

Interactive minds

**Life-span perspectives on the social
foundation of cognition**

Edited by

PAUL B. BALTES

*Max Planck Institute for Human
Development and Education*

URSULA M. STAUDINGER

*Max Planck Institute for Human
Development and Education*



CAMBRIDGE
UNIVERSITY PRESS

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© Cambridge University Press 1996

First published 1996

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Interactive minds: life-span perspectives on the social foundation of cognition /
edited by Paul B. Baltes, Ursula M. Staudinger.

p. cm.

Includes bibliographical references and indexes.

ISBN 0-521-48106-6 (hc). – ISBN 0-521-48567-3 (pb)

1. Cognition and culture. 2. Social perception. I. Baltes, Paul B.

II. Staudinger, Ursula M.

BF311.I564 1996

302'.12 – dc20

95-19302

CIP

A catalog record for this book is available from the British Library.

ISBN 0-521-48106-6 Hardback

ISBN 0-521-48567-3 Paperback

11 Rationality: why social context matters

Gerd Gigerenzer

Abstract

Rationality is commonly identified with axioms and rules, such as consistency, which are defined without reference to context, but are imposed in all contexts. In this chapter, I focus on the social context of rational behavior. My thesis is that traditional axioms and rules are incomplete as behavioral norms in the sense that their normative validity depends on the social context of the behavior, such as social objectives, values, and motivations. In the first part, I illustrate this thesis by showing that social context can determine whether an axiom or rule is satisfied or not. In the second part, I describe an alternative to context-independent rationality: a domain-specific theory of rational behavior derived from the evolutionary theory of cooperation.

I want to argue against an old and beautiful dream. It was Leibniz's dream, but not his alone. Leibniz (1677/1951) hoped to reduce rational reasoning to a universal calculus, which he termed the Universal Characteristic. The plan was simple: to establish characteristic numbers for all ideas, which would reduce every question to calculation. Such a rational calculus would put an end to scholarly bickering; if a dispute arose, the contending parties could settle it quickly and peacefully by sitting down and calculating. For some time, the Enlightenment probabilists believed that the mathematical theory of probability had made this dream a reality. Probability theory rather than logic became the flip side of the newly coined rationality of the Enlightenment, which acknowledged that humankind lives in the twilight of probability rather than the noontime sun of certainty, as John Locke expressed it. Leibniz guessed optimistically of the Universal Characteristic that "a few selected persons

I would like to thank Paul Baltès, Robert Boyd, Valerie Chase, Michael Cole, Lorraine Daston, Berna Eden, Bill Goldstein, Dan Goldstein, Wolfgang Hell, Ralph Hertwig, Amy Johnson, Elke Kurz, Peter Sedlmeier, Anna Senkevitch, Ursula Staudinger, Gerhard Strube, Zeno Swijtink, Elke Weber, and three anonymous reviewers for their helpful comments. I am grateful for the financial support provided by the UCSPMP Fund for Research in Mathematics Education and by an NSF Grant SBR-9320797/GG.

might be able to do the whole thing in five years" (Leibniz, 1677/1951, p. 22). By around 1840, however, mathematicians had given up as thankless and even antimathematical the task of reducing rationality to a calculus (Daston, 1988). Psychologists and economists have not.

Contemporary theories embody Leibniz's dream in various forms. Piaget and Inhelder's (1951/1975) theory of cognitive development holds that, by roughly age 12, human beings begin to reason according to the laws of probability theory; Piaget and Inhelder thus echo the Enlightenment conviction that human rationality and probability theory are two sides of the same coin (Gigerenzer et al., 1989). Neoclassical economic theories center on the assumption that Jacob Bernoulli's expected utility maximization principle or its modern variants, such as subjective expected utility, define rationality in all contexts. Similarly, neo-Bayesians tend to claim that the formal machinery of Bayesian statistics defines rational inferences in all contexts. In cognitive psychology, formal axioms and rules – consistency, transitivity, and Bayes's theorem, for example, as well as entire statistical techniques – figure prominently in recent theories of mind and warrant the rationality of cognition (Gigerenzer, 1991a; Gigerenzer & Murray, 1987).

All these theories have been criticized as *descriptively* incomplete or inadequate, most often by showing that principles from logic or probability theory (such as consistency) are systematically violated in certain contexts. Piaget himself wondered why adults outside of Geneva seemed not to reach the level of formal operations. But even critics have generally retained the beautifully simple principles drawn from logic and probability theory as normative, albeit not descriptively valid – that is, as definitions of how we *should* reason. In this chapter, I will address the question of whether these principles are indeed normative: sufficient for defining rational behavior.

My discussion will challenge one central assumption in the modern variants of Leibniz's dream: that formal axioms and rules of choice can define rational behavior without referring to factors external to choice behavior. To the contrary, I will argue that these principles are incomplete as behavioral norms in the sense that their normative validity depends on the *social* context of the behavior, such as social objectives, values, and motivations.

The point I wish to defend in the first section is that formal axioms and rules cannot be *imposed* as universal yardsticks of rationality independent of social objectives, norms, and values; they can, however, be *entailed* by certain social objectives, norms, and values. Thus, I am not arguing against axioms and rules, only against their a priori imposition as context-independent yardsticks of rationality. In the second section, I will describe one alternative to context-independent rationality in some detail. This alternative account starts with the evolutionary theory of cooperation and puts objectives in social interaction first, formal rules second.

Challenges to a conception of rationality without social context

Leibniz's dream was of a formal calculus of reasonableness that could be applied to everything. Modern variants tend to go one step further and assume that the calculus of rationality has already been found and can be imposed in all contexts. I will focus only on the *social* context in this chapter, arguing that the idea of imposing a context-independent, general-purpose rationality is a limited and confused one. The several examples that follow seek to demonstrate that only by referring to something external to the rules or axioms, such as social objectives, values, and norms, can we decide whether an axiom or choice rule entails rational behavior.

Consistency: Property Alpha

Internal consistency of choice figures prominently as a basic requirement for human rationality in decision theory, behavioral economics, game theory, and cognitive theories. It is often seen as *the* requirement of rational choice. One basic condition of the internal consistency of choice is known as "Property Alpha," also called the "Chernoff condition" or "independence of irrelevant alternatives" (Sen, 1993). The symbols S and T denote two (nonempty) sets of alternatives, and $x(S)$ denotes that alternative x is chosen from the set S :

Property Alpha:

$$x(S) \text{ and } x \in T \subseteq S \Rightarrow x(T).$$

Property Alpha demands that if x is chosen from S , and x belongs to a subset T of S , then x must be chosen from T as well. The following two choices would be inconsistent in the sense that they violate Property Alpha:

1. x is chosen given the options $\{x, y\}$.
2. y is chosen given the options $\{x, y, z\}$.

Property Alpha is violated here because x is chosen when the two alternatives $\{x, y\}$ are offered, but y is chosen when z is added to the menu. (Choosing x is interpreted here as a rejection of y , not as a choice that results from mere indifference.) It may indeed appear odd and irrational that someone who chooses x and rejects y when offered the choice set $\{x, y\}$ would choose y and reject x when offered the set $\{x, y, z\}$.

Property Alpha formulates consistency exclusively in terms of the internal consistency of choice behavior with respect to sets of alternatives. No reference is made to anything external to choice – for instance, intentional states such as people's social objectives, values, and motivations. This exclusion of everything psychological beyond behavior is in line with Samuelson's (1938) program of freeing theories of behavior from any traces of utility and from the priority of the notion of "preference." As Little (1949) com-

mented on the underlying methodological program, Samuelson's "revealed preference" formulation "is scientifically more respectable [since] if an individual's behavior is consistent, then it must be possible to explain the behavior without reference to anything other than behavior" (p. 90). Sen (1993) has launched a forceful attack on internal consistency, as defined by Property Alpha and similar principles, and what follows is based on his ideas and examples.

