Ph.D. Program in Cognitive Sciences
Course: Cognitive Economics
Dottorato di ricerca in Logica, Informatica e Scienze Cognitive
Modulo: Economia cognitiva
http://www.labsi.org/cognitive/

LabSi
EXPERIMENTAL ECONOMICS LABORATORY - University of Siena

Introduction to Cognitive Economics

Alessandro Innocenti
innocenti@unisi.it

Index

1. Cognitive and behavioural economics
2. Experimental economics
   2.1 Laboratory methods
   2.2 Biases in judgment
   2.3 Choice under risk and uncertainty
   2.4 Choice anomalies
3. Constructive reactions
   3.1 Prospect theory
   3.2 Connessionism and neural networks
   3.3 Artificial intelligence
4. Neuroeconomics
   4.1 Definitions and tools
   4.2 Intertemporal Choice and self-control
   4.3 Decision-making under risk and uncertainty
   4.4 Mirror neurons and strategic choice
   4.5 Trust game and brain activation
   4.6 Oxytocin and trusting behavior

Main References

Arthur, W. Brian, "Designing Economic Agents that Act like Human Agents: A Behavioral Approach to Bounded Rationality", *The American Economic Review ,* 81, 1991 pp. 353-391.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. (1997). "Deciding advantageously before knowing the advantageous strategy", *Science, 275,* 1293–1295.

Camerer, Colin F., *Behavioral Game Theory. Experiments in Strategic Interaction*, New York and Princeton University Press, Princeton, 2003.

Camerer, Colin - Loewenstein, George, "Behavioral Economics: Past, Present and Future", 2002.

Camerer, Colin - Loewenstein, George- Prelec, Drazen, "Neuroeconomics: How Neuroeconomics can inform economics", 2003.

Davis, Douglas D. - Holt, Charles A., *Experimental Economics*, Princeton University Press, Princeton, 1993.

Friedman, Daniel - Cassar, Alessandra. *Economics Lab. An intensive course in experimental economics*, Routledge, London and New York, 2004.

Guala, Francesco - Motterlini, Matteo, "L'economia cognitiva e sperimentale", Università di Trento.

Jones, Martin - Sudgen, Robert, "Positive confirmation bias in the acquisition of information",*Theory and Decision*, 50, 2001, 59-99.

Kagel, John H. - Roth, Alvin E., *The Handbook of Experimental Economics*, Princeton University Press, Princeton, 1997.

Kahneman, Daniel - Tversky, Amos, "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, 47, 1979, 263-91.

Kenning, Peter – Plassman, Hilke, "Neuroeconomics", University of Munster, 2005

Kosfeld, Michael - Heinrichs, Markus - Zak, Paul J. – Fischbacher, Urs – Fehr, Ernst, "Oxytocin increases trust in humans" *Nature*, 2005.

McCabe, Kevin - Houser, Daniel - Ryan, Lee - Smith, Vernon – Trouard, Theodore, "A functional imaging study of cooperation in two-person reciprocal exchange" *Proc. Nat. Ac. Sci. USA* 2001.

Rustichini, Aldo, "Neuroeconomics: Present and Future", *Games and Economic Behavior*, 52, 2005, 201-212.

Smith, Vernon, "Economics in the Laboratory", *Journal of Economic Perspectives*, 8, 1994, 113-131.

van Dijk, Eric - van Knippenberg, Daan, "Buying and selling exchange goods: Loss aversion and the endowment effect", *Journal of Economic Psychology*, 17 (1996) 517-524.

1. COGNITIVE AND BEHAVIOURAL ECONOMICS

- BEHAVIOURAL OR COGNITIVE ECONOMICS

DEFINITIONS

Cognitive economics is not a distinct subfield of economics but a school of thought based on the idea that the study of economic behaviour has to be founded on the interdisciplinary approach characterizing cognitive sciences

According to a well known definition, the field of Cognitive Science is formed by the intersection of a variety of different disciplines including **cognitive psychology**, **philosophy of mind**, **linguistics**, **artificial intelligence** and **neuroscience**

Behavioural Economics is a discipline adds to the other ones because it takes into cognitive science a legacy of specific tools and theories

"Because economics is the science of how resources are allocated by individuals and by collective institutions like firms and markets, the psychology of individual behavior should underlie and inform economics, much as physics informs chemistry; archaeology informs anthropology; or neuroscience informs cognitive psychology. However, economists routinely—and proudly—use models that are grossly inconsistent with findings from psychology. A recent approach, "behavioral economics," seeks to use psychology to inform economics, while maintaining the emphases on mathematical structure and explanation of field data that distinguish economics from other social sciences" (Camerer 1999)

Behavioural economics would be a reunification of psychology and economics and it would preserve the distinctive emphasis on formal models and descriptive statistics that characterizes mainstream economics

In fact, there are two key issues that contemporary economics has to deal with:

- first, the inconsistency of the predictions of most economic models with the experimental results of psychology;
- secondly, the rigidity of mathematical structure of that same models joined with the indefiniteness of the theoretical implications of the statistical data collected in the field

METHODOLOGY

A prevalent view among behavioural economists - but not endorsed here - is that behavioural economics emerges as the study of deviations from the paradigm of rational choice.

"Behavioral economics applies models of systematic imperfections in human rationality, to the study and engineering of organizations, markets and policy. These imperfections include limits on rationality, willpower and self-interest and any other behavior resulting from an evolved brain with limited attention." (Camerer 2006)

Behavioural economics would be the result of relaxing the assumption of perfect rationality that pervades mainstream economics.

A different view is that cognitive economics is characterized by a specific methodological approach to the study of economic behaviour.

If the field of cognitive economics is, almost by definition, the analysis of the mental and cognitive processes through which the economic agent collects, processes, interprets and uses information and knowledge to make economic choices, it assumes the role of *trait d'union* between economics and psychology.

Its main object is **to open the black-box** containing all the processes through which preferences are formed and are translated into choices.

In this light cognitive economics is different from behavioural economics, whose methodology is based on the analysis of the effectively exhibited behaviours.

The behaviourism is consonant with the axiom of revealed preferences which allows ignoring any psychological determinant of behaviour in economics

Behavioural economics approach is a clear departure from the "as if" approach endorsed by Milton Friedman.

"F-twist" argument combines two criteria:

1. Theories should be judged by the accuracy of their predictions;
2. Theories should not be judged by the accuracy of their assumptions.

Because theories with patently false assumptions can make surprisingly accurate predictions, economic theories that assume that individual agents are highly rational and wilful, judge probabilities accurately, and maximize their own wealth might prove useful, even though psychology shows that those assumptions are systematically false.

The F twist allowed economists to ignore psychology.

The empirically-driven approach to behavioural economics agrees with criterion (1) and rejects criterion (2).

In fact, criterion 2 is rejected *because of* the primacy of criterion 1, based on the belief that replacing unrealistic assumptions with more psychologically realistic ones should lead to better predictions.

NORMATIVE PURPOSES

Policy consequences: Because rational people make few mistakes, policies aren't necessary to help them.

Relaxing rationality assumptions therefore permits reasoned argument about how people can be helped.

For example, if people weight the future hyperbolically rather than exponentially, they will impulsively buy goods they will soon regret having bought.

A good policy to help those who weight the future hyperbolically is a mandatory "cooling off" period that permits "hot" consumers to renege on purchase decisions for a short period of time, such as 3 days. Cooling-off policies exemplify "conservative paternalism" — they will do much good for people who act impulsively and cause very little harm (an unnecessary 3-day wait) for those who do not act impulsively; thus, even conservatives who resist state intervention should find them appealing.

2. EXPERIMENTAL ECONOMICS

POINTS OF VIEW

PROS

"Would it not be better to leave laboratory experiments to psychologists who are trained to run them properly? Nobody doubts that we have a great deal to learn from psychologists about laboratory technique and learning theory and learning theory, but recent history would nevertheless suggest that the answer is a resounding *no*. Our comparative advantage as economists is that we not only understand the formal statements of economic theory, but we are also sensitive to the economic environments and institutions within which the assumptions from which such statements are deduced are likely to be valid. Just as chemists know not to mix reagents in dirty test tubes, so we know that there is no point in testing economic propositions in circumstances to which they should not reasonably be expected to apply."

(Binmore 1999)

"Once models, as opposed to economies, became the focus of research the simplicity of an experiment and perhaps even the absence of features of more complicated economies became an asset. The experiment should be judged by the lessons it teaches about theory and not by its similarity with what nature might happen to have created."

(Plott 1991)

CONS

1- experimental situations often project a gamelike atmosphere in which a 'subject' may see himself as 'matching wits' against the experimenter

2- experimental subjects are cast in roles and they can act in accordance with his (mis)perceptions of these roles

3- experiments have too short horizons (real world lasts many years and many trials)

4- human beings are capable to control their behavior through the implementation of abstract rules

(Cross 1994)

- The laboratory is not a socially neutral context, but is itself an institution with its own formal or informal, explicit or tacit, rules
- Human agency takes place within a socio-economic world that is structured in the sense that it consists of internally-related positions and systems
- Experimentation in economics is likely to be of limited value, save for situations – such as auctions – that exist in conditions of relative isolation and are characterised by low internal complexity

(Siakantaris 2000)

2.1 LABORATORY METHODS

DATA SOURCES

WHERE? \ HOW?	Happenstance (uncontrolled conditions - ongoing processes)	Experimental (controlled conditions - deliberately created)
Field (naturally occurring environment)	National Accounts Commodity Prices	Income Maintenance Experiments Field Experiments
Laboratory (artificial environment)	Casual Processes in the Lab Discovery of Penicillin	Choice Experiments Auctions Simulation Laboratory Asset Markets

EXPERIMENTAL ECONOMICS
LABORATORY EXPERIMENTS
(artificial environment) (controlled *ad hoc* conditions)

PURPOSES OF THE EXPERIMENT (WHY?)

(Davis-Holt, *Experimental Economics* 1994)

1) Test of Behavioral Hypotheses.
   by constructing a laboratory environment that satisfies as many of the **structural** assumptions of a particular theory, it is possible to verify its **behavioral** implications

2) Theory Stress Tests
   to examine the **sensitivity** of a theory to violations of obviously **unrealistic** assumptions

3) Searching for Empirical Regularities
   heuristic experiments to **discover** and document stylized facts

(Roth 1986)

a) Speaking to Theorists

b) Searching for Facts

c) Whispering in the Ears of Princes

EXPERIMENTAL METHODOLOGY (HOW?)

A) PROCEDURAL REGULARITY
to permit replications that the researcher and observers would accept as being valid
- instructions
- subject pool and methods of recruiting subjects
- experimental physical environment
- computerized or manual

B) MOTIVATION
Induced-value theory: use of a reward medium allows to induce prespecified characteristics in experimental subjects and to make subjects' innate characteristics largely irrelevant
- monotonicity: subjects prefer more reward medium to less and not become satiated
- salience: rewards are explicitly and unambiguously connected to the decisions made
- dominance: changes in subjects' utility from the experiment come mainly from the reward medium and other subjective costs or benefits are rendered negligible by comparison, i.e. others' reward

C) UNBIASEDNESS
Experiments should be conducted in a manner that does not lead participants to perceive any particular behavioural pattern as being correct or expected, unless explicit suggestion is a treatment variable - double blind setting

D) CALIBRATION
The design has to pre-specify and to cleanly separate the experimental predictions of alternative theories.

E) DESIGN PARALLELISM
Results established in the lab hold in other, especially non-lab, real-world situations where similar ceteris paribus conditions hold

Do not replicate in the lab the complexities of a field environment (which has infinite details) or the precise assumptions of a formal model (which usually leave out details)

Vernon Smith's parallelism precept (1982) "Propositions about the behavior of individuals and the performance of institutions that have been tested in laboratory microeconomics apply also to nonlaboratory micro economies where similar ceteris paribus conditions hold" - presumption of external validity

Charles Plott (1982): "While laboratory processes are simple in comparison to naturally occurring processes, they are real processes in the sense that real people participate for real and substantial profits and follow real rules in doing so. It is precisely because they are real they are interesting"

PROFESSIONAL SUBJECTS OR STUDENTS?

Main Subjects pool: undergraduate or MBA students

Advantages

1. readily accessible
2. low opportunity costs
3. steep learning curve
4. they do not know much about experimenter's hypothesis

PhD students: unreliable subjects because they get interested in what are you doing and respond to their understanding of your topic rather than to incentives you have constructed

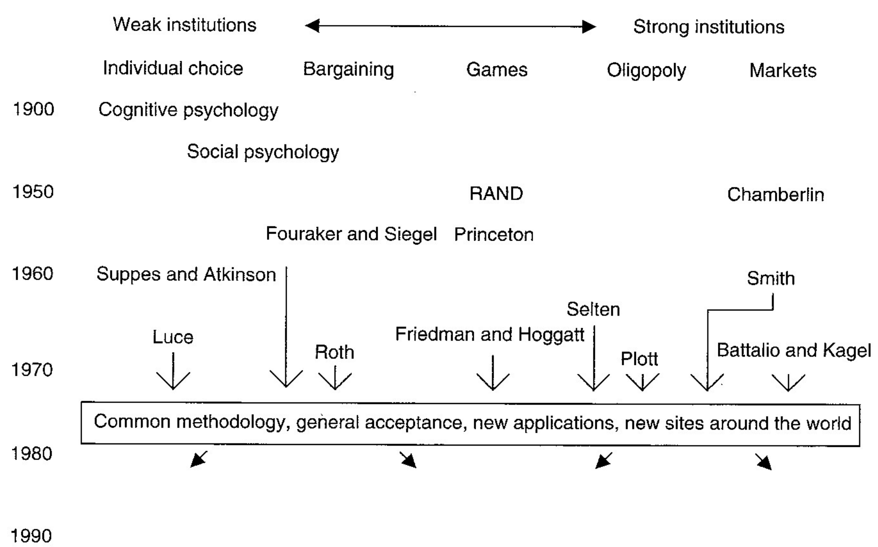
Class students or friends: dominance or salience at risk, conflicts between personal, teaching and scientific aims

Professional subjects: comparisons show that students are more adept at maximizing their profits and learning in the lab – high opportunity costs – prespecified and innate characteristics are too strong: when involved in laboratory markets they attempt to apply rules of thumb, which, valuable for dealing with uncertainty in the parallel natural market, are meaningless guides in the lab.

Burns (1985): professional wool buyers and students in a progressive auction (professionals apply familiar rules and not adjust to design requirements)
Dyer, Kagel, and Levin (1985): bidding behavior of students and construction workers (no difference)
Dejong et al (1988): Businessmen and students in sealed-offer markets (some profits, but higher variance for businessmen)

What about gender, age, risk attitude, experience?

