Deduction or Darwinian Algorithms?

An explanation of the "elusive" content effect on the Wason selection task

A thesis presented

by

Leda Cosmides

to

The Department of Psychology and Social Relations

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Psychology

Harvard University

Cambridge, Massachusetts

July, 1985

Abstract

This thesis develops the idea that natural selection has shaped how humans reason about evolutionarily important domains of human activity. The human mind can be expected to include "Darwinian algorithms" that are specialized for processing information about such domains. Evolutionary principles were heuristically applied to pinpoint social exchange as an adaptively important domain of human activity; these principles were also applied in developing computational theories of how humans process information about social exchange. Evidence is presented supporting the hypothesis that the human mind includes Darwinian algorithms specialized for reasoning about social exchange. This hypothesis both predicts and explains "content effects" on the Wason selection task -- a test of logical reasoning -- better than alternative theories.

Table of Contents

Deduction or Darwinian Algorithms?
An explanation of the "elusive" content effect
on the Wason selection task

Introduction.....1

Chapter 1
Logic and the Study of Human Reasoning

Why were psychologists interested in deductive logic?.....6
What would a logic module be like?.....9
Do humans have a logic module?.....13

Chapter 2
A review of the literature on the "elusive" content effect
on the Wason selection task

Introduction.....23
The Transportation Problem.....24
The Food Problem.....33
The School Problem.....38
Social Contract Problems.....44

Chapter 3
"Differences in Experience":
Proposed explanations for the elusivity of the
content effect on the Wason selection task

Families of explanation.....72
Explanations proposed in the literature.....74

Differential Availability.....77
Memory-cueing/Reasoning by Analogy.....82
Mental Models.....89
Frames and Schemas.....92
Auxiliary Mechanisms.....95
Family 2 Explanations.....100
Summary of Explanations.....103

Chapter 4

Darwinian Algorithms

Another view of human rationality.....106
A brief primer on natural selection.....109
Why should Darwinian algorithms be specialized and domain specific?.....116

Chapter 5

Human Social Exchange

Introduction.....129
Natural selection and social exchange.....130
Social exchange and the Pleistocene environment.....146
A computational theory of social exchange.....149
Human social exchange requires some fundamental cognitive capacities.....152
The grammar of social contracts.....172

Chapter 6

Social contracts and the Wason selection task: Experiments

Introduction.....196

Experiment 1: Unfamiliar Standard Social Contracts (Law).....205
Experiment 2: Unfamiliar Switched Social Contracts (Law).....211
Experiment 3: Unfamiliar Standard Social Contracts (Exchange).221
Experiment 4: Unfamiliar Switched Social Contracts (Exchange).226
Experiment 5: Abstract Standard Social Contract.....237
Experiment 6: Familiar Standard Social Contracts.....244

Chapter 7

Discussion and Conclusions

The social contract hypothesis uniquely accounts for empirical results on the Wason selection task.....256
Are social contract algorithms innate?.....259
The role of evolutionary theory in psychology.....266

Appendix A.....269
Appendix B.....272
Appendix C.....273
Bibliography.....274

Introduction

The equipotentiality assumption has crept, incognito, from the meta-theory of behaviorism* into the meta-theory of cognitive psychology. Behaviorists do not expect the laws of learning to differ from domain to domain; cognitive psychologists do not expect the processes that govern attention, memory, or reasoning to differ from domain to domain.

To the behaviorist, stimuli are stimuli and responses, responses: their content is not supposed to affect how they are paired. When content effects are discovered, the behaviorist speaks of adjusting "parameter values", or of differences in the organism's "experience" with various content domains. To the cognitive psychologist, information is information: the content of the information is not supposed to affect how it is processed. When content effects are discovered, the cognitive psychologist also speaks of differences in the organism's "experience" with various content domains.

Though unspoken, the message is clear: Content is noise. Cognitive processes are content-independent, domain general, equipotential. The human mind is a general purpose information processing system, designed to process any kind of information with equal efficiency. The format of the information -- for example, whether it is imagistic or propositional -- might make a difference in how it is processed, but its content will not. The amount of experience the organism has had with a domain may affect performance, but correct for this and the effect will

---

* for review, see Herrnstein, 1977.

disappear. These claims are rarely tested; they are merely assumed.

The alternative view -- that the human mind includes a number of domain specific, content-dependent, information processing systems -- is seldom entertained (Cf. Chomsky, 1975; Marr & Nishihara, 1978; Fodor, 1983). Although behaviorism came under brisk attack from evolutionary and ethological quarters for assuming that learning was equipotential (Herrnstein, 1977), cognitive psychology seems untouched by this fray and the serious problems it raised. Yet the evolutionary arguments against the equipotentiality assumption in behaviorism apply equally to cognitive psychology.