The last apple. Consider Property Alpha in the context of the social politics at a dinner party. There is one apple left in the fruit basket. Dining alone, Mr. Polite would face no dilemma; but now he must choose between taking the apple (y) or having nothing (x). He decides to behave decently and go without (x). If the basket had contained another apple (z), he could reasonably have chosen y over x without violating standards of good behavior. Choosing x over y from the choice set $\{x, y\}$ and choosing y over x from the choice set $\{x, y, z\}$ violates Property Alpha, even though there is nothing irrational about Mr. Polite's behavior given his scruples in social interaction. If he had not held to such values of politeness, then Property Alpha might have been entailed. But it cannot be imposed independent of his values.

Waiting for dinner. Consider a second example. Mr. Pleasant is invited to his colleague's home on Sunday at 9:00 p.m. Upon arriving, he takes a seat in the living room, and his host offers him crackers and nuts (y). Mr. Pleasant decides to take nothing (x), because he is hungry for a substantial meal and does not want to fill up before dinner. After a while, the colleague's wife comes in with tea and cake (z). The menu has thereby been extended to $\{x, y, z\}$, but there is a larger implication: The new option z also has destroyed an illusion about what the invitation included. Now Mr. Pleasant chooses the crackers and nuts (y) over nothing (x). Again, this preference reversal violates Property Alpha. Given the guest's expectations of what the invitation entailed, however, there is nothing irrational about his behavior.

Tea at night. Here is a final example similar to the last. Mr. Sociable has met a young artist at a party. When the party is over, she invites him to come back to her place for tea. He chooses to have tea with her (x) over returning home (y). The young lady then offers him a third choice – to share some cocaine at her apartment (z). This extension of the choice set may quite reasonably affect Mr. Sociable's ranking of x and y . Depending on his objectives and values, he may consequently choose to go home (y).

All three examples seek to illustrate the same point: Property Alpha will or will not be entailed depending on the social objectives, values, and expectations of the individual making the choice. To impose Property Alpha as a

general yardstick of rational behavior independent of social objectives or other factors external to choice behavior seems fundamentally flawed.

The conclusion is not that consistency is an invalid principle; rather, consistency, as defined by Property Alpha or similar principles, is indeterminate. The preceding examples illustrate different kinds of indeterminateness. With respect to the last apple, social values define what the alternatives in the choice set are and, thereby, *what* consistency is about. If there are many apples in the basket, the choice is between “apple” and “nothing.” If a single apple remains and one does not share the values of Mr. Polite, the alternatives are the same; for Mr. Polite, however, they become “last apple” and “nothing.” In the dinner and tea example, one learns something new about the old alternatives when a new choice is introduced. The fresh option provides new information – that is, it reduces uncertainty about the old alternatives.

To summarize the argument: Consistency, as defined by Property Alpha, cannot be imposed on human behavior independent of something external to choice behavior, such as social objectives and expectations. Social concerns and moral views (e.g., politeness), as well as inferences from the menu offered (learning from one option as to what others may involve), determine whether internal consistency is or is not entailed.

Maximizing: choice under certainty

Maximizing expectation was the cornerstone of the new Enlightenment rationality. The Chevalier de Méré, a notorious gambler, posed the following problem: Can one expect to make more money by betting on the occurrence of at least one “six” in 4 throws of a fair die, or on that of at least one “double six” in 24 throws of a pair of dice? (The first gamble offers the higher expectation.) The correspondence between Blaise Pascal and Pierre Fermat in 1654 over this and similar problems marks the first casting of the calculus of probabilities in mathematical form.

I will turn first to a simpler situation, the choice between alternatives that have certain rather than uncertain monetary payoffs. The choice set is $\{x, y\}$, and the values of x and y are $V(x)$ and $V(y)$ – for instance, the option of either a guaranteed \$200 (x) or a guaranteed \$100 (y). In this simple situation, one can maximize the gain according to the rule:

Maximizing under certainty:

Choose x if $V(x) > V(y)$.

The monetary values V may be replaced by the utilities U , but for all monotone utility functions the outcome is the same. Like Property Alpha, this rule seems to be so trivial that anyone who violates it would appear odd or irrational. Who would choose \$100 instead of \$200?

In many situations, of course, we are quite right to interpret violations of maximization as peculiar. Nevertheless, I seek to make the same point with maximization as with consistency: It cannot be imposed independent of something external to choice behavior. In particular, maximization, like internal consistency, does not capture the distinction between the individual in isolation and the individual in a social context. The following anecdote illustrates my point (Gigerenzer, 1991b).

The village idiot. In a small town there lives a village idiot. He was once offered a choice between one pound (x) and one shilling (y). He took the shilling. Having heard about this phenomenon, all of the townspeople in turn offered him the choice between a pound and a shilling. He always took the shilling.

Seen as a singular choice (as the first time was intended to be), his taking the shilling seems irrational by all strictly monotone utility functions. Yet seen in a social context where that particular choice increases the probability of getting to choose again, the violation of maximization makes sense. Make a "stupid" choice and you may get another chance.

My point here is the same as with Property Alpha. The principle of maximizing under certainty is indeterminate, unless the social objectives, motivations, and expectations are analyzed in the first place.

Maximizing: choice under uncertainty

Now consider a choice between alternatives with uncertain outcomes. The choice set is $\{x, y\}$. Choosing x will lead to a reinforcement with a probability $p(x) = .80$, whereas choosing y will only lead to the same reinforcement with a probability $p(y) = .20$. That is, the utilities of the outcomes (reinforcements) are the same, but their probabilities differ. It is easy to see that when the choice is repeated n times, the expected number of reinforcements will be maximized if an organism always chooses x :

Maximizing with equal utilities:

Always choose x if $p(x) > p(y)$.

Consider a hungry rat in a T-maze where reinforcement is obtained at the left end in 80% of cases and at the right end in 20% of cases. The rat will maximize reinforcement if it always turns left. Imagine students who watch the rat running and predict which side the reinforcement will appear in each trial. They will also maximize their number of correct predictions by always saying "left." But neither rats nor students seem to maximize. Under a wide variety of experimental conditions, organisms choose both alternatives with relative frequencies that roughly match the probabilities (Gallistel, 1990):

Probability matching:

Choose x with probability $p(x)$;

choose y with probability $p(y)$.

In the preceding example, the expected rate of reinforcements is 80% for maximizing, but only 68% for probability matching (this value is calculated by $.80^2 + .20^2 = .68$). The conditions of the seemingly irrational behavior of probability matching are discussed in the literature (e.g., Brunswik, 1939; Estes, 1976; Gallistel, 1990).

Violations of maximizing by probability matching pose a problem for a context-independent account of rational behavior in animals and humans. What looks irrational for an individual, however, can be optimal for a group. Again, the maximizing principle does not capture the distinction between the individual in social isolation and in social interaction. Under natural conditions of foraging, there will not be just one rat but many who compete to exploit food resources. If all choose to forage in the spot where previous experience suggests food is to be found in greatest abundance, then each may get only a small share. The one mutant organism that sometimes chooses the spot with less food would be better off. Natural selection will favor those exceptional individuals who sometimes choose the less attractive alternative. Thus, maximizing is not always an evolutionarily stable strategy in situations of competition among individuals. Given certain assumptions, probability matching may in fact be an evolutionarily stable strategy, one that does not tend to create conditions that select against it (Fretwell, 1972; Gallistel, 1990).

To summarize the argument: The maximization rule cannot be imposed on behavior independent of social context. Whether an organism performs in isolation or in the context of other organisms can determine, among other things, whether maximization is entailed as an optimal choice rule.

Betting against the probabilities

Mr. Smart would like to invest the \$10,000 in his savings account in the hope of increasing his capital. After some consideration, he opts to risk the amount in a single gamble with two possible outcomes, x and y . The outcomes are determined by a fair roulette wheel with 10 equal sections, 6 of them white (x) and 4 black (y). Thus, the probability $p(x)$ of obtaining white is .6, and the probability $p(y)$ of obtaining black is .4. The rules of the game are that he has to bet *all* his money (\$10,000) either on black or on white. If Mr. Smart guesses the outcome correctly, his money will be doubled; otherwise, he will lose three-quarters of his investment. Could it ever be advantageous for Mr. Smart to bet on black?