A bit of history



Kagel, John H. - Roth, Alvin E.
The Handbook of Experimental Economics
Princeton University Press, Princeton, 1997

INDEX

a) public goods (Ledyard)
- cooperation vs. selfishness (social dilemmas, free-riding, institutions)
- what improves cooperation (thresholds, learning)

b) coordination problems (Ochs)
- experiments with overlapping generations
- coordination games with Pareto ranked equilibria
- decentralized matching environments

c) bargaining experiments (Roth)
- agreements
- causes of disagreements and costly delays
- bargaining protocol and preplay communications

d) industrial organization (Holt)
- trading institutions centralized and decentralized
- monopoly regulation and potential entry
- market structure and market power
- collusion factors
- product differentiation and multiple markets

e) experimental asset markets (Sunder)
- informational efficiency of markets
- state-contingent claims and bubbles
- learning and dynamics of adjustment paths
- investment and public policy

f) auctions (Kagel)
- symmetric independent private-values models
- common value auctions
- collusion

g) individual choice behavior

INDIVIDUAL CHOICE BEHAVIOR

I. Judgment

- A. Calibration
    - 1. Scoring Rules
    - 2. Confidence Intervals
- B. Perception and Memory Biases
- C. Bayesian Updating and Representativeness
    - 1. Underweighting of Base Rates
    - 2. Underweighting on Likelihood Information (Conservatism)
    - 3. The Law of Small Numbers and Misperceptions of Randomness
- D. Confirmation Bias and Obstacles to Learning
- E. Expectations Formation
- F. Iterated Expectations and the Curse of Knowledge
    - 1. False Consensus and Hindsight Bias
    - 2. Curse of Knowledge
- G. The Illusion of Control

II. Choice under Risk and Uncertainty

- A. Mounting Evidence of Expected Utility Violation (1965-1986)
    - 1. The Allais Paradoxes
    - 2. Process Violations
    - 3. Prospect Theory
    - 4. Elicitation Biases
- B. Generalizations of Expected Utility and Recent Tests
    - 1. Predictions of Generalized EU Theories
    - 2. Empirical Studies Using Pairwise Choices
    - 3. Empirical Studies Measuring Indifference Curves
    - 4. Empirical Studies Fitting Functions to Individuals
    - 5. Cross-Species Robustness: Experiments with Animals
- C. Subjective Expected Utility
    - 1. The Ellsberg Paradox
    - 2. Conceptions of Ambiguity
- D. Choice over Time
- E. Description Invariance
    - 1. Framing Effects
    - 2. Lottery Correlation, Regret, and Display Effects
    - 3. Compound Lottery Reduction
- F. Procedure Invariance
    - 1. New Evidence of Preference Reversal
    - 2. Arbitrage and Incentives
    - 3. Reversals and Markets
    - 4. Social Comparisons and Reversals
- G. Endowment Effects and Buying-Selling Price Gaps
    - 1. Market Experiments
    - 2. Endowment Effects: Some Psychology and Implications
- K. Search
    - 1. Search for Wages and Prices
    - 2. Search for Information

2.2 BIASES IN JUDGMENT

"People rely on heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors" (Tversky and Kahneman 1974)

- CONFIRMATION BIAS

Once individuals devise a strong hypothesis they will tend to misinterpret or even misread new information unfavourable to this hypothesis

Also production of treatment effects: when a researcher believes a hypothesis is true, he often produces a biased sample of evidence that reinforces his or her belief (unconsciously?)

Consequence is obvious: confirmation bias inhibit learning whether one's underlying belief is false

But also

Fresh thinkers may be better at seeing solutions to problems than people who have meditated at length on the problems, because the fresh thinkers are not overwhelmed by the "interference" of old hypotheses.

Correlated phenomena

- FALSE CONSENSUS: People use their own tastes and beliefs as information in guessing what others like and believe - Application: to put in other people's shoes is not useful to find focal points

- HINDSIGHT BIAS: current recollections of past judgments tend to be biased by what actually happened since then – Application: adaptive expectations vs. rational expectations

AN ILLUSTRATIVE EXPERIMENT

Martin Jones and Robert Sugden "Positive confirmation bias in the acquisition of information", *Theory and Decision*, 50, 2001, 59-99

- **Positive confirmation bias**: tendency, when testing an existing belief, to search for evidence which could confirm that belief, rather than for evidence which could disconfirm it

- Application to economic learning: an agent who repeatedly faces the same set of options might retain the false belief that a particular option was optimal, even after long exposure to evidence which, rationally interpreted, would indicate the contrary

Wason's (1968) *selection task*

Four double-sided cards

Subjects are told that each card has a letter on one side and a number on the other, but they can see only the upper faces of the four cards

Four cards show 'A', 'D', '4' and '7'

Each subject is asked to consider the following *rule*, as applied to the four cards: 'If a card has a vowel on one side, then it has an even number on the other side'

Instruction: 'Your task is to say which of the cards you need to turn over to find out whether the rule is true or false'

Two most common responses are the 'A' card alone, and the 'A' and '4' cards in combination

The correct answer to the question posed is, of course, the combination of 'A' and '7'

The frequently-chosen '4' card can provide no information which is relevant to the issue of whether the rule is true or false

The 'A' and '4' cards are the ones that are capable of providing evidence which *confirms* the rule: by turning over either of these cards, the subject may find a card with a vowel on one side and an even number on the other

In contrast, the '7' card can only *disconfirm* the rule (i.e. by revealing a card which has a vowel on one side but not an even number on the other)

In this sense, the evidence from the selection task can be interpreted as consistent with positive confirmation bias

Criticism: The original selection task was formulated in highly abstract terms

Counterargument: Correct response might be facilitated by adding *thematic* content to the task, i.e. by providing a *cover story* which accounts for the statement and gives some point to the selection task

Jones and Sudgen's design

- Subjects have to pay a fixed *cost* per card turned over
- After she has made this choice, the cards she has chosen are turned over
- She then makes the *judgment* that the statement is 'true' or 'false'
- Finally the remaining cards are turned over and she receives a fixed *reward* if and only if her judgment was in fact correct

Experiment carried out at the University of East Anglia in Norwich
120 students recruited on the campus (wide range of courses)
Computerized experiment
No communication between subjects

Each task is presented by means of a sequence of six screens

The screen presents first the cover story, then the statement and finally four cards to choose

Each object has two characteristics, each of which can take one of two values that correspond with p, ¬p, q, and ¬q (as before vowel and consonant, even and odd)

Each subject perform seven different tasks

<p, ¬q> or <¬q, p>, if turned over, is a *disconfirmation* of the experimental HP
<p> <p, q> and <q, p> are *confirmations*

Exemplificative Tasks

1. Relatives
A survey is taken of 100 people in Los Angeles, Seattle, London and Norwich who have relatives living in other cities. Each person in the survey living in Britain has relatives in Los Angeles or Seattle and each person living in America has relatives in Norwich or London. No one has relatives in more than one city. The details of the survey are written down on report cards by putting the city each person lives in on one side of the card and the city their relatives live in on the other side. A sample of four report cards is selected. Look at whichever cards you wish to test the statement:
[Standard statement] Every person in the sample who lives in London also has a relative who lives in Los Angeles.
[Contraposed statement] Every person in the sample who lives in Seattle also has a relative who lives in Norwich.

2. Drinkers
Only people over the age of eighteen are allowed to drink alcohol in a pub in Britain. A survey is carried out of 100 people in a large public house which identifies their age and whether they are drinking alcohol or a soft drink. Each person's details are put down on a report card with the person's age on one side and their drinking behaviour on the other. A sample of four report cards is selected. To find out if the four people in the sample are obeying the law, look at whichever cards you wish to test the statement:
[Standard statement] Every person in the sample who is drinking alcohol is also over eighteen.
[Contraposed statement] Every person in the sample who is under eighteen is also drinking a soft drink.

Results
In favour of the confirming bias hypothesis: 62% of the choices (445/720)
<No cards> 18%
<p> 14%
<p, q> 18%

Conclusions
Overwhelming evidence that subjects' information-gathering decisions are systematically biased in favour of information which is potentially confirming

But behaviour seems to have been closer to Bayesian rationality than in many other selection task experiments

Especially the drinkers story facilitates Bayesian rationality (why?)

What is the effect of financial incentives?

2.3 CHOICE UNDER RISK AND UNCERTAINTY

What do we mean by rational choice in economics? Lots of formulations, involving assumptions of different strength

Different forms of rationality imply different experiments to test them

- Goal oriented
- Satisficing behaviour
- Maximizing behaviour
- Ordinal utility maximization
- Expected utility maximization
- Subjective expected utility maximization

Experimental economics reveals the hidden or implicit assumption by showing anomalies in the formulation of rationality

Razionalità in un contesto economico:

> gli agenti economici utilizzano l'informazione disponibile in modo da operare la scelta ottimale date le alternative disponibili e gli obiettivi prefissati.

Approccio più usato: massimizzazione dell'utilità (soggettiva) attesa in condizioni di rischio (von Neumann – Morgenstern, 1947)

Quale azione è razionale in condizioni di rischio?

Il profilo di un agente razionale è definito per mezzo degli *assiomi dell'utilità attesa*.

Tre assiomi o principi fondamentali

- assioma di *ordinamento* (A1),
- assioma di *continuità* (A2),
- assioma di *indipendenza* (A3).

X, y, z, w,…	i risultati (detti anche "conseguenze" o "stati del mondo") di una lotteria o prospetto probabilistico;
f	la relazione di preferenza
p	valore di probabilità.
$\Rightarrow$	"se… allora"
(2)	"se e soltanto se"
$\neg$	"non"

(A1) f è una relazione d'ordine:
$(x \; f \; y) \Rightarrow \neg(y \; f \; x)$ [asimmetria]
$(x \; f \; y \;\&\; y \; f \; z) \Rightarrow (x \; f \; z)$ [transitività]

(A2) $(x \; f \; y \; f \; z) \Leftrightarrow [px + (1-p)z \; f \; y \; f \; qy + (1-q)z\,]$
per p e q strettamente fra 0 e 1 [continuità]

(A3) Per qualsiasi p tale che $0 < p = 1$,
$(x \text{ f } y) \Leftrightarrow [px + (1 - p)z] \text{ f } [py + (1 - p)z]$ [indipendenza].

\+ assiomi classici del calcolo della probabilità

Teorema di rappresentazione dell'utilità attesa

se una relazione di ordinamento (f) soddisfa (A1), (A2), (A3), allora esiste una funzione reale di utilità U(.) (definita sui risultati delle lotterie) tale che per tutte le lotterie X e Y:

(1) $X \text{ f } Y \Leftrightarrow EU(X) \text{ f } EU(Y)$,

dove l'utilità attesa *EU* (*Expected Utility*) è data dalla somma delle utilità moltiplicate per le probabilità dei risultati di una lotteria:

(2) $EU = \Sigma_i \, p_i U(x_i)$

quindi in economia un agente è 'razionale' se massimizza *EU*.

Il modello di von Neumann e Morgenstern è applicabile a situazioni di *rischio* e può essere esteso (Savage 1954) alle situazioni di *incertezza*, in cui le probabilità rappresentano gradi di credenza (*beliefs*) individuali, ("teoria dell'utilità attesa soggettiva")

Economia neoclassica come scienza del comportamento razionale.

Egoismo: gli agenti economici massimizzano la propria utilità e sono indifferenti riguardo a quella altrui;

materialismo: l'utilità degli agenti economici dipende soltanto dalla quantità di beni consumati;

utilità decrescente al margine: l'utilità cresce col numero di beni (a più beni corrisponde più utilità che a meno beni) ma diminuisce al margine (al consumo del bene *n*+1 corrisponde meno utilità che al bene *n*)

2.4 CHOICE ANOMALIES

EXPERIMENTS WITHIN SUBJECTS
[*Source:* Tversky and Kahneman 1981, Thaler 1980]

Experiment 1 (certainty effect) Which of the following options do you prefer?
A. A sure win of $30 [78%] EV 30
B. An 80% chance to win $45 [22%] 36
Which of the following options do you prefer?
C. A 25% chance to win $30 [42%] EV 7.5
D. A 20% chance to win $45 [58%] 9

$0{,}20\times U(45) > 0{,}25\times U(30) \Rightarrow U(45)/U(30) > 0{,}25/0{,}20$
$0{,}80\times U(45) < 1\times U(30) \Rightarrow U(45)/U(30) < 1/0{,}80$ but $0{,}25/0{,}20 = 1/0{,}80$

Experiment 2 (loss aversion) Imagine that you face the following pair of concurrent decisions. First examine both decisions; then indicate the options you prefer:
Decision (i). Choose between
A. A sure gain of $240 [84%] EV +240
B. 25% chance to gain $1,000 and 75% chance to lose nothing [16%] +250
Decision (ii). Choose between
C. A sure loss of $750 [13%] -750
D. 75% chance to lose $1,000 and 25% chance to lose nothing [87%] -750

Experiment 3 (mental accounting) Choose between
E. 25% chance to win $240 and 75% chance to lose $760 [0%] -510
F. 25% chance to win $250 and 75% chance to lose $750 [100%] -500
But E = A&D and F = B&C

Experiment 4 (shoes costs) Imagine that you are about to purchase a jacket for ($125)[$15] and a calculator for ($15)[$125]. The calculator salesman informs you that the calculator you wish to buy is on sale for ($10)[$120] at the other branch of the store, a 20-minute drive away. Would you make the trip to the other store?
Yes: 16% No: 84%

Experiment 5 (sunk costs)
Imagine that you have decided to see a play, admission to which is $10 per ticket. As you enter the theater you discover that you have lost a $10 bill. Would you still pay $10 for the ticket to the play? [Yes: 88% No: 12%] Now imagine that you have decided to see a play and paid the admission price of $10 per ticket. As you enter the theatre you discover that you have lost your ticket. The seat was not marked and the ticket cannot be recovered. Would you pay $10 for another ticket? [Yes: 46% No: 54%]

BUYING-SELLING PRICE GAP

A simple class experiment

- Half of you - randomly chosen - is named as "owners" and receive a windfall gift of a classy, stylish, desirable HBS pencil. You are asked to examine it closely. It is yours to keep, or to sell
- The remaining half do not receive a pencil and is refereed to as "non-owners"
- Then each owner is asked to pass his/her pencil to a neighbouring non-owner, so that the non-owners can also fully examine the pencil.
- It may exist some possible gains from trade. In order to assess this, the experimenter wants to elicit from each owner the minimum price at which he/she would be willing to sell the pencil and from each non-owner, the maximum price she/he would be willing to pay to buy the pencil.
- Experimental result:
  Owner prices (WTA) > Non-Owner prices (WTP)

Economic theory predicts that the prices a person will pay to buy and sell an object should be the same.

