology of Chinese and Equivalent English Sayings and Expressions, and an Introduction to Xiehouyu (Chinese Wit)*, University of Queensland Press, (St Lucia), 1981.


Trout, J., ""Positioning" is a game people play in today's me-too marketplace", Industrial Marketing, Vol.54, No.6, (June 1969), pp.51-55.


van Brakel, J., "Whatever seems right to me is right" [Commentary on Palmer], Behavioral and Brain Sciences, Vol.22, No.6, (December 1999), p.973.


Wheeler, L., "Concluding Comment", Personality and Social Psychology Bulletin, Vol.12, No.3, (September 1986), pp.297-299. [Check this is correct; there seems to be something wrong with it.]