An unspoken assumption is an unexamined assumption. Cognitive processes may, in fact, be content-independent; if so, then this should be proved, not presumed. Indeed, when content effects are found, the content-independence of cognitive processes should be a hypothesis of last resort.

Cognitive processes, like electrons, are entities defined solely by input-output relations. An electron gun is fired at a diffraction slit, and then into a cloud chamber: even though the data from the first firing indicates a wave and the data from the second indicates a particle, there are compelling reasons for believing these divergent patterns were created by one and the same entity. It would grossly violate our most basic notions of similarity and causation to categorize two firings of an electron gun as two different "stimuli", just because they were fired at different targets. The same input -- which was, in this case, the very entity physicists were trying to characterize -- yielded

different outputs. The only reasonable theoretical alternative was to complexify the equations defining the electron, and assume that it did not correspond to any ordinary human concept like "particle" or "wave" (Heisenberg, 1971).

But there are no compelling reasons -- other than a misguided sense of parsimony -- for believing that the same cognitive process is involved when manifestly different inputs yield unmistakably different outputs. In fact, because cognitive processes are entities defined by these very input-output relations, the discovery of content effects should be taken as prima facie evidence that the different stimuli tested are accessing different cognitive processes. If response patterns vary with stimulus content, but their variation does not appear to be systematic, then one should rethink one's theory of how to parse the world into content domains. Hand-waving appeals to "differences in experience" -- which are virtually impossible to falsify -- should be explanations of last resort.

When content effects are found, cognitive psychologists should entertain the hypothesis that domain specific, content-dependent, cognitive processes are responsible. Content effects have been found on the Wason selection task, a famous experimental paradigm that tests whether people reason according to the content-independent canons of formal logic. Most attempts at explaining these content effects have appealed to "differences in subjects' experience" with various content domains. A controversy has grown up around these content effects, because, to date, they have eluded prediction.

This thesis uses content effects on the Wason selection task

to test the hypothesis that humans have domain specific, innate mental algorithms specialized for reasoning about social exchange. A computational theory of the functional properties of these algorithms was derived using natural selection theory as a heuristic guide. Critical tests were conducted to choose between the social exchange hypothesis and the hypotheses in the literature that appeal to "differences in experience."

The discovery of systematic variation from domain to domain is strong evidence that domain specific algorithms are at work; so is the discovery of systematic variation within a domain that cannot be easily explained by a content-independent process. Both kinds of evidence are presented in support of the social exchange hypothesis. I argue that no other hypothesis offered so far can predict or explain the experimental results presented herein, and that the social exchange hypothesis best explains the content effects on the Wason selection task that have already been reported in the literature.

The meta-theoretical view entailed by this hypothesis -- of the human mind as a collection of functionally distinct, Chomskian "mental organs" -- also has parsimony on its side. The human mind, like the rest of the body and its functions, was designed by natural selection. The more important the adaptive problem, the more intensely selection will have specialized and improved the performance of information processing mechanisms for solving it. Domain general information processing mechanisms simply cannot insure adaptive responses in evolutionarily important domains of human activity -- domains like social exchange. Reasoning in such domains should be governed by

"Darwinian algorithms": mental algorithms specialized for solving the adaptive problems that define these domains.

With this thesis, I hope to resurrect the arguments against equipotential psychological mechanisms. From the standpoint of evolutionary theory, nothing could be more unparsimonious than the view that the human mind is a general purpose information processor. Yet the application of the Chomskian view has been limited because cognitive psychologists have lacked a systematic heuristic for judging which domains, other than language, were likely to command functionally distinct mental organs. Because it is a theory of function, natural selection theory provides just such a heuristic. Evolutionary principles allow one to pinpoint domains for which natural selection can be expected to have shaped how humans reason. Moreover, they suggest computational theories (sensu Marr) of what their design features are likely to be.

The theory of social exchange developed in this thesis was informed, at every stage, by evolutionary principles. In the study of human reasoning, the search for content-independent inference procedures had generated a confusion of apparently contradictory results; the hypothesis that humans have domain specific Darwinian algorithms for reasoning about social exchange resolves much of this confusion. The heuristic application of evolutionary theory can revolutionize cognitive psychology, allowing it to address issues closer to the heart of what we think of as human nature. This thesis is offered as a small illustration of its potential.