Environmental economists in the 1970s first discovered that this is not true: duck hunters would pay $ 247 to maintain a wetland suitable for ducks but asked $ 1,044 to give up the wetland (Hammack J. and Brown G. M. *Water fowl and wet lands: Toward bio economic analysis*, John Hopkins University Press, 1974)

Students were willing to pay 2.75 on average for college mugs but they asked for 5.25 to sell their mugs (Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler, "Experimental Tests of the Endowment Effect and the Coase Theorem," *JPE* 1990)

EXPLANATIONS

Plott's (1996) discovered preferences hp.: individuals may discover their valuations for unfamiliar items during the elicitation process

Economic factors: income effects and substitution, transaction costs, implied value of the good, profit motivation

Psychological factors: **endowment effect**, legitimacy, ambiguity and moral responsibility

- ENDOWMENT EFFECT

People prefer the things they own, ceteris paribus

(but what about the neighbour's grass is always greener than yours?)

Explanations

- action error (Ritov-Baron 1991): fear of action errors is a bias in favour of inaction

- higher sensitivity to overpaying (out-of-pocket costs) than to selling too cheaply (opportunity costs) (Thaler 1980)

- disposition effect (Weber-Camerer 1992): reluctance to take actions leading to irreversible loss and eagerness to take actions creating gains
  Ex: the volume of houses sold falls when housing prices fall

- status quo bias (Samuelson-Zeckhauser 1988): if you have a current choice you enhance preferences for it

- prospect theory's loss aversion (Tversky-Kahneman 1988): losses are more painful than equally sized gains are pleasurable

- action is different from giving advices: no endowment effect when people advise others (Marshall-Knetsch-Sinden 1988)

Consequences

Invalidates the Coase theorem: the valuation of a property right is not independent of who owns the right – contracting parties allocate efficiently rights and duties if there is no transaction cost

AN ILLUSTRATIVE EXPERIMENT

Eric van Dijk - Daan van Knippenberg, "Buying and selling exchange goods: Loss aversion and the endowment effect", *Journal of Economic Psychology*, 17 (1996) 517-524

The object of the experiment:
to refute Kahneman and Tversky's theory that the endowment effect is not particularly common in markets where goods are specifically bought for exchange

Experimental design

Sixty-six undergraduate students (35 females; 31 males) randomly assigned to the different experimental conditions and paid for their participation.

Half of the participants received a bargaining chip representing money to be exchanged for money at the end of the experiment.

Two treatments:

1) Fixed Exchange Value conditions: participants learned they could exchange the chip with the experimenter for Dfl. 3.50 (1 Dutch Guilder = $.55 US)

2) Uncertain Exchange Value conditions, participants learned they could exchange the chip for an amount of money between Dfl. 1.75 and Dfl. 5.25, depending on a chance procedure.

Participants could trade the bargaining chips among themselves: participants owning a chip (the Sellers) could sell this chip to participants not owning a chip (the Buyers).

On a separate form, prices were listed from Dfl. 0.25 to Dfl. 6.75 (with Dfl. 0.25 intervals).

Sellers were requested to indicate for each price whether or not they would sell at that price.

Buyers indicated for each price whether or not they would buy at that price.

The experimenter would randomly select a price on this form, thus establishing the 'market price' for the chip (procedure to prevent participants from misstating their true values – subjects had to stick to their stated intention)

After the experimenter collected the forms, participants estimated the value of the bargaining chip.

At this point the experiment was ended. Participants were debriefed and all received Dfl. 5.00

RESULTS

The selling price of the sellers (mean = Dfl. 3.76) exceeded the buying price of the buyers (mean = Dfl. 3.05).

As predicted, this main effect was qualified by a significant Position X Uncertainty interaction ($F(1,62) = 4.1$, $p < 0.05$).

Corroborating the findings of Kahneman et al. (1990), no significant endowment effect was observed when the value of the exchange good was fixed ($F(1,62)= 1.2$, $p < 0.3$; overall mean = Dfl. 3.40; mean for sellers = Dfl. 3.56; mean for buyers = Dfl. 3.25).

In agreement with our hypothesis, in the case of an uncertain exchange rate, however, the selling price (mean = Dfl. 3.97) significantly exceeded the buying price (mean = Dfl. 2.87; $F(1,62) = 15.7$, $p < 0.0001$).

These results indicate that exchange goods may, like consumption goods, be susceptible to the endowment effect, provided that exchange rates are uncertain.

IMPLICATIONS

endowment effect on consumption goods (e.g., mugs, chocolate bars) in situations where it may be difficult to compute the net gains and losses of trade: if someone wants to buy your chocolate bar you may perceive giving up the chocolate bar as a loss.

But if someone offered you one and a half chocolate bars for your chocolate bar it is easy to compute the net gains of trade and you would probably not be susceptible to the endowment effect.

Research should not only focus on what is being traded (e.g., exchange goods or consumption goods) but on *what is being traded for what.*

Insights related to the characteristics approach of Lancaster (1971), i.e. goods as bundles of characteristics

Lancaster's theory: goods are more substitutable the more characteristics they have in common

In a similar vein, people would be less subject to the endowment effect the more characteristics the objects traded have in common because it is easier to compute the net gains and losses of trade.

Future research : to investigate the relation between substitutability and the endowment effect by comparing the willingness to trade for goods varying in the number of common characteristics

- PREFERENCE REVERSAL

Prices subjects gave for bets are highly correlated with bet payoffs but choices are more highly correlated with probabilities

Slovic-Lichtenstein 1968

Subjects were offered two bets with the same expected values:

p-bet with high probability and low payoff
$-bet with low probability and high payoff

Subjects choose the *p-bet*, but when asked to state the lowest price at which they would be willing to sell each gamble if they owned it, they put a higher price on the *$-bet*

p-bet 8/9% to win $4
$-bet 1/9% to win $40
Choice *p-bet* [71%]
Price *$-bet* > Price *p-bet* [67%]

Standard analysis of choice assumes *procedure invariance*:
A is preferred to B if A is selected when B is available or if A has a higher reservation price than B. These two procedures have to give rise to the same ordering

In the laboratory, different methods of eliciting preference often give rise to systematically different orderings

Explanations

- violations of transitivity

- violations of procedure invariance (pricing is different from choosing)

- violations of the independence axiom

BDM procedure (Becker, De Groot, Marschak 1964)

After the subject states a selling price for a gamble, an offer is generated by some random process. The subject receives the offer if it exceeds the stated selling price, and plays the gamble if the stated price exceeds the offer. The price stated by the subject, therefore, serves only to determine whether the subject will play the bet or receive the cash, but it does not determine the actual amount. If the subject is an expected utility maximizer, this procedure is incentive compatible: the decision maker has no incentive to state a selling price that departs from his or her actual cash equivalent

AN ILLUSTRATIVE EXPERIMENT

The compatibility hypotheses (Slovic, Griffin and Tversky 1990)

stimulus-response compatibility: the weight of a stimulus attribute in judgment or in choice is enhanced by its compatibility with the response scale

- If there is no compatibility effort and error may reduce the impact of the stimulus
- A response mode tends to focus attention on the compatible features of the stimulus

Experimental design

- Subjects were given two pieces of information about each of 12 large companies, i.e., the company's 1986 **market value** and the company's **rank** among the Top 100 with respect to 1987 profits
- Half of the subjects asked to predict the 1987 market value in billions of dollars whereas the other half were asked to predict the company's rank with respect to its 1987 market value
- Each subject had one predictor measured on the same scale (that is, money or rank) as the dependent variable and one predictor measured on a different scale

Results

- As implied by compatibility each predictor was given more weight when the predicted variable was expressed on the same scale

- As a consequence, the relative weight of the 1986 market value was twice as high for those who predicted in dollars than for those who predicted the corresponding rank

- This effect produced many reversal in which one company was ranked above another but the order of their predicted value was reversed

Other experimental design

- If preference reversal are due primarily to the compatibility of prices and payoffs both expressed in dollars, their incidence should be substantially reduced by the use of nonmonetary outcomes, i.e. a one-week pass for all movie-theatres in town or a dinner for two at a good restaurants
- Results: the prevalence of preference reversal was reduced by nearly 50 percent

Conclusions
The compatibility hypothesis implies that the payoffs, which are expressed in the same units, will be weighted more heavily in pricing bets than in choosing between bets. Furthermore, since the payoffs of L bets are much larger than the payoffs of H bets, the major consequence of a compatibility bias is the overpricing of the low-probability high-payoff bets

3. CONSTRUCTIVE REACTIONS

Main finding of experimental economics: there is a variety of definitions of rational individual (what about decision makers' heterogeneity?)

Risk neutral economic man: never buys insurance, but would be willing to pay any finite amount to participate in Petersburg paradox.

Expected utility maximizing man: buys insurance, but ignores sunk costs, and is immune to framing effects.

Almost rational economic man (e.g. prospect theory man) has malleable reference points and probability perceptions, but still has preferences - comfortable with non-utility Allais choices, but doesn't do preference reversals.

Psychological man doesn't have preferences, has mental processes. Different frames and contexts, and different choice procedures elicit different processes - So he may sometimes exhibit preference reversals because choosing and pricing elicit different mental procedures.

Neurobiological man: doesn't (even) have a fixed collection of mental processes, in the sense of psychological man. He has biological and chemical processes which influence his behaviour. Different blood chemistry leads to different mental processes; e.g. depending on the level of lithium (or Valium or Prozac) in his blood, he makes different decisions (on both routine matters and matters of great consequence - even life and death). An understanding of how chemistry interacts with mental processes has proved to be very useful, for instance in treating depression.

Expected utility theory	→	prospect theory	→	asymmetric response to price increases, downward-sloping labour supply among cab drivers
Exponential discounting	→	hyperbolic discounting	→	addition and procrastination
Self-seeking behaviour	→	social utilities	→	trust and reciprocity in financial relationship
Equilibrium	→	processes of equilibration	→	drift effect, automata
ranked preferences	→	constructed preferences	→	information manipulation in horse race betting, void informational cascades in the artistic markets
Bayesian probability judgments	→	confirmation bias	→	Self-fulfilling expectations in financial markets, focal points

- Alternative theories to explain anomalies

Prospect theory (Tversky and Kahneman 1991)

Conception of rationality alternative to expected utility maximization - Machina's (1989) non expected utility, Gilboa and Schmeidler (2006) case based decision theory

- Attempts to reconcile rational theory and irrationality in experiments

it does not take much rationality to behave nearly optimally in an experimental market - Gode and Sunder's (1993) zero intelligence agents in simulated experimental markets lead to nearly efficient outcomes

artificial intelligence and connessionism

to study learning processes within the experiments – Cox and Grether (2005) endogenous "loss aversion" discovery of preference by watching others

to construct efficient experimental markets in which individual irrationality may persist but is reduced by market forces – Camerer and Weber (1990)

- Heuristic experiments : searching for new facts

heterogeneous agents models: the abandonment of the fictitious construct of representative agent

local network analysis – complex dynamic systems characterized by dispersed interaction among agents acting locally on each other in some space

empirically-driven analysis à la Schelling

neuroeconomics

3.1 PROSPECT THEORY

Kahneman and Tversky *Econometrica* 1979

Tversky and Kahneman *Journal of Risk and Uncertainty* 1992

Experimental evidence

a) people perceive the outcome of a monetary prospect in terms of the variations (positive or negative) related to a non-constant reference level (usually the *status quo*) rather than in terms of absolute levels of wealth

b) people appear to be more adverse to losses, relatively to their reference level, than how they are attracted by the winning of the same value. The disutility of the monetary loss x is lower than the utility of winning the same amount x. Consequently, reaction to losses is stronger than the reaction to winnings.

Prospect Theory postulates the existence of two functions - the value function v and the weight function (or decisions weights) p - such as the decision maker strictly prefers X a Y iff

$$\sum \pi(p_i) v(\Delta x_i) > \sum \pi(q_i) v(\Delta y_i)$$

where $\Delta x_i = x_i - x_0$ is the variation associated to a prospect x_i with respect to a reference point x_0.

Rational Choice and the Framing of Decisions S259

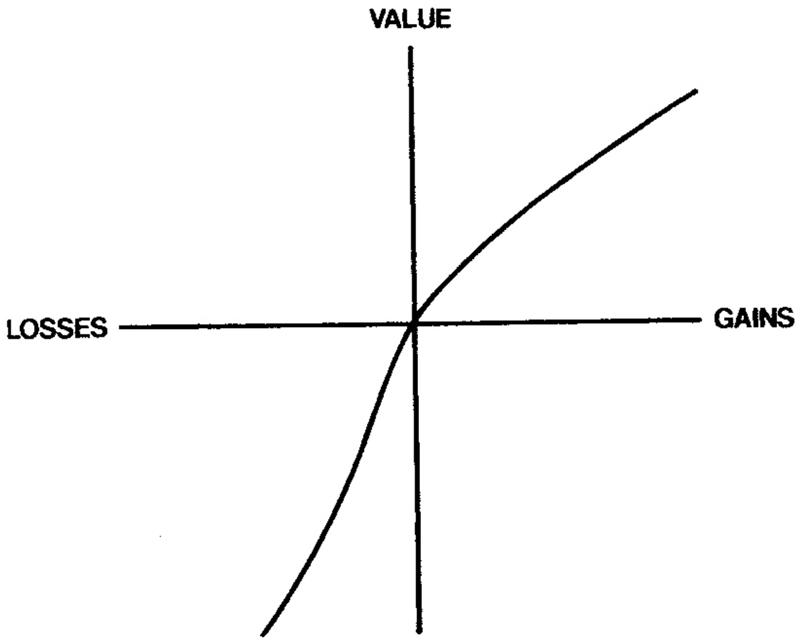


FIG. 1.—A typical value function

As implied by the value function, the majority choice is risk averse in

Differences between prospect theory (PT) and subjective expected utility theory (SUET)

1) the decision maker is not interested in the final status per sé (SUET) but at the change of status (Δx_i) with regard to the reference point (x_0) (PT)

2) the value function v is concave ("risk averse") for gains and convex ("risk seeking") for losses (PT).

3) the value function v is steeper around the reference point for losses than for gains ("loss aversion").