If Mr. Smart bets on white, his expectation is \$20,000 with a probability of .6,

and \$2,500 with a probability of .4. The expected value $E(x)$ is $(.6 \times \$20,000) + (.4 \times \$2,500) = \$13,000$. But if he bets on black, the expected value $E(y)$ is only $(.4 \times \$20,000) + (.6 \times \$2,500) = \$9,500$. Betting on white would give him an expectation larger than the sum he invests. Betting on black, on the other hand, would result in an expectation lower than the sum he invests. A maximization of the expected value implies betting on white:

Maximizing expected value:

Choose x if $E(x) > E(y)$,

where $E(x) = \sum p(x)V(x)$. The principle of maximizing the expected value (or subjective variants such as expected utility) is one of the cornerstones of classical definitions of rationality. Mr. Smart would be a fool to bet on black, wouldn't he?

Let me apply the same argument again. The principle of maximizing the expected value does not distinguish between the individual in social isolation and in social interaction. If many individuals face the same choice, could it be to the benefit of the whole group that some sacrifice themselves and bet on black? Let us first look at an example from biology.

Cooper (1989; Cooper & Kaplan, 1982) discussed conditions under which it is essential for the survival of the group that some individuals bet against the probabilities and do *not*, at the individual level, maximize their expected value. Consider a hypothetical population of organisms whose evolutionary fitness (measured simply by the finite rate of increase in their population) depends highly on protective coloration. Each winter predators pass through the region, decimating those within the population that can be spotted against the background terrain. If the black soil of the organisms' habitat happens to be covered with snow at the time, the best protective coloration is white; otherwise, it is black. The probability of snow when predators pass through is .6, and protectively colored individuals can expect to survive the winter in numbers sufficient to leave an average of two surviving offspring each, whereas the conspicuous ones can expect an average of only 0.25 offspring each. This example assumes a simple evolutionary model with asexual breeding (each offspring is genetically identical to its parent), seasonal breeding (offspring are produced only in spring), and semelparous breeding (each individual produces offspring only once in a lifetime at the age of exactly one year).

Adaptive coin-flipping. Suppose two genotypes, W and WB , are in competition within a large population. Individuals of genotype W always have white winter coloration; that is, W is a genotype with a uniquely determined phenotypic expression. Genotype WB , in contrast, gives rise to both white and black individuals, with a ratio of 5 to 3. Thus, 3 out of 8 individuals with

genotype *WB* are “betting” on the low probability of no snow. Each of these individuals’ expectation to survive and reproduce is smaller than that of all other individuals in both *W* and *WB*.

How will these two genotypes fare after 1,000 generations (1,000 years)? We can expect that there was snow cover in about 600 winters, exposed black soil in about 400 winters. Then, the number of individuals with genotype *W* will be doubled 600 times and reduced to one-fourth 400 times. If n is the original population size, the population size after 1,000 years is

$$n2^{600}\left(1/4\right)^{400} \approx 6n \times 10^{-61}.$$

That is, genotype *W* will have been wiped out with practical certainty after 1,000 years. How does genotype *WB* do? In the 600 snowy winters, 5/8 of the population will double in number and 3/8 will be reduced to 25 percent, with corresponding proportions for the 400 winters without snow. The number of individuals after 1,000 years is then

$$n\left(\frac{5}{8} \times 2 + \frac{3}{8} \times \frac{1}{4}\right)^{600} \left(\frac{5}{8} \times \frac{1}{4} + \frac{3}{8} \times 2\right)^{400} \approx 8n \times 10^{59}.$$

Thus, genotype *WB* is likely to win the evolutionary race easily.¹ (The large estimated number is certainly an overestimation, however, because it does not take account of such other constraints as food resources.) The reason why *WB* has so much better a chance of survival than *W* is that a considerable proportion of the *WB* individuals do not maximize their individual expectations, but “bet” on small probabilities.

This violation of individual maximization has been termed “adaptive coin-flipping” (Cooper & Kaplan, 1982), meaning that individuals are genetically programmed to “flip coins” to adopt phenotypic traits. Thus, the phenotype is ultimately determined by the nature of the coin-flipping process, rather than uniquely specified by the genotype.²

Back to Mr. Smart. Assume he won and wants to try again. So do his numerous brothers, sisters, and cousins, who all are willing to commit their entire investment capital to this gamble. The game is offered every week, and the rules are as before: Each person’s choice is every week to bet *all* his or her investment capital either on black or on white (no hedging of bets). If everyone wanted solely to maximize his or her individual good, his or her money would be better invested in white than in black, since the chances to double one’s assets are 60% for white compared with only 40% for black. Investing in black would appear irrational. But we know from our previous calculations that someone who invests all his or her money every week in white will, with a high probability, lose every dollar of assets in the long run.

If Mr. Smart and his extended family, however, acted as one community rather than as independent individuals – that is, create one investment capital fund in which they share equally – they can quickly increase their capital with a high probability. Every week they would need to instruct $3/8$ of their members to invest in black, the rest in white. This social sharing is essentially the same situation as the “adaptive coin-flipping” example (Cooper, 1989). Thus, Mr. Smart’s betting on black needs to be judged against his motivation: If he is cooperating with others for their common interest, then betting on the wrong side of a known probability is part of an optimal strategy. If he is not cooperating but, rather, investing for his own immediate benefit, then betting on black is the fastest way to ruin.

This example, like the preceding ones, attempts to illustrate that a rule such as maximizing the expected value cannot be imposed on behavior without consideration of the social context. Is this context a single individual wagering all his or her assets at once or a population that risks their collective assets or offspring at regular intervals? It makes all the difference, since individual maximization can lead to the extinction of the genotype.

Conclusions

These examples show that general principles such as consistency and maximizing are insufficient for capturing rationality. I have argued that there is no way of determining whether a behavioral pattern is consistent or maximizes without first referring to something external to choice behavior (Sen, 1993). The external factor investigated in this chapter is the *social context* of choice behavior, including objectives, motivations, and values. I am not arguing against consistency, maximization, or any given rule per se, but against the a priori imposition of a rule or axiom as a requirement for rationality, independent of the social context of judgment and decision and, likewise, of whether the individual operates in isolation or within a social context (Elster, 1990; Gigerenzer, 1991b).

One way to defend general principles against this argument would be to say that maximization poses no restrictions on what individuals maximize, be it their own good (utilities) or the fitness of their genotype. Switching from individual goals to genotypic fitness can save the concept of maximization. Such a defense would imply, however, that maximization cannot be imposed independent of the motivations and goals built into living systems, which is precisely the point I have asserted. By the same token, to claim that consistency poses no restrictions on whatever consistency is about would destroy the very idea of behavioral consistency, because Property Alpha would as a result be open to any external interpretation and would no longer impose any constraint on choice.

More generally, the formal principles of logic, probability theory, rational choice theory, and other context-independent principles of rationality are often rescued and defended by post hoc justifications. Post hoc reasoning typically uses the social objectives, values, and motivations of organisms to make room for exceptions or to reinterpret the alternatives in axioms or rules until they are compatible with the observed result. Contemporary neoclassical economics, for instance, provides little theoretical basis for specifying the content and shape of the utility function; it thus affords many degrees of freedom for fitting any phenomenon to the theory (Simon, 1986). In Elster's (1990) formulation, a theory of rationality can fail through indeterminacy (rather than through inadequacy) to the extent that it fails to yield unique predictions.

The challenge is to go beyond general-purpose principles of rationality that allow context to slip in through the back door. What would a theory of reasoning that lets social context in through the front door look like? In the second part of this chapter I will present and discuss one example of such a "front door" theory. Inspired by evolutionary theory, it appears to be the only theory so far to relate reciprocal altruism to human reasoning.