4) the psychological sensitivity to losses and gains diminishes marginally: incremental winnings/losses give decreasing marginal utility/disutility

5) while in SUET the utility of any possible event is weighted with his probability, in PT the value of any welfare change is multiplied by a "decision weight", that is not a probability but a probability transformation. Probability transformations do not follow probability rules and cannot be interpreted as degree of beliefs. They are obtained by choices and measure the impact of events on prospects' desirability and not the perceived probability of events.

6) the weight function p is monotone, increasing, and discontinuous between 0 and 1, because it systematically overweighs very low probabilities and under weights medium and high probabilities ("certainty effect")

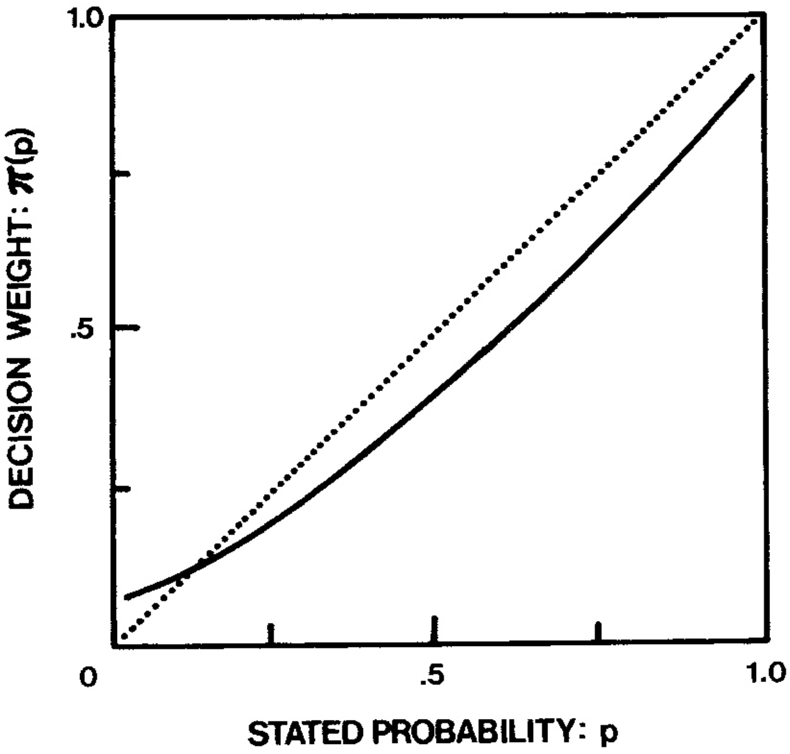


Fig. 2.—A typical weighting function

3.2 CONNESSIONISM AND NEURAL NETWORKS

Approccio computazionale allo studio della mente e del cervello

Costruzione di un modello del cervello in base all'ipotesi che esiste una relazione (isomorfismo) tra la struttura del cervello e la struttura dei processi cognitivi

Connessionismo come simulazione di sistemi complessi formati da insiemi interconnessi di semplici unità di elaborazione

Rete neurale come modello esplicativo approssimativo dei processi di rappresentazione e di apprendimento che avvengono nel cervello e quindi come formalizzazione delle concezioni connessioniste in materia di processi cognitivi

Obiettivo: formazione di rappresentazioni mentali delle azioni economicamente rilevanti e descrizione dei processi di apprendimento

Hebb (1949), *The Organization of Behavior*, New York, Wiley

- spiega il comportamento psicologico in termini di funzionamento del cervello superando il dualismo tra mente e cervello

- processi cognitivi sono il frutto di particolari modalità di connessione tra neuroni formanti catene lungo le quali viaggia un segnale elettrofisiologico

- assemblee cellulari come gruppo di neuroni tra loro interconnessi: l'eccitazione di un solo neurone si comunica a tutti gli altri del gruppo

- fenomeni come percezione, memoria, apprendimento spiegati come processi combinatori di unità elementari e non come rapporto stimolo/risposta (approccio comportamentista)

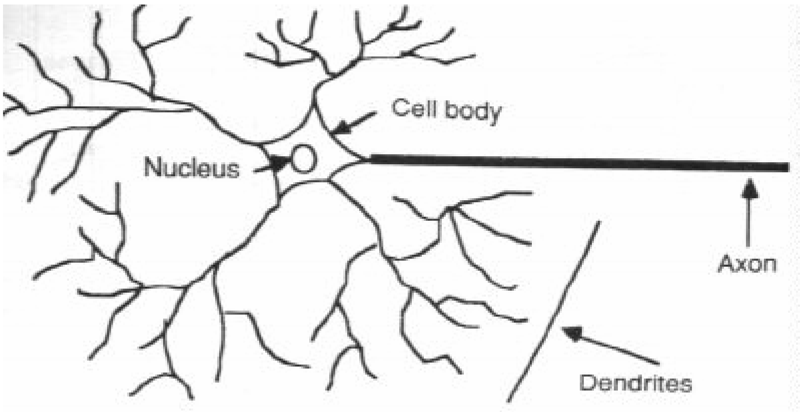
Cenni all'organizzazione sistema nervoso

Sistema nervoso come rete specifica (composta di parte aventi tra di loro legami funzionali) che funziona attraverso impulsi elettrici trasmessi attraverso cellule chiamate neuroni

La rete riceve gli input dai *recettori* (sensi) che convertono gli stimoli provenienti dal corpo e dal mondo esterno in *pattern*, e cioè configurazioni, di impulsi elettrici che trasportano le informazioni nella rete

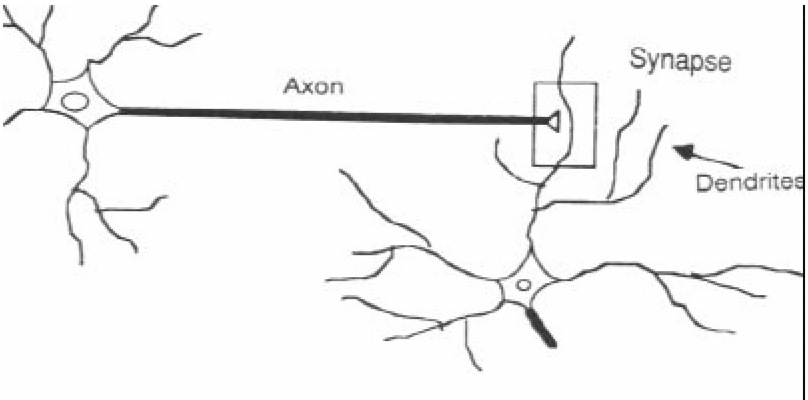
Pattern interagiscono con l'insieme di impulsi che già viaggiono nella rete neurale e provocano l'emisione di impulsi che controllano gli *effettori*, quali i muscoli, le ghiandole, nel determinare le risposte.

Neurone come cellula che contiene *il nucleo* che è portatore cromosomico del patrimonio ereditario.



Dal nucleo partono i *dendriti* (che veicolano i segnali di input) e gli *assoni* che veicolano i segnali di output

Le *sinapsi* sono i punti di contatto tra due neuroni



I nervi conducono gli impulsi dagli organi recettori agli effettori, l'impulso viaggia in un primo neurone lungo un assone e genera un impulso nel secondo neurone attraverso i dendriti.

Le comunicazioni tra neuroni avvengono mediante rilascio di sostanze chimiche dette *neurotrasmettitori*

Un *neurotrasmettitore* è una sostanza che nel sistema nervoso trasporta segnali tra neuroni attraverso le sinapsi chimiche.

I neurotrasmettitori possono essere *eccitatori* o *inibitori*, cioè possono rispettivamente promuovere la creazione di un impulso nervoso nel neurone ricevente o inibire l'impulso.

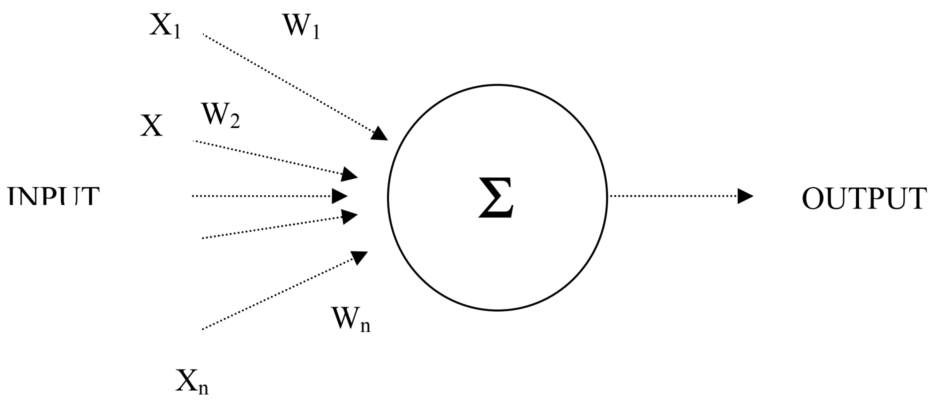
HP di partenza: le scelte economiche sono in gran parte basate su comparazioni tra beni o tipologie (classi di beni) o *patterns*

Processo di apprendimento descritto come modificazione dell'intensità (o dei pesi) delle connessioni sinaptiche, modificazioni in parte genetiche e in parte acquisite.

Problema di *pattern recognition*, categorizzazione come attività primaria attraverso il quale i soggetti danno un significato ai nuovi stimoli e identificano il significato di simboli conosciuti

Funzionamento e organizzazione delle reti neurali

Neuronodo Σ come unità elementare costitutiva della rete neurale

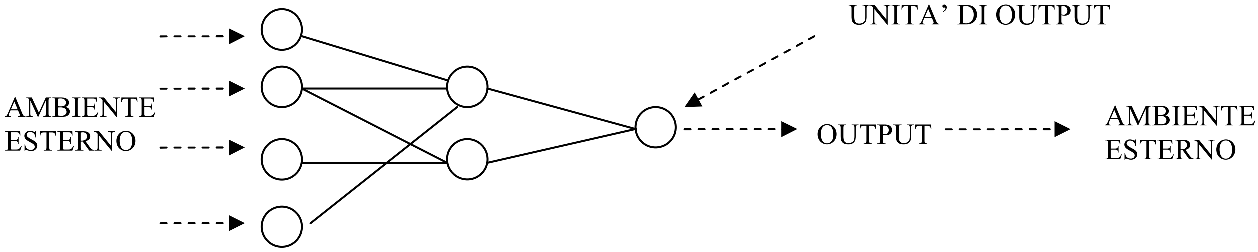


Ogni input X_i ha un certo impatto su un neuronodo Σ misurato da un peso W_i - che nei neuroni è la connessione sinaptica

L'output è pari alla somma dei valori (pesi) degli input che sono espressi vettorialmente.

Gruppi di neuronodi costituiscono una rete neurale artificiale

SCHEMA DI UNA RETE NEURALE ARTIFICIALE



Nella fase di apprendimento i pesi possono essere modificati in risposta ai vari input

Se la somma dei pesi degli input supera un valore predefinito di soglia il neuronodo si attiva

L'apprendimento secondo lo schema neurale consiste nel rafforzare alcune connessioni ed estinguerne altre

La rete elabora le informazioni accettando *patterns* e cioè configurazioni e forme in input ai nodi di input e produce una forma in output

Esempio illustrativo : Teoria del consumatore

Assunzione standard: il consumatore ha un ben definito sistema di preferenze ma può essere incerto sull'utilità da attribuire al consumo di uno o più beni

Assunzione alternativa: il consumatore apprende consumandoi in sequenza un insieme di beni in modo da testare sperimentalmente l'utilità ricavabile da essi

Hp modalità apprendimento: il consumatore costruisce un insieme di classi-tipo di beni e assegna ogni bene specifico ad una determinata **classe**

Ogni bene specifico è una combinazione di **caratteristiche**

Processo di memorizzazione per caratteristiche dei beni

Bene rappresentabile come un vettore:

$$[p_\alpha, p_\beta, \ldots, p_\lambda]\begin{bmatrix}\alpha\\ \beta\\ \\ \lambda\end{bmatrix}$$ dove $p_\alpha, p_\beta, \ldots, p_\lambda$ sono i pesi delle caratteristiche α, β, λ

Il peso di ogni caratteristica varia in funzione delle esperienze di consumo passate che hanno effetto sul valore che il consumatore attribuisce ai vari beni valutati come combinazioni di caratteristiche

Hp: consumatore opera come un *perceptron*, che è la forma più semplice di rete neurale, che impara a classificare le forme di input in categorie di appartenenza e impiega una funzione di attivazione che ha la forma seguente $f(x) = \begin{cases} 1 \\ 0 \end{cases}$

L'attivazione (1) del neuronodo si verifica se i valori del vettore di input superano il livello di soglia *v* e il consumatore giudica che il bene appartiene ad una certa classe, altrimenti il neuronodo rimane inattivo (0)

3.3 ARTIFICIAL INTELLIGENCE

W. Brian Arthur "Designing Economic Agents that Act like Human Agents: A Behavioral Approach to Bounded Rationality", *The American Economic Review* , 81, 1991 pp. 353-391

Standard approach in economics to modelling limited rationality: to lay down axioms and assumptions that suppose limits to economic agents' computational ability or memory, and investigate their consequences

Alternative approach: to develop theoretical (virtual) economic agents that act and choose in the way *actual* humans do

Agents are represented as using parametrized decision algorithms so that the agents' behaviour matches real human observed in the same decision context

Calibrated agents furnish predictions based on *actual agents* rather than idealized behaviour

Modelling gives a repertoire of calibrated algorithms to cover the various contexts that might arise and to reproduce statistically the characteristics of human choice, including the distinctive errors or departures from rationality that humans make

Application: Iterated choice under uncertainty

A decision maker chooses one of *n* possible actions at each time t

Actions give a random payoffs or profits drawn from a stationary distribution that is unknown in advance

Agent chooses one alternative at each time, observes its consequence or payoff and over time updates his choice as a result

In this setting there is a trade-off between exploitation of high-payoffs actions and exploration of seldom-tried actions that potentially may have higher payoffs

Automata's process of learning: learning as the updating of the probabilities of taking each action on the basis of the payoffs or outcomes experienced

Action *i* brings reward $\Phi(i)$, that is unknown to the agent in advance, positive and distributed randomly with a stationary distribution

The artificial agent learns by means of the following simple algorithm: it associates a vector or strengths S_I with the actions I=1,….,N at each time t, where the current sum of these strengths is C_t and p_t is the agent's probabilities of taking actions 1 through N at time t

Four-step decision process (at time t)

1- the agent calculates the probability vector as the relative strengths associated with each action, it sets $p_t = S_t/C_t$

2- chooses one action *j* from the set according to the probabilities p_t and triggers that action

3- observes the payoff received and updates strength by adding the chosen action j's payoff to action j's strength, i.e. it sets the strengths to $S_t+\beta_t$
where $\beta_t = \Phi(j)\, e_j$ (e_j is the jth unit vector)

4- renormalizes the strengths to sum to a value from a prechosen time sequence. In this case it renormalizes strengths to sum to $C_t=Ct^v$

In this way the rate of learning is proportional to $1/(Ct^v)$ and parameters C and v define a two-parameter family of algorithms that can be used to calibrate the automaton

Behavioural interpretation of the algorithm

The strength vector summarizes the current confidence the agent or automaton has learned to associate with actions 1 through N.