Toward models of social rationality

The psychological flip side of Leibniz's dream of a universal calculus of reasonableness is the assumption that there is one – or at most a few – universal mechanisms that govern all of reasoning, learning, memory, inference, imitation, imagery, and so on. I will call these assumed mechanisms *general-purpose* mechanisms, because they have no features specialized for processing particular kinds of content. For instance, when Piaget started to work on mental imagery and memory, he did not expect and search for processes different from logical thinking. Rather, he attempted to demonstrate that at each stage in development, imagery and memory express the same logical structure as the one he had found in his earlier studies on children's thinking (Gruber & Vonèche, 1977). Similarly, B. F. Skinner's laws of operant behavior were designed to be general-purpose: to hold true for all stimuli and responses (the assumption of the *equipotentiality* of stimuli).

John Garcia's anomalous findings (e.g., Garcia & Koelling, 1966) challenged not only the notion of the equipotentiality of stimuli, but also the law of contiguity, which postulates the necessity of immediate reinforcement, independent of the nature of the stimulus and response. For instance, when the taste of flavored water is repeatedly paired with an electric shock immediately after tasting, rats have great difficulty learning to avoid the flavored water. Yet in just one trial the rat can learn to avoid the flavored water when it is followed by experimentally induced nausea, even when the nausea occurs

2 hours later. "From the evolutionary view, the rat is a biased learning machine designed by natural selection to form certain CS-US [conditioned stimulus-unconditioned stimulus] associations rapidly but not others. From a traditional learning viewpoint, the rat was an unbiased learner able to make any association in accordance with the general principles of contiguity, effect, and similarity" (Garcia y Robertson & Garcia, 1985, p. 25). Garcia's evolutionary challenge, however, was not welcomed by mainstream neobehaviorists. In 1965, after 10 years of research, he openly pointed out the clash between the data and the ideal of general-purpose mechanisms – and his manuscripts suddenly began to be rejected by the editors of the APA (American Psychological Association) journals. This pattern continued for the next 13 years until, in 1979, Garcia was awarded the APA's Distinguished Scientific Contribution Award (Lubek & Apfelbaum, 1987). By then, stimulus equipotentiality was driven out of behaviorism, but had found new fertile ground in cognitive psychology.

The view that psychological mechanisms such as those described in the laws of operant behavior are designed for specific classes of stimuli rather than being general-purpose is known as *domain specificity* (e.g., Hirschfeld & Gelman, 1994), *biological preparedness* (Seligman & Hager, 1972), or, in biology, *special-design* theories (Williams, 1966).

Mainstream cognitive psychology, however, still tries to avoid domain specificity. The senses, language, and emotions have occasionally been accepted as domain-specific adaptations (Fodor, 1983). But the "central" cognitive processes that define the rationality of *Homo sapiens* – reasoning, inference, judgment, and decision making – have not. Even such vigorous advocates of domain specificity as Fodor (1983) have held so-called central processes to be general-purpose. Research on probabilistic, inductive, and deductive reasoning tends to define good reasoning exclusively in terms of formal axioms and rules similar to those discussed in the first section. Mental logic, Johnson-Laird's mental models, and Piaget's formal operations all are examples of the hope that reasoning can be understood without reference to its content.

Yet content-dependence is by no means denied; rather, it is generally acknowledged, but only for *behavior* and not for *cognitive processes*, and more as an annoyance than as a challenge to rethink the nature of our theorizing. For instance, in the last pages of their *Psychology of Reasoning: Structure and Content* (1972), Wason and Johnson-Laird conceded that "for some considerable time we cherished the illusion . . . that only the structural characteristics of the problem mattered," finally concluding that "content is crucial" (pp. 244–245). However, neither this classic nor subsequent work on mental models (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991) has found a way to deal effectively with content. A similarly unresolved tension exists in Kahneman and Tversky's work on judgment under uncertainty. They grant that "human

reasoning cannot be adequately described in terms of content-independent rules" (Kahneman & Tversky, 1982, p. 499). Despite this insight, however, they continued to explain reasoning by general-purpose heuristics such as representativeness and availability. Note that the notion of availability assumes that behavior is dependent on the context (such as the ease with which particular examples come to mind), but as with Skinner's laws, the assumption is that the process is general-purpose. Similarly, current research on decision making generally acknowledges that behavior is dependent on content, but it refrains from the assumption that the cognitive processes themselves may be domain-specific (see Goldstein & Weber, *in press*). In the same vein, artificial intelligence systems and research that models expert knowledge usually reduce domain specificity to knowledge and assume a single unified inference system as a working hypothesis. Finally, exemplar models of categorization that posit categories as represented by memory traces of the specific instances experienced nevertheless portray the categorization process itself by employing content-independent, general-purpose laws (Barsalou, 1990).

A glance at textbooks on cognitive psychology reveals how we have bottled our students on the idea that whenever reasoning is the object of our investigation, content does not matter. Typically, a chapter on "deductive reasoning" teaches propositional logic and violations thereof by human reasoning, while a chapter on "probabilistic reasoning" teaches the laws of probability theory and violations thereof by human reasoning. Similarly, "fallacies" of reasoning are defined against formal structure – the base rate fallacy, the conjunction fallacy, and so on. Content is merely illustrative and cosmetic, as it is in textbooks of logic. Whether a problem concerns white and black swans, blue and green taxis, or artists and beekeepers does not seem to matter. Content has not yet assumed a life of its own. For the most part, it is seen only as a disturbing factor that sometimes facilitates and sometimes hinders formal, rational reasoning.

Is there an alternative? In what follows, I shall describe a domain-specific theory of cognition that relates reasoning to the evolutionary theory of reciprocal altruism (Cosmides & Tooby, 1992). This theory turns the traditional approach upside down. It does not start out with a general-purpose principle from logic or probability theory or a variant thereof; it takes social objectives as fundamental, which in turn makes content fundamental, since social objectives have specific contents. Traditional formal principles of rationality are not imposed; they can be entailed or not, depending on the social objectives.

Cheating detection in social contracts

One feature that sets humans and some other primates apart from almost all animal species is the existence of cooperation among genetically unrelated

individuals within the same species, known as *reciprocal altruism* or *cooperation* (see Hammerstein, Chapter 1, this volume). The thesis that such cooperation has been practiced by our ancestors since ancient times, possibly for at least several million years, is supported by evidence from several sources. First, our nearest relatives in the hominid line, chimpanzees, also engage in certain forms of sophisticated cooperation (de Waal & Luttrell, 1988), and in more distant relatives, such as macaques and baboons, cooperation can still be found (e.g., Packer, 1977). Second, cooperation is both universal and highly elaborated across human cultures, from hunter-gatherers to technologically advanced societies. Finally, paleoanthropological evidence also suggests that cooperation is extremely ancient (e.g., Tooby & DeVore, 1987).

Why altruism? Kin-related helping behavior, such as that by the sterile worker castes in insects, which so troubled Darwin, has been accounted for by generalizing "Darwinian fitness" to "inclusive fitness" – that is, to the number of surviving offspring an individual has *plus* the individual's effect on the number of offspring produced by its relatives (Hamilton, 1964). But why reciprocal altruism, which involves cooperation among two or more nonrelated individuals? The now-classic answer draws on the economic concept of trade and its analogy to game theory (Axelrod, 1984; Williams, 1966). If the reproductive benefit of being helped is greater than the cost of helping, then individuals who engage in reciprocal helping can outreproduce those who do not, causing the helping design to spread. A vampire bat, for instance, will die if it fails to find food for two consecutive nights, and there is high variance in food-gathering success. Food sharing allows the bats to reduce this variance, and the best predictor of whether a bat, having foraged successfully, will share its food with a hungry nonrelative is whether the nonrelative has shared food with the bat in the past (Wilkinson, 1990).