Confidence associated with an action increases according to the random payoff it brings in when taken.

S_0 is the initial confidence in the action which represents prior beliefs carried over from past experiences

Connectionistic interpretation of the algorithm

Algorithm as a set of N classifiers each competing to be activated where the classifier J is the simple couple "If is time to act/choose alternative *j"*.

One classifier is triggered on the basis of current strengths and the chosen classifier's strength is updated by the associated reward.

Algorithm nonlinearity

Actions that are frequently taken are further reinforced (to permit exploitation of useful actions)

Algorithm stocasticity

Actions triggered randomly and rewards are drawn randomly from a distribution (to allow for exploration)

What about the *long-run properties* of the system?

There is a tradeoff between two events:
1) If an inferior highpayoff action is triggered early, it gains immediately in strength and may dominate other superior actions
2) If exploration does not disappear too fast, it uncover that there are better actions than the inferior ones

Arthur's (1991) main result: the probability of choosing action *i* grows at a rate proportional to the difference between the expected payoff and the average payoff at current probabilities plus an unbiased perturbation term

Calibration of C and v parameters for three purposes

a- representation of actual human behaviour

b- if the measured value of *v* lies within the range that guarantees asymptotically optimal choices

c- characteristics of learning (speed and ability to discriminate)

Arthur (1991) uses Robillard experiment to calculate v=0.00 C=31.1 and obtains a good fit between automata and human beings (see Figures 1-2 p. 356)

In Herrnstein et al. (1990) the distribution of payoffs is no longer fixed but depends instead on the frequency of actions taken

In this case human behaviour – that is replicated by artificial agents - shows the characteristic of *melioration*: their choices do not converge to optimal frequencies that maximize expected payoff, but to quite different frequencies that equalize expected payoffs of each action

What about optimal convergence?

The experimental and artificial result is that the likelihood of convergence to Nash depends on the difficulty of discrimination among the action payoffs.

Human choice, if captured by the calibration, appears to "discover" and exploit the optimal action with high probability, **as long as it is not difficult to discriminate.**

Beyond a certain perceptual threshold, where differences in alternatives become less pronounced, nonoptimal outcomes become more likely

Economic applications

In simulated financial markets the calibrated agents learn to buy and sell stock appropriately and that stock price indeed converges to small fluctuations around the rational expectations value

But also speculative bubbles and crashes may occur

Conclusions

It is possible to design artificial learning agents and calibrate their "rationality" to replicate human behaviour

This simulation can also reproduce two stylized facts that are well-known to psychologists:

a) with frequency-dependent payoffs humans "meliorate" rather than optimizing

b) there is a threshold in discrimination among payoffs below which humans may lock in to suboptimal choices

The simulation shows that humans may converge to Nash equilibrium but may also select inferior choices and explore less known alternatives.

There is a characteristic learning time for human decisions because behaviour does not settle before 40 to 100 more trials.

4. NEUROECONOMICS

4.1 Definitions and tools

Neuroeconomics is the grounding of microeconomics in details of neural functioning. While the revealed preferences approach has deliberately avoided trying to discover the neural determinant of choices, neuroscience is beginning to allow direct measurement of thoughts and feelings

Methodologically, neuroeconomics is not intended to test economic theory in a traditional way - particularly under the view that utilities and beliefs are only revealed by choices - but to establish the neural circuitry underlying economic decisions, for the eventual purpose of making better predictions.

Starting points

- Much of the brain is constructed to support automatic processes which are faster than conscious deliberation and which occur with little or no awareness or feeling of effort
- Economic behaviour is under the pervasive and often unrecognized influence of finely tuned affective (emotion) systems that are localized in particular brain region
- If affective systems are damaged or perturbed by brain damage, stress, imbalances in neurotransmitters, alcohol or "the heat of the moment" the deliberative system generally is not capable of getting the job done
- Many behaviours that are clearly established to be caused by automatic or affective systems are *interpreted* by human subjects, spuriously, as the product of cognitive deliberation
- The deliberative system, which is the system that is responsible for making sense of behaviour, does not have perfect access to the output of the other systems, and exaggerates the importance of processes it understands when it attempts to make sense of the body's behaviour.

Consequence: behaviour emerges from the interplay between controlled and automatic systems on the one hand and between cognitive and affective systems on the other

Main tools of analysis

Brain imaging

Comparison of people performing different tasks (an "experimental" task and a "control" task) by observing the images of the regions of the brain that are differentially activated by the experimental task.

Changes in electric currents (to measure electrical activity of the brain)	**Changes in metabolism** (to measure neural metabolism processes)
Electro-encephalogram (EEG) Magnetencephalogaphy (MEG)	Positron emission topography (PET) Functional transcranial Doppler-Sonography (FTCD) Functional magnetic resonance imaging (fMRI)

- Electro-encephalogram (or EEG) uses electrodes attached to the scalp to measure electrical activity synchronized to stimulus events or behavioural responses known as Event Related Potentials, or ERPs
(unobtrusiveness and precise temporal sequence but poor spatial resolution)

- Magnetencephalogaphy (MEG) captures magnetic currents running along individual nerve fibers
(good temporal resolution and depict also deeper cortical brain structures)

- Positron emission topography (PET) scanning measures blood flow in the brain, which is a reasonable proxy for neural activity, since neural activity in a region leads to increased blood flow to that region
(poor temporal resolution for stochastic lag of blood, but excellent spatial resolution and use of radioactive contrast substances)

- Functional transcranial Doppler-Sonography (FTCD) measure blood flow velocity within two cerebral arteries and draw inferences on their relative activity level
(activable only on certain areas of the brain)

- Functional magnetic resonance imaging (fMRI) is the most popular technology. It tracks activity in the brain proxied by changes in blood oxygenation. The procedure uses magnetic fields and radio waves in order to depict different kinds of body issue. The strenght of transmitted MR signals varies according to the density of the different kinds of body tissue and the strenght of the magnetic field. The MR signals are captured by detectors and converted by a computer into colored maps
(good spatial and temporal resolution)

Other tools of analysis

A) Single-neuron measurement

Tiny electrodes are inserted into the brain, each measuring a single neuron's firing, but insertion of the wires damages neurons and thus it is largely restricted to animals

Studying animals is informative about humans because many brain structures and functions of non-human mammals are similar to those of humans (e.g., we are more genetically similar to many species of monkeys than those species are to other species).

B) Electrical brain stimulation (EBS)

Other method largely restricted to animals.

In 1954, psychologists James Olds and Peter Milner (Olds and Milner 1954) discovered that rats would learn and execute novel behaviours if rewarded by brief pulses of electrical brain stimulation (EBS) to certain sites in the brain.

Only one EBS study by economists (Green and Rachlin 1991).

C) Psychopathology and brain damage in humans

Chronic mental illnesses (e.g., schizophrenia), developmental disorders (e.g., autism), and degenerative diseases of the nervous system help us understand how the brain works. Most forms of illness have been associated with specific brain areas. Localized brain damage, produced by accidents and strokes, is also a rich source of insight, especially when damage is random (e.g. Damasio 1994).

When patients with known damage to an area X perform a special task more poorly than "normal" patients, and do other tasks equally well, one can infer that area X is used to do the special task.

D) Transcranial magnetic stimulation (TMS)

Use of pulsed magnetic fields to temporarily disrupt brain function in specific regions. The difference in cognitive and behavioural functioning that results from such disruptions provides clues about which regions control which neural functions.

The use of TMS is currently limited to the cortex and is somewhat controversial because it can causes seizures and may have other bad long-run effects.

Basic facts about the brain

The brain is divided into four lobes: frontal, parietal, occipital and temporal

Regions of these lobes are interconnected and create specialized "circuits" for performing various tasks.

Prefrontal cortex seems to assume the role of an important hub. It is the executive region because it draws inputs from almost all other regions and plan actions.

The prefrontal area is the region that has grown the most in the course of human evolution and which, therefore, most sharply differentiates us from our closest primate relatives

The human brain is a primate brain with more neocortex. The fact that many human and animal brain structures are shared means that human behaviour generally involves interaction between "old" brain regions and more newly-evolved ones.

We might learn something about human behaviour from other species.

For example, rats become addicted to all drugs that humans become biologically addicted to, which implies that old reward circuitry shared by rat and human brains is part of human addiction.

While we often think of complex behaviour as deliberate, resources for "executive function" or "cognitive control" are rather scarce. As a result, the brain and body delegate components of complex behaviour into automatic processes.

For example, a student driver is overwhelmed by visual cues, verbal commands, memory required for navigation, and mastery of motor skills.

Many accidents result during this learning process

But within a few years, driving becomes so effortless that drivers can eat and talk (perhaps on a cell phone) while driving safely

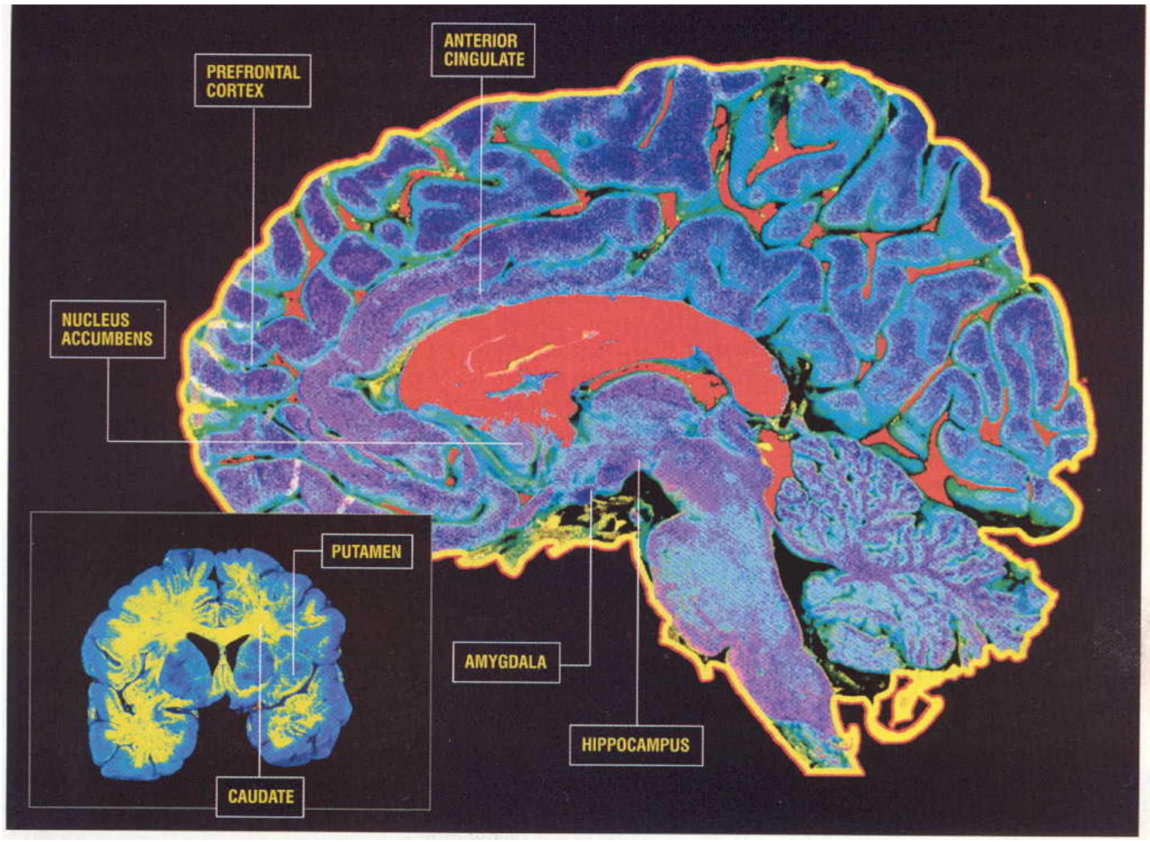


Table 1: A two-dimensional characterization of neural functioning

	Cognitive processes	Affective processes
Controlled Processes	I orbital and prefrontal (front)	II
Automatic Processes	III occipital (back) parietal (top) temporal (side)	IV amygdala (below the cortex)

Quadrant I - deliberate whether to refinance your house, poring over present-value calculations (is the realm of economics)

Quadrant II - used by "method actors" who imagine previous emotional experiences to fool audiences into thinking they are experiencing those emotions

Quadrant III - governs the movement of your hand

Quadrant IV - makes you jump when somebody screws

Controlled processes

(game theory, dynamic programming in economics epitomizes this kind of process)

- conscious and introspectively accessible
- tend to be serial and to use a step-by-step logic,
- tend to be invoked deliberately by the agent when her or she encounters a challenge or surprise
- are often associated with a subjective feeling of effort

Automatic processes

- operate outside of conscious awareness
- tend to operate in parallel
- are often associated with a feeling of effort
- people often have surprisingly little introspective access to automatic choices

Ex. a face is perceived as 'attractive', or a verbal remark as 'sarcastic', automatically and effortlessly

Cognitive processes

- those that answer true/false questions

Affective processes

- those that motivate approach/avoidance behaviour.
- include emotions such as anger, sadness, and shame, as well as "biological affects" such as hunger, pain, and the sex drive.

Quadrants in action

Ex.: a party host approaches you with a plate of sushi.

Quadrant III: try to figure out what is on the plate.

The occipital cortex in the back of the brain is the first on the scene, drawing in signals from your eyes via your optic nerves. It decodes the sushi into primitive patterns such as lines and corners (), then uses a "cascading process" to discern larger shapes (Kosslyn 1994).

Further downstream, in the inferior temporal visual cortex (ITVC), this information becomes integrated with stored representations of objects, which permits you to recognize the objects on the plate as sushi.

Quadrant IV: affect enters the picture.

Neurons in the inferior temporal visual cortex are sensitive only to the identity of an object; they don't tell you whether it will taste good.

Outputs of the inferior temporal visual cortex as well as outputs from other sensory systems feed into the orbitofrontal cortex to determine the food's reward value.

Reward value depends on many factors: personal history (if you got sick on sushi in the past, you will have an automatic aversion to it- the amygdala seems to play a critical role in this kind of long-term learning), current level of hunger (the orbitofrontal cortex and a subcortical region called the hypothalamus are sensitive to the level of hunger).

Quadrants I and II: actions and further processing

If you are hungry, and like sushi, your motor cortex will guide your arm to reach for the sushi and eat it, drawing on automatic quadrant III (reaching) and IV (taste and enjoyment) processes.

higher level processing may enter the picture: explicit memories from the hippocampus (if you saw a recent documentary on the risks of eating raw fish, you may recoil) inputs from the affective system (sometimes referred to as the "limbic system"), and anticipation (planning) from the prefrontal cortex

Automatic processes

Key principles

- **Parallelism**
- **Plasticity**
- **Modularity**
- **Specialization**

Parallelism

- much of the brain's processing involves processes that unfold in parallel and are not accessible to consciousness
- to give rapid responses (connectionist neural network models)
- black box processes

Plasticity

- the brain undergoes physical changes as a result of these processes: when signals are repeatedly conveyed from one neuron to another, the connections between those neurons strengthen (Hebb 1949)
- information processing is unlikely to be reversible because the physiological processes that produce learning are themselves not reversible
- we cannot ignore the effect of useless information or can undo the effect of information that is redundant or discredited (when people form beliefs based on evidence that is later discredited definitively, the belief founded on the discredited evidence persists)

Modularity

- it draws upon multiple modules specialized to perform specific functions

- neurons in different parts of the brain have different shapes and structures (ex. Broca's and Wernicke's areas are involved, in the production and comprehension of language - patients with Wernicke damage can speak sentences of correctly articulated words strung together into ungrammatical gibberish)

- many neuroscientists believe there is a specialized 'mentalizing' module, which controls a person's inferences about what other people believe, feel, or might do - recent exp on monkeys:

  "mirror neurons" in the prefrontal cortex, which fire either when an experimenter performs a physical action (e.g., grasping a peanut) or when the monkey performs ("mirrors") the same action: having such neurons makes learning by imitation easy and supports mind reading by, for example, internally simulating the facial expressions of others

- Economic theory plainly assumes that agents can 'mentalize', i.e., make inferences from the actions of others to their underlying preferences and beliefs. Such mentalizing inferences sustain a Bayesian equilibrium, and are not normally regarded any more or less difficult than any other types of inferences. Neuroscience shows that mentalizing is a special ability and general logical-deductive reasoning can only partially compensate for its absence. Furthermore, it would not be surprising to find normal individuals differing in their mentalizing skills

  McCabe et al. (2001): mentalizing is important in games involving trust and cooperation (players more trusting and cooperative show more brain activity in Brodmann area 10 (thought to be the locus of mentalizing) and more activity in the limbic system which processes emotions)

Specialization

- it figures out how to use existing modules to accomplish new tasks efficiently, whatever functions they originally evolved to perform

- when the brain is confronted with a new problem it initially draws heavily on diverse modules, including, often, the prefrontal cortex, but over time, activity becomes more concentrated in modules that specialized in processing relevant to the task

- for a wide range of problems and tasks, people will rely on cognitive capabilities (modules) that are relatively well developed, such as visual perception and object recognition rather than operations that we are not very good at, like decomposing and then summing up costs and benefits.
  Gobet and Simon (1996): expert chess players were able to store the positions of players almost instantly – but only if they were in positions corresponding to a plausible game. For randomly arranged chess pieces, the experts were not much better than novices

- economics implicitly assumes that people have general cognitive capabilities that can be applied to any type of problem, and hence that they will perform equivalently on problems that have similar structure.
  Automaticity, in contrast, suggests that performance will depend critically on whether a particular problem can be, and is in fact, processed by a module that is well adapted to that form of processing.
  Ex.: there is no module that directly corresponds to time preference, but modules that are responsible for different dimensions of time preference – e.g., for inhibiting emotion-driven behaviour and for thinking about future consequences -> economic agent are heterogeneous

Affective processes

Principles guiding the affective system

Homeostasis
Raw motivation
Collaboration
Competition
Erroneous sense-making

Homeostasis

Affective system involves detectors that monitor when a system departs from a 'set-point' and mechanisms that restore equilibrium when such departures are detected (it is highly attuned to *changes* in stimuli rather than their *levels*)

Some of these mechanisms do not involve deliberate action (when the core body temperature falls below the 98.6F set-point, blood tends to be withdrawn from extremities, and when it rises above the set-point one begins to sweat)

Homeostasis may explain why the evaluation of risky gambles depends on a reference point which encodes whether an outcome is a gain or a loss

Homeostasis poses a fundamental challenge to the economic account of behaviour

For economists preferences are the starting point for human behaviour and behaviour is the ending point.

For neuroscientist explicit behaviour is only one of many mechanisms that the brain uses to maintain homeostasis, and preferences are transient state variables that ensure survival and reproduction.

Human's action to maximize preferences starts in the middle (or perhaps near the end) of the neuroscience account, that views pleasure not as the goal of human behaviour but as a homeostatic cue, i.e. an informational signal

Raw motivation

Economists usually view behaviour as a search for pleasure

Neuroscience and other areas of psychology show that the motivation to take an action is not always closely tied to hedonic consequences

Ex: Ken Berridge (1996) argues that decision making involves the interaction of two separate, though overlapping systems, one responsible for pleasure and pain - the "liking" system -, and the other for motivation - the "wanting" system.

Economics proceeds on the assumption that satisfying people's wants is a good thing.

But if wanting and liking are two separate processes, then it cannot be assumed that satisfying someone's desires necessarily makes them better off.

Welfare economics would need to be augmented by an analysis of when and why wanting and liking do, sometimes, diverge.

Collaboration - Interactions between the systems

Behaviour emerges from a continuous interplay between neural systems supporting activity within each of the four quadrants -> Quadrant I is not able do the job alone

Different processes – most notably affective and cognitive – often drive behaviour in conflicting directions and compete for control of behaviour

The affective system provides inputs into deliberative decision making in the form of crude affective evaluations of behavioural options (Damasio refers to as "somatic markers.")

This approach attributes these patients' inability to make advantageous decisions in real-life to a defect in an emotional mechanism that rapidly signals the prospective consequences of an action, and accordingly assists in the selection of an advantageous response option

Deprived of this emotional signal, these patients rely on a reasoned cost-benefit analysis of numerous and often conflicting options involving both immediate and future consequences.

The impairment degrades the speed of deliberation (e.g., choosing between two brands of cereal may take a patient a very long time because of endless reasoned analyses of the pros and cons of each brand), and also degrades the adequacy of the choice, i.e., patients may choose disadvantageously.

Affect can also distort cognitive judgments:

- Emotions have powerful effects on memory: when people become sad, they tend to recall sad memories
- Emotions affect perceptions of risks : anger makes people less threatened by risks, and sadness makes them more threatened (Lerner and Keltner 2001)
- Emotions also create "motivated cognition": people are good at persuading themselves that what they would *like* to happen is what *will* happen.

Competition

When it comes to spending money or delaying gratification, taking or avoiding risks and behaving kindly or nastily toward other people, people often find themselves of "two minds"

Our affective systems drive us in one direction and cognitive deliberations in another.

Theory of multiple selves

Erroneous sense-making

The brain's powerful drive toward *sense making* leads us to strive to interpret our own behaviour.

Such interpretations use quadrant I mechanisms to make sense of behaviour which is caused by all four quadrants and their interaction.

Since quadrant I often does not have conscious access to behaviour in the other quadrants, it is often tends to over attribute behaviour to itself, i.e. to a deliberative decision process.

Ex.: labour market discrimination

Economic models assume that labour market discrimination against minorities is either a *taste* (a dislike of minorities or a distaste for working with them), or a *belief* (that minority status is a proxy for unobservable differences in skill, or "statistical discrimination").

Neuroscience suggests that automaticity contributes to discrimination because neural networks rapidly spread activation through associated concepts and stereotypes.

Discrimination originated by rapid, automatic, associations between social categories, stereotypes, and affect.

Because people lack introspective access to the processes that produce such biases, they are unable to correct for them even when they are motivated to make impartial judgments and decisions.

Overview of the Economic Applications (2001 – 2004)

Kenning, Peter – Plassman, Hilke, "Neuroeconomics", University of Munster, 2005

Author	Theoretical Background	Problem	Method	Results
Breiter, Aharon, Kahneman, Dale, & Shizgal (2001)	Prospect theory	neural reactions to anticipation and experience of monetary gains and losses	fMRI	Partly differing brain areas for expectations and factual counting of monetary incentives involved as well as partly identical areas of the brain. The latter overlap with regions that respond actively to tasting stimuli and drugs that creates euphoria
Lo/Repin Lo & Repin (2001)	Decisions in financial markets	role of emotions in live decisionmaking processes of stockbrokers	EEG/ E DR	differing activation states depending on degree of market volatility and experience of the dealer
McCabe, Houser, Ryan, Smith, & Trouard (2001)	Game theory, particularly trust and willingness to cooperate	Determination of neural correlates of cooperative behavior	fMRI	Relationships between cooperation as well as willingness to trust and brain activity in areas responsible for emotional processes as well as their integration in decision-making
Erk, Spitzer, Wunderlich, Galley, & Walter (2002)	Choice decisions between different cultural objects or products (automobile)	investigation of neural representations of social incentives	fMRI	Products which symbolize wealth and status lead to increased activity in areas of the brain that are responsible for perceptions of rewards
Kenning et al. (2002)	Preference decisions of consumers with respect to markets	Neural correlates of brands in decisionmaking processes	fMRI	Subjectively strong brands relieve pressure on areas responsible for rational processes and lead to increase activity in those areas responsible for emotional decisions
Smith, Dickhaut, McCabe, & Pardo (2002)	Game theory, in particular decision-making subject to ambiguity, risk, gains and losses	Neural carriers of attitudes about monetary payments (gains or losses) and assumptions about possible outcomes (risk or ambiguity)	PET	Independence between attitudes about payments and assumptions about the probability of outcomes in the form of different neural systems

Sanfey et al. (2003)	Game theory, in particular Ultimatum Game	Investigation of neural carriers of cognitive and emotional decisionmaking processes during the Ultimatum Game	fMRI	Interrelationships between fair and unfair behavior and areas of the brain which are responsible for Processing positive and negative emotional states as well as between decisions to accept or reject
Ambler, Braeutigam, Stins, Rose, & Swithenby (2004)	Purchasing behavior	Comparison of reaction times to complicated (diverse brand) and simple (identical products, but different package sizes) purchasing decisions	MEG	Negative interrelationship between brand familiarity and time required for decisionmaking, negative interrelationship between simple purchasing decision and reaction time
Knutson & Peterson (in press)	Decision-making subject to uncertainty, in particular investor behavior	Determination of neural correlates of expectations benefit theory	fMRI	Significant role of emotions in anticipating stimuli with respect to neurological reconstruction of expected benefit
de Quervain et al. (2004)	Altruism, cooperation	Investigation of neural bases of "altruistic punishment"	PET	Sanctions against defectors activate reward centers in the brain (reward related brain regions)
McClure et al. (2004b)	Neural impact of visual stimuli (brand)	Neural bases for evaluating a soft drink	fMRI	Depending on whether and, if yes, what brand information given to subjects will activate the enjoyment of a soft drink with respect to various different regions of the brain
McClure, Laibson, Loewenstein, & Cohen (2004a)	Temporal preferences for monetary stimuli	Neural bases for discounting alternative premiums	fMRI	Short-term premiums activate limbic regions, long-term premiums are processed in the prefrontal cortex

Some illustrative applications

- **Intertemporal Choice and self-control**
- **Decision-making under risk and uncertainty**
- **Mirror neurons and strategic choice**
- **Trust game and brain activation**
- **Oxytocin and trusting behaviour**

4.2 Intertemporal Choice and self-control

Intertemporal choice as a trade-off of utility at different points in time. Individual differences in the way that people make this trade-off are captured by the notion of a discount rate at which people discount future utilities as a function of when they occur

Empirical research on time discounting challenges the idea that people discount all future utilities at a constant rate

To understand intertemporal choice in humans we need to take account not only of the processes we share with other animals (especially quadrant IV), but also those that are uniquely human (quadrant I)

Collaboration between the two kinds of processes: decisions to delay gratification involve an admixture of affect and cognition.

The cognitive awareness of the delayed benefits in delaying gratification is insufficient alone to motivate delay of gratification, i.e. emotions play a critical role in forward-looking decision making.

"The capacity to experience anxiety and the capacity to plan are two sides of the same coin" (David Barlow 1988)

People only care about the delayed consequences of their decisions to the degree that contemplating such consequences evokes immediate affect

Ex.: frontal lobotomies create a deficiency in areas of the brain that underlie the capacity for images of absent events to generate experiences of pleasure or discomfort (Cottle and Klineberg 1974)

Competition between the two kinds of processes: ubiquity of self-control problems in which one's cognitive judgment of the best course of behaviour departs from the actions one is affectively motivated to take.

The affective system is designed to ensure that certain survival and reproduction functions are met and it achieves this function in part by motivating individuals to take certain actions.

In most animals, emotions and drives motivate behaviours that have short-term goals, such as eating, drinking, copulating, so the automatic system is inherently myopic

Humans are different from other animals in that we worry about or derive immediate pleasure from thinking about delayed consequences, so our affective system can also motivate behaviours that have long-term beneficial consequences.

Indeed, a number of human pathologies, such as anxiety disorders, workaholics, and self-destructive miserliness, seem to be driven by an *excess* of future-mindedness.

An intriguing aspect of self-control is that it is often associated with a subjective feeling of mental effort.

It is tempting to attribute this to the fact that self-control involves the same part of the brain – the executive prefrontal cortex – that is itself associated with feelings of mental effort.

In a typical study, subjects on diets who resisted temptation (by foregoing the chance to grab snacks from a nearby basket) later ate more ice cream in an ice-cream taste test and also quit earlier when confronted with an intellectual problem they couldn't solve. They acted as if their ability to resist temptation was temporarily "used up" by resisting the snacks or, alternatively, that they had "earned" a reward of ice cream by skipping the tempting snacks

The contribution of neuroscience to a better understanding of the problem

1) Neuroscience research points to ways to "unpack" the concept of time preference

2) Because people are likely to make myopic choices when under the influence of powerful drives or emotions this suggests that a key to understanding impulsivity in individuals might be to understand what types of situations get them 'hot'

3) We might be tempted to look for individual differences in what could be called 'willpower' – i.e., the availability of the scarce internal resource that allows people to inhibit viscerally-driven behaviours.

Such a framework might also help to explain why people appear so inconsistent when their behaviour is viewed through the lens of discounted utility.