But "always cooperate" would not be an evolutionarily stable strategy. This can be seen using the analogy of the prisoner's dilemma (Axelrod, 1984). If a group of individuals always cooperates, then individuals who always defect – that is, who take the benefit but do not reciprocate – can invade and outreproduce the cooperators. Where the opportunity for defecting (or cheating) exists, indiscriminate cooperation would eventually be selected out. "Always defect" would not be an evolutionarily stable strategy, either. A group of individuals who always defect can be invaded by individuals who cooperate in a selective (rather than indiscriminate) way. A simple rule for selective cooperation is "Cooperate on the first move; for subsequent moves, do whatever your partner did on the preceding move" (a strategy known as tit for tat). There are several rules in addition to tit for tat that lead to cooperation with other "selective cooperators" and exclude or retaliate against cheaters (Axelrod, 1984).

The important point is that selective cooperation would not work without a cognitive program for detecting cheaters – or, more precisely, a program for directing an organism's attention to information that could reveal that it (or its group) is being cheated (Cosmides & Tooby, 1992). Neither indiscriminate cooperation nor indiscriminate cheating demands such a program. In vampire bats, who exchange only one thing – regurgitated blood – such a program can be restricted to a sole commodity. Cheating, or more generally noncooperation, would mean, “That other bat took my blood when it had nothing, but it did not share blood with me when I had nothing.” In humans, who exchange many goods (including such abstract forms as money), a cheating-detection program needs to work on a more general level of representation – in terms, for example, of “benefits” and “costs.” Human cooperation can have the form

If you take the benefit, then you must pay the cost.

Information that can reveal cheating is of the following kind:

The other party took the benefit, but did not pay the cost.

Benefits, by definition, evolve from cooperation; costs, however, may be but are not necessarily incurred. For instance, the other party may possess the exchanged item in such abundance that no cost is associated with satisfying the requirement. Here, “must pay the cost” can be replaced by the more general term “satisfy the requirement” (Cosmides, 1989). For simplicity, I do not make this distinction here.

To summarize, cooperation between two or more individuals for their mutual benefit is a solution to a class of important adaptive problems, such as the sharing of scarce food when foraging success is highly variable. Rather than being indiscriminate, cooperation needs to be selective, requiring a cognitive program that directs attention to information that can reveal cheating. This evolutionary account of cooperation, albeit still general, has been applied to a specific topic in the psychology of reasoning.

Reasoning about conditional statements

In 1966, Peter Wason invented the “selection task” to study reasoning about conditionals. This was to become one of the most extensively researched subjects in cognitive psychology during the following decades. The selection task involves four cards and a conditional statement in the form “If *P* then *Q*.” One example is, “If there is a ‘D’ on one side of the card, then there is a ‘3’ on the other side.” The four cards are placed on a table so that the subject can read only the information on the side facing upward. For instance, the four cards may read “D,” “E,” “3,” and “4.” The subject's task is to indicate which

Table 11.1. *Three selection tasks*

Numbers-and-letters rule:

If there is a "D" on one side of the card, then there is a "3" on the other side.

Each of the following cards has a letter on one side and a number on the other. Indicate only the card(s) you definitely need to turn over to see if the rule has been violated.

D	E	3	4
---	---	---	---

Transportation rule:

If a person goes in to Boston, then he takes the subway.

The cards below have information about four Cambridge residents. Each card represents one person. One side of the card tells where the person went and the other side tells how the person got there.

Indicate only the card(s) you definitely need to turn over to see if the rule has been violated.

SUBWAY	ARLINGTON	CAB	BOSTON
--------	-----------	-----	--------

Day-off rule:

If an employee works on the weekend, then that person gets a day off during the week.

The cards below have information about four employees. Each card represents one person. One side of the card tells whether the person worked on the weekend, and the other side tells whether the person got a day off during the week.

Indicate only the card(s) you definitely need to turn over to see if the rule has been violated.

WORKED ON THE WEEKEND	DID GET A DAY OFF	DID NOT WORK ON THE WEEKEND	DID NOT GET A DAY OFF
-----------------------------	----------------------	--------------------------------------	-----------------------------

of the four cards need(s) to be turned over to find out whether the statement has been violated. Table 11.1 shows three examples of selection tasks, each with a different content: a numbers-and-letters rule, a transportation rule, and a "day-off" rule.

Because the dominant approach has been to impose propositional logic as a general-purpose standard of rational reasoning in the selection task (independent, of course, of the content of the conditional statements), it is crucial to recall that, according to propositional logic, a conditional "If *P* then *Q*" can be

violated only by " P & not- Q ." In general, that is, the logical falsity of a material conditional is defined within propositional logic in the following way:

Logical falsity:

"If P then Q " is logically false if and only if " P & not- Q ."

Thus, the " P " and "not- Q " cards, and no others, must be selected, since only these can reveal " P & not- Q " instances. In the numbers-and-letters rule, these cards correspond to the " D " and "4" cards; in the transportation problem, to the "Boston" and "cab" cards; and in the "day-off" problem, to the "worked on the weekend" and "did not get a day off" cards.

Wason's results showed, however, that human inferences did not generally follow propositional logic. An avalanche of studies has since confirmed this, reporting that, with numbers-and-letters rules, only about 10% of the subjects select both the " P " and "not- Q " cards, while most select the " P " card and the " Q " card, or only the " P " card. It was soon found that the selections were highly dependent on the content of the conditional statement. This was labeled the "content effect." For instance, about 30% to 40% of subjects typically choose the " P " and "not- Q " cards in the transportation problem (Cosmides, 1989), but 75% in the day-off problem (Gigerenzer & Hug, 1992). These results are inconsistent with Piaget's claim that adults should have reached the stage of formal operations (Legrenzi & Murino, 1974; Wason, 1968; Wason & Johnson-Laird, 1970).

Within a decade it was clear that these results – a low overall proportion of " P & not- Q " answers and the content effect – contradicted the model of human reasoning provided by propositional logic. One might expect that propositional logic was then abandoned; but it was abandoned only as a *descriptive* model of reasoning. Propositional logic was, however, retained as the *normative*, content-independent yardstick of good reasoning, and actual human reasoning was blamed as irrational. The experimental manipulations were evaluated, as is still the case today, in terms of whether or not they "facilitated logical reasoning." Much effort was directed at explaining subjects' apparent irrationality, including their "incurable conviction that they are right when they are, in fact, wrong" (Wason, 1983, p. 356). It was proposed that the mind runs with deficient mental software – for example, confirmation bias, matching bias, and availability heuristic – rather than by propositional logic. Yet these proposals were as general-purpose as propositional logic; they could be applied to any content. It seems fair to say that these vague proposals have not led to an understanding of what subjects do in the selection task.

Only since the mid-1980s have a few dissidents dared to design theories that start with the *content* of the conditional statement rather than with propositional logic (Cheng & Holyoak, 1985; Cosmides, 1989; Light, Girotto, & Legrenzi, 1990; Over & Manktelow, 1993). I will concentrate here

exclusively on Cosmides' proposal, which takes the evolutionary theory of cooperation as its starting point (for a different evolutionary account, see Klux, 1993).

Cosmides' (1989) central point is that selective cooperation demands the ability to detect cheaters. This ability presupposes several others, including that of distinguishing different individuals, recognizing when a reciprocation (social contract) is offered, and computing costs and benefits, all of which I ignore here (Cosmides & Tooby, 1992). Being cheated in a social contract of the type

If you take the benefit, then you have to pay the cost

means that the other party has exhibited the following behavior:

Benefit taken and cost not paid.

The evolutionary perspective suggests that humans, who belong to one of the few species practicing reciprocal altruism since time immemorial, have evolved a cognitive system for directing attention to information that could reveal cheaters. That is, once a cognitive system has classified a situation as one of cooperation, attention will be directed to information that could reveal "benefit taken and cost not paid." Note that cheating detection in social contracts is a domain-specific mechanism; it would not apply if a conditional statement is coded as a threat, such as "If you touch me, then I'll kill you." But how does this help us to understand the content effect in the selection task?

The thesis is that the cheating detection mechanism required by the theory of reciprocal altruism guides reasoning in the selection task:

If the conditional statement is coded as a social contract, and the subject is cued in to the perspective of one party in the contract, then attention is directed to information that can reveal being cheated.

In other words, a subject should select those cards that correspond to "benefit taken" and "cost not paid," whatever the cards' logical status is. This application of the theory of reciprocal altruism to an unresolved issue in human reasoning is, of course, a bold thesis.

Experimental studies

Cosmides (1989) has shown that her results as well as earlier studies corroborated this thesis. If the conditional statement expressed a social contract, then the percentage of "benefit taken" and "cost not paid" selections was very high. For instance, in the day-off problem in Table 11.1, 75% of subjects selected the cards "worked on the weekend" and "did not get a day off" (Gigerenzer &

Hug, 1992). However, this result can also be consistent with competing accounts that do not invoke reciprocal altruism, so we need to look more closely at tests that differentiate between competing accounts. Below is a sample of tests with that aim.

What guides reasoning: availability or cheater detection? The major account of the content effect in the 1970s and 1980s was variously called “familiarity” and “availability” (Manktelow & Evans, 1979; Pollard, 1982), without ever being precisely defined. The underlying idea is that the more familiar a statement is, the more often a subject may have experienced associations between the two propositions in a conditional statement, including those that are violations (“benefit taken” and “cost not paid”) of the conditional statement. In this view, familiarity makes violations more “available” in memory, and selections may simply reflect availability. According to this conjecture, therefore, familiarity and not social contracts accounts for selecting the “benefit taken” and “cost not paid” cards. If familiarity were indeed the guiding cognitive principle, then *unfamiliar* social contracts should not elicit the same results. However, Cosmides (1989) showed that social contracts with unfamiliar propositions elicit the same high number of “benefit taken” and “cost not paid” selections, in contradiction to the availability account. This result was independently replicated by both Gigerenzer and Hug (1992) and Platt and Griggs (1993).

Are people simply good at reasoning about social contracts? The game-theoretic models for the evolution of cooperation require, as argued earlier, some mechanism for detecting cheaters in order to exclude them from the benefits of cooperation. The second conjecture, however, rejects any role of cheating detection in the selection task, claiming that people are, for some reason, better at reasoning about social contracts than about numbers-and-letters problems. Social contracts may be more “interesting” or “motivating,” or people may have some “mental model” for social contracts that affords “clear” thinking. Although this alternative is nebulous, it needs to be taken into account; in her tests, Cosmides (1989) never distinguished between social contracts and cheating detection.

But one can experimentally disentangle social contracts from cheating detection. Klaus Hug and I also used social contracts, but varied whether the search for violations constituted looking for cheaters or not (Gigerenzer & Hug, 1992). For instance, consider the following social contract: “If someone stays overnight in the cabin, then that person must bring along a bundle of wood from the valley.” This was presented in one of two context stories.

The “cheating” version explained that a cabin high in the Swiss Alps serves as an overnight shelter for hikers. Since it is cold and firewood is not otherwise

available at this altitude, the Swiss Alpine Club has made the rule that each hiker who stays overnight in the cabin must bring along a bundle of firewood from the valley. The subjects were cued to the perspective of a guard who checks whether any of four hikers has violated the rule. The four hikers were represented by four cards (similar to those in Table 11.1) that read "stays overnight in the cabin," "does not stay overnight," "carried wood," and "carried no wood." The instruction was to indicate only the card(s) you definitely need to turn over to see if any of these hikers have violated the rule.

In the "no-cheating" version, the subjects were cued to the perspective of a member of the German Alpine Association, visiting the same cabin in the Swiss Alps to find out how it is managed by the local Alpine Club. He observes people carrying firewood into the cabin, and a friend accompanying him suggests that the Swiss may have the same overnight rule as the Germans, namely, "If someone stays overnight in the cabin, then that person must bring along a bundle of wood from the valley." That this is also the Swiss Alpine Club's rule is not the only possible explanation; alternatively, only its members (who do not stay overnight in the cabin), and not the hikers, might bring firewood. The subjects were now in the position of an observer who checks information to find out whether the social contract suggested by his friend actually holds. This observer does not represent a party in a social contract. The subjects' instruction was the same as in the cheating version.

Thus, in the cheating scenario, the observation "benefit taken and cost not paid" means that the party represented by the guard is being cheated; in the no-cheating scenario, the same observation suggests only that the Swiss Alpine Club never made the supposed rule in the first place.

Assume as true the conjecture that what matters is only that the rule is a social contract, making the game-theoretic model (which requires a cheating mechanism) irrelevant. Since in both versions the rule is always the same social contract, such a conjecture implies that there should be no difference in the selections observed. In the overnight problem, however, 89% of the subjects selected "benefit taken" and "cost not paid" when cheating was at stake, compared with 53% in the no-cheating version. Similarly, the averages across all four test problems used were 83% and 45%, respectively, consistent with the game-theoretic account of cooperation (Gigerenzer & Hug, 1992).

Do social contracts simply facilitate logical reasoning? In most of Cosmides' tests, the predicted "benefit taken" and "cost not paid" selections corresponded to the truth conditions of conditionals in propositional logic. Thus, a third conjecture would be that social contracts may somehow facilitate logical

reasoning, which we tested by deducing predictions from the cheating-detection hypothesis that contradicted propositional logic (Gigerenzer & Hug, 1992). The key to these tests is that cheating detection is pragmatic and perspectival, whereas propositional logic is aperspectival. For instance, in the day-off problem in Table 11.1, subjects were originally cued to the perspective of an employee, in which case cheating detection and propositional logic indeed predict the same cards. We switched the perspective from employee to employer but held everything else constant (the conditional statement, the four cards, and the instruction shown in Table 11.1). For the employer, being cheated means "did not work on the weekend and did get a day off"; that is, in this perspective subjects should select the "did not work on the weekend" and the "did get a day off" cards, which correspond to the "not-*P*" and "*Q*" cards. (Note that "not-*P* & *Q*" selections have rarely been observed in selection tasks.) Thus, perspective change can play cheating detection against general-purpose logic. The two competing predictions are: If the cognitive system attempts to detect instances of "benefit taken and cost not paid" in the other party's behavior, then a perspective switch implies switching card selections; if the cognitive system reasons according to propositional logic, however, pragmatic perspectives are irrelevant and there should be no switch in card selections.

The results showed that when the perspective was changed, the cards selected also changed in the predicted direction. The effects were strong and robust across problems. For instance, in the employee perspective of the day-off problem, 75% of the subjects had selected "worked on the weekend" and "did not get a day off," but only 2% had selected the other pair of cards. In the employer perspective, this 2% (who had selected "did not work on the weekend" and "did get a day off") rose to 61% (Gigerenzer & Hug, 1992). The result is consistent with the thesis that attention is directed toward information that could reveal oneself (or one's group) as being cheated in a social contract, but is inconsistent with the claim that reasoning is directed by propositional logic independent of content.³

Thus, social contracts do not simply facilitate logical reasoning. I believe that the program of reducing context merely to an instrument for "facilitating" logical reasoning is misguided. My point is the same as for Property Alpha. Reasoning consistent with propositional logic is entailed by some perspectives (e.g., the employee's), but is not entailed by other perspectives (e.g., the employer's).

Two additional conjectures can be dealt with briefly. First, several authors have argued that the cheating detection thesis can be invalidated because "logical facilitation" (large proportions of "*P* & not-*Q*" selections) has also been found in some conditional statements that were not social contracts (e.g., Cheng & Holyoak, 1989; Politzer & Nguyen-Xuan, 1992). This conjecture

misconstrues the thesis in two respects. The thesis is not about "logical facilitation"; the conjunction "benefit taken and cost not paid" is not the same as the logical conjunction " $P \& \text{not-}Q$," as we have seen. Furthermore, a domain-specific theory makes, by definition, no prediction about performance outside its own domain; it can only be refuted within that domain.