The ability to think about future consequences may not be strongly correlated with the degree to which different experiences produce visceral reactions, and these in turn might not be correlated with an individual's level of willpower.

4.3 Decision-making under risk and uncertainty

The expected utility model views decision making under uncertainty as a trade-off of utility under different states of nature

People react to risks at two different levels.

On the one hand, as posited by traditional economic theories and consistent with quadrant I, people do attempt to evaluate the objective level of risk that different hazards could pose.

On the other hand, and consistent with quadrant IV, people also react to risks at an emotional level, and these emotional reactions can powerfully influence their behaviour

Centrality of the neural processes underlying affective responses to risks.

Much risk adverse behaviour is driven by immediate fear responses to risks, and fear, in turn, seems to be largely traceable to a single small area of the brain called the *amygdala*.

The amygdala constantly scans incoming stimuli for indications of potential threat, and responds to inputs both from automatic and controlled processes in the brain.

The amygdala also receives cortical inputs, which can moderate or even override its automatic quadrant IV response.

Decision making under risk and uncertainty, like intertemporal choice, nicely illustrates both collaboration and competition between systems.

Collaboration: risk taking (or avoiding) behaviour involves an exquisite interplay of cognitive and affective processes.

Illustrative experiment

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. (1997). Deciding advantageously before knowing the advantageous strategy. *Science, 275,* 1293–1295.

Subjects

Patients suffering prefrontal damage, which produces a disconnect between cognitive and affective systems, and normal subjects chose a sequence of cards from four decks whose payoffs the subjects only learned from experience.

Design

Two decks had more cards with extreme wins and losses (and negative expected value)

Two decks had less extreme outcomes but positive expected value.

Results

Both groups exhibited similar skin conductance (sweating judged as an indication of fear) after large-loss cards were encountered

But, compared to normals, prefrontal subjects rapidly returned to the high-paying risky decks after suffering a loss and, as a result went "bankrupt" more often.

Although the immediate emotional reaction of the prefrontal patients to losses was the same as the reaction of normals (measured by skin conductance), the damage patients apparently do not store the pain of remembered losses as well as normals, so their skin conductance rose much less than normals when they resampled the high risk decks.

Conclusions

Insufficient fear can produce nonmaximizing behaviour when risky options have negative value

But, it is also well established that fear can also discourage people from taking advantageous gambles (see, e.g., Gneezy and Potters 1997).

Loewenstein, Shiv, Bechara, Damasio and Damasio (2002): frontal patients actually make more money on a task in which negative emotions cause normal subjects to be extremely risk averse: a series of take-it-or-leave-it choices to play a gamble with a 50% chance of losing $1.00 or gaining $1.50.

Normal subjects and frontal subjects were about equally likely to play the gamble on the first round, but normals rapidly stopped playing when they experienced losses while frontal patients' play was unresponsive to losses.

Having frontal damage undermines the overall quality of decision making; but there are situations in which frontal damage can result in superior decisions.

At a more macro level, emotional reactions to risk can help to explain risk-seeking as well as risk-aversion.

GAMBLING

When gambling is pleasurable a model that incorporates affect naturally predicts that people will be risk-seeking and that self-control will be required to reign in risk-taking

The standard economic explanations for gambling-- convex utility for money or a special taste for the act of gambling-- don't help explain why some gamblers binge and don’t usefully inform policies to regulate availability of gambling

Neuroscience’s contribution

Pathological gamblers tend to be overwhelmingly male, and tend to also drink, smoke, and use drugs much more frequently than average.

Genetic evidence shows that a certain gene allele (D_2A1), which causes gamblers to seek larger and larger thrills to get modest jolts of pleasure, is more likely to be present in pathological gamblers than in normal people (Comings 1998)

One study shows tentatively that treatment with naltrexone, a drug that blocks the operation of opiate receptors in the brain, reduces the urge to gamble (e.g. Moreyra, AIbanez, Saiz-Ruiz, Nissenson and Blanco 2000).

The same drug has been used to successfully treat “compulsive shopping” (McElroy, Satlin, Pope, Keck and al. 1991).

Understanding the affective and cognitive components of reactions to risk is especially important when the two diverge and hence compete for control of behaviour (see Loewenstein, Weber, Hsee and Welch 2001): people are indeed often 'of two minds' when it comes to risks

The divergence between different systems' evaluations of risk can also be seen when it comes to judgments of probability.

Numerous studies by psychologists have observed systematic divergences between explicit judgments of probability in different settings (presumably the product of controlled processing) and implicit judgments or judgments derived from choice.

For example, Kirkpatrick and Epstein (1998) found that people prefer to draw a bean from a bowl containing 10 winning beans and 90 losing beans than from a bowl containing 1 winning bean and 9 losing beans.

Subjects say that they know the explicit probabilities of winning are the same, but they still have an automatic quadrant III preference for the bowl with more winning beans.

Another violation is that subjects often report probabilities which are logically incoherent.

FMRI (Functional magnetic resonance imaging) evidence suggests an explanation for why probability judgments are incoherent, but can be corrected upon reflection: when guessing probabilities, the left hemisphere of the brain is more active; but when answering logic questions, the right hemisphere is more active (Parsons and Osherson 2001).

Enforcing logical coherence requires the right hemisphere to `check the work' of the left hemisphere.

Neural evidence also substantiates the distinction between risk (known probability) and "Knightian" uncertainty, or ambiguity.

Subjects facing ambiguous gambles—knowing they lack information they would like to have about the odds—often report a feeling of discomfort or mild fear.

Brain imaging shows that different degrees of risk and uncertainty activate different areas of the brain (McCabe, Houser, Ryan, Smith and Trouard 2001; Rustichini, Dickhaut, Ghirardato, Smith and Pardo 2002) which corroborates the subjects' self-reports

4.4 Mirror neurons and strategic choice

Rustichini, Aldo, "Neuroeconomics: Present and Future", *Games and Economic Behavior*, 52, 2005, 201-212

A sector of the central premotor cortex of the monkey controls hand and mouth movements.

An important functional property of this area is that most of the neurons are active (in the form of a discharge) in association with actions, such as grasping, holding, tearing, and manipulating objects.

The activity of these neurons is not associated with any single movement constituting the action: for example, it is not associated with the initial movement of the action of grasping, but with the action in its entirety.

Rizzolatti et al. (19) discovered that some of the neurons in this area (F5) discharge both when the subject (a monkey) performs the action and when the same subject observes another monkey performing that action.

In view of this dual property of being active both when the action is performed and also when it is observed, these neurons have been called *mirror neurons*

Other experimental facts

1- Mirror neurons generalize: for example, they discharge in response to the observation of a specific action irrespective of whether the subject performing it is close or far

2- Mirror neurons respond to the action (for example, grasping) not to the object being grasped. The presentation of an object in isolation does not induce any activation, even if it does when being grasped

3- There is a wide agreement that mirror neurons facilitate action understanding, through the simulation of the action in the premotor system of the observer. The fundamental element here is that understanding is produced by simulation: that is, by the activation in the observer of the same brain region that produces the action.

It has also been tested that mirror neurons activity extends to human subjects. (Fadiga et al. 1995)

Consequence: there is an overlap of the brain structures that are devoted to the observation and to the execution of actions.

Another extension of the research has tested whether the same principle applies to activities that are not just motor.

Is there a mirroring of different activities, or even internal states, such as emotions?

An experiment (Wickers et al. 2003), shows that this is the case for a simple emotion, namely physical disgust.

In the experiment, subjects were in the condition either of inhaling odorants (which could in turn be disgusting, pleasurable, or neutral) or of observing someone inhaling the same substances by watching their facial expressions in a short movie clip.

A brain imaging analysis of the activation in the two cases isolated for each case a set of brain regions.

The experimental finding is that the intersection of the two sets was nonempty, and consisted in large part of the anterior insula, that is the region in the brain that is usually associated with feelings of disgust (physical or social).

Conclusion: "as observing hand action activates the observer's motor representation of that action, observing an emotion activates the neuronal representation of that emotion" (Wickers et al., 2003).

Similar experiments have been conducted for pain, see for example, Singer et al. (2004).

In both cases, the results provide striking support for the main hypothesis, that there is a substantial overlap between the areas that are activated when we experience an emotion and when we observe someone experiencing that same emotion.

These facts point to sympathy as a component in a rich system of information processing.

Sympathy is the process by which a subject who is observing a second person can internally reproduce the mental process of the observed person.

This reproduction is possible because the neural structure of the subject who is observing and of the observed is similar.

The purpose is that of extracting information from the observed subject.

The general definition includes both types of sympathy:

> the "Smithian sympathy" that proceeds from the observation of the event and simulates the internal state of the observed individual,
>
> the "mirror sympathy" that proceeds from the observation of the actions or displays of affections of the observed person.

In both cases the intent is to acquire information on the internal state of the observed person.

Sympathy, however, is *not* the only element in the process of understanding of the motivations, intentions, and future actions of the others.

Rather it provides information, and hence it is useful in understanding the environment.

But as with any other piece of information, it is even more useful if it is processed, and combined with a prior assessment of the same environment.

In other words, sympathy may be an affective state, but it is always sophisticated and not naive.

Ex.: consider a player who is deciding what to do at a node in an extensive form game.

His decision is obviously influenced by the expectation of what the other players will do at subsequent nodes.

How can he form such expectations? He formulates in his mind some hypothesis about the possible distribution of strategies in the population; or even more indirectly, about the distribution of parameters (such as preferences, or beliefs) that are relevant to shaping these decisions.

Sympathy suggests a different way: the player may introspectively consider what he would do (or, more indirectly, what he would think) at those nodes, and take that as useful information on what the others are likely to do or think.

"Smithian sympathy": the player is considering the effect of an external event (finding oneself at a node) on another person, and is trying to infer the internal state as useful predictor of the action of the other player in that situation.

This hypothesis suggests that the true thought process of the player is going to be a combination of two elements:

1- introspection, because sympathy is permanently active
2- assessment that the player gives of the general population, discounting the information that introspection gives is not representative of the general population

4.5 Trust game and brain activation

Trust game (or investment game)

5. Two players are paired off anonymously and respectively named as the sender and the responder
6. The sender is given a certain amount of money and told that he or she can keep the entire amount or send some or all of it to the responder.
7. Any money passed from the sender to the responder is tripled by the experimenter and then given to the responder.
8. The responder can keep the entire amount or give back some or all of it to the sender.
9. When the sender receives the amount sent back by the responder the game ends.

It measure of the propensities to trust, which is the proportion of the initial endowment sent by the sender, and to reciprocate, which is the ratio between the amount returned and the amount received by the responder.

Backward induction solution:
the responder will not send any money back
anticipating the responder's decision, the sender will not send any money to the responder.

Results from earlier experiments are inconsistent with the conventional game theory prediction. (Innocenti-Pazienza 2006)

Table 1. Experimental results on the trust game

	Berg, Dickhaut and McCabe (1995)		Buchan, Croson and Johnson (2000)		Schwieren and Sutter (2003)	Burks, Carpenter and Verhoogen (2003)			Chaudhuri and Gangadharan (2003)
	No history	Social history	Only U.S.	All		Single role	Both roles, no prior	Both roles, prior	
Trust	51.6	53.6	49	67	65.7	65.0	65.2	47.3	43.3
Reciprocity	30.1	40.2	22	37	37.6	43.6	25.9	17.1	17.5

Trust = Average fraction sent (Amount sent / Initial endowment); Reciprocity = Average fraction returned (Amount sent back/ Amount received)

Berg, Dickhaut McCabe (1995)

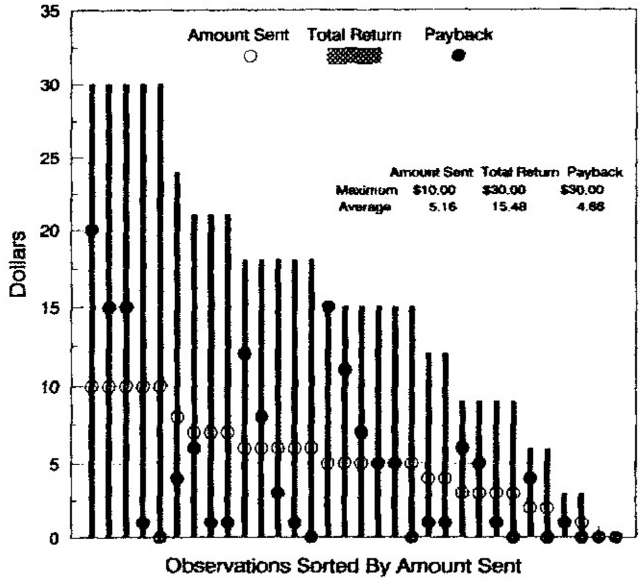


FIG. 2. Trust experiment results showing amount sent (○), total return (■), and payback (●). No history was provided to the subjects.

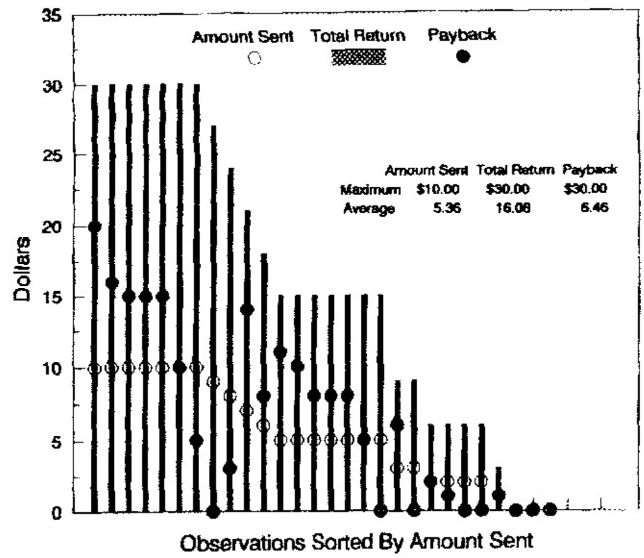
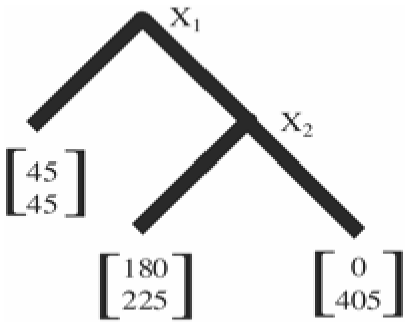


FIG. 3. Trust experiment results showing amount sent (○), total return (■), and payback (●). A social history was provided to the subjects.