The second conjecture also tries to reduce the findings to propositional logic, pointing out that a conditional that states a social contract is generally understood as a biconditional "if and only if." In this case all four cards can reveal logical violations and need to be turned over. However, it is not true that four-card selections are frequent when cheating detection is at stake. We found in about half of the social contract problems (12 problems, each answered by 93 students) that not a single subject had selected all four cards; for the remaining problems, the number was very small. Only when cheating detection was excluded (the no-cheating versions) did four-card selections increase to a proportion of about 10% (Gigerenzer & Hug, 1992). There is, then, no evidence that subjects follow propositional logic even if we assume that they interpret the implication as a biconditional.

Such logical reductionism cannot explain how the mind infers that a particular conditional should be understood as a material implication, a biconditional, or something else. This inference is accomplished, I believe, by coding the specific content of the conditional statement as an instance of a larger domain, such as social contract, threat, and warning (Fillenbaum, 1977).

Conclusions

The evolutionary theory of cooperation illustrates how to begin constructing a theory of cognition situated in social interaction. The idea is to begin with a specific design that a cognitive system requires for social interaction, rather than with a general-purpose, formal system – in other words, to start with the functional and see what logic it entails, rather than to impose some logic a priori. The virtues of this approach are as evident as its unresolved questions. Among these questions are: How can we precisely describe the "Darwinian algorithms" that determine when a social contract is in place? How does the mind infer that the conditional statement "If you touch me, then I'll kill you" does not imply a social contract, but a threat? What are the cues coding this specific statement into the domain of "threats" rather than "social contracts"? Once a statement is categorized into a particular domain, what distribution of attention is implicated by that domain? In a threat, for example, attention needs to be directed to information that can reveal being bluffed or double-crossed rather than cheated (Fillenbaum, 1977). The challenge is to design theoretical proposals for reasoning and inference processes in other domains of human interaction beyond cooperation in social contracts.

To approach reasoning as situated in social interaction is to assume that the cognitive system (a) generalizes a specific situation as an instance of a larger domain, and (b) reasons about the specific situation by applying a domain-specific cognitive module. This raises two questions about the nature of the domains and the design of the modules.

What are the domains and at what level of abstraction are they located? Imagine a vertical dimension of abstraction, in which the specific problem corresponds to the lowest level of abstraction, and some formal representation of the problem, stripped of any content and context, to the highest. Two diametrically opposed views correspond to the ends of this continuum of abstraction. First, it may be argued that the cognitive system operates at the lowest level of abstraction, guided by familiarity and availability of instances in memory (e.g., Griggs & Cox, 1982). Second, it may be argued that the cognitive system generalizes the specific problem to the highest level of abstraction (e.g., propositional logic), performs some logical operations on this formal representation, and translates the result back into application to the specific problem. Variants of the latter view include Piaget's theory of formal operations and mental logics.

The primary challenge of domain-specificity is to find a level of abstraction between the two extremes, where some content is stripped but an adequate amount retained. For instance, the level of social contracts and cheating detection could turn out to be too abstract, because cheating may assume different forms (e.g., in contracts in which both or only one side can be cheated; Gigerenzer & Hug, 1992), requiring different procedures of cheating detection. In contrast, the notions of social contracts and cheating detection may not be abstract enough, needing to be stripped of some content and placed at the more general level of social regulations, such as obligations, permissions, and other kinds of deontic reasoning (Cheng & Holyoak, 1985, 1989; Over & Manktelow, 1993). This focus on level of abstraction parallels Rosch's (1978) concern with "basic level objects."

What is the design of a domain-specific cognitive module? A cognitive module organizes the processes – such as inference, emotion, and the distribution of attention – that have been evolved and learned to handle a domain. To classify a specific situation as an instance of a given domain, a cognitive module must be connected to an inferential mechanism. For instance, David Premack and others assume that humans and primates first classify an encounter as an instance of either social interaction (in the broadest sense) or interaction with the nonliving world. There is evidence that the cues used for this inference involve motion patterns, analyzed by cognitive systems to classify objects in the world as "self-propelled" or not; this analysis is reminiscent of Fritz Heider's and Albert Michotte's work (Premack, 1990;

Sperber, 1994; Thinès, Costall, & Butterworth, 1991). Cognitive modules dealing with something external that has been coded as "self-propelled" attend to information such as whether it is friend or enemy, prey or predator. For a module that deals with inanimate things, no attention needs to be directed to information of this kind. Domain-specific modules can thus distribute attention in a more focused way than a domain-general mechanism. The challenge now before us is to come up with rich and testable models about the design of cognitive modules.

Toward a social rationality

Researchers in several disciplines are converging on a domain-specific program of studying reasoning and inference situated in social interaction. Primatologists have joined philosophers and psychologists in studying "social intelligence" (Kummer, Daston, Gigerenzer, & Silk, in press) and "Machiavelian intelligence" (Byrne & Whiten, 1988). Linguists and philosophers have begun to reinterpret the conclusions of experimental research, in particular the so-called fallacies and biases, by arguing that the interaction between subject and experimenter is constrained by conversational rather than formal axioms (e.g., Adler, 1991; Grice, 1975; Sperber & Wilson, 1986). Social psychologists have tested some of these proposals experimentally, concluding among other things that pervasive reasoning "biases" may not reflect universal shortcomings of the human mind, but instead the application of Gricean conversational principles that conflict with what formal logic seems to dictate (e.g., Schwarz, Strack, Hilton, & Naderer, 1991). Similarly, Tetlock's (1992) concept of "accountability" models the social side of decision making by emphasizing that people do not simply choose the better alternative, but in certain social interactions, choose the alternative they can better justify. Developmental psychologists have departed from Piaget's general-purpose processes and investigate the domain-specific processes and their change during development (Hirschfeld & Gelman, 1994). The convergence of these approaches promises a new vision of reasoning and rationality situated in social context.

I can only hope that this chapter will inspire some readers to rethink the imposition of formal axioms or rules as "rational," independent of context. In my opinion, the challenging alternative is to put the psychological and the social first – and then to examine what formal principles these entail. We need less Aristotle and more Darwin in order to understand the messy business of how to be rational in the uncertain world of interacting human beings. And we may have to abandon a dream. Leibniz's vision of a sovereign calculus, the Universal Characteristic, was a beautiful one. If only it had proved true.

Notes

- 1 I have reported the numbers only for the most likely event (i.e., 600 snowy winters out of 1,000 winters). If one looks at all possible events, one finds that those in which *W* would result in a larger population size than *WB* are extremely rare (Cooper, 1989). Nevertheless, the expected value is larger for *W* than for *WB*, due to the fact that in those very few cases where *W* results in a larger population size, this number is astronomically large. The reader who is familiar with the St. Petersburg paradox will see a parallel (Wolfgang Hell, in a personal communication, has drawn my attention to this fact). The parallel is best illustrated in Lopes's (1981) simulations of businesses selling the St. Petersburg gamble. Although these businesses sold the gamble far below its expected value, most nonetheless survived with great profits.
- 2 Adaptive coin-flipping is a special case of a general phenomenon: In variable environments (in which the time scale of variation is greater than the generation time of the organism, as in the example given), natural selection does not maximize expected individual fitness, but geometric mean fitness (Gillespie, 1977).
- 3 I should mention here that Platt and Griggs (1993) have claimed that they could not replicate Gigerenzer and Hug's (1992) effects of perspective change. Platt and Griggs, however, cued all subjects to the same perspective, *never* changing the perspective from one party in a social contract to the other. What they labeled a "cheating perspective manipulation" involved deleting from the story explicit hints that people might cheat in a social contract and that the contract would be reinforced. But there is no need for such hints, nor should these have any effect, since the thesis is that there is a cognitive program for cheating detection. Therefore, I am not surprised by their results – only by their conclusions.

References

- Adler, J. (1991). An optimist's pessimism: Conversation and conjunction. In E. Eells & T. Maruszewski (Eds.), *Probability and rationality: Studies on L. Jonathan Cohen's philosophy of science* (pp. 251–282). Amsterdam: Rodopi.
- Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic.
- Barsalou, L. W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition: Vol. 3. Content and process specificity in the effects of prior experiences* (pp. 61–88). Hillsdale, NJ: Erlbaum.
- Brunswick, E. (1939). Probability as a determiner of rat behavior. *Journal of Experimental Psychology*, 36, 553.
- Byrne, R., & Whiten, A. (Eds.). (1988). *Machiavellian intelligence*. Oxford: Oxford University Press.
- Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17, 391–416.
- Cheng, P. W., & Holyoak, K. J. (1989). On the natural selection of reasoning theories. *Cognition*, 33, 285–313.
- Cooper, W. S. (1989). How evolutionary biology challenges the classical theory of rational choice. *Biology and Philosophy*, 4, 457–481.
- Cooper, W. S., & Kaplan, R. (1982). Adaptive "coin-flipping": A decision-theoretic examination of natural selection for random individual variation. *Journal of Theoretical Biology*, 94, 135–151.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? *Cognition*, 31, 187–276.
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). Oxford: Oxford University Press.

- Daston, L. J. (1988). *Classical probability in the Enlightenment*. Princeton, NJ: Princeton University Press.
- de Waal, F. B. M., & Luttrell, L. M. (1988). Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethology and Sociobiology*, 9, 101–118.
- Elster, J. (1990). When rationality fails. In K. S. Cook & M. Levi (Eds.), *The limits of rationality* (pp. 19–51). Chicago: University of Chicago Press.
- Estes, W. (1976). The cognitive side of probability learning. *Psychological Review*, 83, 37–64.
- Fillenbaum, S. (1977). Mind your *p*'s and *q*'s: The role of content and context in some uses of *and*, *or*, and *if*. *Psychology of Learning and Motivation*, 11, 41–100.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Fretwell, S. D. (1972). *Populations in seasonal environments*. Princeton, NJ: Princeton University Press.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
- Garcia, J., & Koelling, R. A. (1966). The relation of cue to consequence in avoidance learning. *Psychonomic Science*, 4, 123–124.
- Garcia y Robertson, R., & Garcia, J. (1985). X-rays and learned taste aversions: Historical and psychological ramifications. In T. G. Burish, S. M. Levy, & B. E. Meyerowitz (Eds.), *Cancer, nutrition and eating behavior: A biobehavioral perspective* (pp. 11–41). Hillsdale, NJ: Erlbaum.
- Gigerenzer, G. (1991a). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, 98, 254–267.
- Gigerenzer, G. (1991b). How to make cognitive illusions disappear: Beyond “heuristics and biases.” In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 2, pp. 83–115). New York: Wiley.
- Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition*, 43, 127–171.
- Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.
- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance: How probability changed science and everyday life*. Cambridge: Cambridge University Press.
- Gillespie, J. H. (1977). Natural selection for variances in offspring numbers: A new evolutionary principle. *American Naturalist*, 111, 1010–1014.
- Goldstein, W. M., & Weber, E. (in press). Content and discontent: Indications and implications of domain specificity in preferential decision making. In J. R. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *The psychology of learning and motivation*. New York: Academic.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). New York: Academic.
- Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, 73, 407–420.
- Gruber, H. E., & Vonèche, J. J. (Eds.). (1977). *The essential Piaget*. New York: Basic.
- Hamilton, W. D. (1964). The genetic evolution of social behavior. Parts 1 and 2. *Journal of Theoretical Biology*, 7, 1–52.
- Hirschfeld, L. A., & Gelman, S. A. (Eds.). (1994). *Mapping the mind: Domain specificity in cognition and culture*. Cambridge: Cambridge University Press.
- Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge: Cambridge University Press.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.
- Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 493–508). Cambridge: Cambridge University Press.
- Klix, F. (1993). Erwachendes Denken: Geistige Leistung aus evolutionspsychologischer Sicht [Thought coming to life: Cognitive performance from an evolutionary psychological perspective]. Heidelberg: Spektrum Akademischer Verlag.
- Kummer, H., Daston, L., Gigerenzer, G., & Silk, J. (in press). The social intelligence hypothesis.

- In P. Weingart, P. Richerson, S. Mitchell, & S. Maasen (Eds.), *Human by nature*. Hillsdale, NJ: Erlbaum.
- Legrenzi, P., & Murino, M. (1974). Falsification at the pre-operational level. *Italian Journal of Psychology*, 1, 363-368.
- Leibniz, G. W. (1677/1951). Toward a universal characteristic. In G. W. Leibniz (Ed.), *Selections* (pp. 17-25). New York: Scribner's.
- Light, P., Girotto, V., & Legrenzi, P. (1990). Children's reasoning on conditional promises and permissions. *Cognitive Development*, 5, 369-383.
- Little, I. M. D. (1949). A reformulation of the theory of consumers' behavior. *Oxford Economic Papers*, 1, 90-99.
- Lopes, L. L. (1981). Decision making in the short run. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 377-385.
- Lubek, I., & Apfelbaum, E. (1987). Neo-behaviorism and the Garcia effect: A social psychology of science approach to the history of a paradigm clash. In M. Ash & W. Woodward (Eds.), *Psychology in twentieth-century thought and society* (pp. 59-92). Cambridge: Cambridge University Press.
- Manktelow, K. I., & Evans, J. S. B. T. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology*, 70, 477-488.
- Over, D. E., & Manktelow, K. I. (1993). Rationality, utility and deontic reasoning. In K. I. Manktelow & D. E. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 231-259). London: Routledge.
- Packer, C. (1977). Reciprocal altruism in *Papio annubis*. *Nature*, 265, 441-443.
- Piaget, J., & Inhelder, B. (1951/1975). *The origin of the idea of chance in children*. New York: Norton.
- Platt, R., & Griggs, R. (1993). Darwinian algorithms and the Wason selection task: A factorial analysis of social contract selection task problems. *Cognition*, 48, 163-192.
- Politzer, G., & Nguyen-Xuan, A. (1992). Reasoning about promises and warnings: Darwinian algorithms, mental models, relevance judgments or pragmatic schemas? *Quarterly Journal of Experimental Psychology*, 44A, 402-421.
- Pollard, P. (1982). Human reasoning: Some possible effects of availability. *Cognition*, 12, 65-96.
- Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition*, 36, 1-16.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.
- Samuelson, P. A. (1938). A note on the pure theory of consumers' behavior. *Economica*, 5, 61-71.
- Schwarz, N., Strack, F., Hilton, D., & Naderer, G. (1991). Base rates, representativeness and the logic of conversation: The contextual relevance of "irrelevant" information. *Social Cognition*, 9(1), 67-84.
- Seligman, M. E. P., & Hager, J. L. (Eds.). (1972). *Biological boundaries of learning*. New York: Appleton-Century-Crofts.
- Sen, A. (1993). Internal consistency of choice. *Econometrica*, 61(3), 495-521.
- Simon, H. (1986). Rationality in psychology and economics. In R. Hogarth & M. Reder (Eds.), *Rational choice: The contrast between economics and psychology* (pp. 25-40). Chicago: University of Chicago Press.
- Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain-specificity in cognition and culture* (pp. 39-67). Cambridge: Cambridge University Press.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.
- Tetlock, P. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. *Advances in Experimental Social Psychology*, 25, 331-357.
- Thinès, G., Costall, A., & Butterworth, G. (Eds.). (1991). *Michotte's experimental phenomenology of perception*. Hillsdale, NJ: Erlbaum.
- Tooby, J., & DeVore, I. (1987). The reconstruction of hominid behavioral evolution through strategic modeling. In W. G. Kinzey (Ed.), *The evolution of human behavior: Primate models* (pp. 183-237). Albany: State University of New York Press.

- Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20, 273–281.
- Wason, P. C. (1983). Realism and rationality in the selection task. In J. S. B. T. Evans (Ed.), *Thinking and reasoning: Psychological approaches* (pp. 44–75). London: Routledge & Kegan Paul.
- Wason, P. C., & Johnson-Laird, P. N. (1970). A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology*, 61, 509–515.
- Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.
- Wilkinson, G. S. (1990). Food sharing in vampire bats. *Scientific American* (February), 76–82.
- Williams, G. C. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton, NJ: Princeton University Press.