[Kevin McCabe, Daniel Houser, Lee Ryan, Vernon Smith, and Theodore Trouard
"A functional imaging study of cooperation in two-person reciprocal exchange"
Proc. Nat. Ac. Sci. USA 2001]

Design
Subjects play the trust game both against a human opponent and against a computer program which, they were told, would play a human-like strategy.



Cooperation seems to be associated with the activation of the anterior paracingulate cortex, a brain region associated with interpreting and monitoring the mental state of others

Tool

Functional magnetic resonance imaging (fMRI)

Data analysis examines the bold response one TR (1.5 s) before the results screen, because decision making for cooperation is likely to be salient at this TR independent of the subject's position in the game

Subjects are likely to ask themselves during this wait condition, "What is my counterpart doing?" and begin to form beliefs about what a delay means about their counterpart's desires.

Expectations

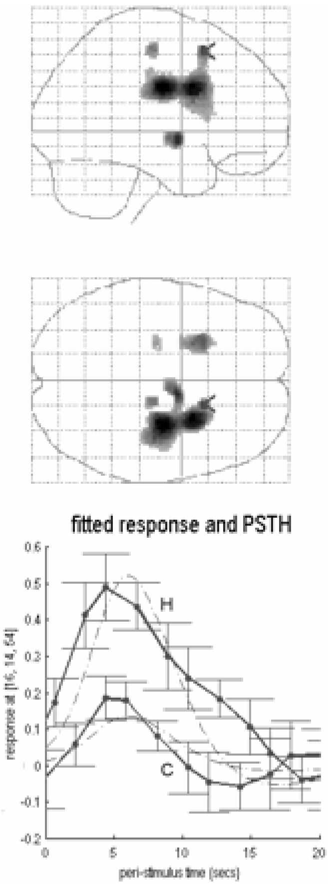
Human and computer treatments to generate differential activations associated with predicting and understanding the cooperative intentions of another human. Our analysis treats the rolling-bars condition as the baseline.

Results

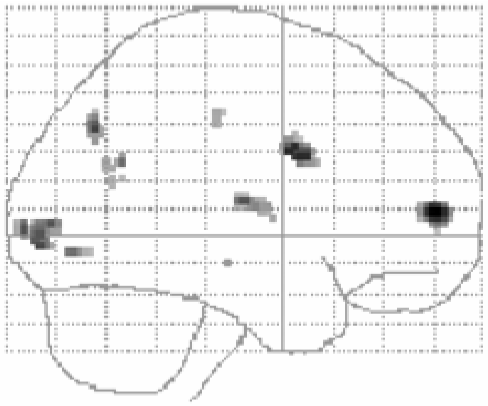
McCabe at al found that subjects were more likely to cooperate with real humans than with computers and that cooperators have significantly different brain activation in the two conditions

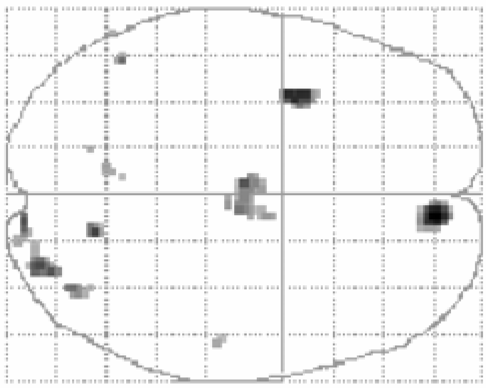
The six subjects with the highest cooperation scores show significant increases in activation in medial prefrontal regions during human-human interactions when compared with human-computer interactions.

The six subjects who received the lowest cooperation scores (22, 10, 18, 21, 11, and 3) did not show significant activation differences in medial prefrontal cortex between the human and computer conditions.



Bold response of a cooperator for the contrast human (H) > computer (C). The blobs on the glass brain are clusters of at least 12 contiguous voxels that show significantly more activation in the human than computer condition (SPM *t* map, *P* < 0.001 uncorrected).





Cooperators have a common pattern of "bold" activation differences.

Cooperation requires an active convergence zone, possibly in prefrontal cortex, that binds joint attention to mutual gains with the inhibition of immediate reward gratification to allow cooperative decisions.

Systematic activation differences are observed in (*i*) the occipital lobe (Brodmann area 17, 18), in which we hypothesize greater visual demands are placed on subjects who are trying to understand both their own payoff/incentives and the payoff/incentives of their counterparts.

Common activation differences are also observed in (*ii*) the parietal lobe (Brodmann area 7), which is part of the "where" pathway for primate vision and (*iii*) the thalamus.

Consistent with our hypothesis that cooperation requires prefrontal control activation, differences are observed in (*iv*) the middle frontal gyrus and (*v*) the frontal pole (Brodmann area 10)

In conclusion, our behavioural data shows that half the subjects in our experiment consistently attempted cooperation with their human counterpart.

Within this group, and within subjects comparison, we find that regions of prefrontal cortex are more active when subjects are playing a human than when they are playing a computer following a fixed (and known) probabilistic strategy. Within the group of noncooperators we find no significant differences in prefrontal cortex between the computer and human conditions.

One possible explanation for our results is that within this class of games, subjects learn to adopt game form-dependent rules of thumb when playing the computer or when playing noncooperatively with a human counterpart.

In comparison, cooperation requires an active convergence zone that binds joint attention to mutual gains with sufficient inhibition of immediate reward gratification to allow cooperative decisions.

4.6 Oxytocin and trusting behavior

Michael Kosfeld, Markus Heinrichs, Paul J. Zak, Urs Fischbacher & Ernst Fehr
"Oxytocin increases trust in humans" *Nature* 2005

In non-human mammals, the neuropeptide oxytocin has a key role in general behavioural regulation, particularly in positive social interactions.

Oxytocin receptors are distributed in various brain regions associated with behaviour, including pair bonding, maternal care, sexual behaviour, and the ability to form normal social attachments.

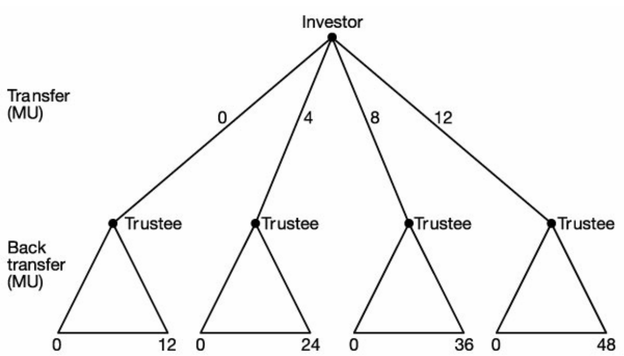
Thus, oxytocin seems to permit animals to facilitate approach behaviour.

HP.: oxytocin might also promote prosocial approach behaviours (such as trust) in humans.

Recent neuroscientific finding: neuropeptides cross the blood-brain barrier after intranasal administration

Double-blind study design to compare:
- trusting behaviour in a group of subjects who receive a dose of intranasal oxytocin
- trusting behaviour in a control group of subjects who receive placebo.

Experiment's object
trust game with real monetary stakes (29+29 subjects)



Hypothesis to test:
oxytocin increases the trusting behaviour of investors $\Rightarrow$ the investors in the oxytocin group will show higher money transfers than those in the placebo group.

Two treatments:

A) standard trust game
B) risk trust game

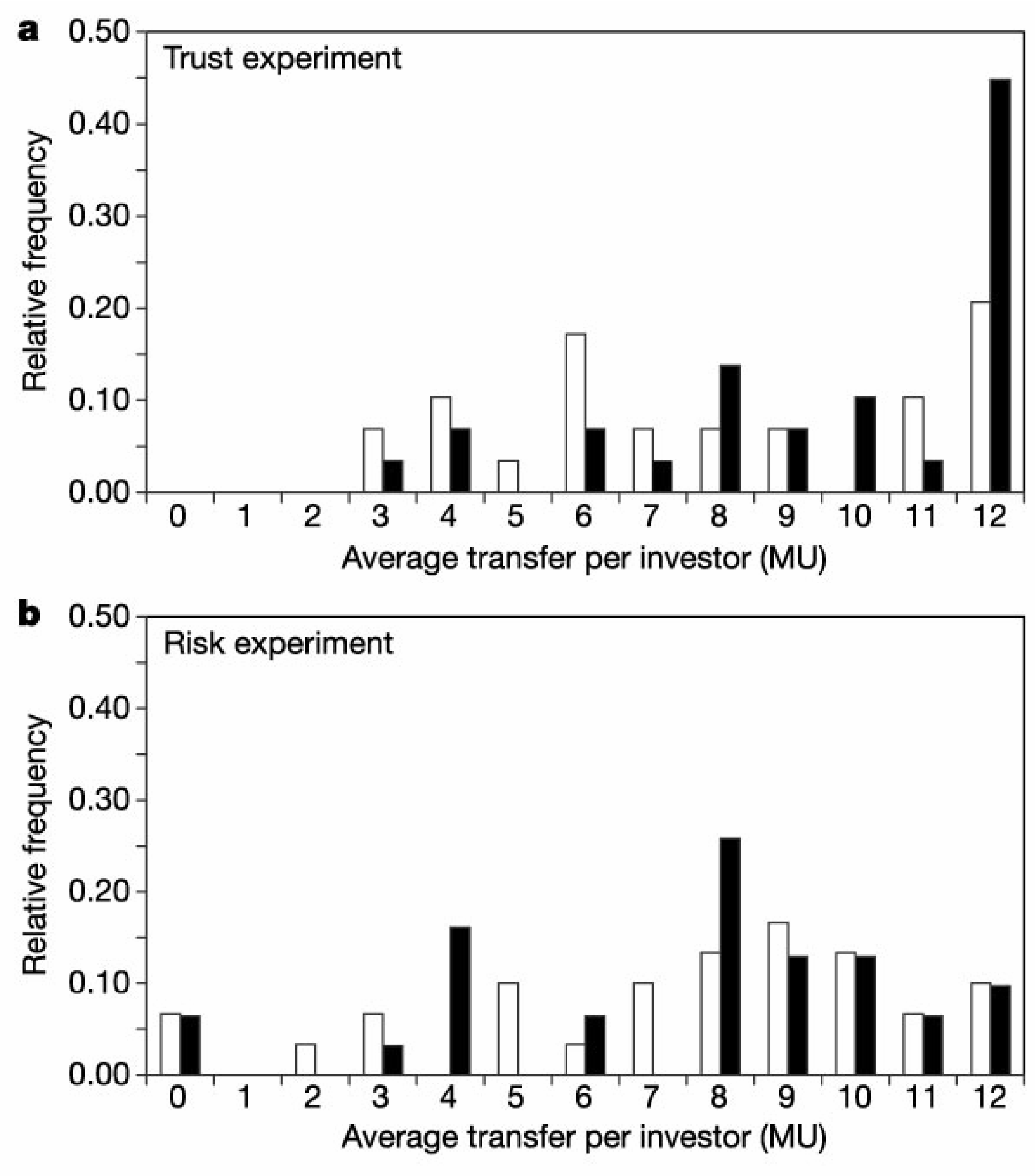
In B the investor faced the same choices as in the trust game but in which a random mechanism, not the trustee's decision, determined the investor's risk.
The random mechanism in the risk experiment replicated the trustees' decisions in the trust experiment.

Therefore, the investors faced exactly the same risk as in the trust experiment; however, their transfer decisions were not embedded in a social interaction because there were no trustees in the risk experiment.

Experimental design

- 194 male students (mean age ±s.d., 22.0 ±3.4 yr) from different universities in Zurich
- 128 participants in the trust experiment and 66 subjects participated in the risk experiment
- Exclusion criteria: medical or psychiatric illness, medication, smoking, drug or alcohol abuse
- Subjects were instructed to abstain from food and drink (other than water) for 2 h before the experiment, and from alcohol, smoking and caffeine for 24 h before the experiment
- Participants were informed at the time of recruitment that the experiment would evaluate the effects of a hormone on decision making
- 16 individuals out of the original sample of 194 were excluded because of incorrect substance administration (7 in the trust experiment, 5 in the risk experiment) or their stated disbelief that the opponent in the trust game was actually a human being (4 participants)
- Subjects received a single intranasal dose of 24 IU oxytocin (Syntocinon-Spray, Novartis; 3 puffs per nostril, each with 4 IU oxytocin) or placebo 50 min before the start of the experiment
- Subjects were randomly assigned to the oxytocin or placebo group
- In order to avoid any subjective substance effects (for example, olfactory effects) other than those caused by oxytocin, the placebo contained all inactive ingredients except for the neuropeptide.

Results



a. Relative frequency of investors' average transfers in oxytocin (filled bars) and placebo (open bars) groups in the trust experiment (n = 58). Subjects given oxytocin show significantly higher transfer levels.
The investors' average transfer is 17% higher in the oxytocin group (Mann-Whitney U-test; z = -1.897, P = 0.029, one-sided).
Median transfer: 10 MU (oxytocin group) > 8 MU (placebo group)

b. Relative frequency of investors' average transfers in oxytocin (filled bars) and placebo (open bars) groups in the risk experiment (n = 61). Subjects in the oxytocin and the placebo group show statistically identical transfer levels.
Median transfer: 8 MU (in both groups)
Average transfer 7.5 MU (in both groups)
(Mann-Whitney U-test; z = 0.022, P = 0.983; two-sided test, n = 31 in oxytocin group, n = 30 in placebo group).

Conclusion

oxytocin increases the investors' transfer levels in the trust experiment but not in the risk experiment ⇒ oxytocin specifically affects trust in interpersonal interactions.

Explanations

a) oxytocin causes a general increase in prosocial inclinations

Oxytocin should affect not only the prosocial behaviour of the investors but also that of the trustees. But trustees given oxytocin do not show more trustworthy behaviour. At every positive transfer level (4, 8 or 12 MU), their back transfers are statistically indistinguishable from those of placebo trustees (Mann Whitney *U*-tests; $P > 0.243$, two-sided tests for each positive transfer level).

b) oxytocin does not increase the general inclination to behave prosocially. Rather, oxytocin specifically affects the trusting behaviour of investors.

c) effect of subjects' beliefs. Oxytocin might render subjects more optimistic about the likelihood of a good outcome.

In order to address this question, we measured the investor's subjective expectation about the trustee's back transfer after every transfer decision. A Mann-Whitney *U*-test indicates that these expectations do not differ significantly between oxytocin and placebo groups at every feasible positive transfer level

d) oxytocin helps subjects to overcome their betrayal aversion in social interactions.

This explanation is consistent with the differing effects of oxytocin across the trust and the risk experiments, and is further supported by the fact that investors faced a considerable betrayal risk.

*****